

INSTITUTO POLITÉCNICO NACIONAL



UNIDAD PROFESIONAL INTERDISCIPLINARIA EN
INGENIERÍA Y TECNOLOGÍAS AVANZADAS

UPIITA

**“SISTEMA DE RECONOCIMIENTO DE VOZ BASADO EN
EL FUNCIONAMIENTO DEL SISTEMA AUDITIVO
HUMANO”**

Trabajo Terminal

Que para obtener el Título de

“Ingeniero en Biónica”

Presentan

Israel Linarez González

José Luis Romero Rodríguez

Asesores

M. en C. Adrián Antonio Castañeda Galván

M. en C. Yesenia Eleonor González Navarro

Dr. Sergio Suárez Guerra



México, D. F., febrero del 2011

INSTITUTO POLITÉCNICO NACIONAL



UNIDAD PROFESIONAL INTERDISCIPLINARIA EN INGENIERÍA
Y TECNOLOGÍAS AVANZADAS

UPIITA

**“SISTEMA DE RECONOCIMIENTO DE VOZ BASADO EN EL
FUNCIONAMIENTO DEL SISTEMA AUDITIVO HUMANO”**

Trabajo Terminal

*Que para obtener el Título de
“Ingeniero en Biónica”*

Presentan

Israel Linarez González
José Luis Romero Rodríguez

Asesores

M. en C. Adrián A. Castañeda Galván

Dr. Sergio Suárez Guerra

M. en C. Yesenia Eleonor González Navarro

Presidente del Jurado

Profesor Titular

M. en C. Álvaro Anzueto Ríos

M. en C. Adrián Morales Blas



México, D. F. febrero del 2011

*“...pero eso si, y en esto soy
irreductible, no les perdono
bajo ningún pretexto que no
sepan volar...”*

A nuestros padres, porque han entregado su vida para ofrecernos la oportunidad de ser dignos y buenos competidores de nosotros mismos.

A nuestros hermanos por que son el máspreciado legado que la naturaleza nos entregó como amigos de toda la vida, compañías en mente y alma... porque este pequeño triunfo también es suyo.

A nuestros abuelos por ser los pilares de esta gran estructura...porque nada podrá derribar el amor de la familia, cuando en práctica se ponen sus enseñanzas.

A nuestros primos y tíos que compartieron enseñanzas y experiencias, a los símbolos de camaradería.

A los primos que llevamos en el corazón, los que nos enseñaron los extremos de la valentía y la fuerza humana cuando uno se aferra a lo que quiere...aun en contra de las leyes de la naturaleza: David, Paty. A ustedes.

A ti mi amigo, porque conocerte fue ya el primero de una serie de interminables triunfos.

AGRADECIMIENTOS

Es difícil expresar la gratitud a este instituto sin que suene simple y limitada; habría que mencionar la magnitud del legado que edificó en nosotros, porque nos construyó con las bases sociales suficientes para demostrar que podemos ejercer nuestra labor profesional comprometidos con los cambios sociales y que nuestra obligación es no olvidar que éste es un instituto del estado que la pagan los trabajadores y que es digno corresponder ofreciendo con calidad nuestro servicio a nuestra patria.

Agradecemos de manera muy especial al Dr. Sergio Suarez Guerra, director del presente trabajo, por facilitarnos sus trabajos y su conocimiento, siendo para nosotros un ejemplo de calidad profesional y humana.

A nuestra asesora, la M. en C. Yesenia Eleonor González Navarro por su gran apoyo, dedicación y presencia, siendo un elemento muy importante para el éxito de este trabajo.

A nuestro asesor el M. en C. Adrián Antonio Castañeda Galván por la atención brindada, no solo en el desarrollo de este trabajo sino a lo largo de nuestra preparación profesional, al cual consideramos un excelente profesor, una gran persona y un ejemplo a seguir.

Al Ing. Carlos Ríos Ramírez porque su desempeño ha sido motivo de tantas generaciones exitosas, que no habría manera de expresar la admiración a una vida dedicada a reducir nuestra ignorancia.

A nuestro amigo Carlitos, por colaborar con la conclusión de este trabajo, sobre todo cuando nuestra mente estaba a punto de desfallecer.

CONTENIDO

CONTENIDO	i
ÍNDICE DE FIGURAS	iv
RESUMEN	1
ABSTRACT	2
GLOSARIO	3
INTRODUCCIÓN	4
Planteamiento del Problema	5
Justificación	5
Objetivos	7
Objetivo General	7
Objetivos Particulares	7
Organización del trabajo	8
ESTADO DEL ARTE	9
Antecedentes	9
Trabajos actuales	11
MARCO TEÓRICO	12
La voz humana	12
Aspecto sonoro del habla	12
Procesamiento matemático de la señal de voz	15
Transformada de Fourier	15
Coeficientes de Predicción Lineal	16
Algoritmo de Levinson-Durbin	18
Extracción de formantes	19
La audición	20
Sistema auditivo humano	20
Sistema auditivo periférico	21
Oído Externo	21
Oído Medio	21
Oído Interno	22
La vía auditiva	24

Localización del sonido	28
Integración en el colículo inferior	31
Sistema auditivo central	32
Corteza cerebral.....	33
Comprensión auditiva de las palabras	35
Lenguaje	36
La experiencia y el lenguaje.....	37
Reconocimiento de voz	38
Redes neuronales artificiales	38
Analogía con las redes neuronales biológicas	39
Función de red	40
Funciones de activación	40
Características.....	40
Arquitectura de redes	41
Aprendizaje.....	41
Redes competitivas o mapas de auto-organización	42
Sistemas de Lógica Difusa (SLD)	43
Solución de problemas por Lógica Difusa	45
Validación cruzada	45
DESARROLLO DE LA PROPUESTA	46
Sistema de adquisición	46
Micrófono	48
Consideraciones biológicas en la elección del micrófono	48
Consideraciones técnicas en la elección del micrófono.....	48
Filtros pasabandas	49
Procesamiento de las señales	51
Lateralización (Diferencia Interaural de Tiempo)	53
Frecuencia fundamental (F_0)	55
Cálculo de la energía	56
Detección de palabra	58
Ventaneo	58
Extracción de parámetros utilizando Coeficientes de Predicción Lineal (LPC).....	59

Identificación del sexo del locutor	61
Localización del sonido.....	62
Reconocimiento de vocales.....	65
RESULTADOS Y DISCUSIÓN.....	66
Interfaz de usuario	66
Validación del sistema.....	67
Resultados en el reconocimiento de vocales.....	68
Resultados para el reconocimiento de sexo	89
Resultados de localización (lateralización)	92
Discusión de resultados.....	100
Conclusiones	101
Trabajos Futuros.....	101
REFERENCIAS	103

ÍNDICE DE FIGURAS

Figura 3. 1 Ubicación de las cuerdas vocales en la laringe y la acción de sus principales músculos. Copyright © (18)	12
Figura 3. 2 Esquema de la producción de voz.	13
Figura 3. 3 Espectro de la palabra <i>uno</i>	13
Figura 3. 4 Comparación entre las señales de sonidos <i>sonoro, sordo y plosivo</i> respectivamente. ...	14
Figura 3. 5 Modelo de predicción lineal para la producción de voz. Copyright © (12)	15
Figura 3. 6 Respuestas de frecuencias $H(w)$ dB del filtro todo polos con $p=6, 15, 24$. Copyright © (16)	19
Figura 3. 7 Oído humano Copyright © (18).....	22
Figura 3. 8 Ondas de propagación a lo largo de la cóclea. Copyright © (18)	23
Figura 3. 9 Potenciales de receptor de una célula ciliada. Copyright © (18)	24
Figura 3. 10 Curvas de sintonización de frecuencias de seis fibras del nervio auditivo.	24
Figura 3. 11 Las vías auditivas más importantes. Copyright © (18)	26
Figura 3. 12 Corteza Auditiva Humana. Copyright © (18).....	27
Figura 3. 13 Localización de sonido por diferencia de tiempo interaural. Copyright © (18)	29
Figura 3. 14 Diferencias de intensidad interaural. Copyright © (18).....	31
Figura 3. 15 Estructuras involucradas en la repetición de una palabra oída, de acuerdo con la conceptualización de Wernicke de la función del lenguaje.	32
Figura 3. 16 Vista lateral izquierda del cerebro en donde se presenta su clasificación anatómica. Copyright © (18)	33
Figura 3. 17 Hemisferio cerebral izquierdo con sus áreas de Brodmann numeradas.	34
Figura 3. 18 Córtex especializada según la citoarquitectura. Copyright © (18)	34
Figura 3. 19 Imágenes PET (Tomografía por Emisión de Positrones). Copyright © (19)	36
Figura 3. 20 Unidad de proceso típica.....	39
Figura 3. 21 Red neuronal artificial perceptrón simple.....	41
Figura 3. 22 Conjuntos Difusos más utilizados en ingeniería	44
Figura 3. 23 Comparación entre conjuntos convencionales y conjuntos difusos.....	44
Figura 3. 24 Comparación entre las operaciones AND y OR para conjuntos binarios y conjuntos difusos.	44
Figura 3. 25 Red neuronal artificial perceptrón simple.....	45
Figura 4. 1 Diagrama general en materia de hardware.	46
Figura 4. 2 Los tres planos característicos para estudiar la localización por parte del ser humano. Copyright © (27)	47
Figura 4. 3 Pabellones auriculares artificiales.	47
Figura 4. 4 Circuito de polarización de un micrófono electret.	49
Figura 4. 5 Filtro activo pasabandas de primer orden.	50
Figura 4. 6 Diagrama a bloques del primer procesamiento en la PC.	52
Figura 4. 7 Diferencias en las distancias que deben recorrer las ondas.	53

Figura 4. 8 Gráfica que describe la diferencia de tiempo interaural dependiente del ángulo de incidencia.	53
Figura 4. 9 Respuesta en frecuencia del filtro digital pasabandas de orden 41 diseñado.	54
Figura 4. 10 Procesamiento aplicado a la señal cuyos resultados son: reconocimiento de vocales, género y localización de fuente.	57
Figura 4. 11 Respuesta del filtro pasa altas de 1er orden diseñado para un método de detección de palabra.	58
Figura 4. 12 Grafica de la función ventana de Hamming en el tiempo y en frecuencia.	59
Figura 4. 13 LPC vocal "a"	59
Figura 4. 14 LPC vocal "e"	60
Figura 4. 15 LPC vocal "i"	60
Figura 4. 16 LPC vocal "o"	60
Figura 4. 17 LPC vocal "u"	60
Figura 4. 18 Universo difuso para el reconocimiento del sexo.	61
Figura 4. 19 Universo difuso para la localización del sonido.	62
Figura 4. 20 Picos de resonancia en las HRTF corresponden a diferentes localizaciones de las fuentes sonoras en el plano medio.	64
Figura 4. 21 Gráfica de la HRTF de frente y atrás.	64
Figura 5. 1 Interfaz de usuario.	66

Sistema de reconocimiento de voz basado en el funcionamiento del sistema auditivo humano

RESUMEN

El presente trabajo describe el diseño y desarrollo de un sistema multilocutor de reconocimiento de voz capaz de reconocer las 5 vocales del alfabeto español de manera aislada para un conjunto de 6 personas (3 hombres y 3 mujeres).

El diseño del sistema contempla la morfología del pabellón auricular y la estructura general de la vía auditiva, con el fin de otorgarle al mismo la capacidad de mostrar algunas de las características cualitativas del sistema auditivo humano tales como la localización de la fuente sonora y la identificación del locutor (si es hombre o mujer).

El sistema utiliza parámetros como los Coeficientes de Predicción Lineal LPC, la energía de segmentos y la transformada de Fourier, que son interpretados y comparados mediante redes neuronales de competencia tipo ganador toma todo WTA.

El trabajo presentado pretende proponer una metodología en la medición y análisis de señales para la caracterización de la respuesta humana ante sonidos provenientes de distintas posiciones en el espacio, además de permear la técnica biónica en la realización de sistemas de reconocimiento de voz, sumándose a la búsqueda de algoritmos más robustos que permitan cubrir exigencias en los estudios referidos al reconocimiento de voz con una proyección a establecer el reconocimiento de voz como interfaz de comunicación hombre-máquina, cuyo impacto estará descrito en el ámbito psicosocial, económico y laboral.

Palabras Clave: Reconocimiento de Voz, sistema auditivo humano, Corteza Auditiva, Corteza Asociativa, Sistemas Neurodifusos, Coeficientes de Predicción Lineal.

ABSTRACT

The present work describes the design and development of a multispeaker voice recognition system capable of recognizing the 5 vowels of the Spanish alphabet in an isolated way by a set of six people (3 males and 3 females).

The design of the system considers the morphology of the pinna and the general structure of the auditory pathway, with the purpose of granting it the capability of showing some of the qualitative characteristics of the human auditory system such as the localization of the sound source and the identification of the speaker (if it is a male or a female).

The system utilizes parameters such as the Linear Prediction Coefficients, the energy of the segments and the Fourier's Transform, which are interpreted and compared by Winner Takes All competence neural networks WTA.

The presented work aims to propose a methodology in the measurement and signal analysis to characterize the human response to sound from different positions in space, as well as permeate the bionic technique in performing voice recognition systems, adding to the robust algorithms search that can meet requirements in the studies related to speech recognition with a projection to set the speech recognition interface man-machine communication, the impact will be described in the psychosocial field, economic and labor.

Keywords: *Voice Recognition, Human Auditory System, Auditory Cortex, Association Cortex, Neurofuzzy Systems, Linear Prediction Coefficients.*

GLOSARIO

Protoplasma: sustancia viva de la célula, se subdivide en dos partes: el citoplasma y el carioplasma; el citoplasma se encuentra desde la membrana celular hasta el núcleo y es el lugar donde ocurre el metabolismo celular, y el carioplasma, el líquido intranuclear, es el sitio donde ocurre el metabolismo de los ácidos nucleicos.

Citoarquitectura de la corteza: es la organización de la corteza según los tejidos que poseen células nerviosas.

Código semiótico: conjunto de signos, estructurados en los que existe una relación entre el significante y el significado.

Matriz de Toeplitz: una matriz de Toeplitz, es una matriz cuadrada con todas sus diagonales de izquierda a derecha paralelas numéricamente.

CAPÍTULO 1. INTRODUCCIÓN

INTRODUCCIÓN

El hombre se ha denominado a sí mismo *Homo Sapiens* (hombre sabio) puesto que sus capacidades mentales tienen una gran importancia para él. Durante miles de años, ha tratado de entender el origen de su pensamiento; es decir, entender como un simple puñado de materia puede percibir, entender, predecir y manipular un mundo mucho más grande y complicado que ella misma (1). El campo de la Inteligencia Artificial (IA) va más allá. Por ahora, es suficientemente claro que el objetivo de la IA es el de entender la naturaleza de la inteligencia a través del diseño de sistemas que la exhiban. En forma más concreta, puede afirmarse que, en lo que ha transcurrido su corta historia, la IA ha estado dirigida por tres objetivos generales (2):

1. El análisis teórico de las posibles explicaciones del comportamiento inteligente
2. La explicación de habilidades mentales humanas
3. La construcción de entidades inteligentes

Con estos propósitos en su agenda de investigación, los estudiosos de la IA han recurrido al uso de cuatro diferentes estrategias metodológicas: el desarrollo de tecnologías útiles en esta área, la simulación, el modelado y la construcción de teoría sobre la inteligencia artificial.

Dentro de la IA, uno de los retos de imitar correctamente una de las capacidades del cerebro humano, es la del reconocimiento de voz. Para alcanzar metas importantes en esta área resulta necesario unificar conocimientos de diferentes disciplinas de la ciencia y la técnica (3).

En el desarrollo de este prototipo se abordan áreas de naturaleza multidisciplinaria como son:

- Neurofisiología
- Psicofisiología
- Informática
- Tratamiento de señales
- Sistemas neuronales y difusos

Así, en el contenido del presente trabajo se explican las etapas involucradas en el reconocimiento de vocales aisladas, fuente sonora de voz y diferenciación entre un hombre y una mujer teniendo como base biológica el sistema auditivo humano.

Al realizar el presente trabajo se propone un sistema que tenga una proyección en diferentes ámbitos, desde un ámbito tecnológico como lo es el área de desarrollo de reconocimiento de voz, hasta un ámbito social y económico.

Con la permanente revolución tecnológica, el desarrollo del presente sistema resulta una base fundamental para la realización de sistemas de mayor exigencia.

Planteamiento del Problema

El prototipo que aquí se presenta contempla el diseño de un sistema que reciba señales de audio (señales de voz digitalizada) a través de dos transductores electroacústicos ensamblados a un par de estructuras basadas en la morfología del pabellón auricular. Posteriormente ambas señales representarán la entrada de un sistema de adquisición que realizará un análisis preliminar para determinar la localización del sonido. Las señales son posteriormente enviadas por la entrada de audio a una computadora donde se extraerán los parámetros de la señal. A continuación se emplearán las técnicas de Cálculo Inteligente antes descritas para determinar la salida del sistema.

Justificación

Las investigaciones desarrolladas en el campo de reconocimiento de voz están en una constante búsqueda de algoritmos más robustos para identificar patrones en señales acústicas, lo cual permitirá cubrir exigencias en los estudios referidos al reconocimiento de voz hablado en lenguaje natural estableciendo sistemas de habla continua, independientes del locutor, con reconocimiento de un vocabulario amplio y una gramática abierta, con una proyección a establecer el reconocimiento de voz como interfaz de comunicación hombre-máquina, cuyo impacto estará descrito en el ámbito psicosocial, económico y laboral.

En este aspecto, el reconocimiento de voz ha tenido un gran avance al incorporar la biónica como solución, mediante el diseño de dispositivos, métodos y algoritmos inspirados en la fisiología humana, como en el caso del análisis de formantes debidos a los resonadores naturales en la cavidad supraglótica y otros analizadores espectrales, recordando la transformación de la energía mecánica de las ondas sonoras en energía nerviosa, llevada a cabo en los órganos de Corti de la cóclea.

La propuesta presentada tiene como objetivo, permear la técnica biónica en la realización de un sistema de procesamiento de voz basado en el funcionamiento del sistema auditivo humano, comparando a su vez la eficiencia de distintos métodos de identificación de patrones de voz, reduciendo el error en su respuesta, sumándose a la búsqueda de algoritmos más robustos en el ámbito del procesamiento de voz.

Sectores que han o pueden incluir el procesamiento de voz como herramienta (4):

-La primera industria que utilizó algún software de reconocimiento de voz, fue la industria de la salud con la visión de reemplazar a las transcripciones médicas tradicionales. Sin embargo, esta idea no fue muy exitosa, al no contar con la confianza de los médicos hacia una computadora en transcripciones críticas, las cuales son hechas a la perfección por los seres humanos.

No obstante, los avances en la informática, hicieron posible, el uso del reconocimiento de voz, en artículos como celulares, automóviles y computadoras personales. El reconocimiento de voz añade

un nivel de simplicidad para todos sus usuarios. Con las fuertes campañas de compañías como Microsoft y otros sistemas operativos móviles, la tecnología de reconocimiento de voz está siendo absorbida lentamente por el estilo de vida actual.

-Hogar: la automatización del hogar se ha desarrollado con la intención de aumentar sus niveles de comodidad y seguridad. El incorporar sistemas de reconocimiento de voz facilita su uso otorgándole mayor comodidad.

-Algunos dispositivos inteligentes de GPS, ya poseen la función de reconocimiento de voz, especialmente, los que se encuentran instalados en los vehículos, permitiéndole al conductor recibir instrucciones, pero al mismo tiempo, se puede concentrar en conducir.

-Educación: los estudiantes con discapacidad, que poseen un control limitado sobre las computadoras, se encuentran en una situación de desventaja. Pero con la tecnología de reconocimiento de voz tienen una herramienta eficaz para controlar el equipo y ser tan productivos como sus compañeros que no poseen ninguna discapacidad.

-Vida cotidiana: adultos que sufren algún tipo de discapacidad pueden manejar una computadora mediante la voz, permitiéndole tener una serie de herramientas a su disposición que pueden mejorar su calidad de vida.

-La industria de la atención de la salud es el principal consumidor de este tipo de tecnología, ya que los beneficios dirigidos a las personas con capacidades diferentes se centran en facilitar tareas cotidianas en las que constantemente necesitan ayuda extra, o en el peor de los casos dependen de otras personas, tareas como encender una luz, mover una cama clínica, apagar la televisión, etc.

Objetivos

Objetivo General

Diseñar un prototipo que, imitando al *sistema auditivo humano* por medio del Cálculo Inteligente (*Soft Computing*) utilizando algoritmos neurodifusos, detecte y analice parámetros en señales de audio portadoras de información consciente (señales de voz digitalizada), los identifique y emita una respuesta describiendo la localización de la fuente de voz, el locutor (género masculino o femenino) y la vocal reconocida.

Objetivos Particulares

- Estudiar y analizar el comportamiento del *sistema auditivo humano*; sus principios fisiológicos y psicológicos.
- Construir un sistema de adquisición de señales basado en la morfología del pabellón auricular humano y las limitaciones en frecuencia del oído.
- Implementar algoritmos neuronales y/o difusos en un sistema multilocutor para:
 - Reconocer vocales aisladas.
 - Localizar la fuente sonora en un plano.
 - Diferenciar una voz masculina de una femenina de entre un conjunto de 3 voces masculinas y 3 femeninas.
- Elaborar un sistema que reconozca las 5 vocales con un porcentaje de acierto del 80% y diferencie en un 90% las voces masculinas de las femeninas de su corpus de entrenamiento.

Organización del trabajo

El presente trabajo se divide en 5 capítulos. A continuación se comenta el contenido de los mismos:

Capítulo 1. Introducción. En este capítulo se presenta la descripción del proyecto. Este incluye el planteamiento del problema, la justificación, el objetivo general y los objetivos particulares.

Capítulo 2. Estado del arte. En éste se presentan los antecedentes y los trabajos en materia de reconocimiento de voz realizados en los últimos años.

Capítulo 3. Marco teórico. Se expone el sistema biológico que comprende la base del sistema biónico: la producción de voz, la localización del sonido, el sistema auditivo humano periférico y central. Así mismo, se expone el procesamiento matemático de la voz y su reconocimiento: transformada de Fourier, Coeficientes de Predicción Lineal, extracción de formantes y redes neuronales artificiales.

Capítulo 4. Desarrollo de la propuesta. En este capítulo se abordan las implicaciones del sistema biológico a su análogo de ingeniería. En esta implicación se lleva a cabo un diseño de cada etapa del reconocimiento, en la cual se detallan los dispositivos empleados, tanto en software como en hardware.

Capítulo 5. Resultados y discusión. Este capítulo hace referencia a los elementos construidos e implementados que constituyen al prototipo final. Para la validación del sistema se estableció que se reconocería en una medida del 90 % la diferenciación entre hombre y mujer, y en un 80% el reconocimiento de vocal. Se presentan las evaluaciones de los resultados obtenidos respecto a los objetivos.

CAPÍTULO 2. ESTADO DEL ARTE

ESTADO DEL ARTE

Antecedentes

A continuación se muestra una breve historia del reconocimiento de voz (5):

-Década de 1870 - Alexander Graham Bell: en el intento de construir un dispositivo que hiciera el habla visible a las personas con problemas auditivos dio origen al teléfono.

-Década de 1880 - Tihmir Nemes: solicita permiso para una patente para desarrollar un sistema de transcripción automática que identificara secuencias de sonidos y los imprimiera (texto). Pero fue rechazado como "Proyecto no Realista"

-30 años después - AT&T Bell Laboratories: construye la primera máquina capaz de reconocer voz (basada en Templates) de los 10 dígitos del inglés. Requería extenso reajuste a la voz de una persona, pero una vez logrado tenía un 99% de certeza. Por lo tanto surge la esperanza de que el reconocimiento de voz sea simple y directo.

-Década de 1960: la mayoría de los investigadores reconoció que era un proceso mucho más intrincado y sutil de lo que habían anticipado. Por lo tanto empiezan a reducir los alcances y se enfocan a sistemas más específicos:

- Dependientes del Locutor.
- Flujo discreto de habla (con espacios / pausas entre palabras)
- Vocabulario pequeño (menor o igual a 50 palabras)

Estos sistemas empiezan a incorporar técnicas de normalización del tiempo (minimizar diferencia en velocidad del habla). Además, ya no buscaban una alta exactitud en el reconocimiento.

IBM y CMV trabajan en reconocimiento de voz continuo pero no se ven resultados hasta la década de 1970.

Hacia los setenta la influencia de los trabajos en inteligencia artificial fue decisiva, centrando su interés en la representación del significado. Como resultado se construyó el primer sistema de preguntas-respuestas basado en lenguaje natural.

De esta época es Eliza, que reproducía las habilidades conversacionales de un psicólogo. Para ello recogía patrones de información de las respuestas del cliente y elaboraba preguntas que simulaban una entrevista.

-Década de 1970: se produce el 1er Producto de reconocimiento de voz, el VIP100 de Threshold Technology Inc. (utilizaba un vocabulario pequeño, dependiente del locutor, y reconocía palabras discretas). Gana el U.S. National Award en 1972.

Nace el interés de ARPA (Advanced Research Projects Agency) del departamento de defensa de EUA, y gracias al lanzamiento de grandes proyectos de investigación y financiamiento por parte del gobierno se precipita la época de la inteligencia artificial.

El proyecto financiado por ARPA busca el reconocimiento de habla continua, de vocabulario grande. Impulsa que los investigadores se enfoquen al entendimiento del habla. Los sistemas empiezan a incorporar módulos de:

- análisis léxico (conocimiento léxico)
- análisis sintáctico (estructura de palabras)
- análisis semántico (significado)
- análisis pragmático (intención)

Este proyecto termina en 1976. CMU, SRI, MIT crean sistemas para el proyecto ARPA SUR (Speech Understanding Research).

En Europa surgen intereses en la elaboración de programas para la traducción automática. Se crea el proyecto de investigación Eurotra, que tenía como finalidad la traducción multilingüe. En Japón aparecen equipos dedicados a la creación de productos de traducción para su distribución comercial.

-1980-1990: surgen los sistemas de vocabulario amplio, que ahora son la norma (más de 1000 palabras). Adicionalmente bajan los precios de estos sistemas.

-Actualidad: los últimos años se caracterizan por la incorporación de técnicas estadísticas y se desarrollan formalismos adecuados para el tratamiento de la información léxica. Se introducen nuevas técnicas de representación del conocimiento cercanas a la inteligencia artificial, y las técnicas de procesamiento utilizadas por investigadores procedentes del área de la lingüística e informática son cada vez más próximas. Surgen así mismo intereses en la aplicación de estos avances en sistemas de recuperación de información con el objetivo de mejorar los resultados en consultas a texto completo.

Las empresas más importantes en el desarrollo de productos con reconocimiento de voz son Philips, Lernout & Hauspie, Sensory Circuits, Dragon Systems, Speechworks, Vocalis, Dialogic, Novell, Microsoft, NEC, Siemens, Intel (apoyo / soporte técnico), entre otros.

Trabajos actuales

- Sistema artificial de audición baural basado en el sistema auditivo humano utilizable en el acondicionamiento de recintos acústicos (2005) (6).

El trabajo presenta el desarrollo de un sistema artificial de audición baural que genera una respuesta auditiva análoga a la obtenida por los oídos humanos, que es utilizada en la detección de fuentes sonoras. Esto se logra mediante técnicas de discriminación temporal, de frecuencia y de intensidad.

El sistema consiste en la elaboración de dos oídos artificiales, cada uno tiene un transductor eléctrico y un conjunto de cavidades mecánicas que tienen la función de simular la respuesta en frecuencia del oído humano a partir de sus impedancias acústicas, así como cabeza y torso artificiales.

- Diseño e implementación de un modelo de localización sonora espacial utilizando técnicas de inteligencia computacional (2006) (7).

El trabajo presenta el diseño, implementación y entrenamiento de un modelo de localización sonora espacial en campo libre para sonidos de banda ancha, inspirado en el sistema auditivo humano e implementado mediante el uso de técnicas de inteligencia computacional.

- Análisis, reconocimiento y síntesis de voz esofágica (8).

El trabajo presenta el diseño, simulación e implementación de un sistema de análisis, reconocimiento y síntesis de voz esofágica. Un algoritmo de extracción de características de la voz utilizando una función wavelet basada en un modelo del oído humano es realizado. Dos sistemas de reconocimiento de voz utilizando redes neuronales y modelos ocultos de Markov fueron diseñados y validados, así como un algoritmo de síntesis de voz esofágica basado en coeficientes de predicción lineal (LPC) y formantes.

- Técnicas para el reconocimiento de voz en palabras aisladas en la lengua náhuatl (9).

El trabajo muestra los resultados obtenidos en la aplicación de técnicas de reconocimiento de voz en la lengua náhuatl. Los principales parámetros que se analizan son los Coeficientes de Predicción Lineal (LPC) y los coeficientes cepstrales en la escala de Mel.

CAPÍTULO 3. MARCO TEÓRICO

MARCO TEÓRICO

La voz humana

La voz humana, una de las formas más complejas de expresión del ser humano, es también un importante instrumento fundamental para la comunicación y la actividad cognoscitiva. Tiene su origen en el sistema fonador, compuesto por el sistema respiratorio, la laringe, las cuerdas vocales y la cavidad bucal. Para llevar a cabo la articulación de palabras, es necesario automatizar y sincronizar una gran cantidad de elementos cuyo logro es parte de una milenaria evolución (8).

Aspecto sonoro del habla

El habla, como manifestación sonora del lenguaje, se desarrolla con el empleo de diversos órganos y funciones anatómicas. La producción y emisión de los sonidos verbales se deben a la acción o funcionamiento secuenciado, sincronizado y automático de los siguientes elementos:

- Una corriente de aire, que es producida por los pulmones y músculos respiratorios.
- Un vibrador sonoro, constituido por las cuerdas vocales de la laringe.
- Un resonador, conformado por la boca, nariz y faringe.
- Articuladores, conformados por los labios, dientes, paladar duro, velo del paladar, mandíbula.

El órgano principal de la producción de la voz es la laringe, un conducto cuya función primordial es la protección de las vías respiratorias de la introducción de cuerpos extraños, que a su vez contiene a las cuerdas vocales, dos pares de pliegues que sirven de acceso a la tráquea. El par más inferior es el que, al ser accionado con la columna de aire proveniente de los pulmones, vibra para producir la voz (9).

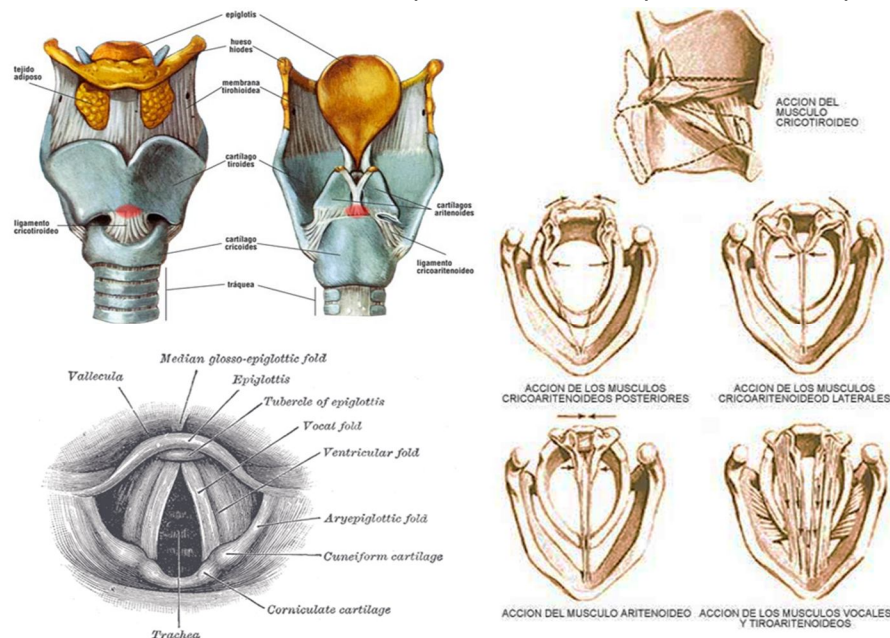


Figura 3. 1 Ubicación de las cuerdas vocales en la laringe y la acción de sus principales músculos. Copyright © (18)

Los pulmones suministran la columna de aire que, atravesando los bronquios y la tráquea, sonorizan las tensadas cuerdas vocales que se encuentran en la laringe. Este es el momento donde se produce la voz. Sin embargo, ésta es modificada en la cavidad supraglótica por la nariz, boca y garganta (faringe) otorgando el timbre a la voz, figura 3.2.

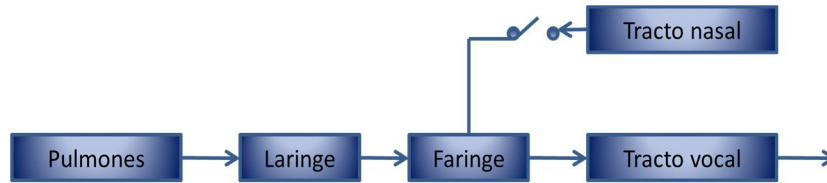


Figura 3. 2 Esquema de la producción de voz.

Las cavidades que conforman la cavidad supraglótica actúan como resonadores acústicos. Si se realiza un análisis espectral del sonido luego de haber atravesado estas cavidades, el efecto de la resonancia produciría un énfasis en determinadas frecuencias del espectro obtenido, a las que se les denomina *formantes* (figura 3.3). Existen tantas formantes como resonadores posee el tracto vocal. Sin embargo se considera que sólo las tres primeras, asociadas a la cavidad oral, bucal y nasal respectivamente, proporcionan la suficiente información para diferenciar los distintos tipos de sonido (3).

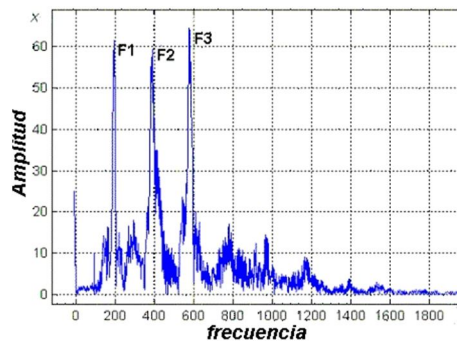


Figura 3. 3 Espectro de la palabra uno

Se aprecian las formantes F1, F2 y F3. Las frecuencias amplificadas dependen del tamaño y la forma que adopta la cavidad oral, y si el aire pasa o no por la nariz.

Finalmente, los labios, lengua, dientes, paladar duro, velo del paladar y mandíbula, modulan las características de dichos resonadores dando origen a los sonidos y articulaciones del habla; es decir fonemas, sílabas y palabras, alternando distintos tipos de sonidos que pueden ser *sonoros*, *sordos* y *plosivos* (figura 3.4).

Los sonidos *sonoros* se generan por la vibración de las cuerdas vocales manteniendo la glotis abierta, lo que permite que el aire fluya a través de ella. Estas señales se caracterizan por tener alta energía y un contenido de frecuencia en el rango de los 300 Hz a 4000 Hz presentando cierta

periodicidad, es decir son de naturaleza cuasiperiódica. El tracto vocal actúa como una cavidad resonante reforzando la energía en torno a determinadas frecuencias (formantes).

Los sonidos *sordos* se caracterizan por tener un comportamiento aleatorio en forma de ruido blanco. Tienen una alta densidad de cruces por cero y baja energía comparadas con las señales sonoras. Durante su producción no se genera vibración de las cuerdas vocales, ya que el aire atraviesa un estrechamiento y genera una turbulencia. Las consonantes que producen este tipo sonidos son la *s*, la *f* y la *z*, entre otras.

Los sonidos *plosivos* se generan cuando el tracto vocal se cierra en algún punto, lo que causa que el aire se acumule para después salir expulsado repentinamente (explosión). Se caracterizan por que la expulsión de aire está precedida de un silencio. La *p*, la *t* y la *k* son ejemplos de consonantes de carácter plosivo.

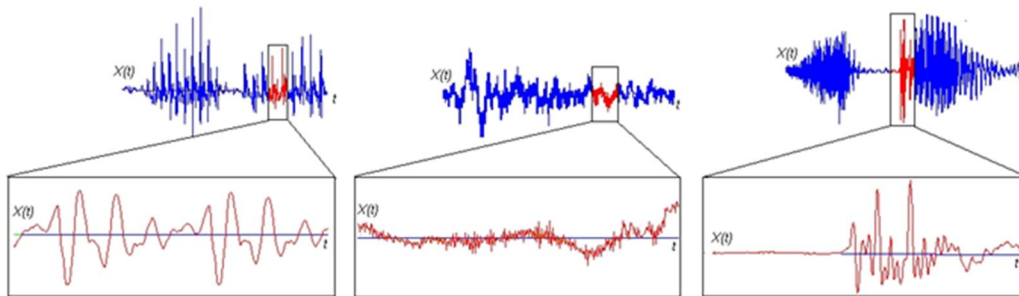


Figura 3. 4 Comparación entre las señales de sonidos *sonoro*, *sordo* y *plosivo* respectivamente.

En la producción sonora, el oído desempeña un papel importante como regulador en el funcionamiento coordinado de los resonadores bucal y faríngeo. La pérdida parcial o total de la audición altera dicho funcionamiento. El tono nasal o total de la audición altera dicho funcionamiento. El tono nasal del lenguaje de los sordos se debe en parte a la falta de control auditivo en la regulación de los movimientos de la lengua y del analizador faríngeo (8).

De lo anterior es posible considerar al tracto vocal como un filtro, cuyos parámetros varían en el tiempo en función de la acción consciente de hablar. Para el mismo se consideran dos entradas que dependerán del tipo de señal a reproducir, *sonora* o *no sonora*. Para señales *sonoras*, la excitación será un tren de impulsos de frecuencia controlada, mientras que para las señales no sonoras la excitación será ruido aleatorio. La combinación de estas señales modela el funcionamiento de la glotis. El espectro de frecuencias de la señal de voz puede obtenerse a partir del producto del espectro de la excitación por la respuesta en frecuencia del filtro (*función de transferencia*).

La figura 3.5 representa un modelo del sistema de producción de voz. El conducto vocal se representa por un sistema lineal que es excitado a través de un interruptor o llave que selecciona entre una fuente de impulsos cuasiperiódicos para el caso de los sonidos *sonoros* (*tonales*), o una fuente de ruido aleatorio para el caso de los sonidos *no sonoros* (*no tonales*). El control de

ganancia G , determina la intensidad de la excitación; es estimada a partir de la señal de voz y la señal escalada es usada como entrada del modelo del conducto (tracto) vocal (12) (11).

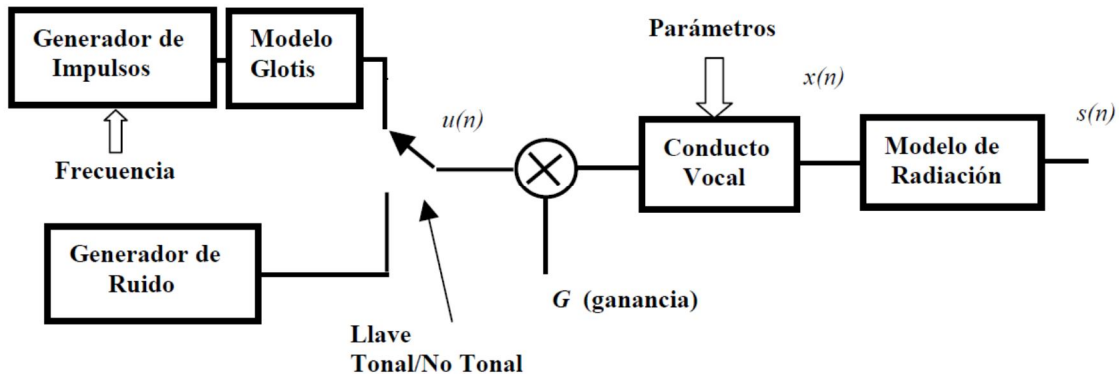


Figura 3. 5 Modelo de predicción lineal para la producción de voz. Copyright © (12)

Los parámetros glotales son los responsables de la forma de la onda de la señal del generador de impulsos (frecuencia fundamental). Éstos son propios de cada persona y dependen de la edad y el sexo. Es posible aproximarlos mediante los modelos polinomiales y la Transformación LF (12).

Procesamiento matemático de la señal de voz

Es necesario saber que la señal de voz se analiza por medio de características espectrales, debido que (13):

- La señal de voz es considerada cuasiestacionaria en periodos cortos de tiempo y puede ser aproximada sumando ondas sinusoidales.
- Las características críticas de la percepción de voz en el oído humano incluyen primordialmente información espectral.

Transformada de Fourier

Es una herramienta muy importante para el análisis de las señales de voz en el dominio de la frecuencia, ya que permite obtener información de las mismas que no son evidentes en el dominio del tiempo. Una de las características de la transformada de Fourier es que permite extraer el contenido de frecuencias de la señal (figura 3.3), lo cual es importante para establecer el muestreo de la misma, el cual debe ser por lo menos dos veces mayor que la frecuencia máxima de la señal (14).

La transformada de Fourier de una señal discreta expresa la señal como una suma infinita de sinusoides, su cálculo se puede efectuar mediante el algoritmo FFT (abreviatura del inglés Fast Fourier Transform) o bien por la Transformada Discreta de Fourier (ecuación 3.1).

$$F(\omega) = \frac{1}{2\pi N} \left| \sum_{n=0}^{N-1} x(n)e^{-j\omega n} \right|$$

Coeficientes de Predicción Lineal

El método de predicción lineal toma como base el modelo del tracto vocal representado como un filtro lineal variable en el tiempo. Según este modelo se distinguen dos elementos separados en la producción de voz: la excitación y el tracto vocal. La onda de voz es el resultado de la convolución entre la excitación y el filtro (tracto vocal). Existen diversos métodos para lograr la separación fuente-filtro, uno de ellos es la predicción lineal (15).

La predicción lineal, caracteriza la forma del espectro de un segmento de voz con un número reducido de parámetros que permiten una codificación eficiente. La Codificación Lineal Predictiva, como también se llama a este método, predice una señal en el dominio del tiempo con base en una combinación de muestras previas, linealmente distribuidas, como se muestra en la ecuación 3.2:

$$\hat{S}(n) = - \sum_{k=1}^p a_k S(n-k)$$

3. 2

Donde $a_k(1 < k < p)$ es un conjunto de constantes reales conocidos como coeficientes de predicción que necesitan ser calculados y p es el orden de predicción.

Ahora, se define el error de predicción en la ecuación 3.3:

$$e(n) = S(n) - \hat{S}(n) = S(n) + \sum_{k=1}^p a_k S(n-k)$$

3. 3

Para obtener las ecuaciones que deben ser resueltas en la determinación de los coeficientes LPC, definimos los segmentos de voz y de error en un tiempo dado n y buscamos minimizar la señal de error mínimo cuadrado, mostrado en la ecuación 3.4:

$$E_n = \sum_m e_n^2(m) = \sum_m \left[S_n(m) + \sum_{k=1}^p a_k S_n(m-k) \right]^2$$

3. 4

La función de autocorrelación resulta una herramienta de gran utilidad para encontrar patrones repetitivos dentro de una señal. La cual está definida por la ecuación 3.5:

$$C_{ss}[n] = \frac{1}{N} \sum_{m=0}^{N-1-|p|} S[m]S[m+|p|]$$

3. 5

Este método sugiere que una manera simple y eficiente de definir los límites de las sumatorias es asumir que los segmentos de voz son nulos fuera del intervalo, lo cual equivale a multiplicar la señal de voz por una ventana de largo finito. Por lo tanto, el error mínimo cuadrado queda descrito por la ecuación 3.6:

$$E_n = \sum_{m=0}^{N-1-p} e_n^2(m)$$

3.6

El problema de la predicción lineal radica en encontrar los coeficientes predictores a_k que minimicen el error e_n . La condición para la minimización del error total cuadrático se obtiene estableciendo la derivada parcial del error total cuadrático E con respecto a cada uno de los coeficientes predictores (ecuación 3.7). Desglosándolo queda como se muestra en la ecuación 3.8.

$$\frac{dE_T^2}{da_k} = 0, \text{ para } k = 1, 2, \dots, p$$

3.7

$$\therefore \sum_{k=1}^p a_k \sum_{n=0}^{N-1} s(n-k)s(n-i) = -\sum_{n=0}^{N-1} s(n)s(n-i), \text{ para } i = 1, 2, \dots, p$$

3.8

Si ponemos los términos de sumatoria de multiplicaciones de muestras en funciones de autocorrelación, tenemos lo siguiente (ecuación 3.9):

$$\sum_{k=1}^p a_k C_{ss}(i-k) = -C_{ss}(i), \text{ para } i = 1, 2, \dots, p$$

3.9

Como la función de autocorrelación es simétrica $C_n(-p) = C_n(p)$ las ecuaciones que deben cumplir los parámetros LPC se pueden expresar como (ecuación 3.10):

$$\begin{bmatrix} C_{ss}(0) & C_{ss}(-1) & \dots & C_{ss}(-(p-1)) \\ C_{ss}(1) & C_{ss}(0) & \dots & C_{ss}(-(p-2)) \\ \vdots & \vdots & \dots & \vdots \\ C_{ss}(p-1) & C_{ss}(p-2) & \dots & C_{ss}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} C_{ss}(1) \\ C_{ss}(2) \\ \vdots \\ C_{ss}(p) \end{bmatrix} \quad 3.10$$

Esta matriz de $p \times p$ valores de autocorrelación es una matriz tipo Toeplitz, la cual se puede resolver mediante el método de Levinson-Durbin. Su costo computacional es de n^2 , una mejora considerable frente a la eliminación de Gauss-Jordan, cuyo costo es de n^3 .

Algoritmo de Levinson-Durbin

Para aplicar este algoritmo, es necesario agregar un orden a la matriz cuadrada, posibilitando con esto la recursividad del cálculo de los resultados. De la siguiente manera (ecuación 3.11):

$$E_T^2 = C_{ss}(0)a_0 + \sum_{k=1}^p a_k C_{ss}(-k) \tag{3.11}$$

Escrita de forma matricial (ecuación 3.12):

$$\begin{bmatrix} C_{ss}(0) & C_{ss}(-1) & \cdots & C_{ss}(-p) \\ C_{ss}(1) & C_{ss}(0) & \cdots & C_{ss}(-(p-1)) \\ \vdots & \vdots & \cdots & \vdots \\ C_{ss}(p) & C_{ss}(p-1) & \cdots & C_{ss}(0) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} E_T^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.12}$$

De esta manera se pueden proponer las ecuaciones 3.13, 3.14, 3.15, 3.16 y 3.17 que de manera iterativa resuelven este sistema de ecuaciones:

$$E^{(0)} = R(0) \tag{3.13}$$

$$K_i = \frac{R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j)}{E^{(i-1)}}$$

Con $1 \leq i \leq p$ 3.14

$$\alpha_i^{(i)} = K_i \tag{3.15}$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - K_i \alpha_{i-j}^{(i-1)}$$

Con $1 \leq j < i$ 3.16

$$E^{(i)} = (1 - K_i^2)E^{(i-1)} \tag{3.17}$$

Una característica de los coeficientes LPC, es que son compatibles con la respuesta de un filtro *todo polos* y que al aplicar la respuesta del filtro al dominio de la frecuencia, se acomoda exactamente a la envolvente de la respuesta de frecuencia del segmento de señal de voz bajo

análisis. A más coeficientes a_p , más empieza a aproximarse la respuesta espectral del filtro a la respuesta de $f/2$ coeficientes espectrales de la señal. A continuación se muestra la respuesta de este filtro para diferentes valores de a_p .

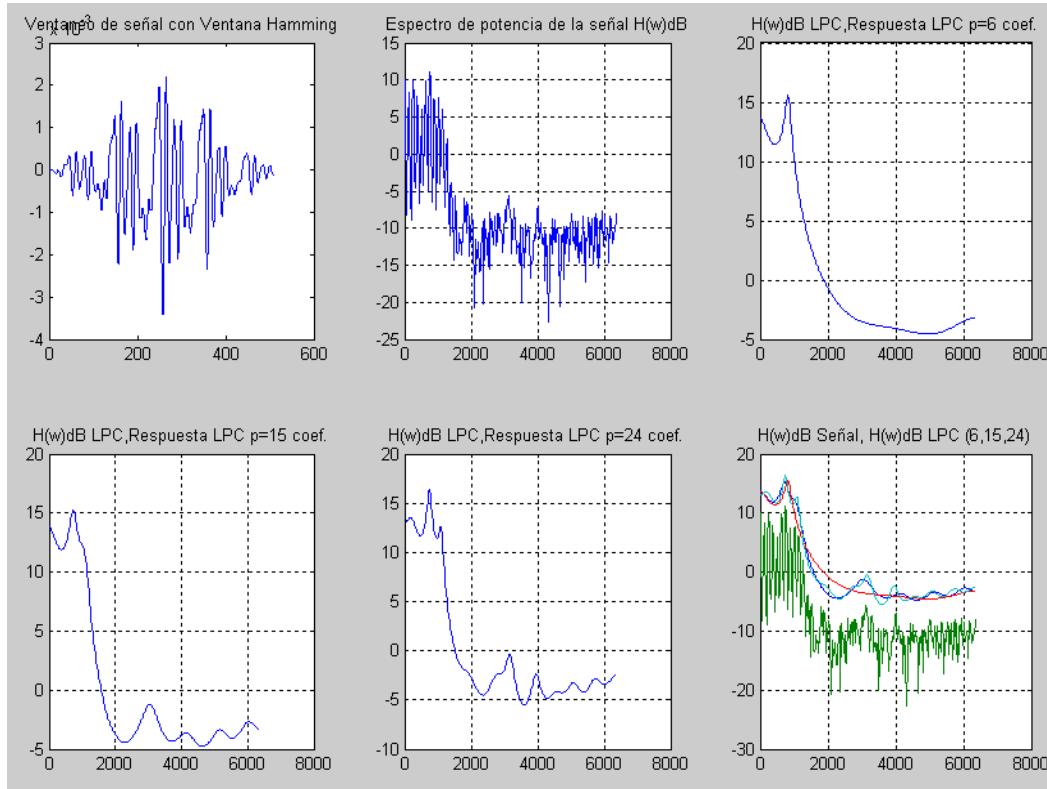


Figura 3. 6 Respuestas de frecuencias $H(w) | \text{dB}$ del filtro todo polos con $p=6, 15, 24$. Copyright © (16)

Extracción de formantes

Los formantes son los picos de la envolvente del espectro de la señal de voz que representan las frecuencias de resonancia del tracto vocal (figura 3.3). Las frecuencias a las que se producen los primeros formantes son muy utilizadas en los sistemas de caracterización de la voz. Las frecuencias formantes pueden variar de una persona a otra por edad, sexo y otros factores pero en sentido general se encuentran dentro de un rango establecido. La tabla 3.1 muestra las frecuencias de los formantes 1 y 2 de las vocales del idioma español pronunciadas por hablantes esofágicos del sexo masculino (15).

	FORMANTE 1 (Hz)	FORMANTE 2 (Hz)
A	550-900	1000-1700
E	350-500	1700-2500
I	140-299	1700-2500
O	350-500	701-1300
U	351-414	400-700

Tabla 3. 1 formantes 1 y 2 de voces esofágicas generadas por hombres.

Existen varios métodos para el cálculo de formantes de la señal de voz, uno de los más utilizados es a partir del filtro de los LPC explicado en el epígrafe anterior. Si el espectro de una señal de voz puede ser aproximado únicamente por sus polos, entonces los formantes pueden ser obtenidos de los polos del filtro LPC de la forma (ecuación 3.13 y ecuación 3.14):

$$H(z) = \frac{1}{A(z)}$$

3. 18

$$A(z) = 1 - \sum_{k=1}^p a_k z^{-k}$$

3. 19

Donde $H(z)$ es la función de transferencia de un filtro que modela el tracto vocal.

Los polos de $H(z)$ pueden ser calculados igualando el denominador de la ecuación 3.13 a "0" y encontrando las raíces. La conversión al plano S se realiza sustituyendo z por $e^{s_k T}$, donde s_k es el polo en el plano s . Las raíces resultantes en su mayoría son pares conjugados. Por lo tanto, las formantes pueden obtenerse de los picos del espectro LPC (15).

La audición

La audición humana constituye los procesos psicofisiológicos que proporcionan al hombre la capacidad de oír. De esta manera, el sonido es una experiencia psicológica creada por el encéfalo en respuesta a la estimulación del sistema auditivo por cambios en la presión del medio (17).

Sistema auditivo humano

El sistema auditivo humano es el conjunto de órganos que hacen posible el sentido del oído en el hombre. El oído humano es extraordinariamente sensible al sonido. En el umbral auditivo, las moléculas de aire son desplazadas un promedio de sólo 10 μm (10^{-11} m). La intensidad de este sonido es aproximadamente un billonésimo de watt por metro cuadrado. Esto significa que un oyente en un planeta sin ruido podría oír una fuente sonora de 1 watt a 3 kHz localizada a más de 450 km de distancia. Incluso los niveles de presión peligrosamente altos (> 100 dB) ejercen una potencia en el tímpano que se encuentra alrededor del miliwatt.

El ser humano puede detectar sonidos en un intervalo de frecuencias de aproximadamente 20 Hz a 20 kHz. Los niños incluso pueden oír frecuencias ligeramente arriba de los 20 kHz, pero pierden algo de la sensibilidad de alta frecuencia a medida que crecen. El límite superior en el adulto promedio es más cercano a 15-17 kHz.

Para su estudio, el sistema auditivo se divide en el sistema auditivo periférico y el sistema auditivo central (18).

Sistema auditivo periférico

El sistema auditivo periférico es el responsable de los procesos fisiológicos de la audición, mismos que permiten detectar el sonido y transformarlo en impulsos eléctricos enviados al cerebro a través de los nervios auditivos. Está constituido por el oído, órgano principal de la audición y del equilibrio. En conjunto, el estudio histoanatómico del oído se divide en tres partes: oído externo, oído medio y oído interno.

Oído Externo

El oído externo está comprendido por el pabellón auricular y el conducto auditivo externo. Su función es la de recoger la energía sonora y dirigirla hacia la caja timpánica. Por sus características, el pabellón tiene una frecuencia de resonancia entre los 4500 Hz y los 5000 Hz, en tanto que la configuración del conducto auditivo externo refuerza selectivamente la presión sonora de 30 a 100 veces para frecuencias de alrededor de 3 kHz. Las características presentadas en el pabellón auricular permiten filtrar las diferentes frecuencias de sonido para brindar señales acerca de la elevación de la fuente sonora. Las circunvoluciones del pabellón presentan su forma característica para que el oído externo transmita mayores frecuencias a mayores alturas del nivel del oído.

Oído Medio

Ubicado en un espacio aéreo dentro del hueso temporal, presenta una parte anterior, representada por la tuba faríngea (trompa de Eustaquio), una parte media o caja del tímpano y una parte posterior, representada por las celdas mastoideas.

La tuba faríngea es un conducto que comunica con la rinofaringe. Las celdas mastoideas representan cavidades dentro del hueso temporal que extienden la caja timpánica hacia posterior.

La caja del tímpano, cerrada por las paredes del hueso temporal y la cara externa del oído interno, contiene a los osteocillos óticos (martillo, yunque y estribo) articulados. El martillo se inserta por su manubrio al tímpano y se articula por su cabeza al yunque, éste se articula con el estribo que a su vez se contacta con la ventana oval del laberinto (oído interno).

La función de este arreglo óseo es transmitir y amplificar las ondas sonoras hacia el oído interno.

Dado que el oído interno está lleno de material linfático, mientras que el oído medio está lleno de aire, debe resolverse un desajuste de impedancias que se produce siempre que una onda pasa de un medio gaseoso a uno líquido. En el pasaje del aire al agua en general sólo el 0.1% de la energía de la onda penetra en el agua, mientras que el 99.9% de la misma es reflejada. En el caso del oído ello significaría una pérdida de transmisión de unos 30 dB.

El oído medio resuelve este desajuste de impedancias por dos vías complementarias. En primer lugar la disminución de la superficie en la que se concentra el movimiento. El tímpano tiene un área promedio de 69 mm^2 , pero el área vibrante efectiva es de unos 43 mm^2 . El pie del estribo, que empuja la ventana oval poniendo en movimiento el material linfático contenido en el oído interno, tiene un área de 3.2 mm^2 . La presión se incrementa en consecuencia en unas 13.5 veces. Por otra parte el martillo y el yunque funcionan como un mecanismo de palanca y la relación entre

ambos brazos de la palanca es de 1.31:1. La ganancia mecánica de este mecanismo de palanca es entonces de 1.3, lo que hace que el incremento total de la presión sea de unas 17.4 veces. El valor definitivo va a depender del área real de vibración del tímpano. Además, los valores pueden ser superiores para frecuencias entre los 2000 Hz y los 5000 Hz, debido a la resonancia del canal auditivo externo y a las frecuencias de resonancia características de los conos asimétricos, como lo es el tímpano. En general entre el oído externo y el tímpano se produce una amplificación de entre 5 dB y 10 dB en las frecuencias comprendidas entre los 2000 Hz y los 5000 Hz, lo que contribuye de manera fundamental para la zona de frecuencias a la que nuestro sistema auditivo es más sensible: La señal de voz contiene frecuencias en el rango de 80 hasta aproximadamente 8000 Hz, estando el contenido de la información de la articulación de voz en el rango de 300 a 4500 Hz aproximadamente.

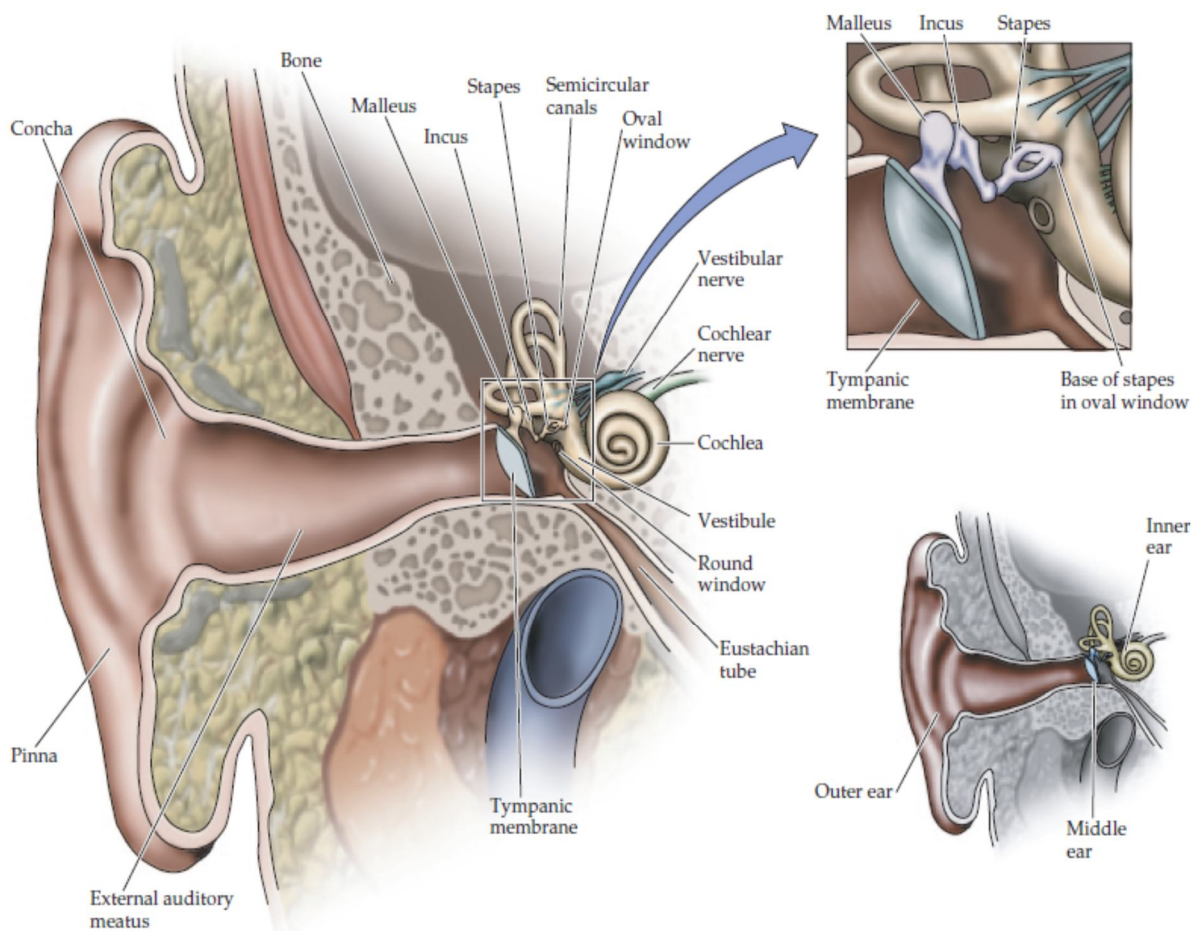


Figura 3. 7 Oído humano Copyright © (18)

Obsérvese la superficie de la membrana timpánica grande con relación a la ventana oval.

Oído Interno

Denominado también como laberinto, es una estructura osteocartilaginosa formada por tres secciones: la cóclea, el vestíbulo y los conductos semicirculares.

Para el caso de la audición, las secciones más importantes son el vestíbulo y la cóclea.

El vestíbulo recibe a través de la ventana oval las vibraciones transmitidas desde el oído medio y las envía a lo largo de la rampa vestibular de la cóclea. Para recibirlas de nuevo en circuito cerrado en su ventana redonda.

La cóclea, del latín *cochlea*, "caracol", es una estructura espiralada de unos 10 mm de ancho, que desenrollada forma un tubo de unos 35 mm de longitud. Está dividida en dos partes por el tabique coclear, una estructura flexible que sostiene la membrana basilar y la membrana tectorial, que a su vez seccionan a la cóclea en tres cámaras: la rampa vestibular, rampa media y rampa timpánica. En el extremo apical de la cóclea, existe un orificio conocido como helicotrema, que sirve de comunicación entre la rampa vestibular y la rampa timpánica, en consecuencia una breve aplicación de presión hacia adentro de la ventana oval hace que la ventana redonda protruya hacia afuera y deforme la membrana basilar.

La forma en que la membrana basilar vibra es una forma para comprender la función coclear. Mediciones de la vibración de diferentes partes de la membrana basilar, así como de las frecuencias de descarga de las fibras individuales del nervio auditivo muestran que ambas características están sintonizadas, es decir, son mayores en respuesta a un sonido de una frecuencia específica. Esta sintonización en la membrana basilar es atribuible a su geometría, que es más ancha y flexible en su extremo apical con respecto a su extremo basal. Georg von Békésy, biofísico Húngaro, mostró que una membrana de ancho y flexibilidad variados genera vibraciones en diferentes posiciones en respuesta a diferentes frecuencias, de tal manera que frecuencias altas generan vibraciones en la base de la cóclea en tanto que frecuencias bajas hacia el vértice. Ver figura 3.8.

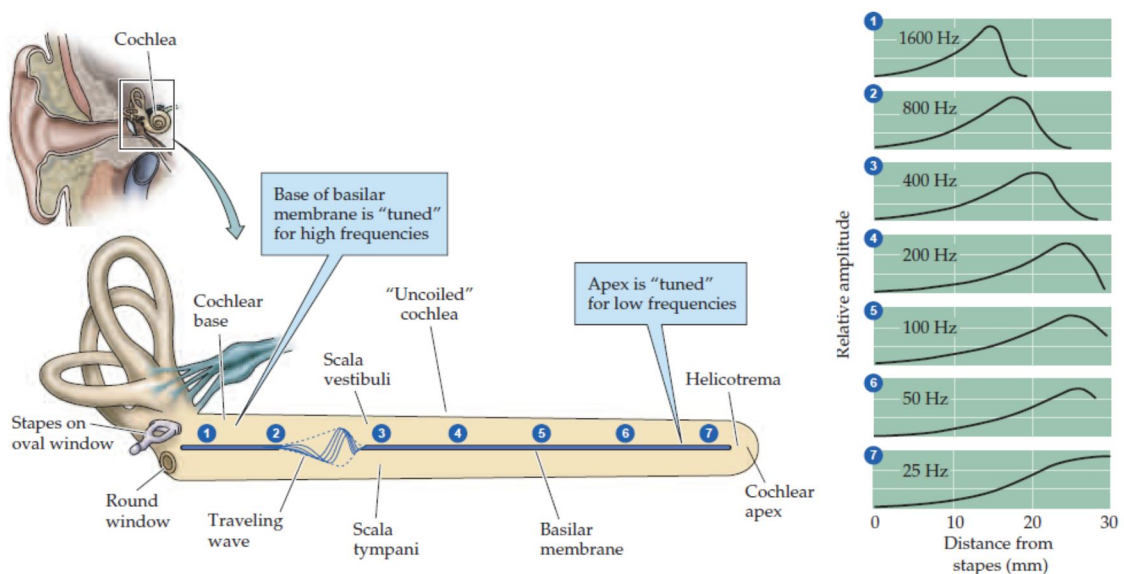


Figura 3. 8 Ondas de propagación a lo largo de la cóclea. Copyright © (18)

A este tipo de distribución espacial de las características en frecuencia se le conoce como tonotopía, y ésta es una cualidad que caracteriza a todo el sistema auditivo a partir de la cóclea.

El modelo de Békésy tiene dos inconvenientes. Primero, una mecánica pasiva como la que propone compromete a las áreas adyacentes al punto excitado. En realidad está claro que la sintonización auditiva es demasiado aguda como para ser explicada únicamente de esta forma. Segundo, con intensidades de sonido muy bajas, la membrana basilar vibra mucho más de lo que se podría predecir por la extrapolación lineal del movimiento medido con altas intensidades. Por lo tanto esto sugiere la existencia de un proceso activo aunado a la mecánica pasiva de Békésy, liderado por las células ciliadas externas.

La vía auditiva

El tiempo rápido de respuesta del aparato de transducción permite convertir ondas mecánicas en impulsos nerviosos en tan solo 10 μ s, sin embargo, los potenciales de receptor de ciertas células ciliadas y los potenciales de acción de las fibras asociadas del nervio auditivo no pueden seguir frecuencias por arriba de 3 kHz (ver figura 3.9). Sin embargo, la tonotopía de la membrana basilar, y de las fibras individuales del nervio auditivo explica la amplitud del intervalo de frecuencias audibles por el ser humano. Cada fibra individual del nervio auditivo responde mejor a una determinada frecuencia, llamada frecuencia característica. Como el orden topográfico de la frecuencia característica de las neuronas es retenido en todo el sistema, también se conserva la información acerca de la frecuencia.

Puesto que las células ciliadas sólo liberan el transmisor cuando están despolarizadas, las fibras del nervio auditivo disparan sólo durante las fases positivas de los sonidos de baja frecuencia.

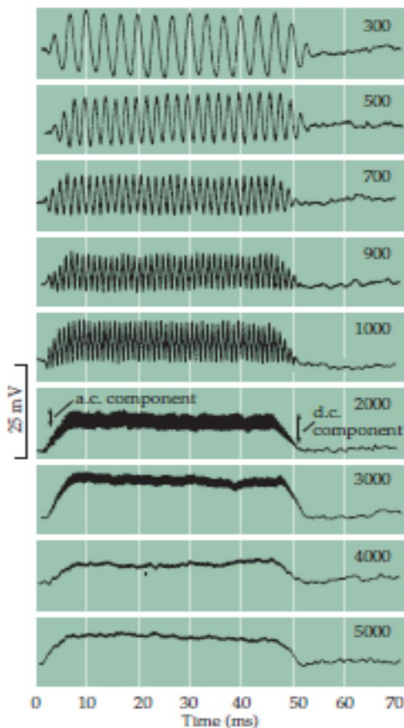


Figura 3. 9 Potenciales de receptor de una célula ciliada. Copyright © (18)

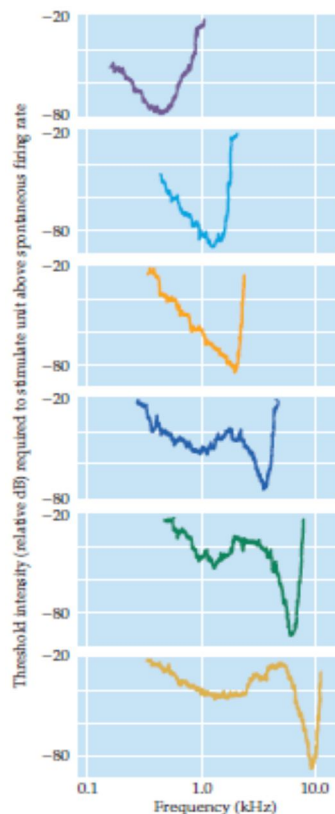


Figura 3. 10 Curvas de sintonización de frecuencias de seis fibras del nervio auditivo.

Copyright © (18)

Cada gráfico exhibe el nivel mínimo de sonido necesario para aumentar la frecuencia de descarga de la fibra por arriba de su nivel de descarga espontáneo. El punto más bajo representa la frecuencia característica.

Un sello del sistema auditivo ascendente es su organización en paralelo. Esta disposición se torna evidente tan pronto como el nervio auditivo ingresa al tronco encefálico, donde se ramifica para inervar las tres ramas del núcleo coclear. En el interior del núcleo coclear, cada fibra del nervio auditivo se ramifica, enviando una rama ascendente hasta el núcleo coclear anteroventral y una rama descendente hasta el núcleo coclear posteroventral y el núcleo dorsal. En todos ellos se mantiene la organización tonotópica, y se procesa la información mediante distintas poblaciones de células con propiedades muy diferentes, sin embargo aún no determinadas.

Tras los procesamientos llevados a cabo en los núcleos cocleares, éstos dan origen a varias vías ascendentes paralelas que se describen más fácilmente por la función que desempeñan. Una de ellas son las aferencias biaurales dirigidas hacia la oliva superomedial. La oliva superomedial contiene células con dendritas bipolares que se extienden tanto en sentido medial como lateral para propiciar la integración de la información proveniente de ambos oídos. En este punto se lleva a cabo un procesamiento que permite la localización de la fuente sonora mediante mecanismos coincidentes.

Un segundo conjunto importante de vías desde el núcleo coclear evita la oliva superior y termina en los núcleos del lemnisco lateral en el lado contralateral del tronco encefálico (vías monoaurales). Algunas células en los núcleos del lemnisco lateral señalan el inicio del sonido, cualquiera que sea su intensidad o frecuencia. Otras células en los núcleos del lemnisco lateral procesan otros aspectos temporales del sonido, como la duración. Aún no se conoce el papel preciso de esas vías en el procesamiento de las características temporales del sonido. Como sucede con las aferencias de los núcleos olivares superiores, las vías desde los núcleos del lemnisco lateral convergen en el mesencéfalo.

Las vías auditivas que ascienden a través de los complejos olivar y lemniscal, así como otras proyecciones que surgen directamente del núcleo coclear, se proyectan al centro auditivo mesencefálico, el colículo inferior. Al examinar estos núcleos en animales como la lechuza, con mayor capacidad para localizar sonidos, se cree que el ser humano podría tener igualmente una representación topográfica del espacio auditivo en este núcleo, dada por neuronas que responden mejor a los sonidos originados en una región específica del espacio.

Otra propiedad importante del colículo inferior es su capacidad para procesar los sonidos con patrones temporales complejos. Muchas neuronas en el colículo inferior responden sólo a sonidos modulados por frecuencias, mientras que otras responden solo a sonidos de duraciones específicas. Estos son parámetros de los sonidos biológicamente relevantes, como los efectuados por predadores o los sonidos de comunicaciones intraespecífica, que en los seres humanos incluyen el habla. El colículo inferior es, evidentemente, la primera etapa en este sistema, continuada en el tálamo auditivo y la corteza, que analiza los sonidos que tienen un significado especial.

A pesar de las vías paralelas en las estaciones auditivas del tronco encefálico y el mesencéfalo, el complejo geniculado medial en el tálamo es una estación obligatoria de toda la información auditiva ascendente destinada a la corteza. La mayoría de las aferencias hacia el complejo

geniculado medial surgen del colículo inferior, aunque algunas fibras auditivas desde el tronco encefálico inferior evitan el colículo inferior para alcanzar directamente el tálamo auditivo. El complejo geniculado medial tiene varias divisiones, que incluyen la división ventral, la cual funciona como principal estación talamocortical, y las divisiones dorsal y medial, que están organizadas como un cinturón alrededor de la división ventral.

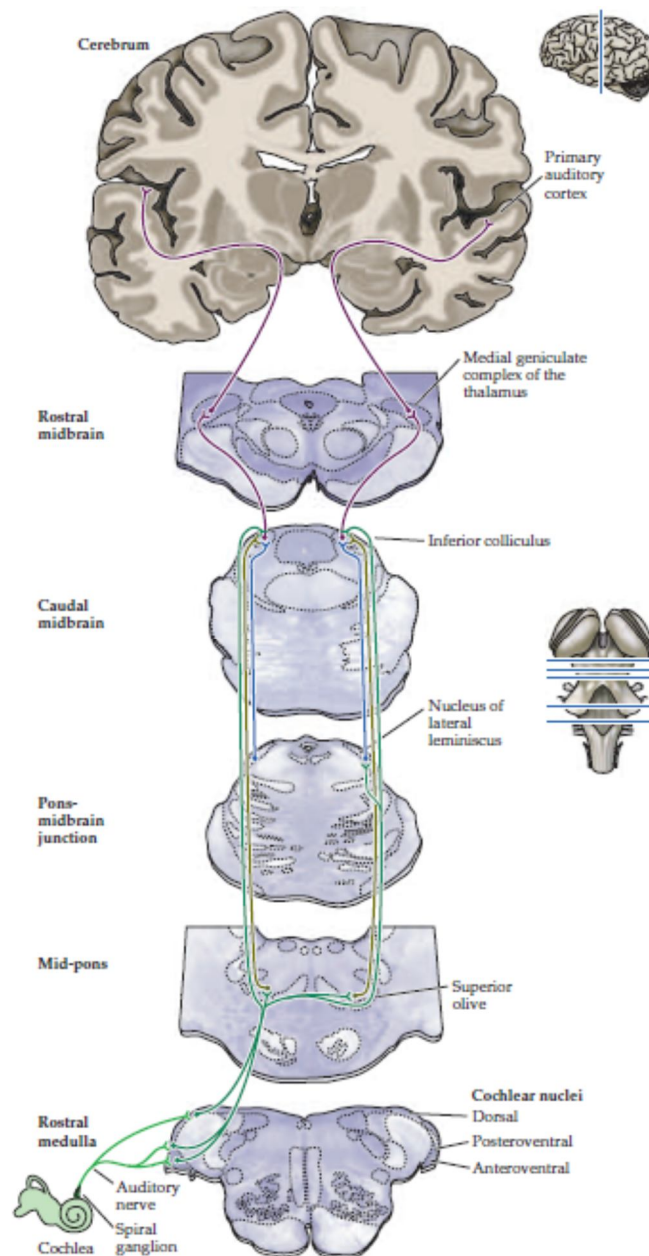


Figura 3. 11 Las vías auditivas más importantes. Copyright © (18)

Se aprecia la organización en paralelo del sistema auditivo y la información proveniente de cada oído que alcanza ambos lados del sistema.

En algunos mamíferos, la tonotopía de las áreas del tronco encefálico inferior mantenida estrictamente es explotada por la convergencia sobre neuronas del complejo geniculado medial, que generan respuestas específicas a ciertas combinaciones espectrales. Las neuronas en el cuerpo geniculado medial reciben aferencias convergentes desde vías separadas de modo espectral y temporal. Este complejo, en virtud de sus aferencias convergentes, media la integración de las características relativas al espectro y a la duración de los sonidos. No se sabe si las células del núcleo geniculado medial humano son selectivas para las combinaciones de sonidos pero, por cierto, el procesamiento de la palabra requiere tanto de sensibilidad combinada espectral como temporal.

El blanco final de la información auditiva ascendente es la corteza auditiva. Si bien la corteza auditiva tiene algunas subdivisiones, se puede efectuar una amplia distinción entre un área primaria y las áreas periféricas o del cinturón. La corteza auditiva primaria (A1) recibe aferencias punto a punto de la división ventral del complejo geniculado medial y por lo tanto contiene un mapa tonotópico preciso. Las áreas del cinturón de la corteza auditiva reciben aferencias más difusas desde las áreas del cinturón del complejo geniculado medial y, por lo tanto, son menos precisas en su organización tonotópica.

La corteza auditiva primaria (A1) tiene un mapa topográfico de la cóclea, al igual que la corteza visual primaria (V1) y la corteza somatosensitiva primaria (S1) tienen mapas topográficos de sus respectivos epitelios sensitivos. Sin embargo, al contrario de los sistemas visual y somatosensitivo, la cóclea ya ha descompuesto el estímulo acústico de modo que está organizado tonotópicamente a lo largo de la membrana basilar.

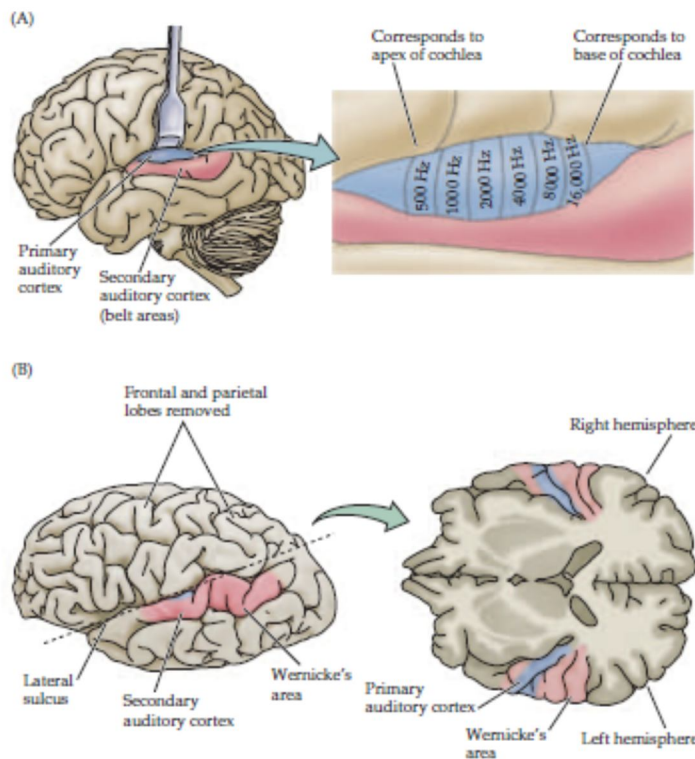


Figura 3. 12 Corteza Auditiva Humana. Copyright © (18)

(A) Encéfalo en vista lateral izquierda. Se aprecia A1 (organizado tonotópicamente) y A2.

(B) Encéfalo en vista lateral izquierda. Se aprecian las áreas corticales relacionadas con el procesamiento de voz. Nótese la ubicación del

Por lo tanto, se dice que A1 representa un mapa tonotópico. En disposición ortogonal al mapa tonotópico existe una organización en bandas de las propiedades binaurales. Las células de una banda son excitadas por ambos oídos mientras que las células en la banda siguiente son excitadas por un oído e inhibidas por el otro. Los tipos de procesamiento sensitivo que se desarrollan en las otras divisiones de la corteza auditiva no se conocen.

Al parecer, algunas áreas están especializadas en el procesamiento de combinaciones de frecuencias, mientras que otras están especializadas en el procesamiento de modulaciones de amplitud o de frecuencia.

Los sonidos que son especialmente importantes a menudo tienen una estructura temporal muy ordenada. En los seres humanos, el mejor ejemplo es la palabra. Parece probable que regiones específicas de la corteza auditiva humana estén especializadas en el procesamiento de sonidos elementales de la palabra, y en otras señales acústicas temporalmente complejas, como la música. En realidad el área de Wernicke, fundamental para la comprensión del lenguaje humano, se ubica entre el área auditiva secundaria y la corteza asociativa del lóbulo temporal.

Localización del sonido

Los seres humanos utilizan como mínimo dos estrategias distintas para localizar la posición horizontal de las fuentes sonoras, según las frecuencias en el estímulo. Para las frecuencias por debajo de 3 kHz (que se puede seguir con el trabado de fase) se utilizan las diferencias de tiempo interaural para localizar la fuente; por arriba de estas frecuencias se usan como señales las diferencias de intensidad interaural. Vías paralelas originadas en el núcleo coclear regulan cada una de estas estrategias para localizar el sonido (18).

La capacidad humana para detectar diferencias de tiempo interaural es notable. La diferencia de tiempo interaural más larga, producidas por los sonidos que nacen directamente laterales, se encuentran en el orden de 700 microsegundos (valor dado por el ancho de la cabeza, dividido por la velocidad del sonido en el aire, unos 340 m/s). En algunos experimentos psicofísicos se observó que en realidad los seres humanos pueden detectar diferencias de tiempo interaural de tan sólo 10 microsegundos; dos sonidos presentados a través de audífonos separados por diferencias de tiempo interaural tan pequeñas se perciben localizados hacia el lado del primer sonido. Esta sensibilidad se traduce en una precisión para localizar el sonido alrededor de 1°.

¿Cómo se logra el cronometrado en el rango de 10 microsegundos por los componentes neuronales que operan en el rango del milisegundo? El circuito neuronal que computa estas diferencias de tiempo interaural tan pequeñas, presenta aferencias binaurales hacia la oliva superior medial que nacen de los núcleos cocleares anteroventrales derecho e izquierdo (figura 3.13; véase también figura 3.10). La oliva superior medial contiene células con dendritas bipolares que se extienden en sentido tanto medial como lateral. Las dendritas laterales reciben aferencias del núcleo coclear anteroventral homolateral y las dendritas mediales reciben aferencias del núcleo coclear anteroventral contralateral (ambas aferencias son excitadoras). Como podría

esperarse, las células de la oliva superior medial funcionan como detectores de coincidencias, responden cuando ambas señales excitadoras llegan al mismo tiempo. Para que un mecanismo de coincidencia sea útil para localizar el sonido, las diferentes neuronas deben ser sensibles al máximo a diferentes retardos del tiempo interaural. Es evidente que los axones que se proyectan desde el núcleo coclear anteroventral varían sistemáticamente en longitud para crear líneas de retardo, (la longitud de un axón multiplicada por su velocidad de conducción es igual al tiempo de conducción). Estas diferencias anatómicas compensan los sonidos que llegan en tiempos ligeramente distintos a los dos oídos, de modo que los impulsos neurales resultantes llegan en forma simultánea a una neurona de la oliva superior medial, lo que torna a cada célula en especial sensible a las fuentes sonoras en un lugar particular. Los mecanismos que permiten que las neuronas de la oliva superior medial funcionen como detectores de coincidencia en el nivel de los microsegundos aún no se conocen con exactitud, pero ciertamente representan una de las especializaciones biofísicas más extraordinarias del sistema nervioso.

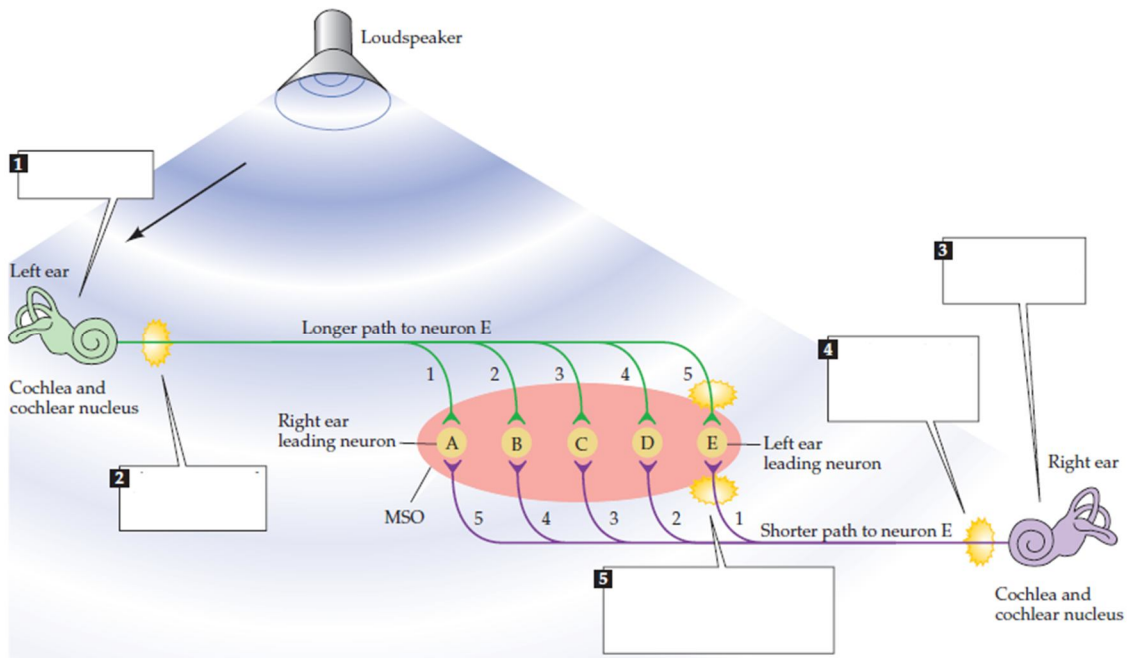


Figura 3. 13 Localización de sonido por diferencia de tiempo interaural. Copyright © (18)

Diagrama que muestra como la oliva superior medial computa la localización de un sonido por las diferencias de tiempo interaural.

- 1) El sonido alcanza primero el oído izquierdo. 2) El potencial de acción comienza a propagarse hacia la oliva superior medial. 3) El sonido alcanza el oído derecho un poco más tarde. 4) El potencial de acción desde el oído derecho comienza a propagarse hacia la oliva superior medial. 5) Los potenciales de acción convergen sobre una neurona de la oliva superior medial que responde más intensamente si su llegada es coincidente.

La localización del sonido percibida sobre la base de las diferencias de tiempo interaural requiere información de trabado de fase desde la periferia, que está disponible en los seres humanos sólo para frecuencias por debajo de 3 kHz. Por lo tanto, un segundo mecanismo debe entrar en juego en frecuencias más altas. En las que se superan los 2 kHz, la cabeza humana comienza a actuar como obstáculo acústico porque las longitudes de onda de los sonidos son demasiado cortas para rodearla. En consecuencia, cuando los sonidos de alta frecuencia están dirigidos a un lado de la cabeza, se crea una "sombra" acústica de menor intensidad en el oído alejado. Estas diferencias de intensidad brindan una segunda señal acerca de la localización de un sonido.

Los circuitos que computan la posición de una fuente sonora sobre esta base se encuentran en la oliva superior lateral y el núcleo medial del cuerpo del trapezoide (figura 3.14). Los axones excitadores se proyectan directamente desde el núcleo coclear anteroventral homolateral hasta la oliva superior lateral (así como la oliva superior medial; véase figura 3.13). Es importante observar que la oliva superior lateral también recibe aferencias inhibitoras desde el oído contralateral, a través de una neurona inhibitora en el núcleo medial del cuerpo trapezoide. Esta interacción excitadora/inhibidora produce una excitación neta de la oliva superior lateral del mismo lado del cuerpo que la fuente sonora.

Para los sonidos que nacen directamente laterales al oyente, las frecuencias de disparo serán máximas en la oliva superior lateral de ese lado; en esa circunstancia la excitación a través del núcleo coclear anteroventral homolateral será máxima y la inhibición desde el núcleo medial del cuerpo trapezoide contralateral mínima. Por el contrario, los sonidos que nacen más cerca de la línea medial del oyente producirán frecuencias de disparo más bajas en la oliva superior lateral homolateral debido al aumento de la inhibición que surge del núcleo medial del cuerpo trapezoide contralateral.

Para los sonidos que surgen en la línea media, o desde el otro lado, el aumento de la inhibición que se origina en el núcleo medial del cuerpo trapezoide es lo suficientemente potente como para silenciar por completo la actividad de la oliva superior lateral, es necesario notar que cada oliva superior lateral sólo codifica sonidos que nacen en el hemicampo homolateral; por lo tanto, para representar toda la gama de posiciones horizontales se necesitan dos olivas superolaterales. En resumen, hay dos vías separadas, y dos mecanismos separados, para localizar el sonido. Las diferencias de intensidad interaural se procesan en la oliva superior medial y las diferencias de intensidad interaural, en la oliva superior lateral. Estas dos vías finalmente se fusionan en los centros auditivos del mesencéfalo.

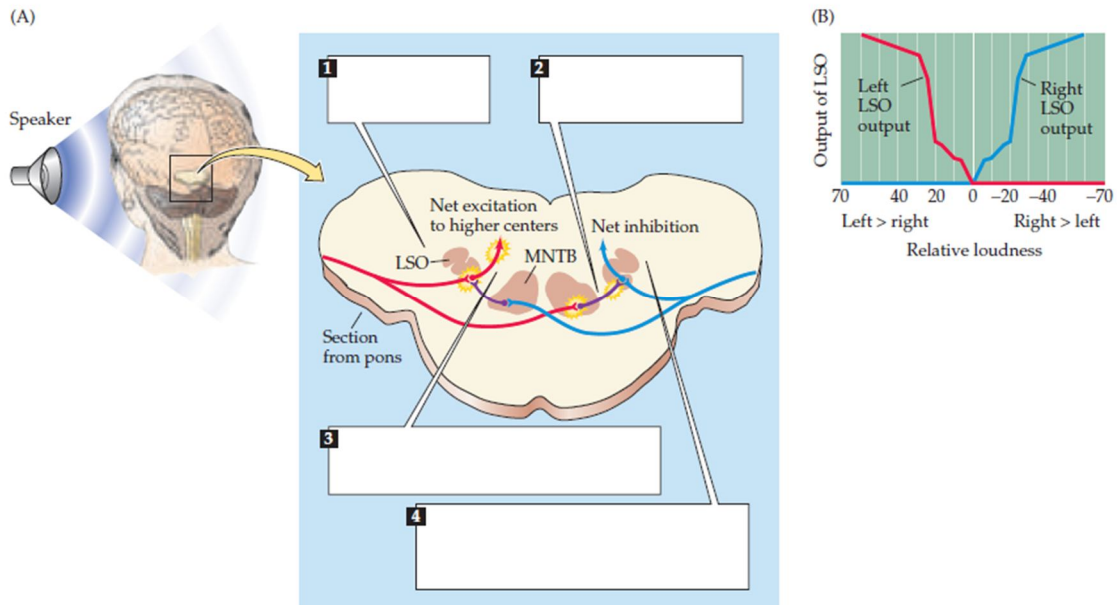


Figura 3. 14 Diferencias de intensidad interaural. Copyright © (18)

A: 1) El estímulo más intenso en el oído izquierdo excita la oliva superior lateral. 2) Este estímulo también inhibe la oliva superior lateral derecha a través de una interneurona del núcleo medial del cuerpo trapezoide. 3) La excitación desde el lado izquierdo es mayor que la inhibición desde el lado derecho, y produce una excitación neta hasta los centros superiores. 4) La inhibición desde el lado izquierdo es mayor que la excitación desde el lado derecho y produce una inhibición neta a la derecha y ninguna señal hacia los centros superiores.

B: Esta disposición de excitación-inhibición hace que las neuronas de la oliva superior lateral disparen con más intensidad en respuesta a los sonidos que nacen directamente en posición lateral al oyente del mismo lado que la oliva superior lateral.

Integración en el colículo inferior

Las vías auditivas que ascienden a través de los complejos olivar y lemniscal, así como otras proyecciones que surgen directamente del núcleo coclear, se proyectan hacia el centro auditivo del mesencéfalo, el colículo inferior. Al examinar cómo se desarrolla la integración en el colículo inferior, nuevamente es instructivo observar el mecanismo auditivo analizado en forma más completa, el sistema biaural para localizar sonido. Como ya se mencionó el espacio no se mapea en la superficie del receptor auditivo; por lo tanto, la percepción del espacio auditivo debe ser sintetizada de alguna forma por el circuito en el tronco del encéfalo inferior y el mesencéfalo. Mediante algunos experimentos en la lechuza, un animal extraordinariamente eficiente para localizar sonidos, mostro que la convergencia de aferencias biaurales en el mesencéfalo produce algo completamente nuevo en relación con la periferia, una representación topográfica computarizada del espacio auditivo. Las neuronas en el interior de este mapa auditivo del colículo responden mejor a los sonidos que se originan en una región específica del espacio y, por lo tanto, tienen tanto una elevación preferida como una localización horizontal preferida o azimut. Si bien, aún no se hallaron mapas comparables del espacio auditivo en los mamíferos, los seres

humanos tienen una percepción clara tanto de los componentes de elevación como azimutales de la localización de un sonido, lo que sugiere que tenemos un mapa similar del espacio auditivo.

Otra propiedad importante del colículo inferior es su capacidad para procesar sonidos con patrones temporales complejos. Muchas neuronas en el colículo inferior responden sólo a sonidos de duraciones específicas. Éstos son componentes típicos de los sonidos biológicamente relevantes, como los efectuados por los predadores, o los sonidos de comunicación intraespecífica, que en los seres humanos comprenden el habla.

Sistema auditivo central

El sistema auditivo central está conformado por los sectores del cerebro dedicados a la audición (corteza auditiva) y las áreas de asociación o córtex asociativo. Su desempeño se centra en la percepción sonora resultado de procesos psicofisiológicos que permiten interpretar los sonidos recibidos (19).

Las áreas involucradas después de que las señales son llevadas a través de los nervios acústicos son la corteza auditiva primaria (compuesta por el área de Heschl) y el área de Wernicke, aunque éste último se complementa con el área de Broca (19).

Generalmente, cuando una persona recibe un estímulo externo proporciona una respuesta motora, en el caso de la audición no es la excepción, más aún cuando la señal acústica contiene información que requiere sea percibida e interpretada y como respuesta motora se requiera una respuesta de voz, misma que si obedece a características muy específicas como lo son forma y estructura, podríamos describir a la respuesta de voz como una palabra, un ejemplo podría ser el repetir una palabra oída. Las estructuras involucradas para la emisión de una palabra a partir de una palabra oída se interrelacionan de la siguiente manera (19): la corteza sensorial se ve asociada con la corteza motora a través de las áreas de asociación, en el caso de la audición, la corteza auditiva se asocia con el área de Wernicke, que a través de un haz de fibras axónicas llamadas fascículo arcuato se relaciona con el área de Broca, dicha área se asocia con la corteza motora la cual se vincula con la musculatura del habla a través del sistema nervioso periférico, tal como se muestra en la figura 3.15.

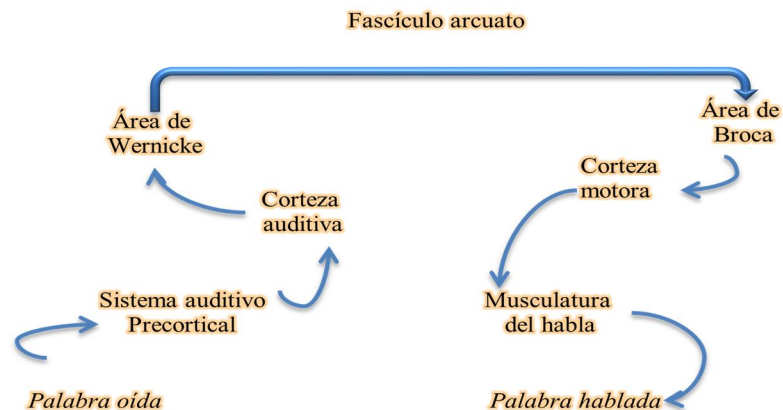


Figura 3. 15 Estructuras involucradas en la repetición de una palabra oída, de acuerdo con la conceptualización de Wernicke de la función del lenguaje.

A continuación se describirán en un aspecto fisiológico las estructuras involucradas en la audición.

Corteza cerebral

La corteza cerebral es el manto de tejido nervioso que cubre la superficie de los hemisferios cerebrales. De acuerdo a su clasificación filogenética, en ella se distinguen tres tipos básicos de corteza (20) :

1. Neocorteza
2. Paleocorteza
3. Arquicorteza

La neocorteza es la zona más evolucionada del cerebro, el lugar donde residen las actividades cerebrales más complejas, como la percepción y la conciencia. El tálamo es la puerta que lleva a la neocorteza. Toda la información sensorial, excepto la olfativa (la paleocorteza comprende el cerebro olfativo), llega a la neocorteza a través del tálamo (21).

Otra forma de presentar a la corteza cerebral es en su división anatómica ya que la corteza cerebral es ante todo una delgada capa de materia gris que está fuertemente circunvolucionada. Las circunvoluciones tienen "crestas" que se llaman giros, y "valles" que se llaman surcos. Algunos surcos son bastante pronunciados y largos, y se usan como límites convenidos entre las cuatro áreas del cerebro llamados lóbulos, como se muestra en la figura 3.16.

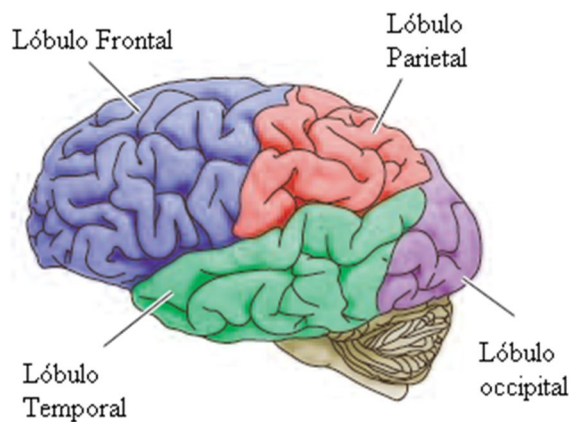


Figura 3. 16 Vista lateral izquierda del cerebro en donde se presenta su clasificación anatómica. Copyright © (18)

Sin embargo, una clasificación que permite exponer a la corteza de una manera más específica es de acuerdo a la citoarquitectura, descrita por las áreas de Brodmann.

En 1878, Brodmann realizó un mapeo histológico del córtex cerebral, dividiéndolo de acuerdo a la citoarquitectura en 52 áreas diferentes como se muestra en la figura 3.17. Cada área tiene una citoarquitectura o distribución neuronal característica, de tal manera que la corteza se puede seccionar en córtex especializados, tal como se muestra en la figura 3.18.

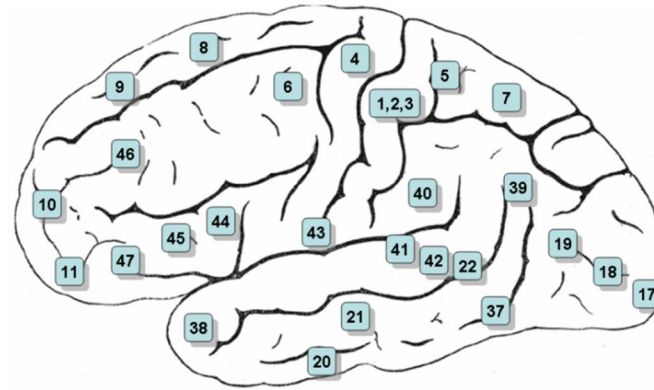


Figura 3. 17 Hemisferio cerebral izquierdo con sus áreas de Brodmann numeradas.

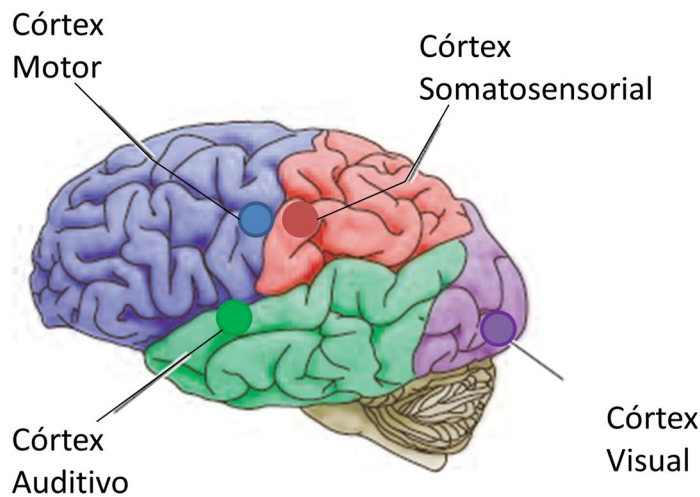


Figura 3. 18 Córtex especializada según la citoarquitectura. Copyright © (18)

Así, se comprobó que las siguientes áreas están relacionadas al sistema auditivo central:

Áreas de Brodmann	Nombre del Córtex
Área 41 y 42	Córtex auditivo primario (Área de Heschl)
Áreas 42 y 22	Córtex auditivo asociativo
Áreas 22, 39 y 40	Área asociativa de Wernicke
Áreas 44 y 45	Área de Broca

Tabla 3. 2 Áreas de Brodmann dedicadas a la audición.

Área de Heschl: El área de Heschl corresponde a las áreas 41 y 42 de la corteza cerebral. Pertenece al área auditiva primaria. La estimulación de esta área produce sensaciones auditivas burdas,

como susurros, zumbidos o golpeteo. Las lesiones pueden producir dificultad en la ubicación del sonido en el espacio y pérdida de la audición (19).

Área de Wernicke: Corresponde a las áreas 22, 39 y 40 de Brodmann. Pertenece a la corteza de asociación o córtex asociativo, específicamente auditiva, situada en la parte posteroinferior de la corteza auditiva primaria.

Su papel fundamental radica en la decodificación auditiva de la función lingüística (se relaciona con la comprensión del lenguaje); función que se complementa con la del área de Broca que procesa la gramática (22).

Área de Broca: Corresponde a las áreas de Brodmann 44 y 45, y se conecta con el área de Wernicke mediante un haz de fibras nerviosas llamado fascículo arcuato.

Es la sección involucrada en la producción del habla, el procesamiento del lenguaje y la comprensión. Aunque tradicionalmente se le ha asociado con la producción del habla, hay evidencia que no es esa su función concreta (22). No hay que olvidar que, pese a la importancia de esta área en el habla, no se puede hablar en términos absolutos.

Comprensión auditiva de las palabras

Los estudios referentes al proceso psicológico y fisiológico de la audición generalmente se exponen bajo evidencias obtenidas a partir de pacientes con lesiones cerebrales o supervisando la actividad cerebral a través de imágenes. Tal como lo hace (19) en su libro *de neuropsicología humana*, en el cual describe la comprensión auditiva como un componente principal de la función del lenguaje:

“La evidencia obtenida de los pacientes con lesiones cerebrales sugiere que la comprensión de una palabra oída requiere al menos 3 tipos de procesamiento: detección de las características acústicas elementales del sonido de la palabra (agudeza auditiva temporal), percepción de los fonemas que componen la palabra (análisis auditivo perceptual) y asignación de significado de la palabra (procesamiento semántico). El primero y más básico de estos procesos parece ser un proceso elemental de la audición que no específico del lenguaje y que es mediado por la corteza auditiva. Por tanto, las lesiones bilaterales de las cortezas auditivas provocan deterioro en la resolución temporal de los estímulos como los que se requieren para la discriminación de dos chasquidos presentados rápidamente de los mostrados de manera simultánea. Estos pacientes requieren una mayor demora entre las presentaciones para discriminarlas. En contraste, los pacientes son capaces de discriminar con agudeza aspectos de los sonidos, como el tono y el volumen, que no requieren un alto nivel de resolución temporal para su percepción.

Al considerar la siguiente etapa del procesamiento, el análisis auditivo perceptual, se encuentra que las lesiones en muchas áreas del hemisferio izquierdo están asociadas con un deterioro en la discriminación de fonemas pero que la región asociada con mayor frecuencia es el área de Wernicke. En contraste, las lesiones en el hemisferio derecho en el área homóloga al área de Wernicke están asociadas con deterioro en la igualación de sonido lingüísticamente no-insignificantes. En particular, Posner y Raichle (1994) han reportado un incremento en la actividad

en el área de Wernicke así como en las cortezas auditivas de ambos hemisferios cuando se escuchan palabras de manera pasiva (figura 3.19).

El procesamiento semántico, la extracción de significado de los fonemas percibidos con precisión, es la siguiente etapa en el proceso. El deterioro en la comprensión de palabras escuchadas, como se valora, por ejemplo, por medio de una prueba de vocabulario auditivo, está asociado con lesiones en varias partes del hemisferio izquierdo pero con más frecuencia con lesiones en el lóbulo temporal izquierdo. De particular importancia es el hallazgo de que los pacientes con comprensión deteriorada de palabra auditiva no necesariamente tienen algún deterioro en el procesamiento de los sonidos de las palabras, cuando se mide, por ejemplo, por medio de una tarea de discriminación de fonemas. Esta disociación sugiere que el proceso de percepción y categorización de sonidos de palabras en preceptos psicológicamente distintos (fonemas) y el proceso de asignar significado a combinaciones de fonemas percibidos (palabras) son procesos separados, aunque el primero es un prerrequisito para el último.”

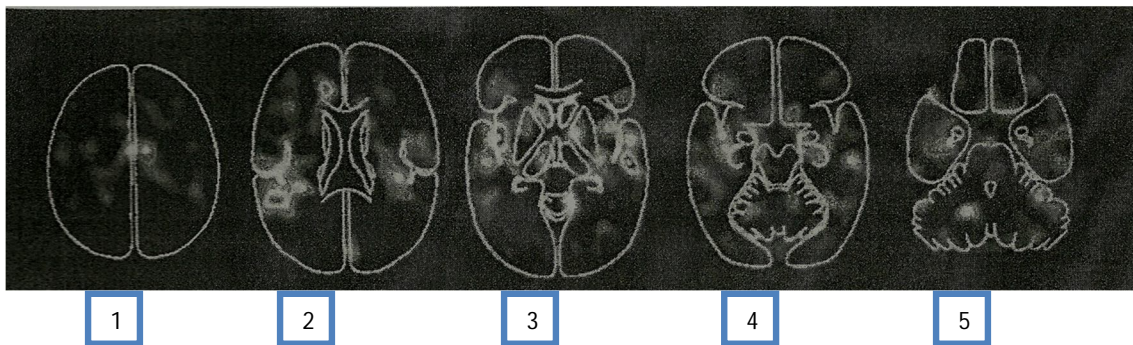


Figura 3. 19 Imágenes PET (Tomografía por Emisión de Positrones). Copyright © (19)

Vistas en sección horizontal, obtenidas cuando sujetos normales escuchaban de manera pasiva algunas palabras. La activación es mayor en los lóbulos temporales de ambos hemisferios (diapositiva 3) y en el área de Wernicke en el hemisferio izquierdo (diapositiva 2).

Lenguaje

Existe evidencia de que el lenguaje no es totalmente innato, ni totalmente aprendido (19). Afirmaciones de conductistas como John Watson y B.F. Skinner, afirman que los niños no aprenden el lenguaje mediante la memorización de lo que han escuchado y luego emiten respuestas condicionadas tras ser expuestos a estímulos condicionados. Ni aprenden el lenguaje, a excepción de ciertas instancias muy especializadas, debido a que sus esfuerzos son recompensados o castigados (19).

Es difícil apreciar que en el pasado se sostuvo una fuerte visión empiricista de la adquisición del lenguaje. Una buena posición que es apoyada por una buena cantidad de evidencias es que, a pesar de que el lenguaje debe ser experimentado para ser aprendido, la capacidad para el lenguaje es una propiedad innata del cerebro humano (19).

La experiencia y el lenguaje

Se ha analizado al cerebro tratando de descubrir la localización estructural de las diferentes cualidades de la experiencia. Lesiones y estimulaciones de sus diversas partes han demostrado que la distribución de los elementos encargados de éstas es más difusa de lo que se predecía, aunque no lo suficiente para tornar imposible su estudio (23).

De todos estos estudios es posible concluir que las diferentes cualidades de la experiencia surgen cuando la información llega a porciones centrales del sistema, las cuales difieren en sus características morfológicas y anatómicas.

En cada milisegundo ocurren millones de despolarizaciones en circuitos neuronales formados por un número astronómico de elementos. Tanto el número de combinaciones de neuronas que pueden encontrarse en un estado simultáneo de activación, como el número de diferentes patrones con los que estas responden, son prácticamente infinitos.

La actividad de un gran conglomerado de circuitos representa información de un objeto, en tanto que la actividad hipercompleja y singular de todo el sistema se asocia con la experiencia. Al percibir un objeto, cambian y se determinan los patrones de actividad de gran parte del cerebro.

En términos generales, concebir una experiencia desde el punto de vista de las unidades de activación del cerebro (potenciales de acción transmitidos a través de canales axónicos), no se explica por qué una pregunta es una pregunta y un color un color. Es necesario observar al cerebro desde de un punto de vista más general para poder siquiera aproximarnos a la contestación de tales preguntas. En este sentido, es posible diferenciar dos fenómenos muy globales de la actividad cerebral: los *contenidos* y las *operaciones*. Con contenidos se quiere denotar las cualidades subjetivas que resultan de la activación del cerebro: dolor, placer, sensación de luz y sonido, etc. Operación significa el manejo lógico al que es sometida tal información al pasar a través de circuitos de convergencia, divergencia, inhibitorios, de interacción entre estructuras, etc.

Los contenidos cualitativos aparecen cuando la activación llega a estructuras o sistemas de estructuras como el límbico, corteza occipital, corteza temporal (sonido), etc. Es esto lo que denominamos experiencia.

La corteza occipital difiere del hipocampo o de la corteza del cíngulo en la velocidad de transmisión de sus fibras, la distribución de sus neuronas, la dirección en que están colocados sus axones y la disposición tridimensional de sus arborizaciones detriticas. La información que fluye a través de tales diferencias morfológicas y anatómicas adquiere una configuración tridimensional y temporal característica de cada estructura. Así pues, tal configuración específica debe ser la que posea o dé lugar a la cualidad subjetiva de la experiencia.

Puesto que no son la actividad de una neurona ni la de un axón lo que explica nuestra experiencia, sino la resultante total de cambios de energía determinados por la disposición global de los elementos a través de los cuales fluye, resulta posible pensar que la cualidad de una experiencia

es independiente (en sentido global) de las operaciones de interacción que sufre la información asociada a ella. Sólo en esta forma es posible explicar por qué la estimulación artificial (con un pulso eléctrico) de la corteza temporal da como resultado la sensación de un sonido.

El lenguaje es, más que contenido, operación. Resulta del manejo de la información de los circuitos encargados de lograr las disposiciones energéticas asociadas a las experiencias cualitativas y además es capaz de reactivar estos mismos. Por ello, el lenguaje al que nos estamos refiriendo no es solamente la salida del sistema, sino además los pasos previos a ésta. Sin embargo, entre la verbalización y los procesos lógicos que la preceden existe una barrera tal, que impide vivir la totalidad de éstos y permite sólo ser conscientes de ella. En otras palabras, somos conscientes de nuestras verbalizaciones, mas no de sus causas.

En conclusión, la cualidad subjetiva de la experiencia debe surgir de la particular disposición energética que ocurre cuando una morfología anatómica específica es activada.

No es en la puesta en marcha de un axón, una sinapsis o una neurona en donde está la diferencia entre una luz, un sonido o una emoción, sino en la actividad global tridimensional de las estructuras o sistemas cerebrales encargados de tal experiencia. La cualidad consciente de una experiencia vivencial depende de un sistema de inclusión que utiliza como unidades de análisis lo que de característico y específico poseen las disposiciones energéticas como totalidades. Es aquí donde la percepción y la experiencia se interconectan.

Reconocimiento de voz

De forma general, las técnicas de reconocimiento de voz se pueden agrupar en cuatro grandes técnicas (15):

- Técnicas topológicas basadas en el cálculo y comparación de distancias como el *“Dynamic Time Warping o DTW”*.
- Técnicas estadísticas o probabilísticas como los *Modelos Ocultos de Markov (HMM)* o *Modelos de Mixturas Gausseanas (GMM)*.
- Sistemas basados en el conocimiento como los *sistemas expertos* o *algoritmos genéticos*.
- *Redes neuronales artificiales*.

Para este trabajo se optó por trabajar con redes neuronales artificiales.

Redes neuronales artificiales

Las redes neuronales artificiales son un paradigma de aprendizaje y procesamiento automático inspirado en la estructura y comportamiento del sistema nervioso humano. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida (24).

Los modelos de las redes neuronales artificiales están basados en los rasgos y características esenciales de las neuronas biológicas y sus conexiones.

Analogía con las redes neuronales biológicas

Las redes neuronales están compuestas por una o más unidades de proceso (neuronas). Cada unidad de proceso se compone de una red de conexiones de entrada (estímulo), una función de red (de propagación del estímulo), encargada de computar la entrada total combinada de todas las conexiones, un núcleo central de proceso (activación), encargado de aplicar la función de activación, y la salida (transmisión del impulso), por donde se transmite el valor de activación a otras unidades (sinapsis) ver figura 3.20.

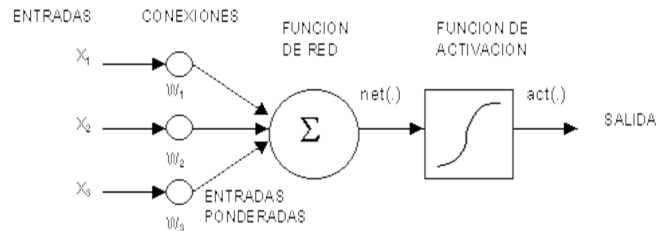


Figura 3. 20 Unidad de proceso típica.

La función de red es típicamente el sumatorio ponderado, mientras que la función de activación suele ser alguna función de umbral o una función sigmoide.

- Función de propagación o de red: calcula el valor de base o entrada total a la unidad, generalmente como simple suma ponderada de todas las entradas recibidas, es decir, de las entradas multiplicadas por el peso o valor de las conexiones. Equivale a la combinación de las señales excitatorias e inhibitorias de las neuronas biológicas.
- Función de activación: es quizá la característica principal o definitoria de las neuronas, la que mejor define el comportamiento de la misma. Se usan diferentes tipos de funciones, desde funciones simples de umbral a funciones no lineales. Se encarga de calcular el nivel o estado de activación de la neurona en función de la entrada total.
- Conexiones ponderadas: hacen el papel de las conexiones sinápticas, el peso de la conexión equivale a la fuerza o efectividad de la sinapsis. La existencia de conexiones determina si es posible que una unidad influya sobre otra, el valor de los pesos y el signo de los mismos definen el tipo (excitatorio/inhibitorio) y la intensidad de la influencia.
- Salida: calcula la salida de la neurona en función de la activación de la misma, aunque normalmente no se aplica más que la función identidad, y se toma como salida el valor de activación. El valor de salida cumpliría la función de la tasa de disparo en las neuronas biológicas.

Redes Neuronales Biológicas	Redes Neuronales Artificiales
Neuronas	Unidades de proceso
Conexiones sinápticas	Conexiones ponderadas
Efectividad de las sinapsis	Peso de las conexiones
Efecto excitatorio o inhibitorio de una conexión	Signo del peso de una conexión
Efecto combinado de las sinapsis	Función de propagación o de red
Activación -> tasa de disparo	Función de activación -> Salida

Tabla 3. 3 Comparación entre las neuronas biológicas y las unidades de proceso artificiales.

Función de red

Se encarga de calcular la entrada total de la neurona como combinación de todas las entradas. La función más utilizada con diferencia es la función lineal de base (LBF), que consiste en el sumatorio ponderado de todas las entradas.

Funciones de activación

Se suele distinguir entre funciones lineales, en las que la salida es proporcional a la entrada; funciones de umbral, en las cuales la salida es un valor discreto (típicamente binario 0/1) que depende de si la estimulación total supera o no un determinado valor de umbral; y funciones no lineales, no proporcionales a la entrada.

Características

1. *Aprendizaje inductivo*: no se le indican las reglas para dar una solución, sino que extrae sus propias reglas a partir de los ejemplos de aprendizaje, modifican su comportamiento en función de la experiencia. Esas reglas quedan almacenadas en las conexiones y no representadas explícitamente como en los sistemas basados en conocimiento (simbólico-deductivos).
2. *Generalización*: una vez entrenada, se le pueden presentar a la red datos distintos a los usados durante el aprendizaje. La respuesta obtenida dependerá del parecido de los datos con los ejemplos de entrenamiento
3. *Abstracción o tolerancia al ruido*: las redes neuronales artificiales son capaces de extraer o abstraer las características esenciales de las entradas aprendidas, de esta manera pueden procesar correctamente datos incompletos o distorsionados.
4. *Procesamiento paralelo*: las neuronas reales trabajan en paralelo; en el caso de las redes artificiales es obvio que si usamos un solo procesador no podrá haber proceso paralelo real; sin embargo hay un paralelismo inherente, lo esencial es que la estructura y modo de operación de las redes neuronales las hace especialmente adecuadas para el procesamiento paralelo real mediante multiprocesadores (se están desarrollando máquinas específicas para la computación neuronal).

5. *Memoria distribuida*: el conocimiento acumulado por la red se halla distribuido en numerosas conexiones, esto tiene como consecuencia la tolerancia a fallos: una red neuronal es capaz de seguir funcionando adecuadamente a pesar de sufrir lesiones con destrucción de neuronas o sus conexiones, ya que la información se halla distribuida por toda la red, sin embargo en un programa tradicional un pequeño fallo en cualquier punto puede invalidarlo todo y dar un resultado absurdo o no dar ningún resultado.

Arquitectura de redes

Para diseñar una red debemos establecer como estarán conectadas unas unidades con otras y determinar adecuadamente los pesos de las conexiones (25). Lo más usual es disponer las unidades en forma de capas, pudiéndose hablar de redes de una, de dos o de más capas, las llamadas redes multicapa como se muestra en la figura 3.21.

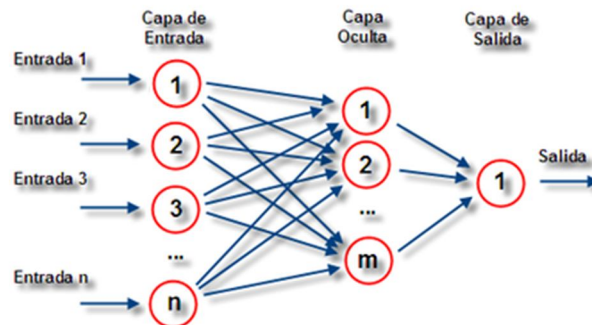


Figura 3. 21 Red neuronal artificial perceptrón simple.

Esta arquitectura presenta n neuronas de entrada, m neuronas en su capa oculta y una neurona de salida.

Otra forma de diseño de redes son las redes recurrentes, estas presentan al menos un ciclo cerrado de activación neuronal. Así que podemos hablar de 3 principales arquitecturas:

- Monocapa (una capa de neuronas)
- Multicapa
- Recurrentes

Aprendizaje

Al igual que las conexiones entre neuronas y sus valores sinápticos constituyen la clave para la codificación de información en el cerebro, el conjunto de valores w_{ji} nos proporciona la información que alberga una RNA.

El proceso de aprendizaje consiste en variar la valores sinápticos w_{ji} siguiendo unas pautas establecidas. En los sistemas biológicos se produce una continua creación y destrucción de conexiones entre las neuronas, en los sistemas que los simulan, la destrucción de una conexión se sigue haciendo que su eso asociado w_{ji} tome el valor cero.

La RNA ha aprendido cuando se han modificado los pesos de tal manera que por cada entrada que se le presente nos proporciona el resultado esperado. Esta finalización en la modificación de los pesos se puede expresar como: $\frac{dw_{ji}}{dt} = 0$, para cada peso de la RNA.

El modo de aprendizaje más sencillo consiste en la presentación de patrones de entrada junto a los patrones de salida deseados (targets) para cada patrón de entrada, por eso se llama aprendizaje supervisado. Si no se le presentan a la red los patrones de salida deseados, diremos que se trata de aprendizaje no supervisado, ya que no se le indica a la red que resultados debe dar, sino que se le deja seguir alguna regla de auto-organización. Un tercer tipo de aprendizaje, a medio camino entre los anteriores, es el llamado aprendizaje reforzado, en este caso el supervisor se limita a indicar si la salida ofrecida por la red es correcta o incorrecta, pero no indica que respuesta debe dar.

Redes competitivas o mapas de auto-organización

Son redes uni o multicapa cuyo común denominador es postular algún tipo de competición entre unidades con el fin de conseguir que una de ellas quede activada y el resto no. Esto se consigue mediante aprendizaje no supervisado, presentando algún patrón de entrada y seleccionando la unidad cuyo patrón de pesos incidentes se parezca más al patrón de entrada, reforzando dichas conexiones y debilitando las de las unidades perdedoras (24).

La competición entre unidades se puede conseguir simulando una característica neurofisiológica del córtex cerebral llamada inhibición lateral. Esto se logra postulando la existencia de conexiones inhibitorias intracapa y conexiones excitatorias intercapa, de tal manera que la presentación de un patrón de entrada tenderá a producir la activación de una única unidad y la inhibición del resto.

Al final se consigue que cada unidad responda frente a un determinado patrón de entrada, y, por generalización, que cada unidad responda frente a patrones de entrada similares, de manera que los pesos aferentes de esa unidad converjan en el centro del grupo de patrones con características similares.

Es usual que haya una capa de neuronas de entrada y una capa de salida. Se usan tantas entradas como dimensiones tenga el espacio vectorial de los patrones de entrada (espacio real o binario), y tantas salidas como clases o categorías se quieren utilizar para clasificar los patrones de entrada, de manera que cada nodo de salida representa una categoría.

Además de las conexiones hacia delante, con función excitatoria, se usa una red intracapa, inhibitoria, simulando el fenómeno neurológico de la inhibición lateral, de ahí que se la denomina capa lateral. La red hacia delante implementa una regla de excitación de aprendizaje de Hebb. Esta regla refuerza las conexiones entre los pares de unidades entrada-salida que se activan

simultáneamente. La red lateral es intrínsecamente inhibidora, realiza la labor de seleccionar al ganador, normalmente mediante un método de aprendizaje competitivo, como el "*Winner Take All (WTA)*" el ganador lo toma todo: la unidad con mayor valor de activación toma el valor máximo, por ejemplo 1 y el resto el mínimo por ejemplo 0.

Las redes competitivas se usan típicamente como clasificadores de patrones, ya que cada unidad responde frente a grupos de patrones con características similares. Para estimar el grado de semejanza de los patrones se utilizan distancias o medidas de similitud, siendo la más común la distancia euclídea (ecuación 3.21):

$$\text{Producto interno} \quad \langle x_i x_j \rangle \equiv x_i^T x_j \equiv \|x_i\| \cdot \|x_j\| \cos(x_i, x_j)$$

3. 20

$$\text{Distancia Euclídea con Pesos} \quad d(x_i x_j) \equiv \sum_k [x_i(k) - x_j(k)]^2$$

3. 21

Podemos comparar estas redes con los métodos estadísticos de análisis de cúmulos, que agrupan los datos en grupos con características similares.

La principal crítica a estos modelos es que no poseen una de las características generales de las redes neuronales: la información no se halla distribuida entre todas las conexiones, la destrucción de una sola unidad provocaría la pérdida de la información relativa a todo un grupo o categoría de patrones. Para solventar este problema se han desarrollado los códigos demográficos, que representan cada categoría o grupo de patrones mediante un conjunto de unidades próximas entre sí, en vez de mediante una sola unidad.

Sistemas de Lógica Difusa (SLD)

La Lógica Difusa es una herramienta que, de acuerdo con (26), permite llevar el conocimiento humano estructurado hacia algoritmos tratables de forma matemática. De forma concreta, la Lógica Difusa es considerada como un sistema lógico dirigido al modelado de formas de razonamiento humano que lejos de ser exactos o absolutos son más bien aproximados, o difusos, y por lo tanto permiten su análisis por métodos clásicos basados en lógica bivalente o teorías probabilísticas. La gran versatilidad que estas técnicas representa se debe a que sus métodos permiten el procesamiento de información tanto simbólica como numérica.

Al tratarse de un método de modelado del entendimiento de los conceptos humanos sobre el mundo, se debe considerar que éste no es precisamente binario. Una persona no se encuentra

usualmente absolutamente sana o terminalmente enferma. Los conceptos no pueden ser totalmente falsos o absolutamente verdaderos. Es así como existe una amplia gama entre los conceptos binarios. Y cada estado particular tiene asociado un nivel o grado de pertenencia a uno o varios estados bien (aproximadamente) definidos dentro de estos límites. Cada uno de estos estados recibe el nombre de conjunto difuso, y tiene asociada una función de pertenencia que determina el grado de pertenencia de un determinado nivel a cada estado.

La figura 3.22 muestra ejemplos de algunas de las funciones de pertenencia más usuales en la ingeniería.

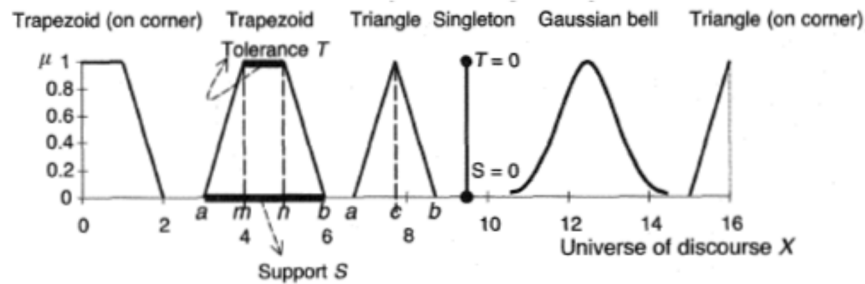


Figura 3. 22 Conjuntos Difusos más utilizados en ingeniería

Utilizando diagramas de Venn, un conjunto difuso puede ser representado como se muestra en la figura 3.23.

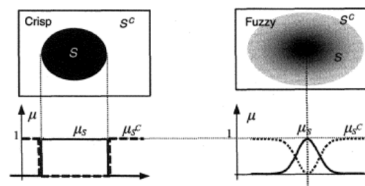


Figura 3. 23 Comparación entre conjuntos convencionales y conjuntos difusos.

De tal manera que las operaciones entre ellos se redefinen de acuerdo a las necesidades del sistema a modelar. La figura 3.24 muestra las principales redefiniciones para las operaciones AND y OR.

AND T-Norm $T(\mu_A(x), \mu_B(x))$	OR S-Norm $S(\mu_A(x), \mu_B(x))$
Minimum $\text{MIN}(\mu_A(x), \mu_B(x))$	Maximum $\text{MAX}(\mu_A(x), \mu_B(x))$
Algebraic product $\mu_A(x)\mu_B(x)$	Algebraic sum $\mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x)$
Drastic product $\text{MIN}(\mu_A(x), \mu_B(x))$ if $\text{MAX}(\mu_A(x), \mu_B(x)) = 1$ 0 otherwise	Drastic sum $\text{MAX}(\mu_A(x), \mu_B(x))$ if $\text{MIN}(\mu_A(x), \mu_B(x)) = 0$ 1 otherwise
Lukasiewicz AND (Bounded Difference) $\text{MAX}(0, \mu_A(x) + \mu_B(x) - 1)$	Lukasiewicz OR (Bounded Sum) $\text{MIN}(1, \mu_A(x) + \mu_B(x))$
Einstein product $\mu_A(x)\mu_B(x)/(2 - (\mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x)))$	Einstein sum $(\mu_A(x) + \mu_B(x))/(1 + \mu_A(x)\mu_B(x))$
Hamacher product $\mu_A(x)\mu_B(x)/(\mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x))$	Hamacher sum $(\mu_A(x) + \mu_B(x) - 2\mu_A(x)\mu_B(x))/(1 - \mu_A(x)\mu_B(x))$
Yager operator $1 - \text{MIN}(1, ((1 - \mu_A(x))^b + (1 - \mu_B(x))^b)^{1/b})$	Yager operator $\text{MIN}(1, (\mu_A(x)^b + \mu_B(x)^b)^{1/b})$

Figura 3. 24 Comparación entre las operaciones AND y OR para conjuntos binarios y conjuntos difusos.

Solución de problemas por Lógica Difusa

El primer paso en el modelado de Sistemas Difusos es la determinación de los diferentes conjuntos difusos que describen al sistema, de tal manera que han de formularse los conjuntos difusos de entrada y los conjuntos difusos de salida. Tras ello se establecen las relaciones entre cada uno de los universos definidos. Las entradas al sistema pasan por una etapa conocida como fusificación donde, en función a estas entradas se obtienen los grados de pertenencia de cada una y estas son procesadas mediante herramientas matemáticas. Finalmente, tras el procesamiento de los datos, éstos pasarán por otro proceso conocido como defusificación donde, a partir de los datos difusos se calcula la salida con dimensiones reales del sistema modelado. La figura 3.25 muestra los métodos de defusificación más importantes:

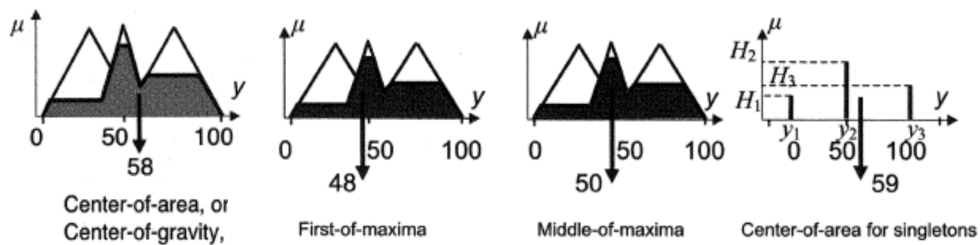


Figura 3. 25 Red neuronal artificial perceptrón simple.

Validación cruzada

La validación cruzada, es una técnica para evaluar cómo los resultados de un análisis estadístico generalizarán a un conjunto de datos independientes. Es principalmente usado en entornos donde el objetivo es predecir, y se quiere estimar con precisión como un modelo funcionará en la práctica.

Una prueba de la validación cruzada consiste en dividir una muestra de datos en subconjuntos complementarios, realizando el análisis sobre un subconjunto (llamado conjunto de entrenamiento), y la convalidación del análisis en el otro subconjunto (llamado conjunto de validación). Para reducir la variabilidad en múltiples pruebas se realizan usando diferentes particiones, y se hace un promedio de los resultados de las pruebas (27).

CAPÍTULO 4. DESARROLLO DE LA PROPUESTA

DESARROLLO DE LA PROPUESTA

Se presenta en la figura 4.1 un diagrama general del sistema propuesto, en donde se muestran los elementos utilizados en materia de hardware. Posteriormente en las figuras 4.6 y 4.10 se presentan diagramas que desglosan los procesos desarrollados en el hardware correspondiente. Finalmente se describe el desempeño de cada bloque detallando su implicación biológica.

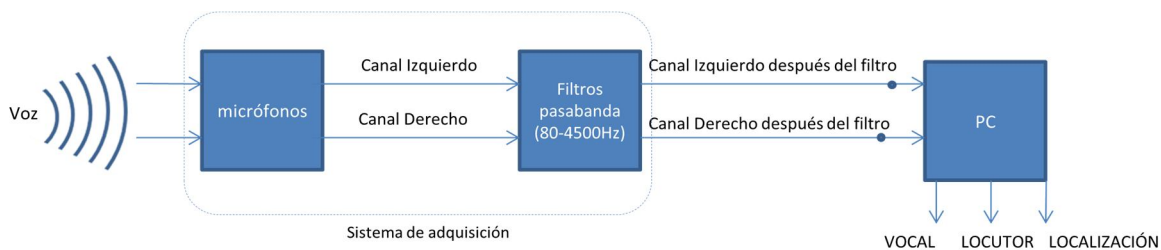


Figura 4. 1 Diagrama general en materia de hardware.

Sistema de adquisición

Para establecer el sistema de adquisición que permitió satisfacer las necesidades planteadas en los objetivos, se tomaron consideraciones teóricas referentes a la localización, a la voz humana y el desempeño del sistema auditivo humano.

Una de las premisas importantes fue la expuesta por (28) donde se menciona que la localización sólo es posible a partir de la audición binaural. Con un solo oído no es posible localizar fuentes sonoras. Además, de acuerdo con (18) las estrategias del sistema auditivo humano para localizar la fuente de sonido están estrechamente relacionadas con la estructura de la cabeza (debido al fenómeno de difracción de la onda) y la distancia entre los pabellones auriculares.

Resumiendo la teoría referente al desempeño de los pabellones auriculares y la cabeza humana se puntualiza lo siguiente (28):

- la importancia de la cabeza y del pabellón auditivo en la localización de sonidos se encuentran en el plano medio (Figura 4.2).
- tanto la cabeza, pero principalmente el pabellón auditivo, modifican el espectro de los sonidos en dependencia del ángulo de incidencia del sonido con respecto a la cabeza.

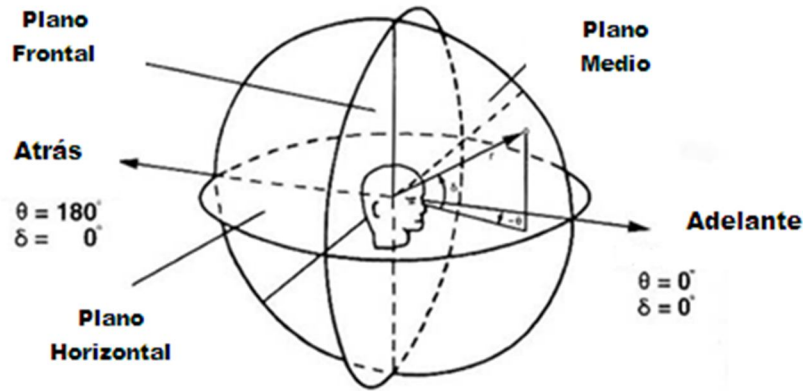


Figura 4. 2 Los tres planos característicos para estudiar la localización por parte del ser humano. Copyright © (27)

De esta forma, se realizó un sistema de adquisición biaural. Para ello se realizaron dos estructuras basadas en la morfología del pabellón auricular (figura 4.3) de un material elástico y blando, procurando que tuviese algunas características del pabellón auricular humano. El material utilizado fue caucho de silicón cuyas especificaciones se encuentran en el anexo I.



Figura 4. 3 Pabellones auriculares artificiales.

Los pabellones auriculares se acoplaron a una estructura con forma de cabeza humana elaborada de poliestireno expandido. La estructura tiene las dimensiones correspondientes a una cabeza humana tomando como base comparativa una persona con una altura de 160 cm, es decir, el tamaño de la cabeza cumple una proporción de un octavo de su altura, que corresponde a una relación antropométrica (29), por lo tanto el tamaño de la cabeza tiene 20 centímetros de largo por 20 de ancho y 20 de alto.

Finalmente, se seleccionaron dos micrófonos electret detallados más adelante, los cuales se acomodaron dentro de su respectivo pabellón auricular correspondiente al área del conducto auditivo externo.

Micrófono

La transducción de voz se lleva a cabo por medio de dos micrófonos electret con características iguales. La determinación del micrófono se basó en el análisis de características técnicas esenciales que emulan y cubren el desempeño del sistema biológico.

Consideraciones biológicas en la elección del micrófono

Una de las consideraciones primordiales es que la respuesta en frecuencia del micrófono cubriera los rangos de frecuencia de la voz humana, que, como se ha mencionado ésta se encuentra en un rango de 80 a 4500 Hz.

Por otro lado, el oído humano tiene una respuesta unidireccional, es decir, presenta una respuesta que no es constante a la direccionalidad del sonido, de hecho, esta cualidad permite una captación localizada del sonido.

Consideraciones técnicas en la elección del micrófono

Un micrófono puede caracterizarse por varios aspectos relacionados con su respuesta a las ondas sonoras. Las de mayor relevancia para su elección fueron:

- **Rango dinámico:** rango de niveles sonoros en los que la señal eléctrica que produce el micrófono es suficientemente alta para ser utilizada. Está relacionado con la amplitud de la onda sonora que llega al micrófono. Es difícil construir micrófonos con un rango dinámico amplio; por un lado deben responder a señales sonoras con gran amplitud sin estropearse, y por otro lado, deben responder correctamente a señales de una intensidad sonora muy baja.
Esto se ve descrito por la sensibilidad del micrófono. La sensibilidad se define como la relación entre la tensión eléctrica expresada en voltios obtenida en los bornes del micrófono en circuito abierto y la presión sonora aplicada expresada en Pascal utilizando una frecuencia de 1000 Hz.
- **Respuesta en frecuencia:** Se caracteriza por la intensidad de la señal eléctrica producida por un micrófono para una amplitud determinada de la presión de la onda sonora, a diferentes frecuencias. La respuesta ideal sería una gráfica completamente plana. En el caso real, para frecuencias bajas, está limitada por la frecuencia de resonancia de la vibración mecánica del diafragma, y para frecuencias altas, decrece rápidamente cuando la longitud de onda de las ondas sonoras es menor que el tamaño de diafragma.
- **Directividad:** Se refiere a la respuesta de sensibilidad dependiendo de la dirección de donde llegue el sonido, estos pueden ser: omnidireccionales, bidireccionales, unidireccionales, parabólicos, de zona de presión y de gradiente de presión.

Los micrófonos seleccionados fueron del tipo electret cuyas especificaciones se muestran en el anexo II. Éstos ofrecen las siguientes ventajas:

- Económico
- Respuesta en frecuencia en el rango audible
- Preamplificado
- Sensible

Como se puede apreciar en el anexo II, el micrófono cuenta con un preamplificador, el cual debe ser polarizado con un filtro RC (figura 4.4) y alimentado con un voltaje que se encuentre en el rango de las especificaciones.

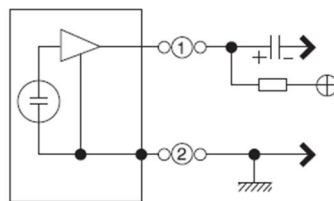


Figura 4. 4 Circuito de polarización de un micrófono electret.

Para polarizar el micrófono se diseñó un circuito impreso basado en el diagrama de la figura 4.4. El diseño del circuito impreso se muestra en el anexo III. De esta manera el sistema de adquisición cumple con una primera etapa como la adecuación de los micrófonos colocados en el conducto auditivo externo de los pabellones auriculares artificiales, que a su vez están acoplados a la estructura con forma de cabeza humana.

Filtros pasabandas

Como se mencionó anteriormente, algunos estudios en seres humanos y animales aportó la idea de que la periferia auditiva está optimizada para transmitir los sonidos vocales típicos de la especie, en lugar de transmitir correctamente todos los sonidos hasta las áreas auditivas centrales (18).

Con esta base, se desarrollo un filtro pasabandas acotado a las frecuencias de voz. Por otro lado, el filtrado de la señal de voz es una etapa fundamental en el acondicionamiento de la señal, uno de los objetivos principales es la eliminación de ruido, sobre todo del ruido ambiente.

Se diseñó un filtro activo pasabandas de primer orden (figura 4.5) con una frecuencia de corte inferior de 80 Hz y una frecuencia de corte superior de 4500 Hz, cuya respuesta se muestra en la figura 4.5 b.

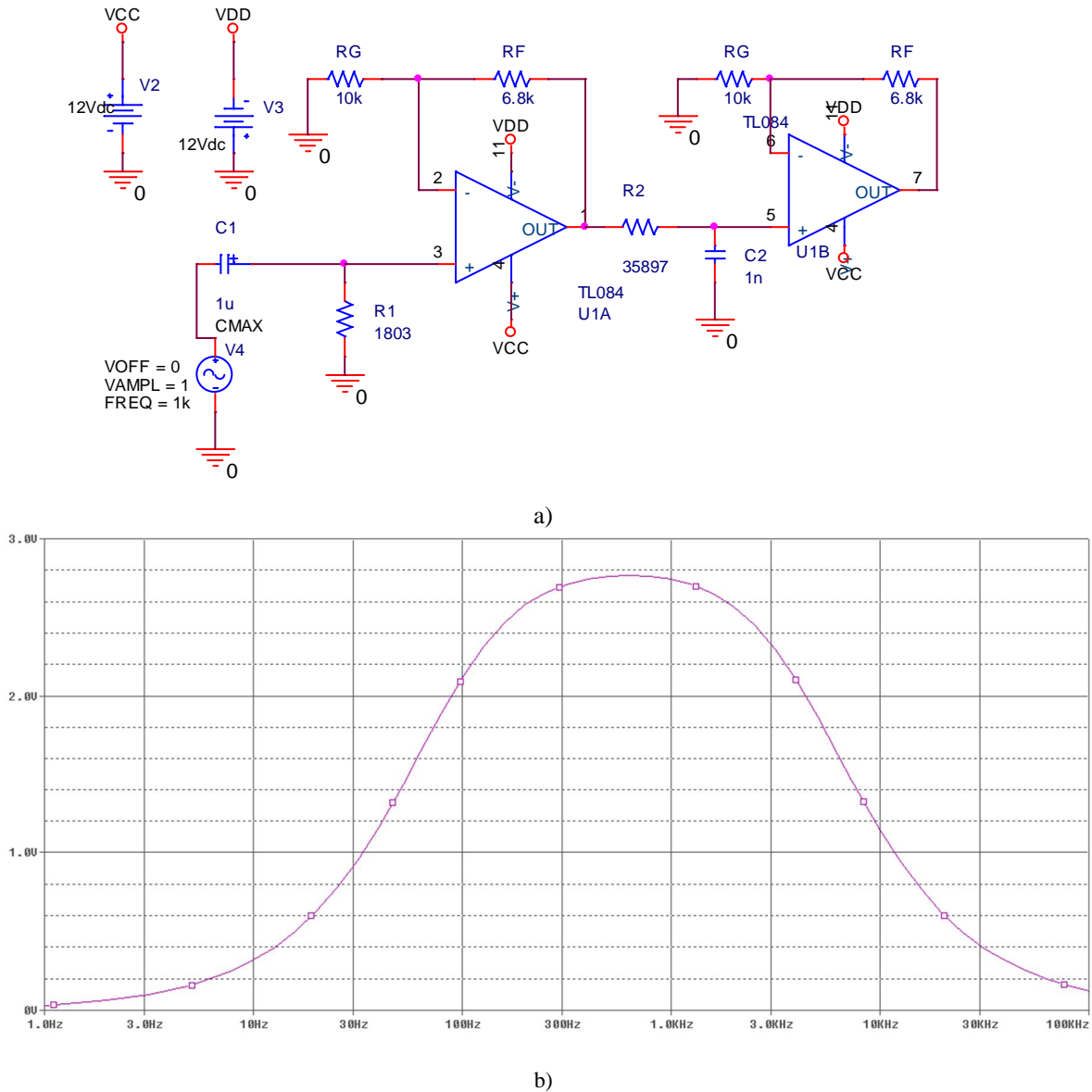


Figura 4. 5 Filtro activo pasabandas de primer orden.

a) Diagrama esquemático del filtro pasabandas diseñado. b) Respuesta en frecuencia del filtro pasabandas acotado a las frecuencias de voz.

Como se puede observar en la figura 4.5 (a) el filtro pasabandas se formó a partir de un filtro pasa-altas en cascada con un filtro pasa-bajas.

La frecuencia de corte inferior está determinada por la frecuencia de corte del filtro pasa-altas, dado por la ecuación 4.1.

$$f_i = \frac{1}{2\pi R_1 C_1}$$

La frecuencia de corte superior está dado por la frecuencia de corte del filtro pasa-bajas dado por la ecuación 4.2

$$f_s = \frac{1}{2\pi R_2 C_2}$$

4.2

La frecuencia central está dada por la ecuación 4.3, siendo ésta de 600 Hz.

$$f_0 = \sqrt{f_s * f_i} = \sqrt{80[Hz] * 4500[Hz]} = 600Hz$$

4.3

El circuito integrado para la elaboración del filtro pasabandas, es el amplificador operacional TL084, el cual se seleccionó porque cumple con los parámetros (ver anexo IV) que se necesitan para el manejo de señales en la frecuencia antes mencionada, sobre todo la relación de producto ganancia-ancho de banda.

El diseño del circuito impreso se muestra en el anexo V.

Procesamiento de las señales

Las señales de voz derivadas del filtro activo pasabandas ingresan a la PC a través de la tarjeta de sonido. La frecuencia de muestreo se estableció en 11025 kHz, ya que es el estándar para grabaciones de voz.

En la figura 4.6 se muestra un diagrama que puntualiza el primer procesamiento aplicado a cada señal, así como las salidas obtenidas, definidas por la energía contenida en cada señal, la frecuencia fundamental de voz (F_0) y la diferencia interaural de tiempo (discreto).

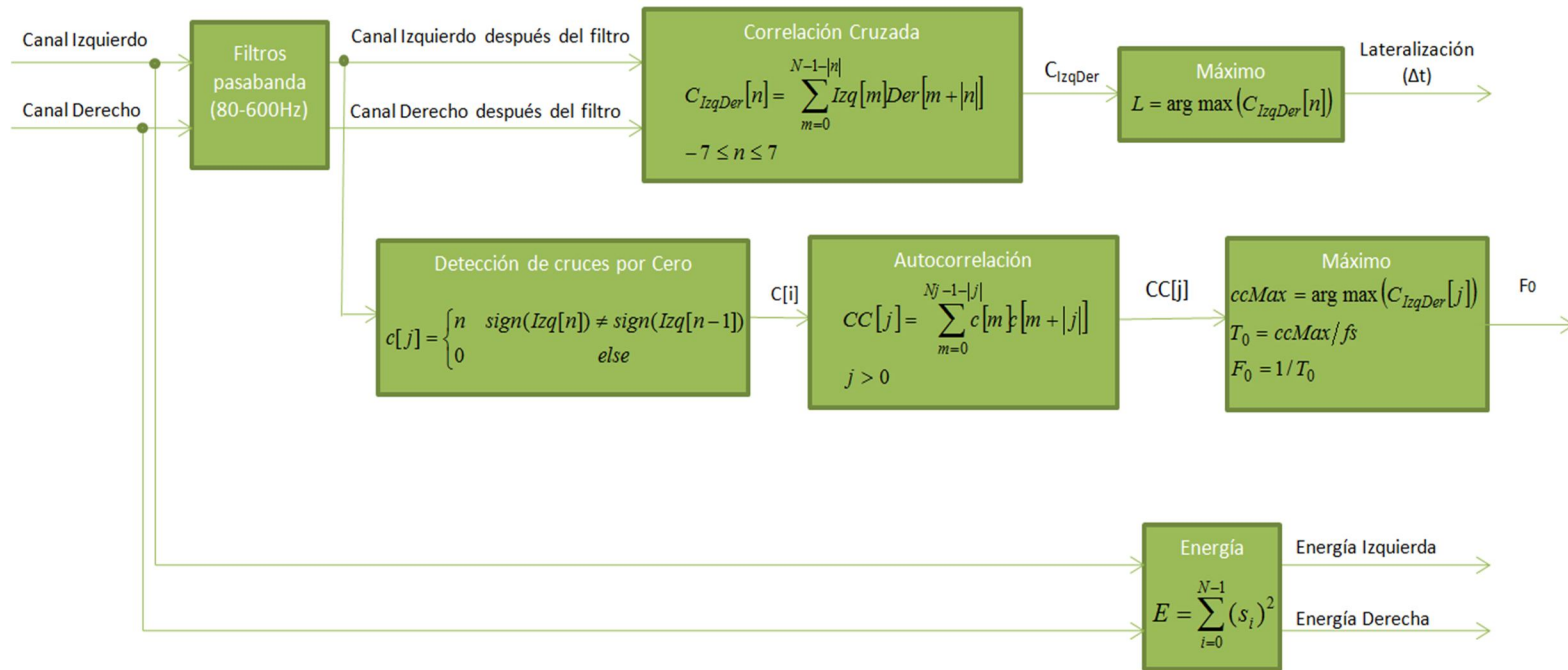


Figura 4. 6 Diagrama a bloques del primer procesamiento en la PC.

Lateralización (Diferencia Interaural de Tiempo)

La primera etapa del reconocimiento de la localización de la fuente sonora considera la Diferencia Interaural de Tiempo (DIT) que para fines de simplicidad se considera el tiempo discreto.

Las DIT pueden calcularse a partir de las diferencias en las distancias que deben recorrer las ondas sonoras (figura 4.7), las cuales serán más grandes mientras más lateralizado se encuentre un sonido.

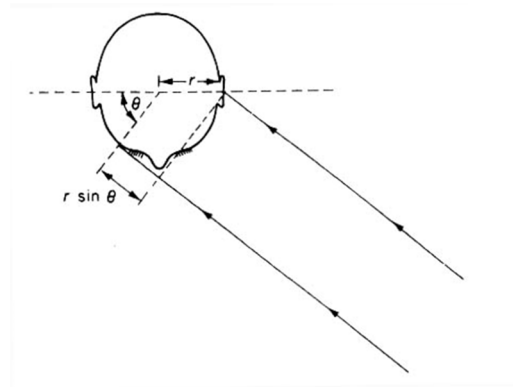


Figura 4. 7 Diferencias en las distancias que deben recorrer las ondas.

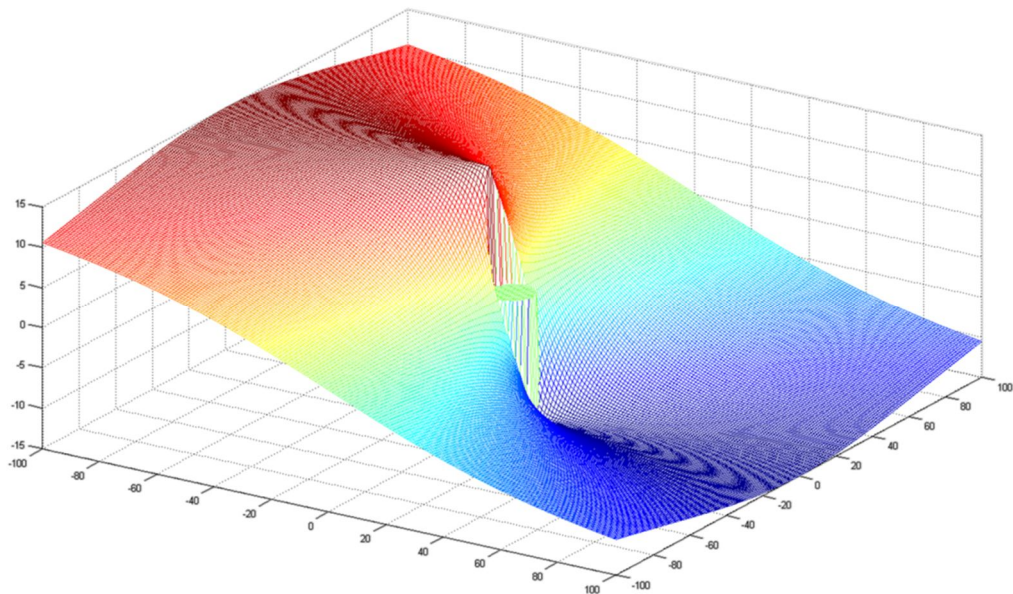


Figura 4. 8 Gráfica que describe la diferencia de tiempo interaural dependiente del ángulo de incidencia.

Una forma de identificar la DIT es mediante el uso de la correlación cruzada entre ambos canales $C_{IzqDer}[n]$ con un corrimiento n acotado en el rango de valores esperados, de tal manera que el valor de n donde $C_{IzqDer}[n]$ es máximo corresponderá a la DIT. El cálculo de la correlación cruzada está dado por la ecuación 4.4.

$$C_{IzqDer}[n] = \sum_{m=0}^{N-1-|n|} Izq[m]Der[m + |n|]$$

Para $-7 \leq n \leq 7$

4.4

Sin embargo, al presentar un sonido más cercano a alguna de las orejas, la modificación de la señal en el oído contralateral, debida a la atenuación que ésta sufre en las altas frecuencias, es tal que la técnica de la correlación cruzada es insuficiente para describir la DIT.

El efecto de la difracción del sonido en los pabellones auriculares es la causa de la atenuación de las altas frecuencias, pero no ocurre así con las frecuencias bajas. Los sonidos de bajas frecuencias tienen longitudes de onda relativamente grandes con respecto a las dimensiones de la cabeza. De acuerdo con (27), el estudio de la difracción determina que cuando la longitud de la onda es suficientemente grande con respecto al obstáculo que encuentra la onda, ésta se difracta fácilmente y no se genera una *sombra acústica* significativa.

Para frecuencias de 570 Hz la longitud de onda del sonido es de unos 60 cm, unas tres veces el diámetro promedio de una cabeza humana. La difracción es poca, por lo que la atenuación de estas frecuencias será despreciable en la segunda oreja.

Por esta razón se diseñó un filtro digital pasabandas de orden 41 con frecuencias de corte en 80 y 600Hz. La grafica 4.9 muestra la ganancia del filtro en la frecuencia.

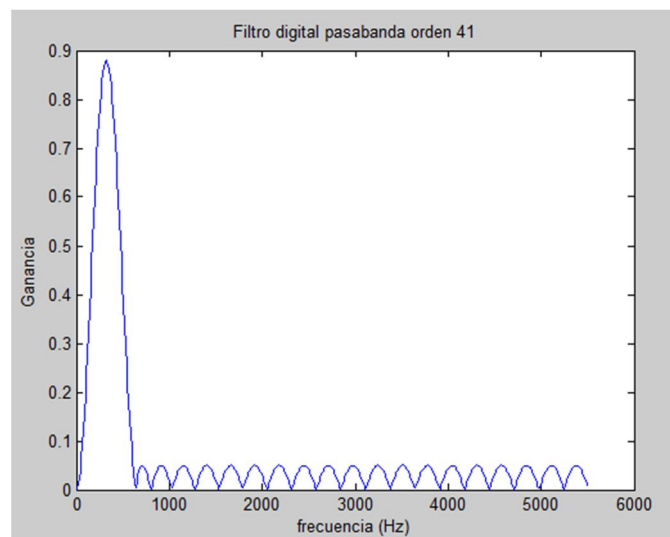


Figura 4.9 Respuesta en frecuencia del filtro digital pasabandas de orden 41 diseñado.

Frecuencia fundamental (F_0)

La frecuencia fundamental es la frecuencia de la excitación que se produce sobre las cuerdas vocales cuando hay voz sonorizada.

Debido a la naturaleza cuasiperiódica de la señal fue posible calcular la frecuencia fundamental por medio de la autocorrelación ponderada, cuya expresión está dada por la ecuación 4.5.

$$CC[n] = k \sum_{m=0}^{N-1-|n|} C[m]C[m + |n|] \tag{4.5}$$

Donde

$$k = E[C[i]] * E[C[j]] \ni \{i \in [0, N - 1 - |n|], \{j \in [m, N - 1]\} \} \tag{4.6}$$

Y

$$E = \sqrt[2]{\text{energía}} \tag{4.7}$$

La detección de n donde $CC[n]$ es máximo, corresponde al periodo fundamental, por lo que es posible calcular la frecuencia fundamental bajo la consideración descrita en la ecuación 4.8:

$$F_0 = 1/T_0 \tag{4.8}$$

Sin embargo, el número de elementos en la autocorrelación de la señal es muy grande, lo que compromete el tiempo disponible de proceso antes de la llegada del siguiente segmento, por lo que se ha propuesto utilizar esta técnica en otra señal generada a partir de los cruces por cero de la señal. La expresión que detecta un cruce por cero es descrita por la ecuación 4.9.

$$\text{cruce por cero} = \begin{cases} 1 & \text{sign}(Izq[n]) \neq \text{sign}(Izq[n - 1]) \\ 0 & \text{para todos los demás casos} \end{cases} \tag{4.9}$$

Entonces generamos la siguiente señal (ecuación 4.10):

$$c[i] = cx0[m] - cx0[m - 1] \tag{4.10}$$

Donde

$$cx0[m] = n \quad \text{sign}(Izq[n]) \neq \text{sign}(Izq[n - 1]) \tag{4.11}$$

Esta técnica reduce considerablemente el costo computacional, ya que convierte a la señal $l_{zq}[n]$ (con 1024 elementos) típicamente en una señal $c[i]$ de 50 elementos o menos.

Cálculo de la energía

El cálculo de la energía de cada canal se obtiene con la ecuación 4.12, que corresponde a la energía de una señal digital.

$$E = \sum_{i=0}^{N-1} s_i^2$$

4. 12

Una vez obtenidos los resultados de los procesamientos descritos anteriormente se procede a realizar lo descrito por la figura 4.10.

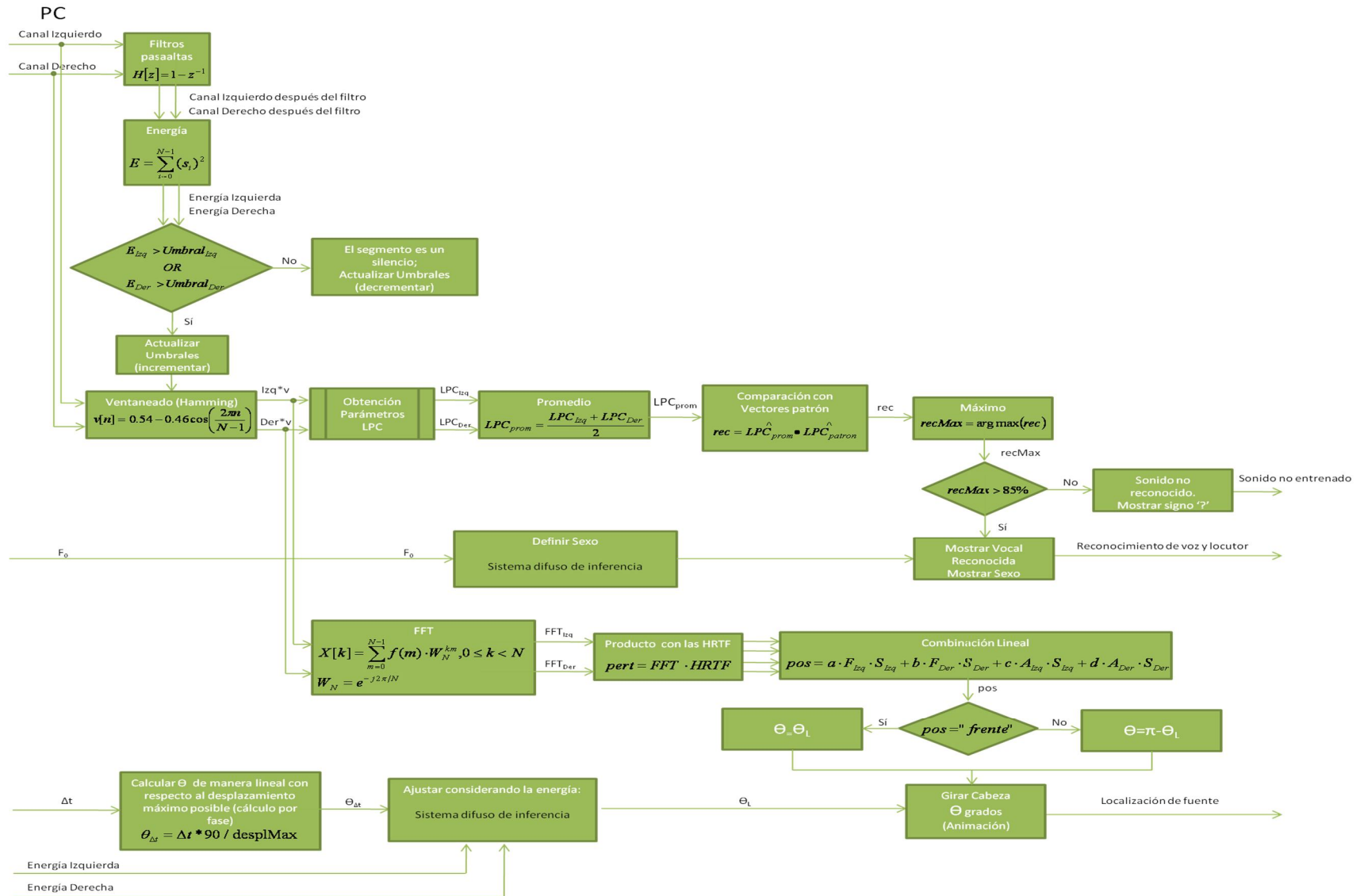


Figura 4. 10 Procesamiento aplicado a la señal cuyos resultados son: reconocimiento de vocales, género y localización de fuente.

Detección de palabra

La detección de inicio y fin de palabras se realiza típicamente tomando en cuenta la actividad de energía y cruce de ceros de la señal S_n , con respecto a los valores que se tienen en condiciones de silencio, ruido ambiente. El comportamiento del aumento de la energía o del cruce de ceros, nos indica que se ha producido una señal. Este aumento en cualquiera de los casos anteriores también puede ser observado en el aumento de la diferencia entre dos muestras consecutivas de la señal. Al sumar los cuadrados de estas diferencias se obtiene un valor susceptible de comparación con un valor de umbral dinámico de tal manera que si éste es superado el segmento se considerará como un segmento de la señal de voz.

Estas operaciones también pueden ser consideradas como el cálculo de la energía de la señal después del filtro pasaaltas de la figura 4.11.

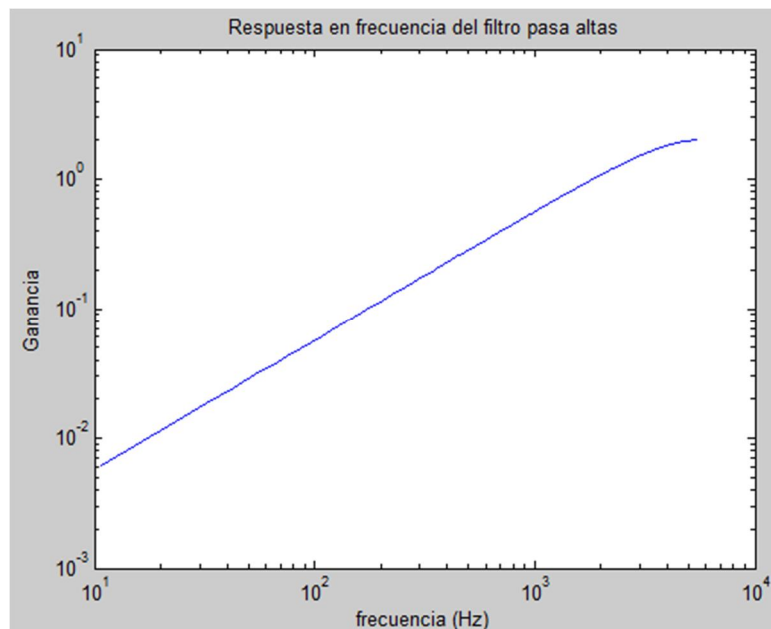


Figura 4. 11 Respuesta del filtro pasaaltas de 1er orden diseñado para un método de detección de palabra.

El umbral de comparación es de carácter dinámico con la finalidad de simular la adaptación del oído a entornos más ruidosos o más silenciosos, de tal manera que ante segmentos que han sido considerados como señales, el umbral crecerá, y ante segmentos de silencio decrecerá de igual manera aunque a un ritmo más acelerado. De esta manera se evita la necesidad de calibrar el sistema para cada equipo y/o entorno.

Ventaneo

Para evitar la aparición de frecuencias no existentes debido a las discontinuidades al principio y fin de cada segmento en la extracción y cálculo de los coeficientes, fue necesario aplicar una ventana no rectangular del mismo tamaño del segmento a analizar. De acuerdo con (12) para el caso de la voz, la ventana de Hamming (figura 4.12) es la más apropiada (ecuación 4.13) y se tuvo esa base para su elección.

$$H_w(n) = 0.54 - \cos\left(\frac{2\pi}{N}n\right)$$

Para $0 \leq n \leq N - 1$

4. 13

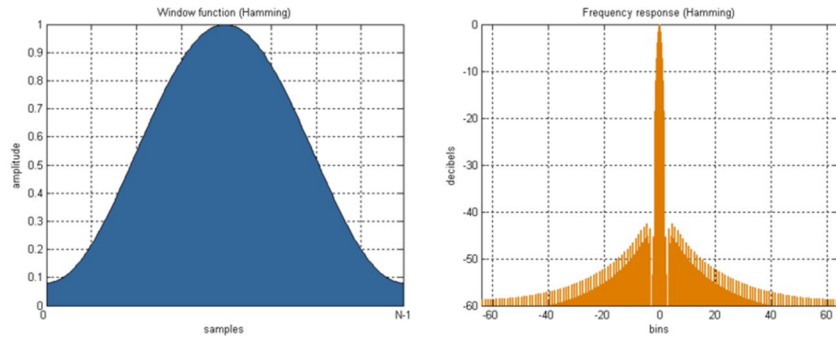


Figura 4. 12 Grafica de la función ventana de Hamming en el tiempo y en frecuencia.

Extracción de parámetros utilizando Coeficientes de Predicción Lineal (LPC)

Para la extracción de parámetros se utilizó el método de LPC descrito por el conjunto de ecuaciones que se exponen en el capítulo 3.

El algoritmo desarrollado toma cada segmento de 10 milisegundos y calcula sus coeficientes. El orden de los coeficientes de predicción se hizo a partir de la frecuencia de muestreo f_s y la ecuación 4.14:

$$p = 4 + \frac{f_s}{1000}$$

4. 14

En las figuras 4.13-4.17 se presentan los LPC obtenidos de las vocales del alfabeto español.

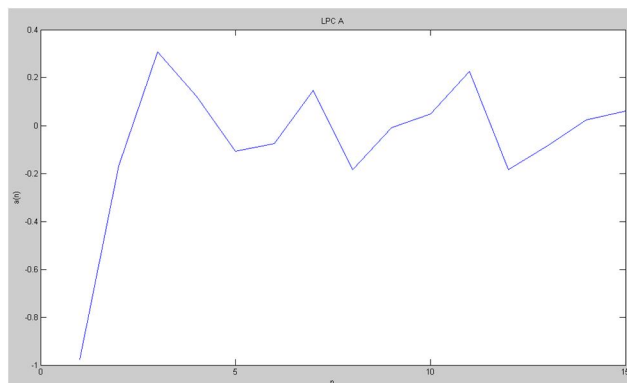


Figura 4. 13 LPC vocal "a"

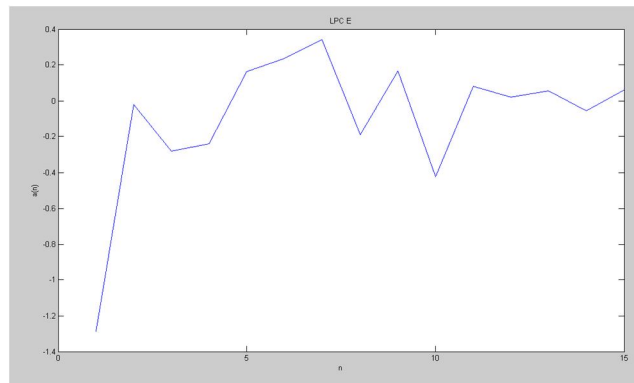


Figura 4. 14 LPC vocal "e"

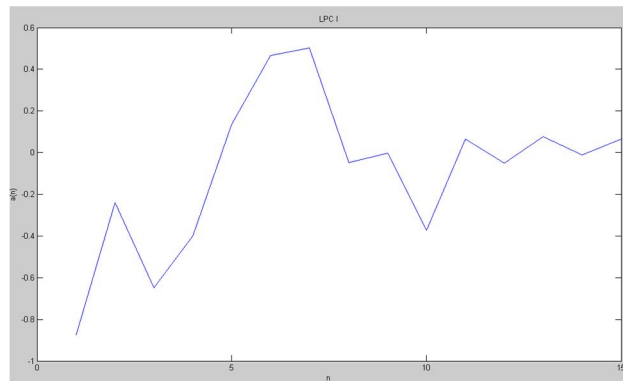


Figura 4. 15 LPC vocal "i"

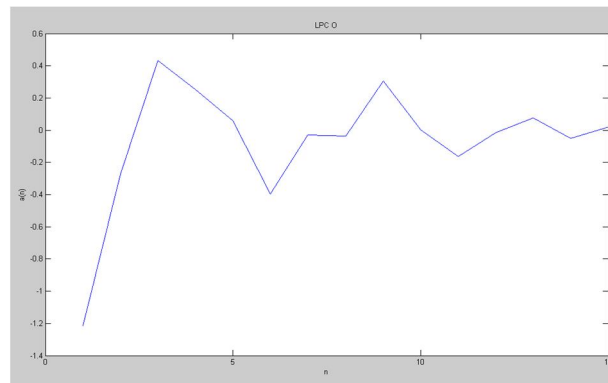


Figura 4. 16 LPC vocal "o"

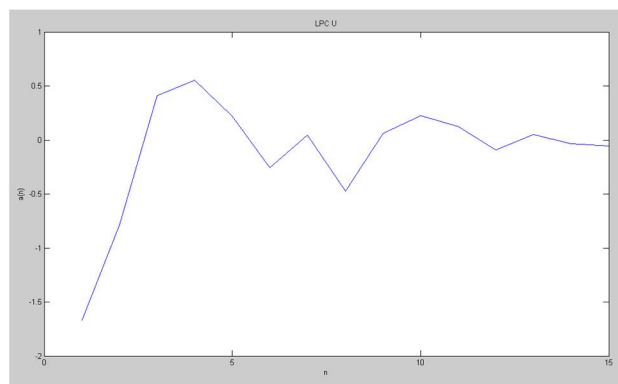


Figura 4. 17 LPC vocal "u"

Identificación del sexo del locutor

La frecuencia fundamental aporta el parámetro de identificación del sexo del locutor. De acuerdo con (12), la frecuencia fundamental de una voz hablada suele variar entre 80 y 300 Hz encontrándose las voces masculinas típicamente en el rango de 80-150Hz, las femeninas en el rango de 150-250 Hz y las voces infantiles hacia los 300 Hz. Estos datos han permitido la construcción del sistema difuso de inferencia mostrado en la figura 4.18.

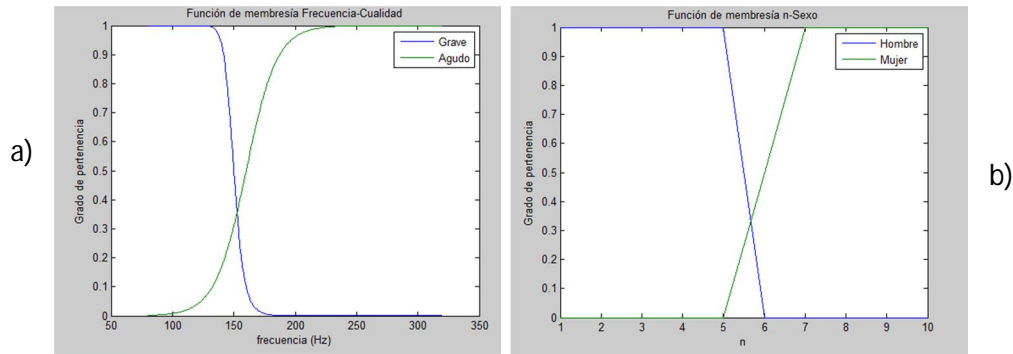


Figura 4. 18 Universo difuso para el reconocimiento del sexo.

a) Variable de entrada Frecuencia-Cualidad. b) Variable de salida n-sexo.

El conjunto *Grave* está comprendido por una función de membresía tipo campana cuya expresión está dada por la ecuación 4.15 con $a=40$, $b=6$ y $c=110$.

$$Campana(x; a, b, c) = \frac{1}{1 + \left| \frac{x - c}{a} \right|^{2b}}$$

4. 15

El conjunto *Agudo* está comprendido por una función de membresía tipo sigmoide cuya expresión está dada por la ecuación 4.16 con $a=.08$ y $c=160$.

$$Sigmoide(x; a, c) = \frac{1}{1 + e^{-a(x-c)}}$$

4. 16

De esta manera se generan las siguientes reglas difusas:

- Si frecuencia es *grave*, entonces sexo igual a *hombre*.
- Si frecuencia es *aguda*, entonces sexo igual a *mujer*.

La defusificación se lleva a cabo por el método de centroide del área, cuya expresión está dada por la ecuación 4.17

$$centroide = \frac{\sum_N i * m(i)}{\sum_N m(i)}$$

4. 17

Localización del sonido

Habiendo considerado las dos técnicas que utiliza el cerebro humano para localizar una fuente sonora, la diferencia interaural de tiempo DIT y la energía de ambos canales E_{Izq} y E_{Der} son los parámetros que permiten la lateralización del sonido, es decir, el ángulo con que incide el sonido a la izquierda o la derecha. Este ángulo es independiente de la posición anterior o posterior de la fuente sonora. Por lo tanto, estos parámetros únicamente permiten la lateralización del sonido. Para modelar este proceso fue definido el sistema difuso de inferencia que arroja la posición de la fuente de sonido como variable de salida en un rango de -90° a $+90^\circ$ (figura 4.19).

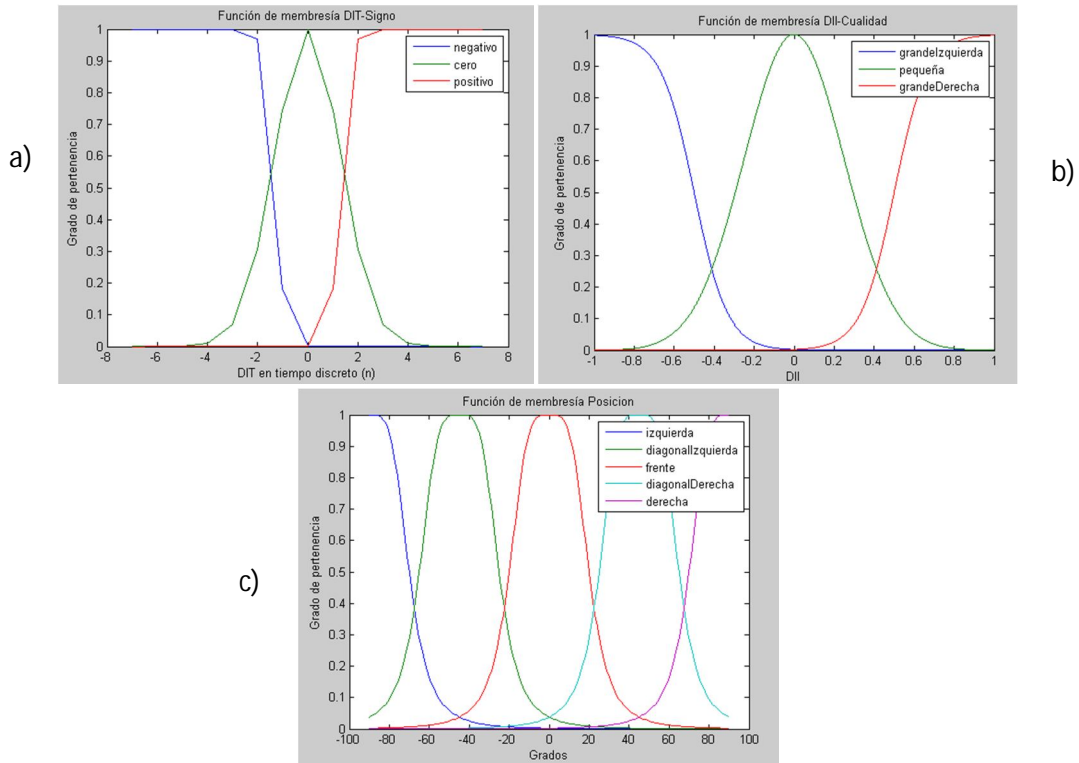


Figura 4. 19 Universo difuso para la localización del sonido.

a) Variable de entrada DIT-Signo. b) Variable de entrada DII-Cualidad. c) Variable de salida Posición.

El conjunto *negativo* está comprendido por una función de membrecía tipo sigmoide (véase ecuación 4.16) con $a=-5$ y $c=-1.3$.

El conjunto *cero* está comprendido por una función de membrecía tipo gaussiana cuya expresión está dada por la ecuación 4.18 con $c=0$ y $\sigma=1.3$.

$$Gaussiana(x; c, \sigma) = e^{-\frac{1}{2} \left(\frac{x-c}{\sigma} \right)^2}$$

4. 18

El conjunto *positivo* está comprendido por una función de membrecía tipo sigmoide (véase ecuación 4.16) con $a=5$ y $c=1.3$.

El conjunto *grandelzquierda* está comprendido por una función de membrecía tipo sigmoide (véase ecuación 4.16) con $a=-12$ y $c=-0.5$.

El conjunto *pequeña* está comprendido por una función de membrecía tipo gaussiana (véase ecuación 4.18) con $c=0$ y $\sigma=0.25$.

El conjunto *grandeDerecha* está comprendido por una función de membrecía tipo sigmoide (véase ecuación 4.16) con $a=12$ y $c=0.5$.

El conjunto *izquierda* está comprendido por una función de membrecía tipo campana (véase ecuación 4.15) con $a=20$, $b=2$ y $c=-90$.

El conjunto *diagonallzquierda* está comprendido por una función de membrecía tipo campana (véase ecuación 4.15) con $a=20$, $b=2$ y $c=-45$.

El conjunto *frente* está comprendido por una función de membrecía tipo campana (véase ecuación 4.15) con $a=20$, $b=2$ y $c=0$.

El conjunto *diagonalDerecha* está comprendido por una función de membrecía tipo campana (véase ecuación 4.15) con $a=20$, $b=2$ y $c=45$.

El conjunto *derecha* está comprendido por una función de membrecía tipo campana (véase ecuación 4.15) con $a=20$, $b=2$ y $c=90$.

De esta manera se generan la siguiente matriz de reglas difusas:

DI\DIT	NEGATIVO	CERO	POSITIVO
<i>grandelzquierda</i>	izquierda	diagonallzquierda	Frente
<i>pequeña</i>	diagonallzquierda	frente	diagonalDerecha
<i>grandeDerecha</i>	Frente	diagonalDerecha	Derecha

La defusificación se lleva a cabo por el método de centroide del área (véase ecuación 4.17).

Para resolver el problema de diferenciar un sonido proveniente de atrás de la cabeza, de uno originado al frente de la misma, se ha propuesto una técnica basada en la modificación que sufre el espectro por efecto de la difracción en la cabeza y orejas en el plano medio, las HRTF (27):

“Las diferencias espectrales entre el sonido original y el sonido medido junto al tímpano, dieron lugar a las HRTF (Head-Related Transfer Functions) o funciones de transferencia relativas a la cabeza.

Las modificaciones espectrales producidas por el pabellón y la cabeza también son usadas por nuestro sistema auditivo para determinar la localización de una fuente sonora. En este caso, es importante que el

sonido tenga energía espectral a lo largo de un amplio rango de frecuencias. Las frecuencias superiores a los 6 kHz son particularmente importantes, dado que es en esa región en la que las longitudes de onda se hacen suficientemente pequeñas como para interactuar eficazmente con el pabellón.

Los distintos picos de resonancia en las HRTF corresponden a diferentes localizaciones de las fuentes sonoras en el plano medio (figura 4.20).



Figura 4. 20 Picos de resonancia en las HRTF corresponden a diferentes localizaciones de las fuentes sonoras en el plano medio

Si se presenta un sonido de banda limitada con frecuencias centrales de 300 Hz o 3 kHz la imagen sonora siempre se formará delante del sujeto. Si la frecuencia central es de 8 kHz la imagen estará siempre arriba. Y si la frecuencia central es de 1 o 10 kHz la imagen se formará siempre detrás. "

Para encontrar las HRTF de al frente y atrás se realizaron grabaciones con ruido blanco desde dichas posiciones y se obtuvo su transformada de Fourier FFT. Las figuras 4.21 y 4.22 muestran las HRTF adquiridas.

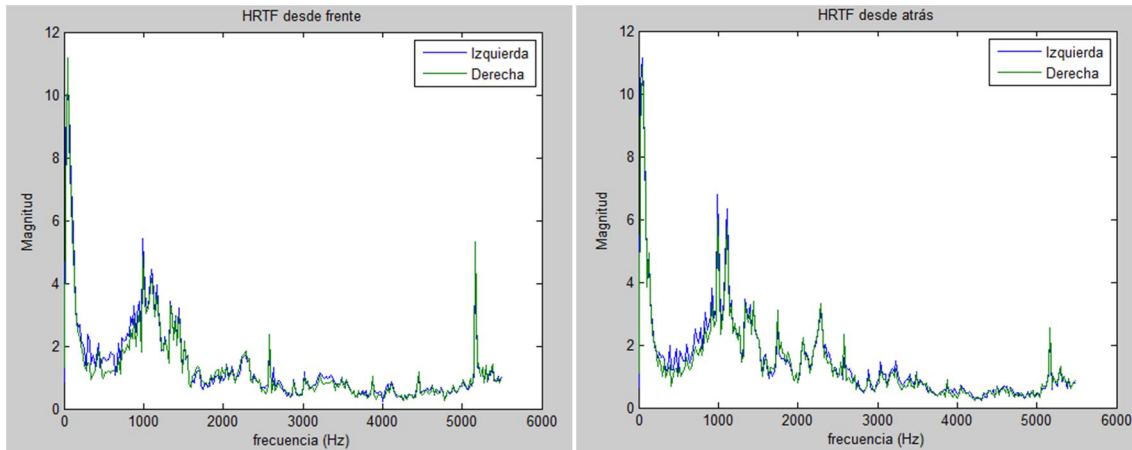


Figura 4. 21 Gráfica de la HRTF de frente y atrás.

Se observa el pico de mayor magnitud en 1kHz para la HRTF de atrás.

De esta manera, al multiplicar el espectro de la señal por las HRTF se obtiene una medida de la pertenencia a estas posiciones. Esta operación resulta en 4 coeficientes (Izquierda-Frente, Derecha-Frente, Izquierda-Atrás, Derecha- Atrás) con los cuales se efectúa una transformación lineal (suma ponderada) para determinar la posición (ecuación 4.19)

$$pos = a \cdot F_{Izq} \cdot S_{Izq} + b \cdot F_{Der} \cdot S_{Der} + c \cdot A_{Izq} \cdot S_{Izq} + d \cdot A_{Der} \cdot S_{Der} \tag{4.19}$$

Reconocimiento de vocales

El reconocimiento de vocales se realiza mediante una red neuronal de competencia tipo ganador toma todo (WTA) la cual tiene como entradas el vector de coeficientes de predicción lineal obtenidos a partir de la señal ventaneada.

Se optó por usar como reconocedor de vocales una red neuronal, ante la posibilidad de otro clasificador más robusto como los Modelos Ocultos de Markov (HMM), dado el tiempo de desarrollo de los algoritmos de reconocimiento y dado que en este caso sólo se reconocerían las vocales del español y no comandos (palabras), para lo cual los HMM son más eficientes.

El número de coeficientes LPC empleados fue 15, debido a la mejor caracterización observada de las señales.

A causa de la existencia de dos señales, y por lo tanto dos conjuntos de coeficientes LPC, se decidió promediar dichos vectores en uno solo para comparar con los vectores prototipo de la red neuronal.

Para estimar el grado de semejanza de los vectores fue utilizado el producto interno como medida de similitud, cuya expresión está dada por la ecuación 3.20.

Para mejorar el reconocimiento de manera multilocutor se ha desarrollado un código demográfico, representando cada vocal mediante un conjunto de unidades o vectores prototipo.

El reconocimiento se lleva a cabo cuando el grado de semejanza es mayor al 85%. Por esta razón es necesario normalizar los vectores prototipo al igual que las entradas de la red neuronal.

Los vectores prototipo, así como las HRTF, fueron almacenados en una base de datos con formato .acddb.

CAPÍTULO 5.RESULTADOS Y DISCUSIÓN

RESULTADOS Y DISCUSIÓN

Interfaz de usuario

La interfaz de usuario es una aplicación informática a pantalla completa que permite el procesamiento de la señal de voz en tiempo real (ver anexo VI). Ésta presenta 3 secciones (figura 5.1):

1. Reconocimiento de vocal. Ubicada en la sección superior izquierda de la pantalla, se encuentra visible únicamente cuando se ha detectado una señal. Muestra la vocal reconocida o un signo de interrogación cuando el grado de semejanza no ha superado el umbral del 85%.
2. Reconocimiento del sexo. Ubicada en la sección superior derecha de la pantalla, se encuentra visible únicamente cuando se ha detectado una señal periódica. Muestra una silueta correspondiente con el sexo del locutor.
3. Localización de la fuente. Ubicada en la sección inferior de la pantalla, se encuentra visible en todo momento. Se trata de una animación 3D con forma de cabeza que girará hacia la posición donde detecte la fuente de sonido.

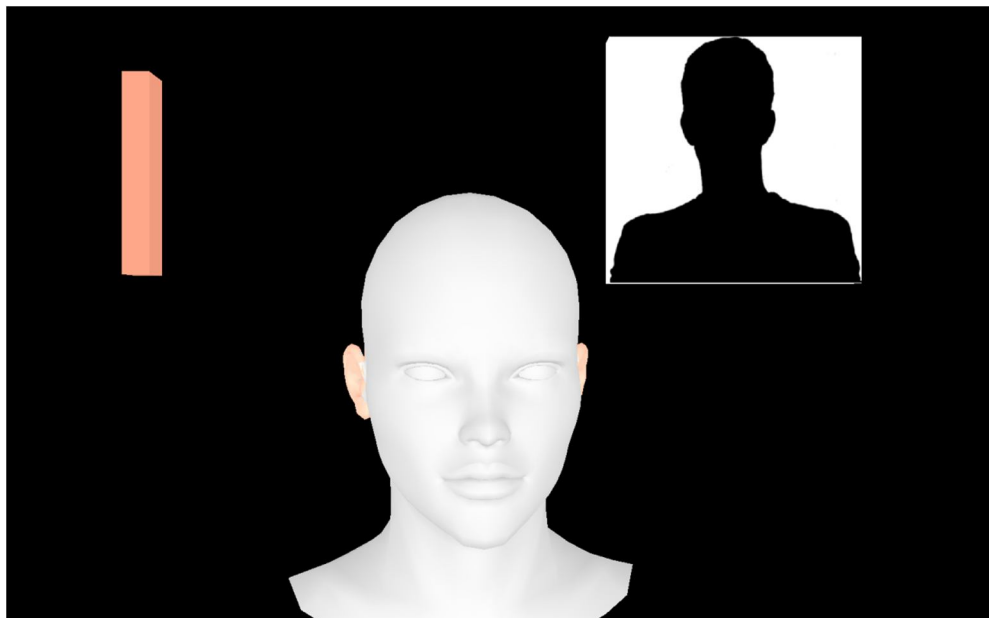


Figura 5. 1 Interfaz de usuario.

Validación del sistema

Para validar el sistema se utilizó un método de validación cruzada llamado k fold cross validation con $k=20$, donde k es el conjunto de validación (véase validación cruzada) y su complemento n ($n=80$) es el conjunto de entrenamiento. De esta forma, se capturaron por persona 10 emisiones de cada vocal en cada posición, siendo las posiciones frente-derecha, frente-izquierda, atrás-izquierda y atrás-derecha, obteniendo un total de 200 archivos de audio por persona, que corresponden, para el caso de una sola vocal:

	FRENTE-DERECHA	FRENTE-IZQUIERDA	ATRÁS-IZQUIERDA	ATRÁS-DERECHA
A	10 EMISIONES	10 EMISIONES	10 EMISIONES	10 EMISIONES

Tabla 5. 1 Número de emisiones de la vocal "a" en las 4 diferentes posiciones.

Por lo tanto dos de cada una de estas emisiones corresponden al conjunto de validación y el resto al conjunto de entrenamiento.

Las emisiones fueron tomadas de un grupo de 6 hombres y 6 mujeres, con las que se obtuvieron 2400 emisiones, con una duración aproximada de 2006 segundos (33.4367 minutos).

Para la presentación de resultados del reconocimiento de vocales se describe la tabla de la siguiente manera (tabla 5.2): en la primera fila se nombra al locutor y se describe su sexo, en la segunda fila se menciona la vocal a verificar. La intersección AI,I (véase X en la tabla 5.2) indica el porcentaje de ocasiones (X%) con que la vocal a verificar en la posición indicada fue reconocida como "I". Finalmente Z indica el porcentaje promedio de reconocimiento, el cual se puede leer como: el porcentaje en que la vocal a verificar fue reconocida fue de Z%.

Nombre del locutor (sexo)					
Vocal a verificar					
	FD	FI	AI	AD	PROMEDIO (%)
A					Z
E					
I			X		
O					
U					

Tabla 5. 2 Tabla donde se ejemplifica el vaciado de resultados

Resultados en el reconocimiento de vocales.

LOCUTOR 1

LOCUTOR 1 (FEMENINO)					
VOCAL A VERIFICAR: <i>a</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	78.5714	88.23	70.58	83.33	80.17785
E	21.4285	5.88	23.52	16.66	16.872125
I	0	0	0	0	0
O	0	0	0	0	0
U	0	0	0	0	0
?	0	5.88	5.88	0	2.94

LOCUTOR 1 (FEMENINO)					
VOCAL A VERIFICAR: <i>e</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	6.25	11.76	20	11.76	9.005
E	93.75	76.47	80	88.23	90.99
I	0	11.76	0	0	0
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 1 (FEMENINO)					
VOCAL A VERIFICAR: <i>i</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	13.33	0	3.3325
I	100	100	86.66	94.11	95.1925
O	0	0	0	5.88	1.47
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 1 (FEMENINO)					
VOCAL A VERIFICAR: O					
	FI	FD	AI	AD	PROMEDIO (%)
A	15	0	19.04	10.52	11.14
E	0	0	0	0	0
I	0	0	0	0	0
O	80	90.47	71.42	78.94	80.2075
U	5	9.52	9.52	10.52	8.64
?	0	0	0	0	0

LOCUTOR 1 (FEMENINO)					
VOCAL A VERIFICAR: u					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	5.55	1.3875
E	0	0	0	0	0
I	0	0	0	0	0
O	0	0	5.55	0	1.3875
U	100	100	94.44	94.44	97.22

LOCUTOR 2

LOCUTOR 2 (MASCULINO)					
VOCAL A VERIFICAR: a					
	FI	FD	AI	AD	PROMEDIO (%)
A	52.63	100	100	100	88.1575
E	0	0	0	0	0
I	0	0	0	0	0
O	47.37	0	0	0	11.8425
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 2 (MASCULINO)					
VOCAL A VERIFICAR: <i>e</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	42.85	66.66	100	100	77.3775
I	33.33	28.57	0	0	15.475
O	0	0	0	0	0
U	0	0	0	0	0
?	23.8	4.76	0	0	7.14

LOCUTOR 2 (MASCULINO)					
VOCAL A VERIFICAR: <i>i</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	0	0	0
I	100	100	100	100	100
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 2 (MASCULINO)					
VOCAL A VERIFICAR: <i>o</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	4	0	56.25	0	15.0625
E	0	0	0	0	0
I	0	0	0	0	0
O	72	100	43.75	93.54	77.3225
U	0	0	0	6.45	1.6125
?	24	0	0	0	6

LOCUTOR 2 (MASCULINO)					
VOCAL A VERIFICAR: <i>u</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	53.125	6.25	14.84375
E	0	0	0	0	0
I	0	0	0	0	0
O	13.33	30.3	43.75	81.25	42.1575
U	86.66	69.69	3.125	12.5	42.99375
?	0	0	0	0	0

LOCUTOR 3

LOCUTOR 3 (MASCULINO)					
VOCAL A VERIFICAR: <i>a</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	100	100	96.55	100	99.1375
E	0	0	0	0	0
I	0	0	0	0	0
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	3.45	0	0.8625

LOCUTOR 3 (MASCULINO)					
VOCAL A VERIFICAR: <i>e</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	100	100	100	69.23	92.3075
I	0	0	0	0	0
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	30.76	7.69

LOCUTOR 3 (MASCULINO)					
VOCAL A VERIFICAR: <i>i</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	22.58	28.57	89	48.27	47.105
I	67.74	71.42	10	41.37	47.6325
O	0	0	0	0	0
U	0	0	0	0	0
?	9.67	0	0	10.34	5.0025

LOCUTOR 3 (MASCULINO)					
VOCAL A VERIFICAR: <i>o</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	40	21.42	57.57	68.75	46.935
E	0	0	0	0	0
I	0	0	0	0	0
O	50	78.57	33.33	31.25	48.2875
U	0	0	0	0	0
?	10	0	9.09	0	4.7725

LOCUTOR 3 (MASCULINO)					
VOCAL A VERIFICAR: <i>u</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	3.45	25	7.1125
I	0	0	0	0	0
O	0	7.14	0	7.14	3.57
U	100	92.85	96.55	67.85	89.3125
?	0	0	0	0	0

LOCUTOR 4

LOCUTOR 4 (FEMENINO)					
VOCAL A VERIFICAR: <i>a</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	100	100	100	100	100
E	0	0	0	0	0
I	0	0	0	0	0
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 4 (FEMENINO)					
VOCAL A VERIFICAR: <i>e</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	86.66	93.75	85.72	100	91.5325
I	13.33	6.25	14.28	0	8.465
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 4 (FEMENINO)					
VOCAL A VERIFICAR: <i>i</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	0	0	0
I	94.11	100	100	100	98.5275
O	0	0	0	0	0
U	0	0	0	0	0
?	5.88	0	0	0	1.47

LOCUTOR 4 (FEMENINO)					
VOCAL A VERIFICAR: <i>o</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	26.31	4.76	9.09	5.26	11.355
E	0	0	0	0	0
I	0	0	0	0	0
O	57.89	61.9	86.36	94.74	75.2225
U	15.78	33.33	4.54	0	13.4125
?	0	0	0	0	0

LOCUTOR 4 (FEMENINO)					
VOCAL A VERIFICAR: <i>u</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	29.41	12.5	23.07	75	34.995
E	0	0	0	0	0
I	0	0	0	0	0
O	17.64	18.75	23.07	0	14.865
U	52.94	68.75	53.84	25	50.1325
?	0	0	0	0	0

LOCUTOR 5

LOCUTOR 5 (MASCULINO)					
VOCAL A VERIFICAR: <i>a</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	56	66.66	80	93.93	74.1475
E	0	0	15	3.03	4.5075
I	0	0	0	0	0
O	44	33.34	5	3.03	21.3425
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 5 (MASCULINO)					
VOCAL A VERIFICAR: <i>e</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	100	100	94.73	100	98.6825
I	0	0	0	0	0
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	5.26	0	1.315

LOCUTOR 5 (MASCULINO)					
VOCAL A VERIFICAR: <i>i</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	7.69	18.18	6.4675
I	100	100	84.61	78.78	90.8475
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	7.69	3.04	2.6825

LOCUTOR 5 (MASCULINO)					
VOCAL A VERIFICAR: <i>o</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	0	0	0
I	0	0	0	0	0
O	25	87.5	100	100	78.125
U	75	8.33	0	0	20.8325
?	0	4.16	0	0	1.04

LOCUTOR 5 (MASCULINO)					
VOCAL A VERIFICAR: <i>u</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	0	0	0
I	0	0	0	0	0
O	0	0	0	0	0
U	100	100	100	100	100
?	0	0	0	0	0

LOCUTOR 6

LOCUTOR 6 (FEMENINO)					
VOCAL A VERIFICAR: <i>a</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	81.81	60	91.66	100	83.3675
E	0	0	0	0	0
I	0	0	0	0	0
O	18.18	40	8.33	0	16.6275
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 6 (FEMENINO)					
VOCAL A VERIFICAR: <i>e</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	27.23	20	81.81	80	52.26
I	72.72	80	18.19	20	47.7275
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 6 (FEMENINO)					
VOCAL A VERIFICAR: <i>i</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	16.66	0	0	4.165
I	100	83.33	100	100	95.8325
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 6 (FEMENINO)					
VOCAL A VERIFICAR: <i>o</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	0	0	0
I	0	0	0	0	0
O	90	1	90	70	62.75
U	10	0	10	30	12.5
?	0	0	0	0	0

LOCUTOR 6 (FEMENINO)					
VOCAL A VERIFICAR: <i>u</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	0	0	0
I	0	0	0	0	0
O	25	0	0	0	6.25
U	75	100	100	100	93.75
?	0	0	0	0	0

LOCUTOR 7

LOCUTOR 7 (MASCULINO)					
VOCAL A VERIFICAR: <i>a</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	85.71	100	61.9	58.33	76.485
E	0	0	0	0	0
I	0	0	0	0	0
O	14.28	0	38.1	29.16	20.385
U	0	0	0	0	0
?	0	0	0	12.5	3.125

LOCUTOR 7 (MASCULINO)					
VOCAL A VERIFICAR: <i>e</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	100	70.37	96.55	100	91.73
I	0	29.63	3.44	0	8.2675
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 7 (MASCULINO)					
VOCAL A VERIFICAR: <i>i</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	0	0	0
I	100	90.9	100	80.76	92.915
O	0	0	0	0	0
U	0	0	0	0	0
?	0	9.09	0	19.24	7.0825

LOCUTOR 7 (MASCULINO)					
VOCAL A VERIFICAR: <i>o</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	17.85	0	77.41	0	23.815
E	0	0	0	0	0
I	0	0	0	0	0
O	75	100	19.35	87.5	70.4625
U	0	0	3.22	12.5	3.93
?	7.14	0	0	0	1.785

LOCUTOR 7 (MASCULINO)					
VOCAL A VERIFICAR: <i>u</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	3.57	36.66	10.0575
E	0	0	0	0	0
I	0	0	0	0	0
O	80.76	69.56	42.85	20	53.2925
U	19.23	30.43	53.57	43.33	36.64
?	0	0	0	0	0

LOCUTOR 8

LOCUTOR 8 (MASCULINO)					
VOCAL A VERIFICAR: <i>a</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	86.66	87.5	95	93.75	90.7275
E	0	0	0	0	0
I	0	0	0	0	0
O	13.34	6.25	5	6.25	7.71
U	0	6.25	0	0	1.5625
?	0	0	0	0	0

LOCUTOR 8 (MASCULINO)					
VOCAL A VERIFICAR: <i>e</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	93.75	100	100	100	98.4375
I	6.25	0	0	0	1.5625
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 8 (MASCULINO)					
VOCAL A VERIFICAR: <i>i</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	0	0	0
I	100	100	100	100	100
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 8 (MASCULINO)					
VOCAL A VERIFICAR: <i>o</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	0	0	0
I	0	0	0	0	0
O	100	93.33	88.23	100	95.39
U	0	6.66	11.76	0	4.605
?	0	0	0	0	0

LOCUTOR 8 (MASCULINO)					
VOCAL A VERIFICAR: <i>u</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	6.66	0	6.66	3.33
E	0	0	0	0	0
I	0	0	0	0	0
O	0	60	87.5	73.33	55.2075
U	100	33.33	12.5	20	41.4575
?	0	0	0	0	0

LOCUTOR 9

LOCUTOR 9 (MASCULINO)					
VOCAL A VERIFICAR: <i>a</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	6.67	29.03	63.33	30	32.2575
E	0	0	0	0	0
I	0	0	0	0	0
O	76.66	67.74	36.67	63.33	61.1
U	16.66	3.22	0	6.67	6.6375
?	0	0	0	0	0

LOCUTOR 9 (MASCULINO)					
VOCAL A VERIFICAR: <i>e</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	100	96.42	100	92	97.105
I	0	3.58	0	8	2.895
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 9 (MASCULINO)					
VOCAL A VERIFICAR: <i>i</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	3.57	0	0.8925
I	100	100	96.43	100	99.1075
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 9 (MASCULINO)					
VOCAL A VERIFICAR: <i>o</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	3.44	0	0.86
E	0	0	0	0	0
I	0	0	0	0	0
O	100	96.66	96.56	96.66	97.47
U	0	0	0	3.34	0.835
?	0	3.33	0	0	0.8325

LOCUTOR 9 (MASCULINO)					
VOCAL A VERIFICAR: <i>u</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	4	0	0	0	1
E	0	0	0	0	0
I	0	0	0	0	0
O	84	56.66	56.66	42.3	59.905
U	12	43.33	43.34	57.7	39.0925
?	0	0	0	0	0

LOCUTOR 10

LOCUTOR 10 (FEMENINO)					
VOCAL A VERIFICAR: <i>a</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	28.58	33.34	41.66	23.52	31.775
E	71.42	66.66	58.34	76.48	68.225
I	0	0	0	0	0
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 10 (FEMENINO)					
VOCAL A VERIFICAR: <i>e</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	2.5	0.625
E	100	100	100	97.5	99.375
I	0	0	0	0	0
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 10 (FEMENINO)					
VOCAL A VERIFICAR: <i>i</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	8.1	0	5	10.25	5.8375
I	91.89	100	87.5	82.05	90.36
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	7.5	7.69	3.7975

LOCUTOR 10 (FEMENINO)					
VOCAL A VERIFICAR: <i>o</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	40.47	64.1	32.5	58.53	48.9
E	0	0	0	0	0
I	0	0	0	0	0
O	57.14	35.89	67.5	41.46	50.4975
U	2.39	0	0	0	0.5975
?	0	0	0	0	0

LOCUTOR 10 (FEMENINO)					
VOCAL A VERIFICAR: <i>u</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	0	0	0
I	0	0	0	0	0
O	0	0	0	0	0
U	100	100	100	100	100
?	0	0	0	0	0

LOCUTOR 11

LOCUTOR 11 (FEMENINO)					
VOCAL A VERIFICAR: <i>a</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	50	55	68	25	49.5
E	50	45	32	75	50.5
I	0	0	0	0	0
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 11 (FEMENINO)					
VOCAL A VERIFICAR: <i>e</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	13.04	28.57	66.66	95.23	50.875
I	86.95	71.42	33.33	4.76	49.115
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 11 (FEMENINO)					
VOCAL A VERIFICAR: <i>i</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	0	0	0	0	0
I	100	80.95	100	100	95.2375
O	0	0	0	0	0
U	0	0	0	0	0
?	0	19.04	0	0	4.76

LOCUTOR 11 (FEMENINO)					
VOCAL A VERIFICAR: <i>o</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	22.72	0	0	0	5.68
E	0	0	0	0	0
I	0	0	0	0	0
O	68.18	81.81	100	100	87.4975
U	9.1	18.19	0	0	6.8225
?	0	0	0	0	0

LOCUTOR 11 (FEMENINO)					
VOCAL A VERIFICAR: <i>u</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	25	0	25	0	12.5
E	5	0	0	0	1.25
I	0	0	0	0	0
O	15	20	5	22.72	15.68
U	50	80	70	77.28	69.32
?	5	0	0	0	1.25

LOCUTOR 12

LOCUTOR 12 (FEMENINO)					
VOCAL A VERIFICAR: <i>a</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	100	94.44	100	100	98.61
E	0	5.56	0	0	1.39
I	0	0	0	0	0
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 12 (FEMENINO)					
VOCAL A VERIFICAR: <i>e</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	10	2.5
E	95.23	95	90.9	90	92.7825
I	4.77	5	9.1	0	4.7175
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 12 (FEMENINO)					
VOCAL A VERIFICAR: <i>i</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	0	0	0	0	0
E	6.25	5	0	0	2.8125
I	93.75	95	100	100	97.1875
O	0	0	0	0	0
U	0	0	0	0	0
?	0	0	0	0	0

LOCUTOR 12 (FEMENINO)					
VOCAL A VERIFICAR: <i>o</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	68.42	5	20	47.36	35.195
E	5.26	0	0	0	1.315
I	0	0	0	0	0
O	26.31	45	70	42.1	45.8525
U	0	45	10	10.52	16.38
?	0	5	0	0	1.25

LOCUTOR 12 (FEMENINO)					
VOCAL A VERIFICAR: <i>u</i>					
	FI	FD	AI	AD	PROMEDIO (%)
A	23.08	12.5	0	35.2	17.695
E	0	0	0	0	0
I	0	0	0	0	0
O	23.08	25	36.84	0	21.23
U	53.84	62.5	63.15	64.7	61.0475
?	0	0	0	0	0

Evaluación del sistema

EVALUACIÓN DEL SISTEMA PARA HOMBRES					
NOMBRE DEL LOCUTOR	VOCAL A RECONOCER				
	A	E	I	O	U
LOCUTOR 2 (MASCULINO)	88.1575	77.3775	100	77.3225	42.99375
LOCUTOR 3 (MASCULINO)	99.1375	92.3075	47.6325	48.2875	89.3125
LOCUTOR 5 (MASCULINO)	74.1475	98.6825	90.8475	78.125	100
LOCUTOR 7 (MASCULINO)	76.485	91.73	92.915	70.4625	36.64
LOCUTOR 8 (MASCULINO)	90.7275	98.4375	100	95.39	41.4575
LOCUTOR 9 (MASCULINO)	32.2575	97.105	99.1075	97.47	39.0925
% DE RECONOCIMIENTO	76.81875	92.6066667	88.4170833	77.8429167	58.249375

EVALUACIÓN DEL SISTEMA PARA MUJERES					
NOMBRE DEL LOCUTOR	VOCAL A RECONOCER				
	A	E	I	O	U
LOCUTOR 1 (FEMENINO)	80.17785	90.99	95.1925	80.2075	97.22
LOCUTOR 4 (FEMENINO)	100	91.5325	98.5275	75.2225	50.1325
LOCUTOR 6 (FEMENINO)	83.3675	52.26	95.8325	62.75	93.75
LOCUTOR 10 (FEMENINO)	31.775	99.375	90.36	50.4975	100
LOCUTOR 11 (FEMENINO)	49.5	50.875	95.2375	87.4975	69.32
LOCUTOR 12 (FEMENINO)	98.61	92.7825	97.1875	45.8525	61.0475
% DE RECONOCIMIENTO	73.9050583	79.6358333	95.3895833	67.0045833	78.5783333

EVALUACIÓN DEL SISTEMA					
NOMBRE DEL LOCUTOR	VOCAL A RECONOCER				
	A	E	I	O	U
LOCUTOR 1 (FEMENINO)	80.17785	90.99	95.1925	80.2075	97.22
LOCUTOR 2 (MASCULINO)	88.1575	77.3775	100	77.3225	42.99375
LOCUTOR 3 (MASCULINO)	99.1375	92.3075	47.6325	48.2875	89.3125
LOCUTOR 4 (FEMENINO)	100	91.5325	98.5275	75.2225	50.1325
LOCUTOR 5 (MASCULINO)	74.1475	98.6825	90.8475	78.125	100
LOCUTOR 6 (FEMENINO)	83.3675	52.26	95.8325	62.75	93.75
LOCUTOR 7 (MASCULINO)	76.485	91.73	92.915	70.4625	36.64
LOCUTOR 8 (MASCULINO)	90.7275	98.4375	100	95.39	41.4575
LOCUTOR 9 (MASCULINO)	32.2575	97.105	99.1075	97.47	39.0925
LOCUTOR 10 (FEMENINO)	31.775	99.375	90.36	50.4975	100
LOCUTOR 11 (FEMENINO)	49.5	50.875	95.2375	87.4975	69.32
LOCUTOR 12 (FEMENINO)	98.61	92.7825	97.1875	45.8525	61.0475
% DE RECONOCIMIENTO	75.3619042	86.12125	91.9033333	72.42375	68.4138542

Resultados para el reconocimiento de sexo

LOCUTOR 1 (FEMENINO)		
	MUJER	HOMBRE
A	39.11	60.89
E	60.4525	39.5475
I	95.67	4.33
O	52.0475	47.9525
U	94.19	5.81
PROMEDIO	68.294	31.706

LOCUTOR 2 (MASCULINO)		
	MUJER	HOMBRE
A	16.1375	83.8625
E	62.0025	37.9975
I	39.6	60.4
O	15.5325	84.4675
U	97.63	2.37
PROMEDIO	46.1805	53.8195

LOCUTOR 3 (MASCULINO)		
	MUJER	HOMBRE
A	12.465	87.535
E	40.825	59.175
I	73.745	26.255
O	11.2925	88.7075
U	75.0625	24.9375
PROMEDIO	42.678	57.322

LOCUTOR 4 (FEMENINO)		
	MUJER	HOMBRE
A	29.0625	70.9375
E	50.4375	49.5625
I	88.8025	11.1975
O	64.6975	35.3025
U	87.255	12.745
PROMEDIO	64.051	35.949

LOCUTOR 5(MASCULINO)		
	MUJER	HOMBRE
A	12.94	87.06
E	31.47	68.53
I	26.31	73.69
O	43.9	56.1
U	68.6275	31.3725
PROMEDIO	36.6495	63.3505

LOCUTOR 6 (FEMENINO)		
	MUJER	HOMBRE
A	16.1725	83.8275
E	61.82	38.18
I	46.3675	53.6325
O	45	55
U	77.0825	22.9175
PROMEDIO	49.2885	50.7115

LOCUTOR 7 (MASCULINO)		
	MUJER	HOMBRE
A	9.765	90.235
E	8.565	91.435
I	83.1075	16.8925
O	15.5125	84.4875
U	91.545	8.455
PROMEDIO	41.699	58.301

LOCUTOR 8 (MASCULINO)		
	MUJER	HOMBRE
A	25.8325	74.1675
E	50.8925	49.1075
I	17.41	82.59
O	62.615	37.385
U	66.04	33.96
PROMEDIO	44.558	55.442

LOCUTOR 9 (MASCULINO)		
	MUJER	HOMBRE
A	15.7	84.3
E	43.1825	56.8175
I	24.0075	75.9925
O	62.49	37.51
U	22.6775	77.3225
PROMEDIO	33.6115	66.3885

LOCUTOR 10 (FEMENINO)		
	MUJER	HOMBRE
A	35.5925	64.4075
E	71.075	28.925
I	91.74	8.26
O	63.775	36.225
U	94.7175	5.2825
PROMEDIO	71.38	28.62

LOCUTOR 11 (FEMENINO)		
	MUJER	HOMBRE
A	46.475	53.525
E	41.51	58.49
I	88.365	11.635
O	66.135	33.865
U	96.75	3.25
PROMEDIO	67.847	32.153

LOCUTOR 12 (FEMENINO)		
	MUJER	HOMBRE
A	60.405	39.595
E	62.185	37.815
I	88.8425	11.1575
O	74.21	25.79
U	97.06	2.94
PROMEDIO	76.5405	23.4595

En la siguiente evaluación del sistema para el reconocimiento de sexo, se resume el porcentaje de las veces en que éste acertó, para ambos casos a reconocer.

EVALUACIÓN DEL SISTEMA	
SEXO A RECONOCER	% DE RECONOCIMIENTO
HOMBRE	59.60175
MUJER	66.2335

Resultados de localización (lateralización)

LOCUTOR 1 (FEMENINO)			
FRENTE IZQUIERDA			
	Izquierda	Centro	Derecha
	69.046	9.138	21.816

LOCUTOR 1 (FEMENINO)		
FRENTE DERECHA		
Izquierda	Centro	Derecha
16.518	4.574	78.908

LOCUTOR 1 (FEMENINO)		
ATRÁS DERECHA		
Izquierda	Centro	Derecha
1.112	1.176	97.712

LOCUTOR 1 (FEMENINO)		
ATRÁS IZQUIERDA		
Izquierda	Centro	Derecha
82.122	14.482	3.398

LOCUTOR 2 (MASCULINO)		
FRENTE IZQUIERDA		
Izquierda	Centro	Derecha
82.694	3.622	13.684

LOCUTOR 2 (MASCULINO)		
FRENTE DERECHA		
Izquierda	Centro	Derecha
2	1	97

LOCUTOR 2 (MASCULINO)		
ATRÁS DERECHA		
Izquierda	Centro	Derecha
0	0	100

LOCUTOR 2 (MASCULINO)		
ATRÁS IZQUIERDA		
Izquierda	Centro	Derecha
83.342	2.678	13.982

LOCUTOR 3 (MASCULINO)		
FRENTE IZQUIERDA		
Izquierda	Centro	Derecha
69.95	19.312	10.738

LOCUTOR 3 (MASCULINO)		
FRENTE DERECHA		
Izquierda	Centro	Derecha
8.02	2.168	89.812

LOCUTOR 3 (MASCULINO)		
ATRÁS DERECHA		
Izquierda	Centro	Derecha
2.646	1.876	95.48

LOCUTOR 3 (MASCULINO)		
ATRÁS IZQUIERDA		
Izquierda	Centro	Derecha
88.394	7.526	4.08

LOCUTOR 4 (FEMENINO)		
FRENTE IZQUIERDA		
Izquierda	Centro	Derecha
57.368	26.342	16.288

LOCUTOR 4 (FEMENINO)		
FRENTE DERECHA		
Izquierda	Centro	Derecha
2.858	0	97.142

LOCUTOR 4 (FEMENINO)		
ATRÁS DERECHA		
Izquierda	Centro	Derecha
13.066	1.666	85.268

LOCUTOR 4 (FEMENINO)		
ÁTRAS IZQUIERDA		
Izquierda	Centro	Derecha
86.188	4.394	9.414

LOCUTOR 5 (MASCULINO)		
FRENTE IZQUIERDA		
Izquierda	Centro	Derecha
64.794	21.064	14.142

LOCUTOR 5 (MASCULINO)		
FRENTE DERECHA		
Izquierda	Centro	Derecha
6.36	3.384	90.256

LOCUTOR 5 (MASCULINO)		
ATRAS DERECHA		
Izquierda	Centro	Derecha
2.626	1.606	95.77

LOCUTOR 5 (MASCULINO)		
ATRAS IZQUIERDA		
Izquierda	Centro	Derecha
87.858	5.538	6.602

LOCUTOR 6 (FEMENINO)		
FRENTE IZQUIERDA		
Izquierda	Centro	Derecha
70.182	20	9.818

LOCUTOR 6 (FEMENINO)		
FRENTE DERECHA		
Izquierda	Centro	Derecha
3.666	0	96.334

LOCUTOR 6 (FEMENINO)		
ATRAS DERECHA		
Izquierda	Centro	Derecha
0	0	100

LOCUTOR 6 (FEMENINO)		
ATRAS IZQUIERDA		
Izquierda	Centro	Derecha
94.848	0	5.152

LOCUTOR 7 (MASCULINO)		
FRENTE IZQUIERDA		
Izquierda	Centro	Derecha
86.544	8.828	4.628

LOCUTOR 7 (MASCULINO)		
FRENTE DERECHA		
Izquierda	Centro	Derecha
4.316	0	95.684

LOCUTOR 7 (MASCULINO)		
ATRAS DERECHA		
Izquierda	Centro	Derecha
3.062	2.27	94.672

LOCUTOR 7 (MASCULINO)		
ATRAS IZQUIERDA		
Izquierda	Centro	Derecha
92.88	1.904	5.214

LOCUTOR 8 (MASCULINO)		
FRENTE IZQUIERDA		
Izquierda	Centro	Derecha
92.416	5.084	2.5

LOCUTOR 8 (MASCULINO)		
FRENTE DERECHA		
Izquierda	Centro	Derecha
0	0	100

LOCUTOR 8 (MASCULINO)		
ATRAS DERECHA		
Izquierda	Centro	Derecha
1.25	1.25	97.5

LOCUTOR 8 (MASCULINO)		
ATRAS IZQUIERDA		
Izquierda	Centro	Derecha
78.834	5.956	15.21

LOCUTOR 9 (MASCULINO)		
FRENTE IZQUIERDA		
Izquierda	Centro	Derecha
92.596	2.03	5.374

LOCUTOR 9 (MASCULINO)		
FRENTE DERECHA		
Izquierda	Centro	Derecha
2.026	2.74	95.234

LOCUTOR 9 (MASCULINO)		
ATRAS DERECHA		
Izquierda	Centro	Derecha
1.466	1.334	97.2

LOCUTOR 9 (MASCULINO)		
ATRAS IZQUIERDA		
Izquierda	Centro	Derecha
87.528	9.708	2.76

LOCUTOR 10 (FEMENINO)		
FRENTE IZQUIERDA		
Izquierda	Centro	Derecha
55.762	5.894	38.342

LOCUTOR 10 (FEMENINO)		
FRENTE DERECHA		
Izquierda	Centro	Derecha
23.376	2.04	74.586

LOCUTOR 10 (FEMENINO)		
ATRAS DERECHA		
Izquierda	Centro	Derecha
5.716	3.49	90.794

LOCUTOR 10 (FEMENINO)		
ATRAS IZQUIERDA		
Izquierda	Centro	Derecha
73.934	2.582	23.484

LOCUTOR 11 (FEMENINO)		
FRENTE IZQUIERDA		
Izquierda	Centro	Derecha
67.912	12.334	19.756

LOCUTOR 11 (FEMENINO)		
FRENTE DERECHA		
Izquierda	Centro	Derecha
8.528	1.904	89.566

LOCUTOR 11 (FEMENINO)		
ATRAS DERECHA		
Izquierda	Centro	Derecha
4.546	5.456	90

LOCUTOR 11 (FEMENINO)		
ATRAS IZQUIERDA		
Izquierda	Centro	Derecha
61.832	11.134	27.034

LOCUTOR 12 (FEMENINO)		
FRENTE IZQUIERDA		
Izquierda	Centro	Derecha
69.084	7.464	23.452

LOCUTOR 12 (FEMENINO)		
FRENTE DERECHA		
Izquierda	Centro	Derecha
19	6.25	74.75

LOCUTOR 12 (FEMENINO)		
ATRAS DERECHA		
Izquierda	Centro	Derecha
22.834	1	76.166

LOCUTOR 12 (FEMENINO)		
ATRAS IZQUIERDA		
Izquierda	Centro	Derecha
62.376	15.286	22.338

En la siguiente evaluación del sistema se resumen los datos que describen el porcentaje de acierto del total de veces en que se le habló en la zona lateral correspondiente.

EVALUACIÓN DEL SISTEMA	
LATERALIZACIÓN	84.54825 %

Discusión de resultados

- El porcentaje total de reconocimiento de vocales del sistema es del 78.85%, lo cual supone una alta valoración para el sistema, considerando su carácter multilocutor, pues su eficiencia es similar a otros sistemas monolocator.
- Este trabajo propone una técnica para la determinación de la frecuencia fundamental de señales periódicas o cuasiperiódicas en el dominio del tiempo, la cual tiene la ventaja de reducir el costo computacional con respecto a la técnica de la autocorrelación. Esta técnica sin embargo requiere de señales con gran amplitud para asegurar que los cruces por cero de la señal a analizar se mantengan constantes (periódicos).
- Los parámetros LPC pueden ser herramientas útiles para el reconocimiento de señales de voz, observándose los mejores resultados en señales del tipo sonoras.
- En el reconocimiento de voz en modo multilocutor, la construcción de cúmulos resulta muy conveniente para mejorar la eficiencia del sistema, debido a la diferencia existente entre los parámetros de diferentes locutores. Esto nos indica también que los LPC pueden ser útiles en aplicaciones de certificación de locutor.
- Durante el desarrollo de este trabajo se notó una mejoría en el reconocimiento al sumar los coeficientes consecutivos del vector de parámetros, esto sin embargo incrementa la cantidad de datos a almacenar.
- Un filtro acotado a las frecuencias de la voz dificulta la localización del sonido en diferentes puntos del espacio, esto es porque el oído utiliza todo el espectro audible para manifestar su respuesta.

Conclusiones

- Se estudió y analizó el comportamiento del *sistema auditivo humano*; sus principios fisiológicos y psicológicos, para el diseño y desarrollo de un sistema de reconocimiento de voz.
- Se construyó un sistema de adquisición de señales basado en la morfología del pabellón auricular humano y las limitaciones en frecuencia del oído, mediante la construcción de 2 orejas artificiales, acopladas a una cabeza de unicel donde fueron instalados 2 micrófonos electret cuya señal es filtrado en el rango de la voz humana.
- Se implementaron algoritmos neuronales y difusos en un sistema multilocutor para:
 - Reconocer vocales aisladas con un índice de reconocimiento del 78.85%
 - Localizar la fuente sonora en 180° en el plano horizontal con un índice de reconocimiento del 84.55%
 - Diferenciar una voz masculina de una femenina de entre un conjunto de 6 voces masculinas y 6 femeninas con un índice de reconocimiento del 62.92%
- Se diseñó y desarrolló un sistema de reconocimiento de vocales capaz de detectar y analizar, en tiempo real, parámetros en señales de voz, identificarlos y emitir una respuesta describiendo la localización de la fuente de voz en 180°, el locutor (género masculino o femenino) y la vocal reconocida.

Se propone una técnica de cálculo de la frecuencia fundamental F_0 que aunque reduce el costo computacional en aproximadamente 20 veces, requiere de un nivel de señal que asegure cruces por ceros periódicos. En tiempo real, son condiciones que no pueden asegurarse, lo cual compromete la eficiencia del algoritmo. Para aplicaciones donde la velocidad de procesamiento no es una limitación, para mejorar el rendimiento se recomienda el uso de la autocorrelación con el segmento entero, o el uso de técnicas como el recorte central con tres niveles (véase (14)).

Trabajos Futuros

Un trabajo para un futuro inmediato tiene que ver con la localización de fuente, que consistiría en un sistema de localización azimutal. Así se continuaría con las bases antes expuestas, pero con el reto de solucionar la localización de atrás y adelante.

Otra propuesta para trabajo futuro, consiste en la aplicación de una técnica de reconocimiento de voz, diferente a las redes neuronales artificiales, siendo las cadenas ocultas de Markov una opción para obtener resultados que permitan comparar la eficacia de ambas técnicas, y visualizar su eficiencia ante un sistema de reconocimiento multilocutor.

El desarrollo de filtros digitales es otra opción, teniendo como base la caracterización de la respuesta del oído, con la finalidad de modificar archivos en tiempo real para crear la ilusión de que éstos provienen de cualquier punto en el espacio (véase audio 3D).

Finalmente, se propone realizar el desarrollo de hardware para la respuesta final del sistema; es decir; que una estructura con forma de cabeza humana rote perpendicular al plano horizontal para ubicar la fuente de sonido.

REFERENCIAS

1. **Russell, S., & Norving, P.** *Inteligencia Artificial: Un Enfoque Moderno*. 2ª. s.l. : Pearson Prentice Hall, 2004.
2. **ELGUEA, JAVIER.** Inteligencia artificial y psicología: la concepción contemporánea de la mente humana. [En línea] http://biblioteca.itam.mx/estudios/estudio/estudio10/sec_14.html.
3. **Bernal Bermudez, J., Bobadilla Sancho, J., & Gómez Vilda, P.** *Reconocimiento de voz y fonética acústica*. 1ª. México, D. F. : Alfaomega, 2000.
4. Reconocimiento de Voz. [En línea] [Citado el: 23 de febrero de 2010.] http://www.articulosinformativos.com/Reconocimiento_de_Voz-a963743.html#8230320.
5. **Informática, Departamento de Electrónica e.** Breve historia del reconocimiento de voz. [En línea] www.dei.uc.edu.py/tai2000/reconocedor/Historia.htm.
6. **Arias Paredes, Flores Ortega, Reyes Agustín.** *Sistema artificial de audición binaural basado en el sistema auditivo humano utilizable en el acondicionamiento de recintos acústicos. (Tesis de licenciatura-UPIITA)*. México D.F. : s.n., 2005.
7. **Venegas Castillo, Rodolfo G.** *Diseño e implementación de un modelo de localización sonora espacial utilizando técnicas de inteligencia computacional. (Tesis de licenciatura-Universidad Tecnológica de Chile)*. Santiago de Chile : s.n., 2006.
8. **Mantilla Caeiros, Alfredo V.** *Análisis, reconocimiento y síntesis de voz esofáfica. (Tesis de Doctorado-ESIME Culhuacán)*. s.l. : [En línea]. Disponible en http://itzamna.bnct.ipn.mx:8080/dspace/bitstream/123456789/2945/1/2401_2007_ESIME-CUL_DOCTORADO_mantilla_caeiros_alfredovicigor.pdf.
9. **Flores Paulín, Juan C.** *Técnicas para el reconocimiento de voz en palabras aisladas en la lengua náhuatl (Tesis de maestría-Cento de Investigación en Computación)*. s.l. : [En línea]. Disponible en http://itzamna.bnct.ipn.mx:8080/dspace/bitstream/123456789/7909/1/2408_tesis_Diciembre_2010_1532426908.pdf.
10. **Castañeda, Pablo Félix.** *El Lenguaje verbal del niño : ¿cómo estimular, corregir y ayudar para que aprenda a hablar bien?* Lima : s.n., 1999.
11. **GUYTON, Arthur C. y HALL, John E.** *Tratado de Fisiología Moderna*. 15ª. 2001.
12. **Gómez, Juan Carlos.** Procesamiento Digital de Señales de Voz. Modelos de Producción de Voz. 2001, pág. 10.
13. **Alexandre Cortizo, Enrique.** Modelo de producción. *Teoría de la Señal y Comunicaciones. Universidad de Alcalá*. [En línea] <http://agamenon.tsc.uah.es/Asignaturas/master/ms/apuntes/2-ModeloVoz.pdf>.

-
14. **Suárez Guerra, S.** *Parámetros de la voz*. Centro de Investigación en Computación IPN. México, D.F. : s.n., 2007.
 15. *Diseño de un sistema de reconocimiento del habla para controlar dispositivos eléctricos*. **Salcedo, Dayana. Texera, Alejandro.** 10, Caracas, Venezuela : s.n., 2007, Tekhne. Revista de la Facultad de Ingeniería.
 16. **Suárez Guerra, S.** *Una metodología para realizar trabajos de reconocimiento de voz*. Centro de Investigación en Computación IPN. México, D. F. : s.n., 2004.
 17. **Popock, Guillian & Christopher D.** *Fisiología Humana: La base de la Medicina (2° edición)*. España : Editorial Masson, (2005).
 18. **Purves, D.** *Invitación a la neurociencia (neuroscience)*. s.l. : Editorial Médica panamerica, 2001.
 19. **Rains, Dennis G.** *Principios de neuropsicología humana*. . México : Mc Graw Hill , 2003.
 20. **Escarabajal Arrieta, María Dolores.** *Fundamentos de psicobiología: libro de práctica I*. Madrid España : Delta Publicaciones Universitarias, S.L, 2005.
 21. **Castro Alamancos, Dr. Manuel A.** *Dinamismo talamocortical: ¿cómo se comunican el tálamo y la neocorteza durante los estados de procesamiento de información?* Philadelphia, EUA : s.n., 2003.
 22. **Goodglass, H. & Geschwind.** *Language disorders. En E. Carterette y M.P. Friedman (eds.) Handbook of Perception: Language and Speech. Volumen II* . N.New York : Academic Press, 1976.
 23. **Zylberbaum, Jacobo Grinberg.** *Nuevos principios de psicología fisiológica: la expansión de la conciencia*. s.l. : Editorial Trillas, 1976.
 24. Redes Neuronales Artificiales. [En línea] . [Citado el: 23 de Febrero de 2010.] [http://www.redes-neuronales.tk/..](http://www.redes-neuronales.tk/)
 25. **Martin T. Hagan, Howar B. Demuth, Mark Beale.** *Neural Network Design*. China : Thomson, 1996.
 26. **Kecman, V.** *Learning and Soft Computing. Support Vector Machines, Neural Networks and Fuzzy Logic*. 2001.
 27. **Pierre Duchesne, Bruno Rémillard.** *Statistical modeling and analysis for complex data problems*. s.l. : Springe, 2005.
 28. LOCALIZACIÓN. [En línea] [Citado el: 24 de diciembre de 2010.] <http://www.eumus.edu.uy/docentes/maggiolo/acuapu/loc.html>.
 29. **Panero, Julius. Zelnik Martín.** *Las dimensiones humanas en los espacios interiores: Estándares antropométricos séptima edición*. México D.F : Ediciones G. Gili , 1996. .

-
30. **J.A.F., Tresguerres.** *Fisiología humana*. España : McGraw-Hill interamericana, 1999.
31. **Castro-Alamancos, Dr. Manuel A.** *Dinamismo talamocortical: ¿cómo se comunican el tálamo y la neocorteza durante los estados de procesamiento de información?* Philadelphia, EUA : s.n., 2003.
32. [En línea] http://www.articulosinformativos.com/Reconocimiento_de_Voz-a963743.html#8230320.
33. **Pierre Duchesne, Bruno Rémillard.** *Statistical modeling and analysis for complex data problems*. USA : Springer, 2005.

ANEXO I

ESPECIFICACIONES TÉCNICAS DEL CAUCHO DE SILICÓN

CAUCHO DE SILICÓN

PROPIEDADES DEL MATERIAL LÍQUIDO

ESPECIFICACIÓN	VALOR	MÉTODO DE PRUEBA
Viscosidad @ 25° C, Brookfield LVF, Aguja # 4 a 6 r.p.m. (cps)	160000 ± 10000	EQPP-CC-002
Densidad @ 25 ° C (g/ml)	1.3 ± 0.2	EQPP-CC-008
Color	Gris	EQPP-CC-007

PROPIEDADES DEL MATERIAL VULCANIZADO


ESPECIFICACIÓN	VALOR	MÉTODO DE PRUEBA
Tiempo de gel @ 25 ° C (min. , seg.) 100 gr. de producto + 3 % de catalizador TP	6 ± 3	EQPP-CC-004
Tiempo de curado @ Min	15 ± 8	EQPP-CC-005
Dureza Shore "A"	30 ± 5	EQPP-CC-025
Alargamiento de ruptura (DIN 53504-S-3A)	> 250 %	EQPP-CC-035
Resistencia al desgarre (ASTM D-624 B) N/mm	> 50	EQPP-CC-024
Resistencia a la tensión (ASTM D-638) N/mm ²	> 8	EQPP-CC-022

CARACTERÍSTICAS

- Viscosidad ajustable con diluyentes.
- Vulcanizado a temperatura ambiente.
- Resistencia a temperaturas elevadas..
- Reproduce piezas que no requieren detalles profundos.

ANEXO II

HOJA DE ESPECIFICACIONES DEL MICRÓFONO UTILIZADO.



KOBITONE™
AUDIO COMPANY

Electret Condenser Microphone

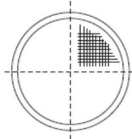
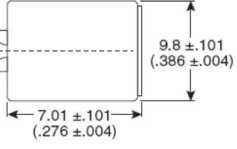
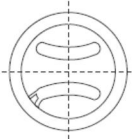

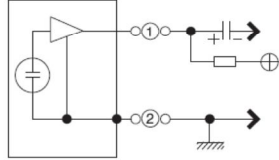
Date: 4/26/05

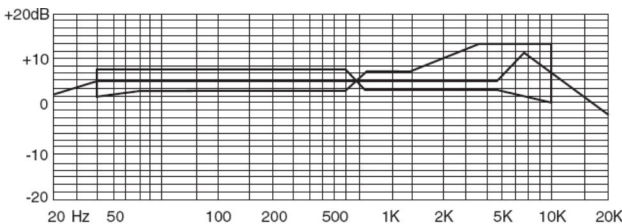
Dimensions: mm (IN)

Part Number:

25LM043

Electret Condenser Microphone





Frequency (Hz)	Sensitivity (dB)
20	0
50	5
100	5
200	5
500	5
1K	5
2K	10
5K	10
10K	5
20K	0

KT-400003

Electrical specifications:

- Sensitivity (0dB=1V/ μ bar @ 1KHz, $R_L=1K\Omega$, $V_{cc}=4.5V$): -65 \pm 3dB
- Impedance: low
- Standard voltage: 4.5V
- Range of operating voltage: 2V to 10V
- Current drain: 0.4mA max.
- Self noise level: less than 34dB SPL
- Operation voltage: 1.5 to 15VDC
- Current consumption: 0.5mA or less (Supply voltage 6V)
- Frequency resonance: 20-12K Hz
- Signal to noise ratio: 40dB

Available from Mouser Electronics

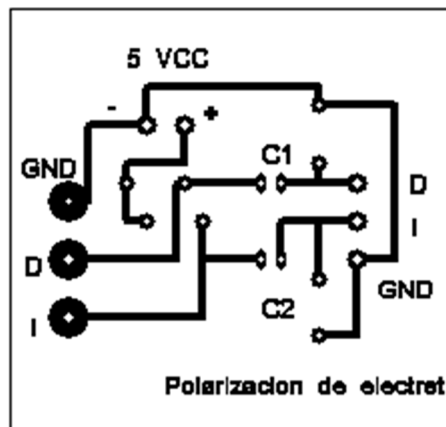
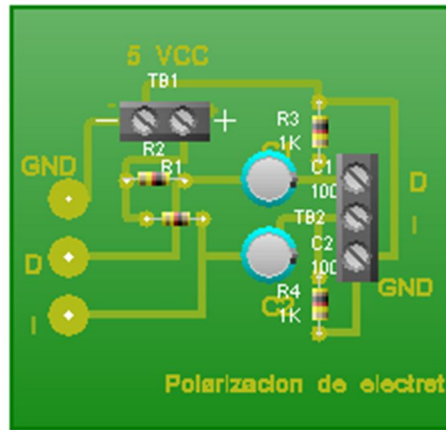
www.mouser.com

(800) 346-6873

Specifications are subject to change without notice. No liability or warranty implied by this information. Environmental compliance based on producer documentation.

ANEXO III

CIRCUITO IMPRESO PARA LA POLARIZACIÓN DE UN MICRÓFONO ELECTRET



ANEXO IV

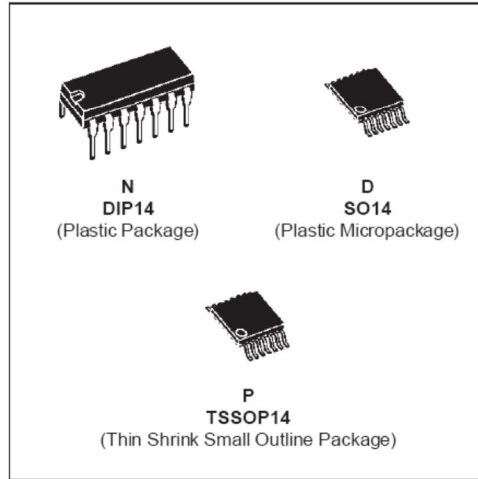
HOJAS DE ESPECIFICACIONES DEL CIRCUITO INTEGRADO TL084



TL084
TL084A - TL084B

GENERAL PURPOSE J-FET
QUAD OPERATIONAL AMPLIFIERS

- WIDE COMMON-MODE (UP TO V_{CC}^+) AND DIFFERENTIAL VOLTAGE RANGE
- LOW INPUT BIAS AND OFFSET CURRENT
- OUTPUT SHORT-CIRCUIT PROTECTION
- HIGH INPUT IMPEDANCE J-FET INPUT STAGE
- INTERNAL FREQUENCY COMPENSATION
- LATCH UP FREE OPERATION
- HIGH SLEW RATE : 16V/ μ s (typ)



DESCRIPTION

The TL084, TL084A and TL084B are high speed J-FET input quad operational amplifiers incorporating well matched, high voltage J-FET and bipolar transistors in a monolithic integrated circuit.

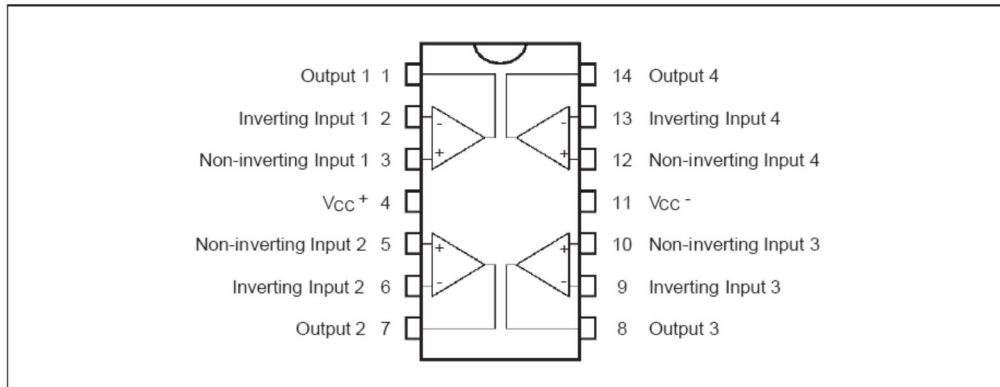
The devices feature high slew rates, low input bias and offset currents, and low offset voltage temperature coefficient.

ORDER CODES

Part Number	Temperature Range	Package		
		N	D	P
TL084M/AM/BM	-55°C, +125°C	•	•	•
TL084I/AI/BI	-40°C, +105°C	•	•	•
TL084C/AC/BC	0°C, +70°C	•	•	•

Examples : TL084CN, TL084CD

PIN CONNECTIONS (top view)



ELECTRICAL CHARACTERISTICS

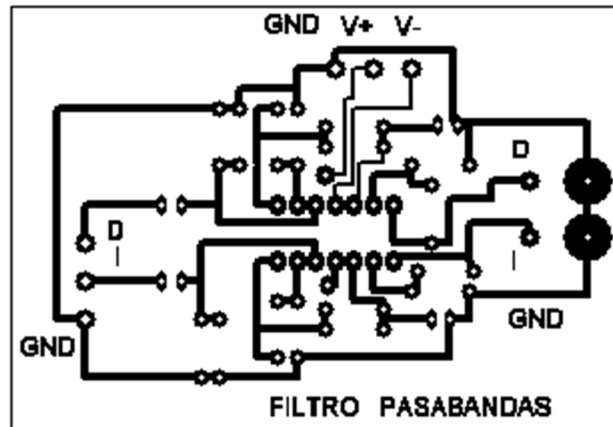
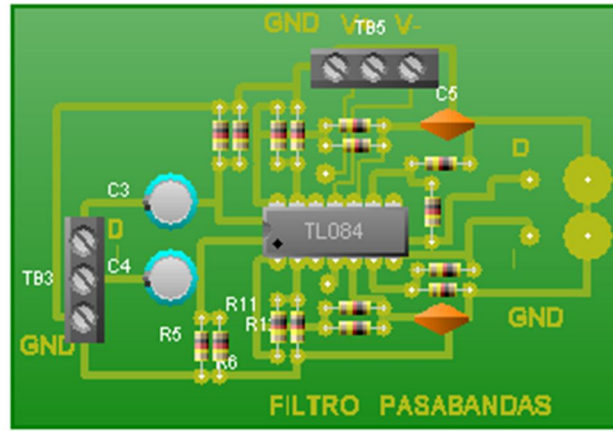
$V_{CC} = \pm 15V$, $T_{amb} = 25^{\circ}C$ (unless otherwise specified)

Symbol	Parameter	TL084I,M,AC,AI, AM,BC,BI,BM			TL084C			Unit
		Min.	Typ.	Max.	Min.	Typ.	Max.	
V_{io}	Input Offset Voltage ($R_S = 50\Omega$) $T_{amb} = 25^{\circ}C$ $T_{min.} \leq T_{amb} \leq T_{max.}$ TL084 TL084A TL084B TL084 TL084A TL084B		3 3 1	10 6 3 13 7 5		3	10 13	mV
DV_{io}	Input Offset Voltage Drift		10			10		$\mu V/^{\circ}C$
I_{io}	Input Offset Current * $T_{amb} = 25^{\circ}C$ $T_{min.} \leq T_{amb} \leq T_{max.}$		5	100 4		5	100 4	pA nA
I_{ib}	Input Bias Current * $T_{amb} = 25^{\circ}C$ $T_{min.} \leq T_{amb} \leq T_{max.}$		20	200 20		30	400 20	pA nA
A_{vd}	Large Signal Voltage Gain ($R_L = 2k\Omega$, $V_O = \pm 10V$) $T_{amb} = 25^{\circ}C$ $T_{min.} \leq T_{amb} \leq T_{max.}$	50 25	200		25 15	200		V/mV
SVR	Supply Voltage Rejection Ratio ($R_S = 50\Omega$) $T_{amb} = 25^{\circ}C$ $T_{min.} \leq T_{amb} \leq T_{max.}$	80 80	86		70 70	86		dB
I_{CC}	Supply Current, per Amp, no Load $T_{amb} = 25^{\circ}C$ $T_{min.} \leq T_{amb} \leq T_{max.}$		1.4	2.5 2.5		1.4	2.5 2.5	mA
V_{icm}	Input Common Mode Voltage Range	± 11	+15 -12		± 11	+15 -12		V
CMR	Common Mode Rejection Ratio ($R_S = 50\Omega$) $T_{amb} = 25^{\circ}C$ $T_{min.} \leq T_{amb} \leq T_{max.}$	80 80	86		70 70	86		dB
I_{OS}	Output Short-circuit Current $T_{amb} = 25^{\circ}C$ $T_{min.} \leq T_{amb} \leq T_{max.}$	10 10	40	60 60	10 10	40	60 60	mA
$\pm V_{OPP}$	Output Voltage Swing $T_{amb} = 25^{\circ}C$ $T_{min.} \leq T_{amb} \leq T_{max.}$ $R_L = 2k\Omega$ $R_L = 10k\Omega$ $R_L = 2k\Omega$ $R_L = 10k\Omega$	10 12 10 12	12 13.5		10 12 10 12	12 13.5		V
SR	Slew Rate ($V_{in} = 10V$, $R_L = 2k\Omega$, $C_L = 100pF$, $T_{amb} = 25^{\circ}C$, unity gain)	8	16		8	16		V/ μs
t_r	Rise Time ($V_{in} = 20mV$, $R_L = 2k\Omega$, $C_L = 100pF$, $T_{amb} = 25^{\circ}C$, unity gain)		0.1			0.1		μs
K_{OV}	Overshoot ($V_{in} = 20mV$, $R_L = 2k\Omega$, $C_L = 100pF$, $T_{amb} = 25^{\circ}C$, unity gain)		10			10		%
GBP	Gain Bandwidth Product ($f = 100kHz$, $T_{amb} = 25^{\circ}C$, $V_{in} = 10mV$, $R_L = 2k\Omega$, $C_L = 100pF$)	2.5	4		2.5	4		MHz
R_i	Input Resistance		10^{12}			10^{12}		Ω
THD	Total Harmonic Distortion ($f = 1kHz$, $A_V = 20dB$, $R_L = 2k\Omega$, $C_L = 100pF$, $T_{amb} = 25^{\circ}C$, $V_O = 2V_{PP}$)		0.01			0.01		%
e_n	Equivalent Input Noise Voltage ($f = 1kHz$, $R_S = 100\Omega$)		15			15		$\frac{nV}{\sqrt{Hz}}$
ϕ_m	Phase Margin		45			45		Degrees
V_{O1}/V_{O2}	Channel Separation ($A_V = 100$)		120			120		dB

* The input bias currents are junction leakage currents which approximately double for every $10^{\circ}C$ increase in the junction temperature.

ANEXO V

CIRCUITO IMPRESO DE UN FILTRO PASABANDAS ACOTADO A LAS FRECUENCIAS DE VOZ.



ANEXO VI

PSEUDOCÓDIGO DE LA INTERFAZ GRÁFICA

```
1  Inicializar micrófono
2  Configurar Tarjeta de Video
3  escrito <- WrittenBytes(BUFFER)
4  SI escrito/2 > tamSegmento ENTONCES continuar SI NO regresar a 3
5  olzquierdo <- BUFFER(1:tamSegmento)
6  oDerecho <- BUFFER(tamSegmento+1:2*tamSegmento)
7  olzquierdo <- olzquierdo*Hamming
8  oDerecho <- oDerecho *Hamming
9  elzq <- olzquierdo · olzquierdo
10 eDer <- oDerecho · oDerecho
11 SI elzq>umbralIzq OR eDer>umbralDer ENTONCES continuar SI NO regresar a 3
12 olzq <- olzquierdo * FIR
13 oDer <- oDerecho * FIR
14 correlacion <- olzq * oDer
15 iMax <- MAXPOS(correlacion)
16 POR CADA Sgn(olzqn)<> Sgn(olzqn-1) HACER crucesXCero(i) <- {n - (n-1)}
17 LPCIzq <- LPC(olzquierdo)
18 LPCDer <- LPC(oDerecho)
19 LPCIzq <- LPCIzqn - LPCIzqn-1
20 parametrosLPC <- LPCDer + LPCIzq
21 POR CADA VectPrototipo HACER Semejanzas(i) <- VectPrototipo · parametrosLPC
22 semejanza <- MAXPOS(Semejanzas)
23 SI semejanza>umbralSemejanza ENTONCES continuar SI NO regresar a 3
24 autocorrelacion <- crucesXCero * crucesXCero
25 cMax <- MAXPOS(autocorrelacion)
26 F0 <- 1/{(cMax/size(crucesXCero))*(tamSegmento/Fs)}
27 Actualizar posición de la animación en base a iMax
28 SI F0>150 ENTONCES mostrar silueta femenina SI NO mostrar silueta masculina
29 Mostrar letra reconocida en Semejanza
30 Regresar a 3
```