CLASIFICACIÓN DE VOCES NORMALES Y PATOLOGICAS EMPLEANDO LA TRANSFORMADA WAVELET

REINALDO RODRIGUEZ VILLALOBOS

DIRECTORA
SONIA HELENA CONTRERAS ORTIZ

UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR
FACULTAD DE INGENIERÍA ELECTRICA Y ELECTRONICA
CARTAGENA DE INDIAS D. T. Y C.
2006

CLASIFICACIÓN DE VOCES NORMALES Y PATOLOGICAS EMPLEANDO LA TRANSFORMADA WAVELET

REINALDO RODRIGUEZ VILLALOBOS

Trabajo de grado presentado como requisito para optar al titulo de Ingeniero Electrónico

DIRECTORA
SONIA HELENA CONTRERAS ORTIZ
MAGISTER EN POTENCIA ELÉCTRICA

UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR
FACULTAD DE INGENIERÍA ELECTRICA Y ELECTRONICA
CARTAGENA DE INDIAS D. T. Y C.
2006

Articulo 107

La Universidad Tecnológica de Bolívar se reserva el derecho de propiedad de los trabajos de grado aprobados y no pueden ser explotados comercialmente sin autorización.

Nota de aceptación

_____

_____

_____

_____

_____

Firma del presidente del jurado

_____

Firma del Jurado

_____

Firma del Jurado

Cartagena D.T y C, Mayo de 2006

Cartagena D. T. Y C., Mayo de 2006

Señores

COMITÉ DE EVALUACIÓN DE PROYECTOS
Programa de Ingeniería Eléctrica y Electrónica

UNIVERSIDAD TECNOLOGICA DE BOLIVAR
La ciudad

Respetados señores:

Con toda atención me dirijo a ustedes con el fin de presentarles a su consideración, estudio y aprobación de la Tesis de Grado titulada CLASIFICACIÓN DE VOCES NORMALES Y PATOLOGICAS EMPLEANDO LA TRANSFORMADA WAVELET como requisito parcial para optar al titulo de ingeniero electrónico.

Atentamente

_____

REINALDO RODRIGUEZ V.

Cartagena D. T. Y C., Mayo de 2006

Señores

COMITÉ DE EVALUACIÓN DE PROYECTOS
Programa de Ingeniería Eléctrica y Electrónica
UNIVERSIDAD TECNOLOGICA DE BOLIVAR
La ciudad

Cordial saludo:

A través de la presente me permito entregar la Tesis de Grado titulada CLASIFICACIÓN DE VOCES NORMALES Y PATOLOGICAS EMPLEANDO LA TRANSFORMADA WAVELET para su estudio y evaluación la cual fue realizada por el estudiante REINALDO RODRIGUEZ VILLALOBOS, de la cual acepto ser su director.

Atentamente,

_____

SONIA HELENA CONTRERAS ORTIZ
Magíster en Potencia Eléctrica

AUTORIZACIÓN

Yo REINALDO RODRIGUEZ VILLALOBOS, identificado con la cedula de ciudadanía número 73.201.297 de Cartagena, autorizo a la universidad tecnológica de Bolívar, para hacer uso de mi trabajo de grado y publicarlo en el catalogo on-line de la biblioteca

_____

REINALDO RODRIGUEZ VILLALOBOS

AGRADECIMIENTOS

Agradezco el apoyo incondicional de mi madre a lo largo de mis estudios universitarios, quien con sus consejos y orientaciones, ha hecho de mí una persona de bien.

Además resalto el apoyo de Sonia Contreras, quien con sus orientaciones, fue pilar fundamental en el desarrollo de este trabajo.

CONTENIDO

Pág.

LISTA DE TABLAS

Pág.

LISTA DE FIGURAS

Pág.

LISTA DE ANEXOS

Pág.

RESUMEN

En este trabajo se presenta el desarrollo de la metodología de un sistema de clasificación de voces normales y patológicas utilizando como herramienta de análisis la Transformada Wavelet, ya que se constituye en una herramienta apropiada para el análisis de señales no estacionarias como la voz cuyo contenido espectral varía con el tiempo. Se plantea el uso de redes neuronales tipo perceptrón multicapa utilizando la técnica de aprendizaje backpropagation y el metodo de optimización de Levenberg-Marquardt como estrategia para la etapa de clasificación de voces normales y patológicas.

Se describe el marco experimental realizado para cada una de las etapas que conforman la metodología propuesta y con el cual se obtuvieron resultados satisfactorios al momento de clasificar las voces normales y patológicas. Además se presentan detalladamente los algoritmos y la interfaz gráfica desarrollada bajo la plataforma de **MATLAB 5.3**.

INTRODUCCIÓN

La presencia de patologías en las cuerdas vocales, definidas como cambios o variaciones fuera de los límites determinados como normales de las cualidades de la voz (timbre, intensidad, altura tonal), y de los niveles anatómico-fisiológicos que intervienen en la producción vocal como es el nivel respiratorio, de resonancia, auditivo, emisor, hormonal y de comando reflejan el deterioro en la calidad de voz producida. Existen numerosas afecciones en la laringe y en las cuerdas vocales, cuyos métodos de detección tradicionales son de carácter invasivo o requieren de un análisis especializado sobre un conjunto de parámetros acústicos de la voz.

El desarrollo de sistemas computarizados orientados al análisis y evaluación del mecanismo de producción de la voz, en los que se emplean modernas técnicas de procesamiento digital de las señales, ha tenido un importante desarrollo durante la última década, con diversas aplicaciones en áreas como Bioacústica, Criminología, Lingüística, Fonética, Psicolinguística, entre otras. En particular, se han creado sistemas computarizados que sirven de soporte y facilitan el trabajo de un Fonoaudiólogo o el de un Otorrinolaringólo en el estudio de la normalidad o disfuncionalidad de la voz, dado que complementan los procedimientos tradicionales en los que la valoración subjetiva se apoya con métodos que brindan un alto potencial cuantitativo basándose en el análisis de los parámetros acústicos de la señal de la voz.

De esta manera se hace necesario evaluar los métodos de extracción de características de las señales de voz mas eficaces que suministren

información relevante de los parámetros de la voz para su análisis y posterior valoración en el estudio de voces normales y patológicas.

En general, las técnicas utilizadas en la representación de señales de voz, asumen cierta estacionalidad por ejemplo, la Transformada de Fourier en Tiempo Corto (STFT) no puede ser aplicada con el objeto de obtener información precisa de cuando o donde las diferentes componentes de frecuencia se encuentran en la señal. En otras palabras, la transformada de Fourier posee una muy pobre resolución en tiempo, es decir, cualquier intento de obtener una resolución (que determina la cantidad de información presente en una señal) más fina en dominio temporal lleva consigo una pérdida de información en el dominio de la frecuencia y viceversa. Por otra parte existe otra técnica conocida como la Transformada Wavelet que permite obtener una representación tiempo-frecuencia de la señal que se constituye en una herramienta apropiada para el análisis de señales no estacionarias como la voz cuyo contenido espectral varía con el tiempo. Usando la transformada Wavelet, una señal de voz puede ser analizada en una escala especifica correspondiente al rango del habla; de aquí la relación entre la transformada Wavelet y como el sistema auditivo humano procesa el sonido.

OBJETIVOS

OBJETIVO GENERAL

Desarrollar una metodología de extracción de características en señales de voz, utilizando como herramienta de análisis la Transformada Wavelet, orientada a la clasificación de voces normales y patológicas de la población de la ciudad de Cartagena de Indias.

OBJETIVOS ESPECÍFICOS

- Construir una base de datos que represente las características representativas de las voces normales y patológicas de la población de Cartagena de Indias.
- Estudiar las formas de implementar la transformada wavelet que permitan una mejor extracción de información de una señal de voz.
- Implementar un clasificador que permita diferenciar claramente una voz normal de una voz patológica
- Evaluar el desempeño del sistema, mediante el análisis de un experto, para comprobar la validez de los resultados obtenidos.

1 ESTADO DEL ARTE

A nivel mundial y nacional se han realizado gran cantidad de investigaciones tendientes al análisis y extracción de características de la voz con el objetivo de caracterizar este tipo de señales de origen biológico y detectar patologías. A continuación se mencionan algunos de los trabajos encontrados que abarcan el procesamiento de señales de voz:

- "Transformada wavelet aplicada a la extracción de información en señales de voz". Es una tesis doctoral en la que se detalla el uso de la transformada wavelet en la parametrización de señales de voz.
- "Applying wavelet analysis to speech segmentation and classification" Presenta el estudio realizado en la selección de la wavelet más apropiada para la clasificación de señales de voz en: voz sonora, oclusiva, fricativa y silencio.
- "The use of wavelet transforms in phoneme recognition". Investiga la utilidad de la transformada wavelet en la etapa de extracción de características para el reconocimiento de fonemas. Se emplean además, modelos ocultos de Markov (HMM) para el sistema de reconocimiento independiente del locutor.
- "Robust Classification of Speech based on the Dyadic Wavelet Transform with application to Celp Coding". Utiliza la transformada wavelet para aplicaciones de codificación de voz.
- "Discrete wavelet transform techniques in speech Processing". Presenta un método para evaluar el uso de la transformada wavelet en el análisis y síntesis de la voz, distinguiendo entre voz y no voz con la determinación del pitch.

- En la Universidad Nacional, sede Manizales se han realizado varios trabajos en el campo, entre los cuales están:
  - “Clasificación automatizada de las características acústicas de la voz normal en la ciudad de Manizales.”
  - ”Diseño y desarrollo de un sistema interactivo de análisis acústico de la voz y el habla para la ciudad de Manizales”.
  - “Extracción de Características usando Transformada Wavelet en la Identificación de Voces Patológicas”.

Las referencias completas se listan en la bibliografía al final del trabajo.

2 ANATOMIA DE LA VOZ

2.1 Mecanismo de producción de la voz.

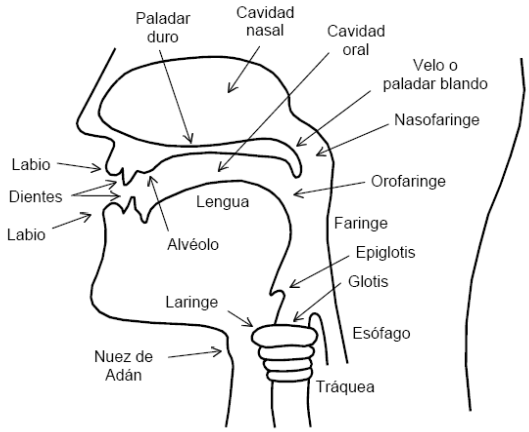
El proceso básico de producción de la voz es el mismo para hablar y cantar. La producción de la voz comienza en el cerebro con la conceptualización de la idea que se desea transmitir. Esta idea se asocia a una estructura lingüística, seleccionando las palabras adecuadas y ordenándolas de acuerdo con unas reglas gramaticales. A continuación el cerebro envía señales a través del sistema nervioso central a los músculos de la laringe, cuello y tórax acompañado de un flujo de aire a través del tracto fonatorio obteniendo finalmente la voz. La onda de voz se propaga a través del aire realimentándose en el oído del hablante y llegando al oído del interlocutor.

La voz se define estrictamente como la producción de sonidos por las cuerdas vocales, en un proceso de conversión de energía aerodinámica generado en el tórax, el diafragma y la musculatura abdominal, a una energía acústica originada en la glotis. El principio fundamental en la producción de la voz es la vibración de las cuerdas vocales, debido a un acoplamiento y modulación del flujo de aire que pasa a través de ellas generando su movimiento. La eficacia en la transformación de energía esta dada por la tensión y la configuración glótica[1].

En la figura 1 pueden observarse los componentes del sistema fonador humano: pulmones, traquea, laringe y cavidades oral y nasal.

[1] JIANG, Jack,and LIN, Emily. Fisiología de las cuerdas vocales. Clínicas de Norteamérica de Otorrinolaringología. 2002. p. 647-665.

Figura 1. Diagrama esquemático del sistema fonador.



El proceso físico de producción de la voz puede dividirse en 3 etapas principales: generación, articulación y radiación[2].

2.1.1 Generación

La generación de la voz se lleva a cabo por medio de la expulsión de aire de los pulmones y su paso por la glotis. De acuerdo a la manera como son generados, los diferentes fonemas pueden clasificarse en sonoros y sordos.

En la producción de los fonemas sonoros como las vocales, el flujo de aire que proviene de los pulmones ejerce una presión sobre las cuerdas vocales haciendo que éstas se separen, sin embargo, los tejidos y músculos de la

[2] FURUI Sadoaki. Digital speech processing, synthesis and recognition. New york: Marcel Dekker Inc.1989.

laringe y cuerdas vocales tienen una elasticidad natural que provoca que éstas se cierren, una vez la presión del aire ha sido contrarrestada.

Este proceso (ciclo glotal) se repite periódicamente con una frecuencia que se denomina frecuencia fundamental, debido a que establece la base para los armónicos mayores generados por las resonancias del tracto vocal. La frecuencia fundamental es el principal aspecto que contribuye a la percepción de la altura (pitch) de los sonidos. Su valor depende principalmente de la presión del aire expelido de los pulmones, la tensión de las cuerdas vocales y su masa, por lo cual varía para los hombres, mujeres y niños. La Tabla 1 resume estas diferencias

Tabla 1. Valores promedio, mínimos y máximos de la frecuencia fundamental para hombres, mujeres y niños.

	F_0 promedio (Hz)	**F_0 mínima (Hz)**	**F_0 máxima (Hz)**
Hombres	125	80	200
Mujeres	225	150	350
Niños	300	200	500

En resumen El proceso de generación de la voz es complejo y secuencial, requiere un acoplamiento de componentes estructurales, flujo de aire y presión, los cuales crean como producto final la voz, la cual es modificada por la interacción de las características intrínsecas de las cuerdas, la función pulmonar y las estructuras de resonancia de la vía aérea superior.

2.1.2 Articulación

En esta etapa, la fuente de sonido se modifica al pasar por el tracto vocal, que va desde la laringe hasta los labios, y la cavidad nasal.

Los órganos articuladores pueden clasificarse en activos y pasivos; los activos son los que cambian su forma o posición para producir los diferentes sonidos, estos son: los labios, la lengua, los dientes inferiores y el velo del paladar. Los pasivos, por su parte, intervienen en la articulación sin realizar ningún movimiento, estos son: los dientes superiores, los alvéolos superiores y el paladar.

2.1.3 Radiación

La onda sonora se propaga a través del aire a una velocidad de 340 m/s, se realimenta en el oído del hablante y llega al oído del interlocutor.

2.2 LA LARINGE

La Laringe, es una estructura móvil que actúa normalmente como una válvula que impide el paso de los elementos deglutidos y cuerpos extraños hacia el tracto respiratorio inferior. Además permite el mecanismo de la fonación diseñado específicamente para la producción de la voz. La emisión de sonidos está condicionada al movimiento de las cuerdas vocales.

Son los movimientos de los cartílagos de la laringe los que permiten variar el grado de apertura entre las cuerdas y una depresión o una elevación de la estructura laríngea, con lo que varía el tono de los sonidos producidos por el paso del aire a través de ellos. Esto junto a la disposición de los otros elementos de la cavidad oral (labios, lengua y boca) permite determinar los diferentes sonidos que emitimos.

La laringe se encuentra situada en la porción anterior del cuello (ver figura 1) y su longitud es más corta en las mujeres y en los niños.

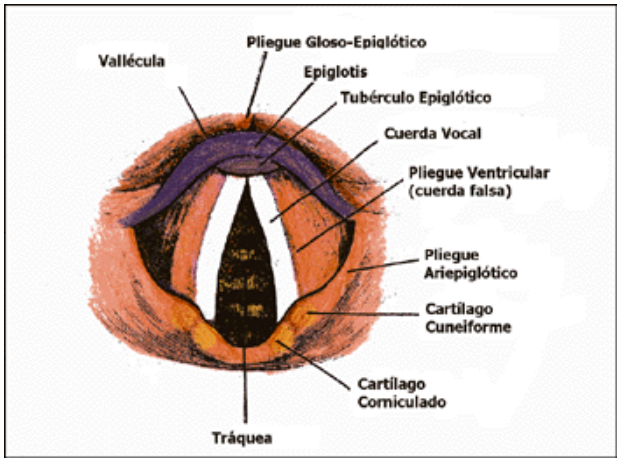
Tabla 2. Dimensiones de la laringe

				Edad Adulta	
		Infancia	**Pubertad**	**Varones**	**Mujeres**
Cuerda vocal	Longitud	6-8 mm	12-15 mm	17-23 mm	12,5 mm
	Porción membranosa	3-4 mm	7-8 mm	11,5-16 mm	8-11,5 mm
	Porción cartilaginosa	3-4 mm	5-7mm	5,5-7mm	4,5-5,5 mm
Glotis	Anchura en reposo	3 mm	5mm	8 mm	6 mm
	Máxima	6 mm	12mm	19 mm	13 mm
Infraglotis	Sagital	5-7 mm	15 mm	25 mm	18 mm
	Transversal	5-7 mm	15 mm	24 mm	17 mm

Su estructura está constituida por un esqueleto cartilaginoso al cual se unen un grupo importante de estructuras musculares recubiertas por una mucosa.

El esquema de la laringe se muestra en la siguiente figura.

Figura 2. Esquema de la laringe.



2.2.1 Funciones de la laringe

Las funciones básicas de la laringe en orden de importancia son tres: Protección, Respiración y Fonación[3].

Protección

La protección es la función más importante de la laringe, actúa como orificio evitando la entrada de cualquier cosa, excepto aire hacia pulmón. En este proceso están implicados los siguientes mecanismos:

- Cierre de la entrada de la laringe: Cuando se tragan los alimentos.
- Cierre de la glotis: Se produce al mismo tiempo que el cierre de la entrada de la laringe.

[3] BALLANTYNE Jhon and GROVES Jhon. Manual de Otorrinolaringología. Barcelona: Salvat Editores, 1982.

- Cese de la respiración. Es automático.
- Tos refleja: Si cualquier partícula entra en la tráquea o en los bronquios es generalmente expulsada por la tos.

Respiración

La laringe interviene en el mecanismo de la respiración mediante los ajustes reflejos de la apertura de la glotis. En este proceso las cuerdas vocales se apartan en forma activa, esto contribuye a la regulación del intercambio gaseoso con el pulmón y la mantención del equilibrio ácido-base.

Fonación

La fonación se desarrolla tardíamente en la evolución de la laringe. Los cambios en la tensión y longitud de las cuerdas vocales, ancho de la hendidura glótica e intensidad del esfuerzo espiratorio provocan variaciones en el tono de voz. Este tono formado por la vibración de las cuerdas vocales en la laringe es modificado por los movimientos de la faringe, lengua y labios para formar el habla.

2.3 Características de la voz

La emisión de la voz se mide mediante tres parámetros básicos: Intensidad o volumen de la voz, Tono o frecuencia de la voz y timbre o calidad de la voz[4].

[4] BALLENGER, Jhon. Enfermedades de la nariz, garganta y oído. Barcelona: Jims, 1981. p. 624-643.

2.3.1 Intensidad o volumen de la voz

El volumen de la voz se relaciona con la presión del sonido creada por la liberación de pulsaciones, es decir, la presión del sonido es directamente proporcional al volumen y a la velocidad de la corriente de aire en la glotis. Si se incrementan el volumen y la velocidad aumenta la amplitud de las ondas sonoras, produciendo mayores excursiones del mecanismo del oído y una sensación de sonido voluminoso. Son los decibeles (db) los que miden dicha intensidad y con esto se determinan las voces fuertes o flojas.

2.3.2 Tono o frecuencia de la voz

El tono de la voz esta directamente relacionado con la frecuencia de vibración de las cuerdas vocales, que esta determinada sobre todo por la masa y elasticidad de los pliegues vocales (cuerdas) en relación con su longitud. Cuando se acortan los pliegues de cualquier laringe normal, su sección se incrementa y pierden elasticidad. Por el contrario, cuando se alargan las cuerdas vocales, se tornan más delgadas y su elasticidad es mayor, y como resultado, cuando la corriente de aire expelido por los pulmones produce su separación, vuelve con más rapidez a sus posiciones de aproximación. Esta reacción origina un aumento de la frecuencia y, por lo tanto un tono más alto. Otro factor que influye sobre el tono es el mayor tamaño de las cueras vocales. Los pliegues de mayores dimensiones producen un tono menor que los pliegues de menor tamaño debido a que los primeros vibran con más lentitud.

2.3.3 Timbre o calidad de la voz

La calidad de la voz se determina por la vibración del pliegue vocal y por la resonancia. El aspecto fonatorio comprende la forma de liberación de la onda aérea en relación con el tipo vibratorio de las cuerdas vocales. La abertura, el cierre y las fases cerradas del ciclo glótico pueden variar en sus relaciones entre sí, y la forma del movimiento vocal es más o menos única en cada laringe. Cada uno de estos factores puede influir en el número y la intensidad relativa de los componentes parciales del sonido vocal complejo y, por consiguiente, en la calidad de la voz. La modificación del sonido a medida que discurre a través de la faringe, boca y nariz, resulta de cuidadosa selección de los hipertonos y de los otros componentes parciales en el sonido complejo generado en la laringe.

2.4 Clasificación vocal

Para una mayor precisión en el estudio de la voz, ésta puede ser clasificada de acuerdo al género, edad y niveles de empleo vocal[5].

2.4.1 Género

La voz de cualquier persona se halla condicionada por sus características anatómico-fisiológicas propias y en particular de las dimensiones de la laringe en cada una de las etapas del ser humano y del sexo (ver tabla 2).

[5] MEJIA, Gloria S, BOTERO, Libia M, CASTRO Luz L. La voz hablada y cantada. Normalidad y Patología. Módulo, Universidad Católica de Manizales. 1999

Estas diferencias marcadas generan discriminación en valor promedio de los parámetros acústicos para cada género.

2.4.2 Edad

La laringe es un órgano con características sexuales secundarias, cuya maduración corre paralela a la del diencéfalo. La estructura laríngea y las características de la voz son un fiel reflejo de la edad del individuo, del género y del estado de salud. Esta maduración se prolonga a largo de los distintos períodos vitales que determinan modificaciones estructurales y fónicas notables. Los períodos que se consideran en este sentido son la niñez (neonatal, primera infancia, segunda infancia) la pubertad, la juventud y la senectud. Las modificaciones mas marcadas de la voz se dan en el varón en la pubertad y en la mujer en la senectud. Los grupos de edad son: segunda infancia, 6-11 años; pubertad, 12-18 años; periodo de estabilización de la voz, 19-50 años; senectud, de 59 en adelante.

Las características fónicas de cada período son:

- Neonatal: Se caracteriza por las altas frecuencias. El ataque del sonido es brusco, de fuerte intensidad y modulación muy reducida.
- Primera Infancia: El ataque de sonido se hace menos brusco. A los 18 meses aparece la modulación vocal.
- Segunda Infancia: Las variaciones vocales llegan hasta una octava y media de extensión.
- Pubertad: La mutación vocal se produce en el varón entre los 13 y 14 años, y en la mujer, entre los 14 y 15 años.

Al cumplirse el descenso laríngeo se hace notable la disminución de las frecuencias de los sonidos producidos, la pérdida de los armónicos, las resonancias de cabeza y faciales; pasan a predominar los armónicos y la resonancia pectoral.

- La senilidad vocal: es más precoz en la mujer que en el hombre y se presenta más marcada en la voz cantada que en la hablada (60 - 70 años). Se produce una pérdida de los sonidos agudos, disminución de la extensión, pérdida de potencia y disminución de los armónicos. En la mujer ocurre una agravación del tono de la voz.

Uso de la voz: En la práctica se adoptan cuatro niveles de empleo vocal:

- Usuario selecto: Corresponde al usuario selecto o especial, una persona en quien aún una ligera aberración vocal tendrá consecuencias desastrosas (por ejemplo, la mayoría de cantantes y actores).
- Profesional de la voz: se refiere a personas en quienes una moderada disfunción vocal impediría un adecuado desempeño laboral, (por ejemplo, la mayoría de sacerdotes, conferencistas y operadores de teléfonos, fonoaudiólogos, profesores, locutores).
- Profesionales no vocales: Se refiere a profesionales no vocales, como maestros, médicos, abogados, hombres de negocios o recepcionistas, es decir, aquellas personas que no podrán desempeñar de manera adecuada su trabajo si sufren disfonía grave.
- No profesionales no vocales: Se refiere al trabajador que no da a su voz un uso profesional entre las cuales se contarían obreros, oficinistas, etc. Si bien si algunos de los sujetos este grupo padecieran de algún trastorno vocal, éste no les impediría realizar su trabajo.

Paralelamente, se han formado otros criterios de clasificación de la voz, orientados a la voz cantada: tesitura y timbre[6].

2.4.3 Tesitura

La tesitura clasifica la voz por su amplitud tonal adecuada, en la que el cantante se mueve a su comodidad sin apurar las notas extremas. Un sentido correcto de interpretar la tesitura, es el que sitúa el conjunto de sonidos, en los que la voz se adapta mejor, es decir la parte de la gamma vocal, en que el cantante se siente cómodo, sin ningún tipo de fatiga. Se considera que las voces masculinas y femeninas se distribuyen en 6 tipos principales:

Tabla 3. Clasificación de las voces según su tesitura

Genero	***Aguda***	***Media***	***Grave***
Femenino	Soprano	Mezo - Soprano	Contralto
Masculino	Tenor	Barítono	Bajo

2.4.4 Timbre

El timbre, que se puede definir como la cualidad que permite diferenciar dos sonidos, que acusen una misma intensidad y frecuencia. Los sonidos no son puros, es decir, no tienen un movimiento armónico simple, sino que provienen de movimientos vibratorios complejos.

[6] Ibid., p. 14.

El timbre corresponde al número de armónicos que conforman el sonido. El timbre, en parte depende, del tipo de cuerdas vocales del individuo, del modo de vibración, y de las cajas de resonancia (senos paranasales, cavidades supralaríngeas, cavidad orofaríngea). Se han distinguido dos timbres en cada voz humana: Timbre vocálico y timbre extravocálico. El timbre vocálico corresponde a circunstancias fisiológicas condicionables, incluyendo aquí todas las técnicas de aprendizaje; mientras, el timbre extravocálico depende en exclusividad de la constitucionalidad laríngea, y es el que caracteriza la voz de cada individuo.

2.5 Estudio funcional de la voz

La vibración normal de los pliegues vocales consta de movimientos relativamente regulares, repetitivos, simultáneos y sincrónicos, a los que sigue en cada secuencia un periodo en el cual las cuerdas vocales entran en contacto entre sí para cerrar la glotis e interrumpir por un momento la corriente de aire. La vibración anormal del pliegue vocal adopta muchas formas, algunas de las cuales no son apreciables a simple vista. Las películas de movimiento ultralento de los pliegues vocales, así como la estroboscopia (examen que se explica con mayor detalle mas adelante), han revelado que una cuerda vocal puede moverse más de prisa que la otra, que a veces falta el cierre glótico, que el cierre puede ocurrir en posición paramedia, que los ciclos vibratorios no siempre son semejantes en diferentes regiones a lo largo de uno o ambos pliegues, y que los periodos y amplitudes vibratorias varían entre dos aperturas glóticas consecutivas.

De todo esto se desprende que la complejidad potencial de los tipos vibratorios resultantes de la combinación de estas anormalidades físicas e irregularidades secuenciales es casi interminable. Por consiguiente, si se supone que la ronquera, en sus diversas formas, obedece a una vibración anormal de las cuerdas vocales, su origen puede hallarse en una o varias de las desviaciones que se han mencionado[7].

2.5.1 Clasificación de los trastornos de la voz

Los trastornos de la voz afectan a los componentes del habla mencionados anteriormente (Tono, volumen y Calidad de la voz).
Cuando, en la voz del individuo, uno o varios de estos factores difieren notoriamente de los que se escuchan en las voces de la mayoría de las personas de sexo, edad y grupo cultural idénticos, se considera que su voz es defectuosa.

Trastornos en el tono de la voz

Los trastornos o patologías en el tono de la voz aparecen cuando la voz es mucho más alta o más baja (en relación con la escala musical) de lo que podría determinarse como normal para un individuo determinado, y cuando el sonido es tembloroso, monótono o grotesco. Las causas frecuentes son los desequilibrios hormonales, las membranas laríngeas o la aproximación anormal de las cuerdas vocales.

[7] BALLENGER, OP. cit., p. 12.

Trastornos en el volumen de la voz

Los trastornos o patologías en el volumen o intensidad de la voz pueden clasificarse en tres categorías, paralelas a las usadas para las desviaciones del tono. La voz puede ser demasiado voluminosa, o carecer del volumen suficiente en relación con el lugar y las circunstancias, o pueden existir variaciones que no son apropiadas para el significado de la expresión. Estas diferencias vocales se reflejan en conductas de tipos de personas tales como personas agresivas, tímidas o inestable social que se clasifican de índole funcional.

Existen también causas orgánicas importantes que pueden influir en el volumen vocal: la alteración de la audición, que conlleva a que el individuo hable con mayor volumen de los que exigen las circunstancias y la parálisis en los músculos laríngeos.

Trastornos en la calidad de la voz

Los trastornos de la calidad de la voz son los más comunes y complejos de los problemas vocales. Abarcan los componentes de resonancia y de fonación, que pueden mezclarse en una gran variedad de formas y se complican, además, por continuas variaciones en el grado de intensidad. Las diversas desviaciones fonatorias son catalogables en la escala auditiva que se extiende desde la afonía (o ausencia de sonido de fonación revelada en la producción de una voz cuchicheada) a la ronquera, con intervención de las calidades intermedias tales como sibilancia y rudeza.

Los trastornos de resonancia son también, pero están compuestos primordialmente de hipernasalidad y de hiponasalidad. Estos problemas suelen afectar más que los trastornos fonatorios la inteligibilidad del habla y constituyen trastornos del habla que revelan claves diagnósticas en potencia significativas que con frecuencia se ignoran.

2.5.2 Patologías de la voz

Las patologías de la voz se pueden agrupar en dos categorías: aquellas que generan disfonía y las que llevan a la afonía.

Disfonía

La voz se produce por la vibración de las cuerdas vocales al pasar por ellas el aire expelido de los pulmones. Cuando existe un problema en las cuerdas vocales, esta vibración es defectuosa y la voz sale con alteraciones (voz ronca, rasposa, apagada, entrecortada etc.). Este cambio de voz anormal se conoce como disfonía. La disfonía, es por tanto, un término general que describe un cambio anormal de la voz producida por muchos tipos de enfermedades.

Alguna de las causas de la disfonía son:

Laringitis aguda. Es la causa más frecuente de disfonía y ocurre por una inflamación de las cuerdas vocales debido a una infección viral o a un uso excesivo de la voz.

Nódulos de cuerdas vocales. Aparecen en personas con un mal uso vocal, que hablan muy alto, durante demasiado tiempo, o con mala técnica de emisión vocal.

Pólipos de cuerdas vocales. Las causas son las mismas que para los nódulos, pero aquí el componente inflamatorio es mayor.

Reflujo gastroesofágico. El reflujo de material gástrico, sobre todo durante la noche, puede producir irritación de las cuerdas vocales y disfonía. Ocurre con mayor frecuencia en personas mayores.

Cáncer de laringe. Esta causa de disfonía justifica por si sola la identificación de otras causas de alteración de la voz aunque sean aparentemente banales. Hay que sospecharla sobre todo ante un paciente fumador. El tabaco es la principal causa de cáncer de laringe.

Parálisis de cuerdas vocales. Por afectación del nervio recurrente debido a cirugía del tiroides o compresión consecuencia de tumoraciones, o sin causa aparente.

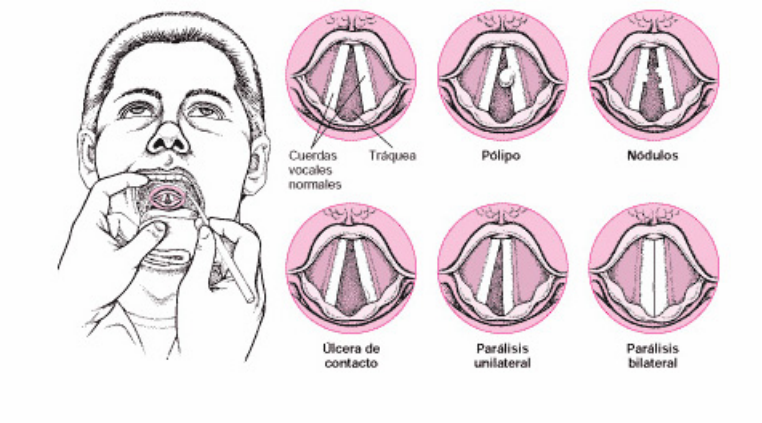
Afonía

Las razones de su aparición se deben a problemas en el sistema nervioso central, así como en patologías de la laringe. En este caso la voz no se genera, y el habla es posible solo en forma de susurro. La afonía se debe a la parálisis y/o cortes en los músculos de la laringe por la afección de la corteza cerebral o del cerebelo; o como causa de infecciones y traumas del nervio inferior de la laringe o en alguna de sus ramificaciones. Como

resultado de la parálisis en los músculos que sirven para la contracción y dilatación de la laringe, las cuerdas vocales no se cierran completamente y la voz desaparece.

En la siguiente figura se muestran algunas de las patologías de la laringe que suelen presentarse:

Figura 3. Patologías de la laringe



2.5.3 Evaluación clínica de la voz.

La historia de la medicina ha demostrado que el tratamiento médico y quirúrgico permite eliminar los trastornos de la voz, pero es de manifiesto también que con este tratamiento no siempre se logra el restablecimiento de la función normal. Pueden ser necesarias medidas no médicas, de rehabilitación, para ayudar a compensar las alteraciones anatómicas y fisiológicas, y los procedimientos reeducativos están casi siempre indicados en el tratamiento de los trastornos habituales o funcionales.

Los objetivos principales del examen clínico de la voz son[8]:

a) Hacer un diagnóstico.
b) Determinar el grado y extensión de la enfermedad
c) Evaluar el grado y la naturaleza de la disfonía.
d) Determinar el pronóstico.
e) Monitorear sus cambios.

El estudio de la voz no se hace solamente para determinar un diagnóstico de enfermedad etiológica sino, para obtener información de cómo se encuentran cada uno de los aspectos evaluados que intervienen en la emisión vocal precisando así las cualidades de la voz: intensidad, altura tonal, timbre, duración.

Se considera que para obtener un estudio completo de la voz se requiere:

- Valoración del Otorrinolaringólogo.
- Entrevista estructurada.
- Análisis acústico de la voz.
- Examen respiratorio y de órganos fonoarticuladores.
- Examen de la postura corporal.

Dichos exámenes aclaran el panorama para tener una adecuada orientación en el trabajo de entrenamiento ó de reeducación; De allí nace el trabajo conjunto de los otorrinolaringolos y los fonoaudiólogos, estos últimos son los encargados de realizar los procedimientos terapéuticos.

[8] TOBON, Libia y CASTELLANOS, Germán. Diseño y desarrollo de un sistema interactivo de análisis acústico de la voz y el habla para la ciudad de Manizales. Manizales.

Entre los diversos procedimientos de rutina para la examinación de la laringe con propósitos clínicos o de investigación, los cuales incluyen:

- Laringoscopia fibroscópica rígida y flexible: examinación con un instrumento de fibra óptica.
- Estroboscopia laríngea: iluminación estroboscópica de la laringe, útil para la visualización de movimientos.
- Electromiografía: Observación indirecta del estado funcional de la laringe.
- Análisis Acústico.

Estroboscopia laríngea

La estroboscopia laríngea es valorada por los especialistas en voz como el más importante procedimiento diagnóstico disponible en la actualidad para la evaluación de pacientes con trastornos de la voz. La vibración de los pliegues vocales durante la fonación es un fenómeno muy complejo y difícil de apreciar a simple vista bajo luz normal, debido a que bajo condiciones normales estos vibran a frecuencias muy altas. Las frecuencias de vibración de un pliegue vocal en un tono conversacional, y bajo condiciones normales es de aproximadamente 120-140 Hz en hombres, de aproximadamente 200-220 Hz en mujeres y de 280 o más Hz aproximadamente en niños.

La luz estroboscópica (luz de flash) proporciona una imagen de aparente lentitud de las vibraciones para que estas puedan ser observadas y analizadas. Con este fin se diseñó un aparato llamado Estroboscopio.

La luz de flash de alta intensidad que este emite es disparada a la misma frecuencia de vibración de los pliegues vocales del paciente. Su objetivo es tomar imágenes individuales fijas de diferentes puntos de ciclos vibratorios sucesivos, y estas imágenes son presentadas como un movimiento vibratorio del pliegue vocal a una frecuencia mucho más baja. En la siguiente figura se ilustra la forma como se realiza un examen de estroboscopia laríngea.

Figura 4. Examen de Estroboscopia Laríngea.

Para mayor información sobre los métodos de evaluación, procedimientos de diagnóstico y tratamientos de las patologías de la laringe se puede consultar la página del Doctor Luís Humberto Jiménez Fandiño[9].

Análisis Acústico

El *Análisis Acústico* de la voz se ha convertido en los últimos años como una alternativa en la determinación de voces normales y patológicas.

[9] Medico Cirujano y Otorrinolaringólogo. Universidad Javeriana. Bogotá D.C http://www.laringeyvoz.com/cirugias.htm

Este tipo de análisis demuestra grandes ventajas sobre los métodos tradicionales debido a su naturaleza no invasiva (no se necesita introducir algún elemento en el tracto vocal) y a su potencial para proveer una medida cuantitativa acerca del estado clínico del funcionamiento de la laringe y el tracto vocal. De este modo, un sistema automático, confiable, preciso y no invasivo para el reconocimiento y monitoreo de anormalidades del habla es una herramienta necesaria en su valoración y evaluación.

Actualmente, existe la tecnología que permite evaluar de manera objetiva la acústica y fisiología del fenómeno, además, provee la retroalimentación visual de los mecanismos de producción vocal, para comprobar el diagnostico realizado con pruebas subjetivas. El empleo de sistemas computarizados en la caracterización acústica y representación de la voz, genera la posibilidad de analizar indicadores imperceptibles al oído humano, lo que ha permitido adoptarlos como una herramienta de apoyo al diagnostico con una amplia y creciente aceptación.

En este sentido, se han diseñado varios procedimientos, basados en la medición de parámetros o características acústicas de la voz y de los fenómenos aerodinámicos que intervienen en la emisión vocal. En la descripción del análisis acústico de la voz se utilizan los indicadores físicos del sonido como son la frecuencia, la intensidad, la composición espectral, las variaciones del sonido modificadas por la resonancia que actúan originando el producto sonoro percibido siendo importante resaltar el pitch (definido como la frecuencia fundamental percibida de la señal de voz), la sonoridad, el timbre y los formantes (que son regiones de énfasis llamadas resonancias o regiones de desénfasis llamadas antiresonancias que se encuentran en el espectro de la señal de voz).

Estos procedimientos permiten establecer el diagnostico de la alteración vocal, siendo interesantes en múltiples aspectos: proporcionan una imagen inicial de algunas deficiencias que manifiesta el malestar vocal y que permiten que la persona comprenda mejor su trastorno; en ocasiones orientan la reeducación al sugerir la aplicación de técnicas especializadas, según las deficiencias que se hayan encontrado, y facilitan, asimismo, el seguimiento de la evolución durante el tratamiento demostrando, por ejemplo, la existencia de la mejoría de un parámetro cuya valoración subjetiva por parte del paciente o del terapeuta puede ponerse en tela de juicio. Por ultimo, estos métodos pueden utilizarse para detectar a personas con riesgo, a las que podría aplicarse provechosamente una pedagogía preventiva.[10]

[10] LE HUCHE, Fracois, Patología Vocal: Semiología y disfonías disfuncionales. MASON, 1994, vol. 2.

3 REPRESENTACIÓN DE LA SEÑAL DE VOZ

En 1807 Jean Baptiste Fourier propuso que una señal periódica $x(t)$ con periodo T_0 se puede representar como una sumatoria de señales exponenciales complejas (o lo que es lo mismo, senos y cósenos), con Amplitudes y Frecuencias que varían armónicamente, en función de la frecuencia fundamental de la señal:

$$w_0 = \frac{2\pi}{T_0}$$

La representación en series de Fourier para una señal continua es de la forma:

$$x(t) = \sum_{k=-\infty}^{+\infty} a_k e^{jkw_0 t} = \sum_{k=-\infty}^{+\infty} a_k e^{jk\left(\frac{2\pi}{T}\right)t}$$

$$a_k = \frac{1}{T}\int_T x(t)e^{-jkw_0 t}dt = \frac{1}{T}\int_T x(t)e^{-jk\left(\frac{2\pi}{T}\right)t}dt$$

A la primera ecuación se le conoce con el nombre de ecuación se síntesis y a la segunda como ecuación de análisis. El conjunto de coeficientes $\{a_k\}$ se conoce a menudio como coeficientes de la serie de Fourier o coeficientes espectrales de $x(t)$. Estos coeficientes complejos miden la porción de la señal $x(t)$ que está en cada armónica de la componente fundamental.

En el año de1829 P.L. Dirichlet suministra las bases matemáticas precisas, bajo las cuales una señal $x(t)$, puede representarse en una serie de Fourier (condiciones de Dirichlet), para un Periodo de la señal se debe cumplir que[11]:

[11] OPPENHEIM, Alan and WILLSKY, Alan. Señales y Sistemas. Segunda ed. México:Prentice Hall,1998.

1. $x(t)$ ha de ser integrable y esa integral sea finita, es decir:

$$\int_T |x(t)dt| < \infty$$

2. Ha de existir un número finito de máximos y mínimos.

3. En cualquier intervalo de tiempo hay sólo un número finito de discontinuidades. Además, cada una de estas discontinuidades debe ser finita.

3.1 Transformada Contínua de Fourier

El desarrollo de la transformada de Fourier es también una de las contribuciones mas importantes realizada por Fourier y es una extensión de los estudios realizados para las series aplicándolo a señales aperiódicas. En particular, Fourier razonó que una señal aperiódica puede considerarse como una señal periódica con un periodo infinito. De manera más precisa, en la representación en serie de Fourier de una señal periódica, conforme el periodo se incrementa, la frecuencia fundamental disminuye y las componentes relacionadas armónicamente se hacen más cercanas en frecuencia. A medida que el periodo se hace infinito, las componentes de frecuencia forman un continuo y la suma de la serie se convierte en una integral. Dado lo anterior se obtiene:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(jw) e^{jwt} dw$$

$$X(jw) = \int_{-\infty}^{+\infty} x(t) e^{-jwt} dt$$

A estas ecuaciones se les conoce como el par de transformadas de Fourier, cuya función $X(jw)$ se conoce como transformada de Fourier o integral de fourier de $x(t)$ y a la primera ecuación se le conoce como la transformada inversa de Fourier[12].

3.2 Transformada de Fourier en Tiempo Corto (STFT)

Como ya es sabido, la transformada de Fourier constituye una herramienta mediante la cual se puede conocer todas las componentes de frecuencia existentes en la señal. La resolución representa la cantidad de información presente en una señal y cuando se hace referencia a la Transformada de Fourier se dice que esta posee una muy buena resolución en frecuencia ya que al evaluarla sobre una señal que contiene varias componentes de frecuencia, sus valores se visualizan de forma clara en el espectro. La Transformada de Fourier es una herramienta muy útil para el análisis de señales estacionarias. Por otra parte, esta transformada no puede ser aplicada con el objeto de obtener información precisa de cuando o donde se encuentran las diferentes componentes de frecuencia existentes en la señal como es el caso de señales no estacionarias cuyo contenido espectral varía con el tiempo. En otras palabras, la transformada de Fourier posee una muy pobre resolución en tiempo.

Una solución propuesta al problema de la localización tiempo-frecuencia proviene de los trabajos de Denis Gabor, quien adaptó la transformada utilizando un procedimiento llamado ventanamiento (**windowing**).

[12] Ibid., p. 29.

Este procedimiento consiste en dividir una señal $f(t)$ en pequeños segmentos a través del tiempo de tal manera que podamos asumir que para cada segmento la señal es estacionaria y así calcular la Transformada de Fourier clásica para cada porción de la señal[13].

La forma de dividir la señal se realiza mediante una función tiempo-ventana $h(t)$ cuyo ancho o soporte corresponde a la longitud de cada segmentación de la señal. Con la función ventana encuadramos la señal alrededor de un instante de tiempo τ y calculamos su transformada de Fourier, luego trasladamos la función ventana sin sobreponerse al espacio anterior cubriendo una nueva porción de la señal a la cual se vuelve a calcular su transformada de Fourier. Este proceso es repetido hasta que se ha cubierto la totalidad de la señal.

El resultado de lo expresado anteriormente se define en forma matemática de la siguiente manera:

$$STFT(t,w) = \int_{-\infty}^{\infty} f(t)h(t-\tau)e^{(-iwt)}dt$$

Esta expresión calcula el producto interno entre la señal y la función tiempo-ventana trasladada.

[13] FAUNDEZ, Pablo y FUENTES, Álvaro. Procesamiento Digital de Señales Acústicas utilizando Wavelets. Instituto de Matemáticas UACH.

3.2.1 Resolución Tiempo – Frecuencia

El ancho o soporte de la ventana constituye un parámetro de gran importancia ya que permite establecer el grado de resolución tanto de tiempo como de frecuencia. La resolución determina la cantidad de información presente en una señal; a mayor información mayor resolución tendrá una señal. Si la ventana es muy angosta, esta analizará una porción muy pequeña de la señal lo que permite tener una buena resolución en tiempo pero una mala resolución en frecuencia ya que se analizará sólo una mínima fracción del espectro total existente en la señal. Por otro lado, si la ventana en muy ancha se tendrá buena resolución en frecuencia pero una mala resolución en tiempo (una ventana de ancho infinito equivale a la transformada de Fourier clásica). Por lo tanto un defecto de la STFT es que no puede entregar una buena resolución tanto en tiempo como en frecuencia de manera instantánea ya que el soporte de la ventana es fijo. La raíz de este problema se basa en el principio de incertidumbre de Heisenberg, el cual establece que es imposible conocer una representación exacta tiempo - frecuencia de una señal, es decir, no se puede saber con certeza que valor de frecuencia existe en un instante de tiempo determinado, sólo se puede conocer que componentes de frecuencia existen dentro de un intervalo de tiempo determinado[14].

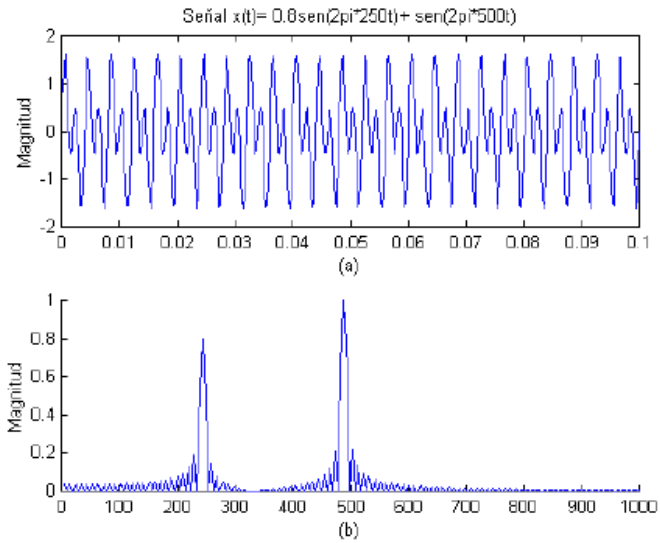
Para ilustrar lo anterior se plantea el siguiente ejemplo: Se tiene una señal $x(t)$ compuesta por sólo dos frecuencias dentro de un intervalo de tiempo igual a una décima de segundo.

$$x(t) = 0.8sen(2\pi 250t) + sen(2\pi 500t)$$

[14] Ibid., p. 32.

Al calcular la transformada de Fourier de esa señal se muestra resolución perfecta en frecuencia como lo ilustra la siguiente figura.

Figura 5. A) grafica de la señal x(t). B) espectro de la señal obtenido mediante la transformada de Fourier

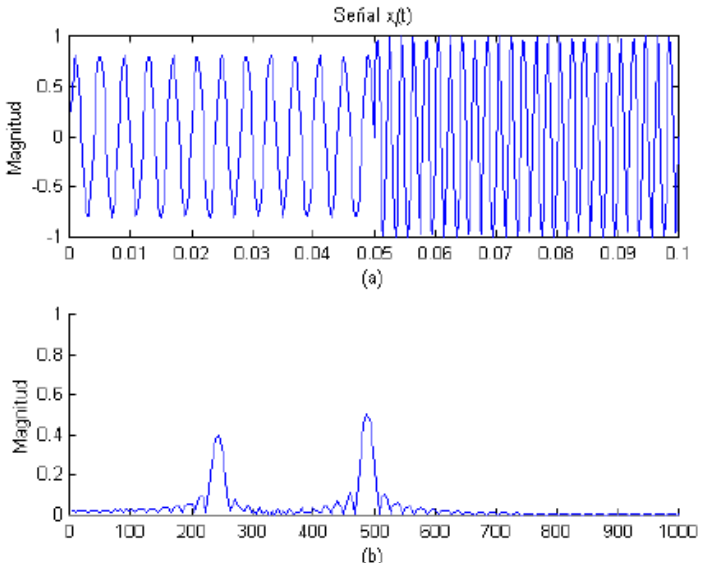


Ahora, si se tiene otra señal $x_1(t)$ con las mismas componentes de frecuencia sobre el mismo intervalo de tiempo, pero con la diferencia que las primeras 5 centésimas de segundo contienen a la frecuencia de 250 Hz y las otras 5 centésimas de segundo restante contienen a la frecuencia de 500 Hz, lo que se define como:

$$x_1(t) = \begin{cases} 0.8sen(2\pi 250t) & 0 \le t \le 0.05 \\ sen(2\pi 500t) & 0.05 \le t \le 0.1 \end{cases}$$

Al aplicar la Transformada de Fourier sobre $x_1(t)$, se observa que también se puede obtener las frecuencias de la señal pero con una amplitud igual a la mitad de la amplitud real debido a que cada componente de frecuencia se encuentra sólo la mitad del tiempo de análisis de la señal como se ilustra en la siguiente figura.

Figura 6. A) grafica de la señal x1(t). B) espectro de la señal obtenido mediante la transformada de Fourier



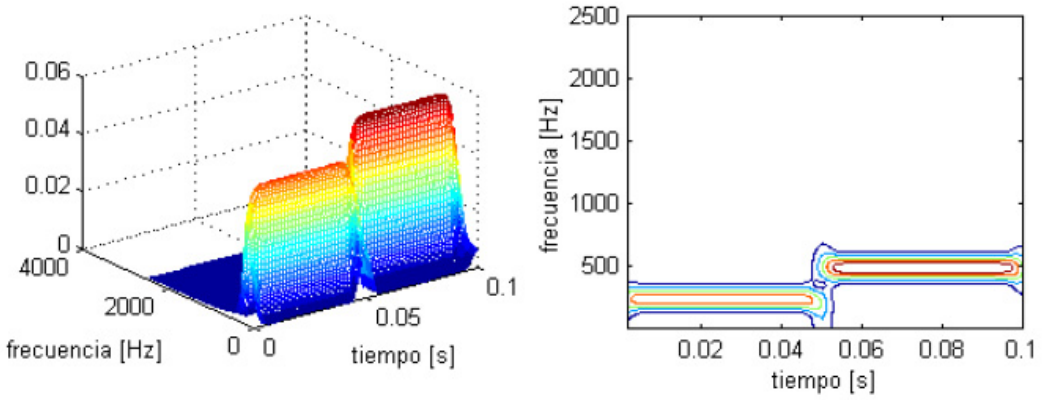
Esta información errónea se debe a que la transformada de Fourier, como se expresó en un principio, no puede determinar en que momento dentro de la señal se encuentra una respectiva componente de frecuencia.

Ahora si se analiza la función anterior con la transformada corta de Fourier (STFT), empleando la función gaussiana como función tiempo – ventana, que se expresa como:

$$h(t-\tau) = e^{-\frac{\pi}{a^2}(t-\tau)^2}$$

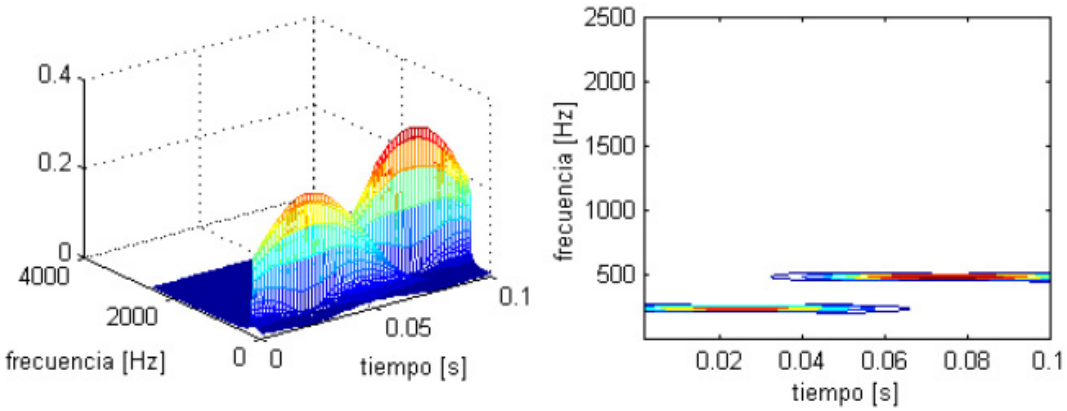
Donde a es el factor que controla el ancho o soporte de h. El primer análisis para un valor de **a=20** se ilustra en la siguiente figura.

Figura 7. Representación Tiempo – Frecuencia con buena resolución en tiempo y mala resolución en frecuencia.



Al ser angosta la ventana utilizada podemos observar que la resolución en el tiempo es buena ya que se diferencia claramente la posición en el tiempo de cada componente de frecuencia. Sin embargo, la resolución en frecuencia es bastante pobre ya que para cada componente se observa un ancho de banda amplio lo que impide una detección precisa del valor real de la frecuencia existente en el intervalo de tiempo donde se encuentra. El segundo análisis se efectúa para un valor de **a=250** y se ilustra en la siguiente figura.

Figura 8. Representación Tiempo – Frecuencia con buena resolución en Frecuencia y mala resolución en tiempo.



Este aumento de **a** significa que la función tiempo – ventana es más ancha y por lo tanto se ha mejorado la resolución en frecuencia ya que el ancho de banda de cada componente ha disminuido permitiéndonos identificar claramente cada frecuencia. Por otro lado la resolución en tiempo se ha empobrecido producto de la mejora en la resolución en frecuencia ya que no se observa una clara separación de la ubicación de cada componente en su respectivo intervalo de tiempo.

3.3 Transformada Continúa Wavelet (CWT)

La transformada wavelet constituye una técnica relativamente nueva que ha sido propuesta por los investigadores como una poderosa herramienta en el análisis sobre el comportamiento local de una señal.

Al igual que la STFT, esta transformada utiliza una función ventana que encuadra una señal dentro de un intervalo y focaliza el análisis sólo en ese segmento de la señal. La ventaja de la transformada wavelet es el empleo de una función ventana que tiene la capacidad de cambiar su soporte o ancho en forma automática dependiendo del contenido espectral del segmento de la señal analizado, ya que una situación ideal de análisis sería tener una buena resolución en tiempo para frecuencias altas y una buena resolución en frecuencia frente a contenido de frecuencias bajas.

La transformada continua wavelet intenta expresar una señal $x(t)$ continua en el tiempo, mediante una expansión de términos o coeficientes proporcionales al producto interno entre la señal y diferentes versiones escaladas y trasladadas de una función prototipo $\psi(t)$ más conocida como wavelet madre.

Asumiendo que tanto la señal como la nueva función $\psi(t)$ son de energía finita, entonces se puede definir transformada continua Wavelet mediante la siguiente expresión:

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi(\frac{t-b}{a})dt$$

Ahora utilizando el teorema de Parseval podemos escribir la ecuación anterior en términos de la Transformada de Fourier de $x(t)$ y $\psi(t)$ como:

$$CWT(a,b) = \frac{1}{2\pi\sqrt{a}} \int_{-\infty}^{\infty} X(w)\Psi(aw)e^{-iwb}dt$$

Se resalta que en las ecuaciones anteriores aparecieron dos nuevas variables: La variable a controla el ancho o soporte efectivo de la función ψ, y la variable b nos da la ubicación en el dominio del tiempo de ψ.

Ahora, para que este análisis sea posible y además para poder lograr una perfecta reconstrucción de la señal a partir de la transformada, la función $\psi(t)$ debe cumplir con la condición de admisibilidad, de la cual se desprende que:

$$\Psi(0) = 0$$

Donde $\Psi = \Psi(w)$ corresponde a la transformada de Fourier de $\psi(t)$. El cumplimiento de esta condición significa que el valor medio de ψ es igual a 0, lo que a su vez implica obligatoriamente que ψ tenga valores tanto positivos como negativos, es decir, que ψ sea una onda. Además como es una función que "enventana" la señal sobre un intervalo de tiempo dado por a alrededor de un punto $t = b$ se observa intuitivamente que ψ es de soporte compacto, es decir, ψ es una onda definida sobre un intervalo de tiempo finito, y esto es el porque de su nombre wavelet o ondita[15].

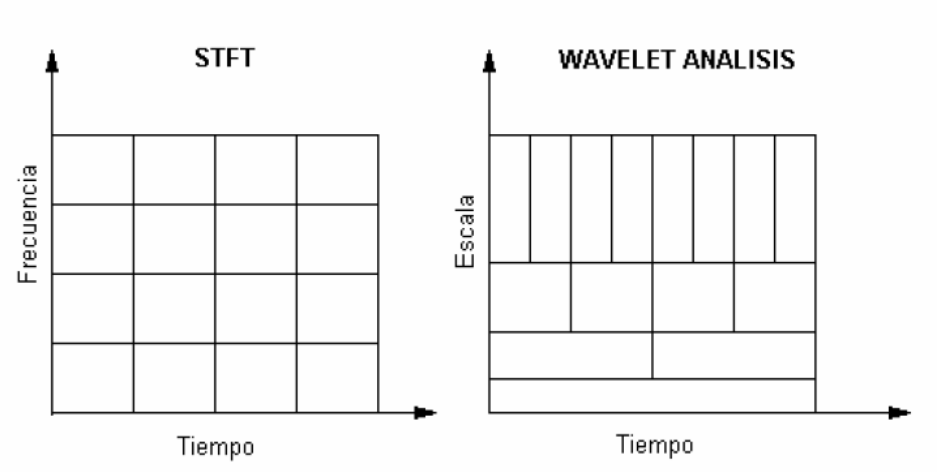
3.3.1 Variables de escala *a* y traslación *b*

Mediante la variable de escala se puede comprimir ($|a| < 1$) o dilatar ($|a| > 1$) la función $\psi(t)$, lo que determina el grado de resolución con el cual se esté analizando la señal.

[15] Ibid., p. 32.

Por definición la Transformada Continua Wavelet es mas una representación tiempo-escala que una representación tiempo- frecuencia. En particular, para valores pequeños de a la CWT obtiene información de $x(t)$ que está esencialmente localizada en el dominio del tiempo mientras que para valores grandes de a la CWT obtiene información de $X(w)$ que está localizada en el dominio de la frecuencia. En otras palabras, para escalas pequeñas la CWT nos entrega una buena resolución en el dominio del tiempo mientras que para escalas grandes la CWT nos entrega una buena resolución en el dominio de la frecuencia. Cuando a cambia, tanto la duración como el ancho de banda de la wavelet ψ cambian pero su forma se mantiene igual. En lo anterior se encuentra la diferencia principal entre la CWT y la STFT, ya que la primera ocupa ventanas de corta duración para altas frecuencias y ventanas de larga duración para bajas frecuencias mientras que la STFT ocupa una sola ventana con la misma duración tanto para altas frecuencias como para bajas frecuencias.

Figura 9. Diferencia Tiempo – Frecuencia v/s Tiempo – Escala entre la STFT y la CWT



Aunque la CWT trabaja con el término escala en vez de frecuencia, es posible mediante una constante $c > 0$ realizar un cambio de variable de una escala a una frecuencia w de la forma:

$$a \to w = \frac{c}{a}$$

Donde c recibe el nombre de constante de calibración en Hz. Con este cambio de variable se puede observar que la CWT localiza tanto la señal $x(t)$ en el dominio del tiempo como su espectro $X(w)$ en el dominio de la frecuencia en forma simultánea.

La variable b controla la ubicación de la función en el espacio de tiempo permitiendo deslizar $\psi(t)$ sobre el intervalo de tiempo en el que se haya definido $x(t)$. Un punto importante es que la función wavelet ψ se traslada cubriendo toda la señal para cada valor de a, es decir, si la escala escogida es pequeña habrán más traslaciones de ψ que si la escala escogida es grande. Por lo tanto, la variable b nos da la cantidad por la cual $\psi(\frac{t}{a})$ ha sido trasladada en el dominio del tiempo.

3.4 Transformada Discreta Wavelet (DWT)

La continuidad de la CWT reside en que tanto la variable de escala como la variable de traslación varían en forma continua. La carga computacional de la CWT es enorme, debido a la alta redundancia en la descomposición. Por lo tanto es imprescindible discretizar la transformada.

Para tal propósito se utiliza la Transformada Discreta Wavelet (DWT) la cual emplea una red diádica para discretizar los valores de a y b. es decir $a = 2^{-j}$ y $b = k2^{-j}$ con $k, j \in Z$, obteniendo el conjunto de funciones:

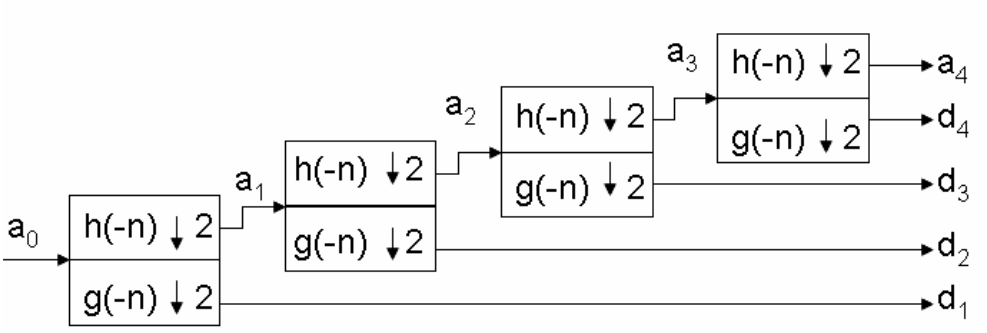
$$\psi_{j,k}(t) = 2^{\frac{i}{2}}\psi(2^j t - k) \quad j,k \in Z$$

Donde j representa la escala y k la traslación temporal.

Esta ecuación permite obtener lo que se denomina un análisis multirresolución. Una vía para implementar este esquema usando filtros fue desarrollada por Mallat, cuyo algoritmo es en efecto un esquema clásico conocido como banco de filtro de octavas.

En el proceso de descomposición o análisis wavelet, una señal $f(t)$ se representa como una serie de aproximaciones (baja frecuencia, alta escala) y detalles (alta frecuencia, baja escala) en diferentes resoluciones. En cada etapa, un par de filtros FIR $h(-n)$ (pasa bajo) y $g(-n)$ (pasa alto), son aplicados a la señal de entrada para producir una señal de aproximación y una de detalle respectivamente. La señal de detalle, representa la información perdida desde una resolución alta, hasta una mas baja. La etapa de filtrado es seguida por una operación de submuestreo por un factor de 2 que realiza un diezmado de la señal original, es decir, toma una señal x_n y produce una salida $y_n = x_{2n}$ tomando todos los valores de índice impar. Este proceso se ilustra en la siguiente figura.

Figura 10. Etapa de análisis o descomposición.



El algoritmo rápido para calcular los coeficientes Wavelet, está dado por la siguiente expresión:

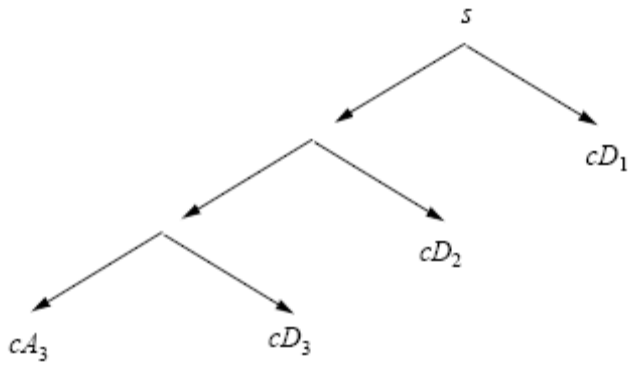
$$ca_{j,k} = \sum_m h[2k-m]\, ca_{j-1}[m]$$
$$cd_{j,k} = \sum_m g[2k-m]\, ca_{j-1}[m]$$

Los filtros h y g son llamados filtros espejo en cuadratura y satisfacen la siguiente propiedad:

$$g[n] = (-1)^{1-n} h[1-n]$$

La representación Wavelet es entonces, el conjunto de coeficientes de detalle en todas las resoluciones y los coeficientes de aproximación en la resolución más baja, es decir, $[ca_j, cd_j, ..., cd_1]$

Figura 11. Estructura de la descomposición Wavelet: árbol Wavelet



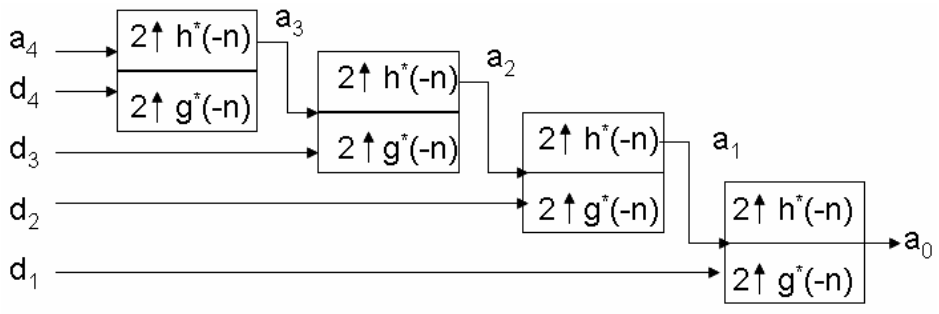
Ahora si se quiere recuperar la señal original sin pérdida de información a partir de las componentes obtenidas durante el análisis o descomposición se realiza el proceso llamado síntesis o reconstrucción y corresponde a la inversa de la transformada discreta wavelet (IDWT). En otras palabras, lo que se desea hacer es poder representar los coeficientes escala en un nivel de resolución más alto mediante una combinación de los coeficientes escala y wavelets en un nivel de resolución más bajo.

Este proceso se realiza mediante la siguiente ecuación:

$$ca_{j-1,k} = 2\sum_{m}(ca_{j,k}[m]h[k-2m] + cd_{m,k}[m]g[k-2m])$$

La estructura de reconstrucción se muestra en la siguiente figura.

Figura 12. Etapa de síntesis o reconstrucción.



Así como en el análisis se hace un filtrado y un submuestreo, en la síntesis se realiza un supmuestreo y posteriormente un filtrado. El supmuestreo es una operación que inserta ceros entre cada muestreo con el fin de aumentar al doble la longitud de las componentes de entrada (coeficientes de aproximación o escala y coeficientes de detalle o wavelet) de tal manera que la señal obtenida después del filtrado tenga la misma longitud que la señal original.

4 REDES NEURONALES ARTIFICIALES

Las Redes Neuronales Artificiales (RNA) son redes conformadas por múltiples nodos no lineales interconectados en paralelo y con organización jerárquica, las cuales intentan interactuar con el mundo real como lo hace una red neuronal biológica. El comportamiento conjunto de la red demuestra la habilidad para aprender y generalizar a partir de los patrones o secuencias de entrenamiento[16].

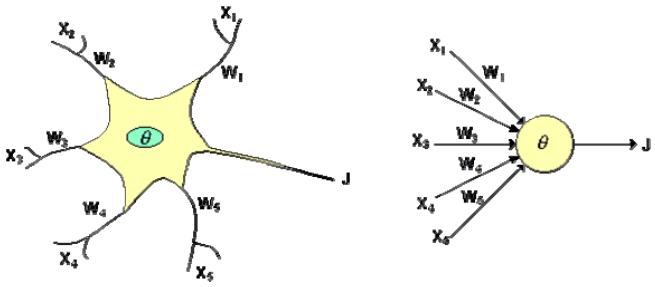
La información presentada a continuación fue extraída del trabajo,**"Tutorial sobre redes neuronales aplicadas a la ingeniería eléctrica y su implementación en un sitio web". Realizado por Maria Isabel Acosta, y Camilo Zuloaga de la Universidad Tecnológica de Pereira.**

4.1.1 Estructura de una red neuronal

El modelo de una neurona artificial es una imitación del proceso de una neurona biológica. Existen varias formas de nombrar una neurona artificial, la cual es conocida como nodo, neuronodo, celda, unidad o elemento de procesamiento (PE). En la siguiente figura se observa un PE en forma general y su similitud con una neurona biológica.

[16] Hilera, José y Martínez Víctor. Redes Neuronales Artificiales, Fundamentos, Modelos y aplicaciones. Madrid: Alfaomega, 1995.

Figura 13. De la neurona biológica a la neurona artificial



Las señales de entrada a una neurona artificial X1, X2,.., Xn son variables continuas en lugar de pulsos discretos, como se presentan en una neurona biológica. Cada señal de entrada pasa a través de una ganancia o peso, los cuales pueden ser positivos (excitatorios), o negativos (inhibitorios), el nodo sumatorio acumula todas las señales de entradas multiplicadas por los pesos o ponderadas y las pasa a la salida a través de una función umbral o función de transferencia.

La entrada neta a cada unidad puede escribirse de la siguiente manera:

$$neta_i = \sum_{i=1}^{n} W_i X_i = \vec{X}\vec{Y}$$

Donde,

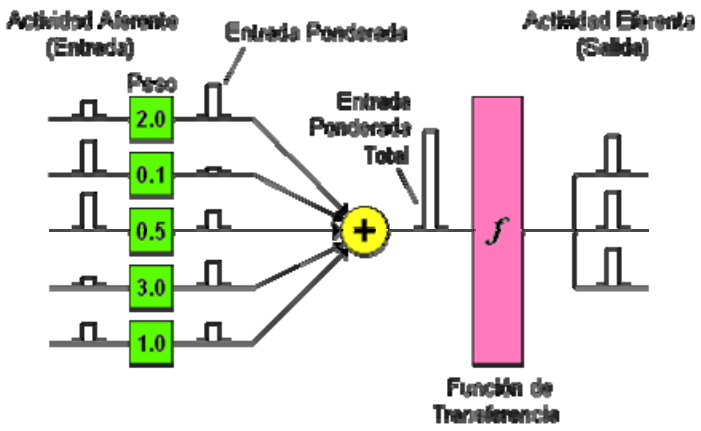
W_i: i-ésimo peso

X_i: i-ésima entrada

$neta_i$: Salida neta de la red

Una idea clara de este proceso se muestra en la figura 15, en donde puede observarse el recorrido de un conjunto de señales que entran a la red.

Figura 14. Proceso de una red neuronal



Una vez que se ha calculado la activación del nodo, el valor de salida equivale a:

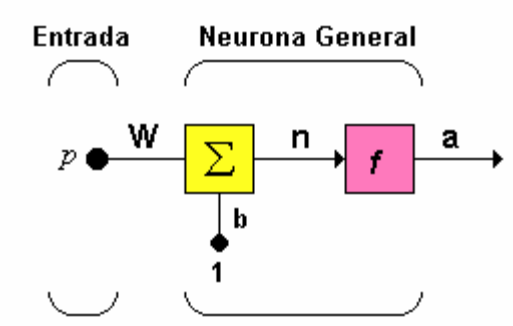
$$X_i = f_i(neta_i)$$

Donde f_i representa la función de activación para esa unidad, que corresponde a la función escogida para transformar la entrada $neta_i$ en el valor de salida X_i y que depende de las características específicas de cada red.

4.1.2 Funciones de transferencia

El modelo de una red neuronal artificial se ilustra en la siguiente figura:

Figura 15. Neurona de una sola entrada



Las entradas a la red serán ahora presentadas en el vector p, que para el caso de una sola neurona contiene solo un elemento, w sigue representando los pesos y la nueva entrada b es una ganancia que refuerza la salida del sumador n, la cual es la salida neta de la red; la salida total está determinada por la función de transferencia, la cual puede ser una función lineal o no lineal de n. Además este modelo puede ser descrito por medio de la ecuación:

$$a = f(wp + b)$$

Cabe notar que W y b son parámetros ajustables de la neurona. La idea central de las redes neuronales es que dichos parámetros pueden ser ajustados tal que la red exhiba algún comportamiento deseado. De este modo, se pueden entrenar redes para realizar un trabajo particular ajustando estos parámetros, o tal vez la propia red pueda ajustarse para alcanzar alguna salida deseada.

La siguiente tabla hace una relación de las principales funciones de transferencia empleadas en el entrenamiento de redes neuronales.

Tabla 4. Funciones de transferencia

Nombre	Relación Entrada /Salida	Icono	Función
Limitador Fuerte	$a = 0\ n < 0$ $a = 1\ n \geq 0$		*hardlim*
Limitador Fuerte Simétrico	$a = -1\ n < 0$ $a = +1\ n \geq 0$		*hardlims*
Lineal Positiva	$a = 0\ n < 0$ $a = n\ 0 \leq n$		*poslin*
Lineal	$a = n$		*purelin*
Lineal Saturado	$a = 0\ n < 0$ $a = n\ 0 \leq n \leq 1$ $a = 1\ n > 1$		*satlin*
Lineal Saturado Simétrico	$a = -1\ n < -1$ $a = n\ \ -1 \leq n \leq 1$ $a = +1\ n > 1$		*satlins*
Sigmoidal Logarítmico	$a = \frac{1}{1+e^{-n}}$		*logsig*
Tangente Sigmoidal Hiperbólica	$a = \frac{e^{n} - e^{-n}}{e^{n} + e^{-n}}$		*tansig*
Competitiva	$a = 1$ *Neurona con n max* $a = 0$ *El resto de neuronas*		*compet*

4.1.3 Topología de una red.

Dentro de una red neuronal, los elementos de procesamiento se encuentran agrupados por capas, una capa es una colección de neuronas; de acuerdo a la ubicación de la capa en la RNA, esta recibe diferentes nombres:

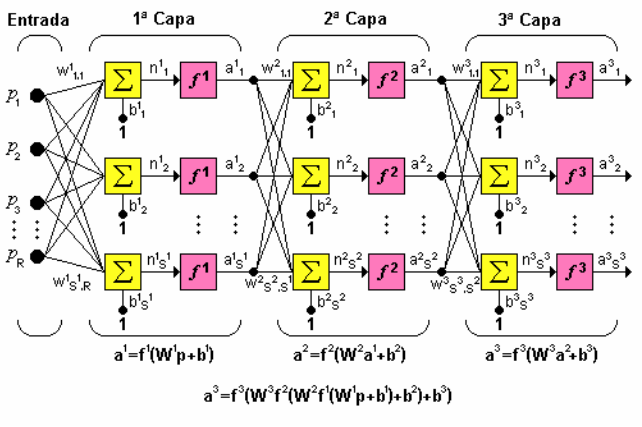
Capa de entrada: Recibe las señales de la entrada de la red, algunos autores no consideran el vector de entrada como una capa pues allí no se lleva a cabo ningún proceso.

Capas ocultas: Estas capas son aquellas que no tienen contacto con el medio exterior, sus elementos pueden tener diferentes conexiones y son estas las que determinan las diferentes topologías de la red

Capa de salida: Recibe la información de la capa oculta y transmite la respuesta al medio externo

En una red con varias capas, o red multicapa, cada capa tendrá su propia matriz de peso W, su propio vector de ganancias b, un vector de entradas netas n, y un vector de salida a. La topología de una red neuronal de tres capas se muestra a continuación:

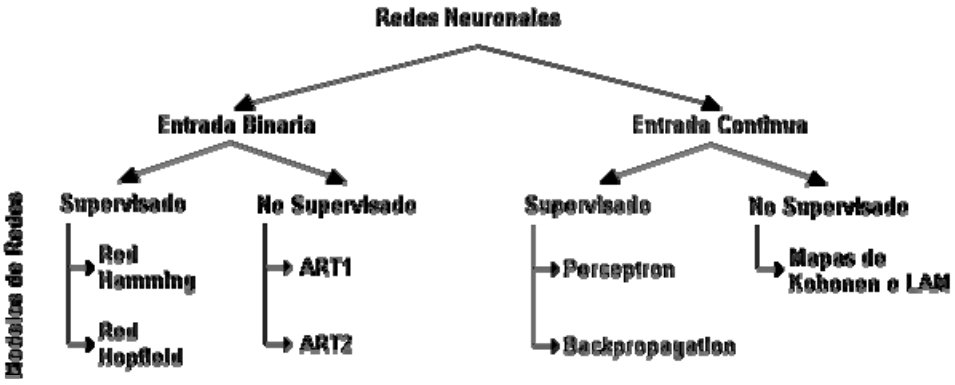
Figura 16. Red Neuronal de tres capas



4.1.4 Clasificación de las Redes Neuronales

En general las redes neuronales se pueden clasificar de diversas maneras, según su topología, forma de aprendizaje (supervisado o no supervisado), tipos de funciones de activación, valores de entrada (binarios o continuos). En la siguiente figura se puede observar de la clasificación de la redes.

Figura 17. Clasificación de las redes neuronales



4.1.5 Redes Backpropagation

En 1986, los investigadores, David Rumelhart, Geoffrey Hinton y Ronald Williams, basando en los trabajos de Paul Werbos y David Parker Formalizaron un método para que una red neuronal aprendiera la asociación que existe entre los patrones de entrada a la misma y las clases correspondientes, utilizando más niveles de neuronas que los que utilizó Rosenblatt para el Perceptron. Este método, conocido como backpropagation (propagación del error hacia atrás), esta basado en la generalización de la regla delta, y a pesar de sus limitaciones, ha ampliado de forma considerable el rango de aplicaciones de las redes neuronales.

La Backpropagation es un tipo de red de aprendizaje supervisado, que emplea un ciclo propagación – adaptación de dos fases. Una vez que se ha aplicado un patrón a la entrada de la red como estímulo, este se propaga desde la primera capa a través de las capas superiores de la red, hasta generar una salida. La señal de salida se compara con la salida deseada y se calcula una señal de error para cada una de las salidas.

Las salidas de error se propagan hacia atrás, partiendo de la capa de salida, hacia todas las neuronas de la capa oculta que contribuyen directamente a la salida. Sin embargo las neuronas de la capa oculta solo reciben una fracción de la señal total del error, basándose aproximadamente en la contribución relativa que haya aportado cada neurona a la salida original. Este proceso se repite, capa por capa, hasta que todas las neuronas de la red hayan recibido una señal de error que describa su contribución relativa al error total.

Basándose en la señal de error percibida, se actualizan los pesos de conexión de cada neurona, para hacer que la red converja hacia un estado que permita clasificar correctamente todos los patrones de entrenamiento.

La importancia de este proceso consiste en que, a medida que se entrena la red, las neuronas de las capas intermedias se organizan a sí mismas de tal modo que las distintas neuronas aprenden a reconocer distintas características del espacio total de entrada. Después del entrenamiento, cuando se les presente un patrón arbitrario de entrada que contenga ruido o que esté incompleto, las neuronas de la capa oculta de la red responderán con una salida activa si la nueva entrada contiene un patrón que se asemeje a aquella característica que las neuronas individuales hayan aprendido a reconocer durante su entrenamiento. Y a la inversa, las unidades de las capas ocultas tienen una tendencia a inhibir su salida si el patrón de entrada no contiene la característica para reconocer, para la cual han sido entrenadas.

Varias investigaciones han demostrado que, durante el proceso de entrenamiento, la red Backpropagation tiende a desarrollar relaciones internas entre neuronas con el fin de organizar los datos de entrenamiento en clases. Esta tendencia se puede extrapolar, para llegar a la hipótesis consistente en que todas las unidades de la capa oculta de una Backpropagation son asociadas de alguna manera a características específicas del patrón de entrada como consecuencia del entrenamiento. Lo que sea o no exactamente la asociación puede no resultar evidente para el observador humano, lo importante es que la red ha encontrado una representación interna que le permite generar las salidas deseadas cuando se le dan las entradas, en el proceso de entrenamiento.

Esta misma representación interna se puede aplicar a entradas que la red no haya visto antes, y la red clasificará estas entradas según las características que compartan con los ejemplos de entrenamiento.

Algoritmos de entrenamiento

El algoritmo backpropagation para redes multicapas es una generalización del algoritmo LMS (Least Mean Square), ambos algoritmos realizan su labor de actualización de pesos y ganacias con base en el error medio cuadrático. Hay muchas variaciones del algoritmo backpropagation, y todas persiguen los mismos propósitos: ser más rápido y tratar de llegar al mínimo error global evitando mínimos locales. La plataforma Matlab 5.3 se utilizo en este trabajo para la implementación del algoritmo de entrenamiento, pero cabe mencionar los algoritmos de entrenamiento que se encuentran en esta plataforma; los más importantes son:

- Adaptativo incremental
- Gradiente descendiente
- Gradiente descendiente con momento
- Gradiente descendiente por lotes
- Resilient
- Gradiente conjugado y sus variaciones
- Método Quasi – Newton
- Levenberg Marquardt

Cada uno de estos algoritmos poseen características particulares que los hacen apropiados para un determinado tipo de redes, es decir, no existe un algoritmo universal que se adapte a todas las topologías de redes existentes, por eso, el diseñador es quien decide al final el empleo de un algoritmo particular para una aplicación especifica.

El algoritmo de entrenamiento empleado en este trabajo es el Levenberg – Marquardt, caracterizado por ser un método de optimización diseñado para minimizar funciones que sean la suma de los cuadrados de otras funciones no lineales, y es por ello que este algoritmo tiene un excelente desempeño en el entrenamiento de redes neuronales donde el rendimiento de la red esté determinado por el error medio cuadrático.

4.1.6 Generalización Mejorada

Unos de los problemas que ocurren durante el entrenamiento de redes neuronales es el sobreajuste, esto es cuando el error utilizado en el set de entrenamiento es de un valor muy pequeño, pero cuando los nuevos datos son presentados a la red, el error tiende a ser muy grande. La red ha memorizado los ejemplos de entrenamiento, pero no aprendió a generalizar las nuevas situaciones. Podrá haber casos muy especiales donde se necesite una red neuronal con sobreajuste, así que esto no ya no será un problema sino una necesidad. Existen varias soluciones para corregir el problema de sobreajsute, MATLAB 5.3 brinda dos de ellas, estas son la regularización y la validación[17].

[17] ACOSTA, Luis y TINOCO, Orlando. Reconocimiento de formas irregulares empleando redes neuronales artificiales. Tesis de grado. 2003.

Regularización

Esta técnica involucra la modificación de la función de ejecución, la cual es normalmente seleccionada para ser la suma de cuadrados de los errores de la red en el set de entrenamiento. MATLAB ha incluido dos rutinas las cuales configuran automáticamente la más óptima función de ejecución para ejecutar la mejor generalización. La regularización puede ser una ayuda al momento de la escogencia del número de neuronas en una red sin que esta se sobreentrene.

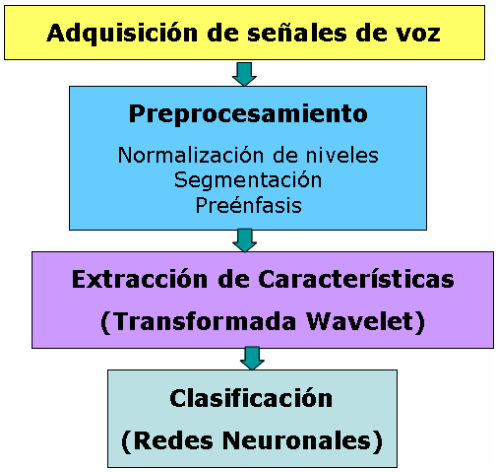
Validación

Esta es una técnica que se basa en la división del conjunto de datos en tres subconjuntos. El primer subconjunto es el set de entrenamiento usado para la computación del gradiente y actualizar los pesos y bias de la red neuronal. El segundo subconjunto es el set de validación, el error sobre este set es monitoreado durante el proceso de entrenamiento. El error de validación normalmente se decrementa durante la fase inicial del entrenamiento, así como también lo hace el error de entrenamiento. Sin embargo cuando la red comienza a sobreajustar los datos, el error comienza a incrementarse. Cuando el error empieza a incrementarse para un número de iteraciones especificado, el entrenamiento se detiene, y los valores de los pesos y bias retornan a los del mínimo error de validación. El tercer subconjunto es le set de prueba y no se usa durante el entrenamiento, pero es usado para compara diferentes modelos.

5 MODELO PROPUESTO PARA LA CLASIFICACIÓN DE VOCES

La metodología empleada para abordar el problema de clasificación de voces normales y patológicas esta compuesta de las etapas que se ilustran en la siguiente figura.

Figura 18. Etapas del sistema de clasificación de voces.



5.1 Adquisición de señales de voz

La base de datos utilizada en este trabajo, corresponde a un conjunto de muestras de señales de voz tomadas a un grupo de Mujeres adultas de la ciudad de Cartagena de Indias cuyas edades oscilan entre 17 y 50 años. Estas señales de voz fueron evaluadas previamente por una fonoaudióloga[18] y con esto se clasificará las voces como normales y patológicas.

[18] Doctora Katia Margarita Africano Rambao. Programa Magisterio

Las muestras de voz fueron adquiridas bajo las siguientes condiciones:

- Micrófono dinámico unidireccional (SHURE PG48)
- Frecuencia de muestreo 11025Hz, resolución de 16 bits por muestra.
- Emisión de las vocales /a/, /e/, /i/, /o/, /u/.

Las grabaciones de las muestras de voz fueron hechas utilizando el programa Cool Edit Pro 2.0 y fueron guardadas en formato .WAV. La captura de las señales se realizo bajo un ambiente de ruido moderado en las cuales parte de las muestras se grabaron en un salón asilado y cerrado con poco ruido, mientras que la parte restante de las muestras se grabaron en el estudio de grabación de la facultad de comunicación social de la Universidad Tecnológica de Bolívar. Esto se hizo con el fin de disminuir el efecto del ruido al momento de procesar la señal. Por ultimo al momento de la grabación la persona recitaba los cinco fonemas vocálicos uno seguido de otro de forma pausada ubicando el micrófono a una distancia promedio de 3cm de la boca y manteniendo una posición erguida.

La distribución de las muestras de voz normales y patológicas empleadas en este trabajo se muestra en la siguiente tabla.

Tabla 5. Base de datos de voces normales y patológicas.

MUESTRAS	
Normales	21
Patológicas	22
Total	43

5.2 Preprocesamiento

5.2.1 Normalización de niveles

En la toma de muestras de voz hay muchos factores que pueden interferir con la señal de entrada o alterar sus propiedades. Estos factores están relacionados con el entorno o con el mismo hablante y pueden afectar seriamente el desempeño de un sistema de análisis, reconocimiento o clasificación de voz. Entre estos factores pueden mencionarse: ruido aditivo, proveniente de señales interferentes; ruido debido a las características acústicas del salón o al equipo de grabación y variaciones en el nivel de la voz provenientes del hablante o debidas a cambios en la orientación o distancia del micrófono o a una atenuación desconocida en el canal. La normalización se realiza para disminuir el efecto de estos factores.

5.2.2 Segmentación

El objetivo de la segmentación de voz es separar los eventos de interés, (la voz que va a procesarse) de otras partes de la señal (ruido de fondo), evitando así cálculos innecesarios al procesar únicamente las partes de la señal que corresponden a voz. Esta se realiza en dos fases: la selección y estimación de un conjunto de parámetros que representan las clases voz y no voz; y la clasificación por algún principio discriminante empleando ya sea métodos estadísticos o métodos de inteligencia artificial[19].

[19] CONTRERAS, Sonia y CASTELLANOS, Germán. Detección Activa de voz orientada a la clasificación de fonemas aislados. Bucaramanga, 2003. p 45-53. Tesis de Maestría. Universidad Industrial de Santander. Facultad de Ingeniería Físico- Mecánicas.

En este trabajo en las dos fases para realización el proceso de segmentación de la voz se utilizaron como parámetros representativos de la señal de voz la energía y la tasa de cruces por cero de la señal y como clasificador se empleó el método estadístico denominado clasificador bayesiano.

Parámetros de segmentación

Como parámetros para la segmentación de voz de utilizaron:

- Tasa de cruces por cero: es un indicador de la frecuencia en la cual hay mayor concentración de energía en el espectro. Se calcula empleando la siguiente ecuación:

$$Z_n = \sum_{i=0}^{l} \left| \operatorname{sgn}\left[s(n+i+1)\right] - \operatorname{sgn}\left[s(n+i)\right] \right|$$

- Energía de la señal: la energía promedio de los segmentos sonoros usualmente es mayor al silencio. Se calcula empleando la siguiente ecuación:

$$Es = 10 \log\left(\varepsilon + \frac{1}{N} \sum_{n=1}^{N} s^2(n) \right)$$

Clasificador Bayesiano

La teoría de decisiones de Bayes constituye la base de los métodos estadísticos de reconocimiento de patrones. Para observar sus principios

básicos se va a partir de que se tienen *K* clases, denominadas w_k y un vector de parámetros de entrada **X**; además se define[20] :

- $P(w_k/\mathbf{X})$*:* es la probabilidad de que la clase correcta sea w_k, dado un vector de parámetros como entrada. Esta probabilidad se denomina probabilidad a *posteriori*, ya que puede ser estimada únicamente después de que los datos de entrada han sido observados.
- $P(w_k)$*:* es la probabilidad de la clase w_k. Se conoce como probabilidad a *priori*, porque puede ser evaluada antes de que **X** sea observado.
- $P(\mathbf{X}/w_k)$*:* es la distribución de probabilidad de **X** condicionada a una clase particular.

Empleando la regla de Bayes, la probabilidad a *posteriori* $P(w_k/\mathbf{X})$ puede calcularse con base en la probabilidad a *priori* y la probabilidad condicional $P(\mathbf{X}/w_k)$ de la siguiente manera:

$$P(w_k / X) = \frac{P(X / w_k) \cdot P(w_k)}{P(X)}$$

Como regla de decisión para un clasificador, puede establecerse que la clase correcta es la que presenta la mayor probabilidad a posteriori. Es decir, el clasificador le asignará un vector **X** a la clase *k***,** si para todo $i \neq k$:

$$g_k(X) > g_i(X),$$
$$g_k(X) = P(w_k / X) = \frac{P(X / w_k) \cdot P(w_k)}{P(X)}$$
$$g_k(X) = P(X / w_k) \cdot P(w_k)$$

[20] HUANG, Xuedong and ACERO, Alex. Spoken Language Processing "A guide to Theory, Algorithm, and System Development". New Jersey: Prentice Hall PTR, 2001

Esta simplificación puede hacerse porque *P(**X**)* es constante para todas las clases.

Como función discriminante se empleó:

$$g_k(X) = P(X / w_k) \cdot P(w_k)$$

El valor de $P(\mathbf{X}/w_k)$ para cada clase se calculó multiplicando las probabilidades condicionales estimadas de los dos parámetros.

Las probabilidades a *priori* para voz y silencio se asumieron iguales, con el objetivo de que la detección de voz se realizara únicamente con la información de las probabilidades condicionales $P(\mathbf{X}/w_k)$.

5.2.3 Filtro de Pre-énfasis

Este filtro se aplica a la señal de voz para aumentar la energía relativa de las componentes de alta frecuencia en el espectro de la señal de voz esto debido a que el modelo del tracto vocal utilizado no filtra suficientemente las componentes de alta frecuencia. Típicamente se utiliza para su implementación un filtro digital de primer orden cuya función de transferencia está dada por:

$$H(z) = 1 - \alpha z^{-1}$$

Un óptimo pre-énfasis tiene la ventaja de garantizar que la salida espectral del bloque de los datos sea tan plana como sea posible, en el sentido de

equiparar al máximo la medida espectral. Los objetivos de utilizar el filtro pre-énfasis son[21]:

- Reducir el efecto de la pendiente espectral de -20dB presente en los segmentos de voz.
- Amplificar la zona del espectro superior a 1kHz donde la percepción auditiva se hace sensible.

El valor optimo del coeficiente de pre-énfasis α, esta dado en función de la señal de entrada; sin embargo, se escoge un valor constante cercano a la unidad con el fin que la estructura de los formantes mayores sea acentuada, por lo que es razonable escoger un valor entre 0.9 y 0.95.

5.3 Extracción de Características

La extracción de características es un proceso que convierte la señal de voz en una representación parametrizada. El proceso de extracción realiza una primera reducción de datos pasando de la señal original de voz a una representación en características de voz derivadas de un adecuado análisis de la señal. La dificultad que existe en este proceso radica en determinar cual es la representación adecuada que permite caracterizar con un mínimo número de parámetros la información existente y discriminante de los sonidos pronunciados.

[21] OJEDA, Fabián y CATELLANOS Germán. Extracción de Características usando Transformada Wavelet en la identificación de voces Patológicas. Manizales, 2003.

Como se expuso anteriormente las funciones Wavelet poseen la capacidad de proveer información localizada en tiempo y frecuencia de una señal, y que al ser funciones bases localizadas que corresponden a versiones dilatadas y trasladadas de alguna función madre fija, se constituyen en una herramienta apropiada para el análisis o extracción de características de señales de naturaleza no estacionaria como lo son las señales de voz cuyo contenido espectral varía con el tiempo. La información que resulta de un análisis Wavelet es generalmente útil en tareas de clasificación.

Para desarrollar un esquema de extracción de características basado en la transformada wavelet, es necesario conocer las características que debe cumplir la función wavelet madre a emplearse de tal forma que esta se adapte a la forma de la señal de la voz humana. Estas características son:

1. *Similitud entre la Wavelet madre y las muestras de voz:* Muchos autores coinciden en que la mayor similitud entre Wavelets y la señal de voz se da para las Symlets y Daubechies, sin embargo estas primeras resultan ser simétricas lo cual no coincide con la característica real de una señal de voz que es de tipo asimétrico, es por tal razón que en este punto las Daubechies resultan atractivas[22].
2. *Ortogonalidad y soporte compacto:* Más que un criterio es un requisito, pues con esto se garantiza la posibilidad de empleo de la DWT.
3. *Orden de la Wavelet:* Un orden alto en la Wavelet se traduce en una longitud grande para las respuestas al impulso de los filtros g y h y por consiguiente representa mayor tiempo de cálculo de la transformada.

[22] KRISHNAN, M., NEOPHYTOU, C. and PRESCOTT, G. "Wavelet Transform Speech Recognition Using Vector Quantization, Dynamic Time Warping and Artificial Neural Networks". Preprint. 1994.

Luego de conocer la función madre a ser empleada se define el nivel de descomposición j para el cálculo del análisis multiresolución, del cual se obtiene como resultado un vector de características (conjunto de coeficientes Wavelet) que tiene la siguiente estructura $\left[ca_j, cd_j, ..., cd_1 \right]$ y que representa el conjunto de coeficientes de detalle en todas las resoluciones y el coeficiente de aproximación en la resolución más baja.

5.4 Clasificador

En este modelo de clasificación de voces normales y patológicas se emplean las redes neuronales artificiales en la determinación de la normalidad o anormalidad de una señal de voz.

Es útil establecer alguna notación para describir la tarea de extracción de características y el problema de clasificación. Un patrón puede consistir de N variables x_i. Sin embargo, es conveniente agruparlas y denotarlas por un único vector $X = \left[x_1 x_2 \ldots x_N \right]$. Cada patrón x_i puede pertenecer a una de las R clases, denotadas como y_R. Se puede entonces plantear que, $x \in X$ es el espacio de señales de entrada y $y \in Y = \left[y_1 y_2 \ldots y_R \right]$ es el espacio de salida, el cual es una colección de R etiquetas de clase. Por tanto, X × Y es el conjunto de todos los pares de señales de entrada y las correspondientes etiquetas de clase (x, y). La clasificación de estos patrones puede definirse como una función $d : X \to Y$ que asigna una etiqueta de clase a cada patrón de entradas $x \in X$.

El vector de características X en este trabajo representará el vector obtenido del cálculo de la energía de cada uno de los coeficientes Wavelet a un nivel de descomposición j especificado, el cual tendrá la estructura $\left[ca_j, cd_j, ..., cd_1\right]$.El espacio de salida Y representará a las etiquetas de clases, que se definen de forma tal que a las voces normales se le asigna +1 y a las voces patológicas se le asigna -1.

6 DISEÑO DEL PROGRAMA DE CLASIFICACION DE VOCES

6.1 Características del programa

El programa diseñado fue desarrollado bajo la plataforma de MATLAB 5.3 y se creó con el fin de integrar cada una de las etapas que conforman la metodología implementada en el desarrollo del sistema de clasificación de voces normales y patológicas ya que los algoritmos están escritos también en esta plataforma.

El código desarrollado presenta las siguientes características:

- Orientado a objetos: la programación utiliza objetos, ligados mediante mensajes, para la solución del problema. Es una extensión de la programación estructurada que permite potenciar los conceptos de modularidad y reutilización de código.
- Objetos: los objetos están representados por los controles de la interfaz gráfica. Estos objetos poseen propiedades y eventos definidos, cuya interacción definen el programa.
- Mensajes: cuando se ejecuta el programa, los objetos están interpretando y respondiendo a mensajes.
- Modularidad: el programa se ha dividido en funciones, pueden agregarse, moverse y quitarse funciones sin afectar el código restante.
- Reutilización: el código define la manera como se modifica las propiedades de los objetos a partir de mensajes o eventos, esto hace que pueda tomarse una función definida para un control e insertarse sin problemas en un nuevo código.

6.2 Descripción

La interfaz realizada para llevar a cabo la clasificación de voces normales y patológicas se compone de tres grandes bloques: Extraer Características, Entrenar red y Clasificación.

6.2.1 Extraer características

En esta etapa consta de dos partes: Voces Normales y Voces Patológicas, y se encarga de realizar la extracción de parámetros de un conjunto de voces normales y patológicas previamente seleccionado. A este conjunto de voces se le aplica la metodología de desarrollo planteada en el trabajo que va desde el preprocesamiento (Normalización de niveles, Segmentación y Preénfasis) hasta el cálculo de la Transformada Wavelet, de la cual resulta el vector de características conformado por los coeficientes Wavelet.

6.2.2 Entrenar Red

En esta etapa se cargan las matrices de características obtenidas de la extracción de características realizada en la etapa anterior que serán las entradas a 5 redes neuronales correspondientes a los 5 fonemas vocálicos. Como salida se observa la curva de entrenamiento de las redes y la medida del desempeño de la red que corresponde al error medio cuadrático.

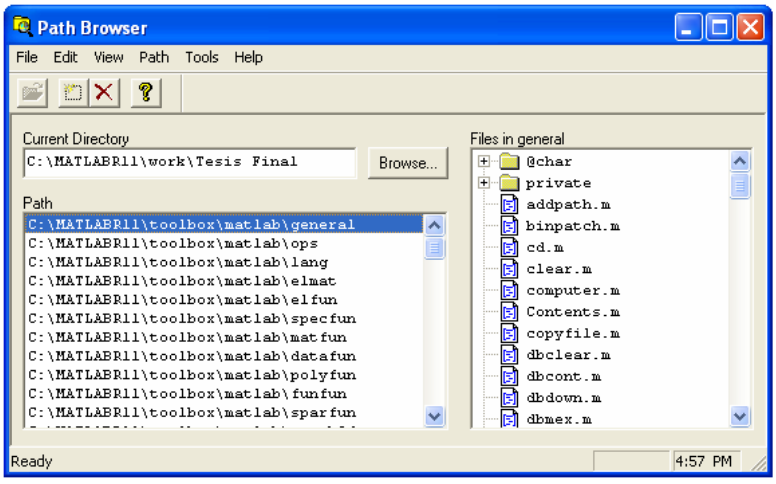
6.2.3 Clasificación

En esta etapa se procede a clasificar las voces normales y patológicas, para esto se ejecuta en su totalidad la metodología desarrollada obteniendo como salida un mensaje en el workspace de MATLAB que indica si una vocal es normal o patológica.

6.3 Instalación

Como se ha mencionado a lo largo del trabajo, la metodología propuesta para la clasificación de voces normales y patológicas esta desarrollada en su totalidad en la plataforma de MATLAB 5.3. Lo primero que se debe hacer es copiar la carpeta que se llama **TESIS FINAL,** que se encuentra en el cd anexo, en la carpeta **work.** Una vez que se ejecuta MATLAB se debe definir la ruta de ubicación de los programas. Esto se realiza en el menú **file** y luego en **Set path**, con lo que aparece la siguiente ventana.

Figura 19. Path Browser de MATLAB



La ruta especificada en Current Directory es la ruta válida para que la aplicación se ejecute correctamente.

Después de realizado lo anterior se digita en el workspace de MATLAB la palabra **TESIS,** con lo que aparece la ventana de la aplicación

Figura 20. Ventana principal de la aplicación en MATLAB

6.4 Operación

La secuencia de ejecución del programa es:

1. Extraer Características
2. Entrenar red
3. Clasificación

6.4.1 Extraer características

En este bloque se compone de dos partes: Voces Normales y Voces patológicas. En cada parte se ejecuta la metodología de extracción de parámetros como se dijo anteriormente a un conjunto de 20 señales de voz de la base de datos previamente seleccionadas, que corresponden a un conjunto de 10 voces normales y 10 voces patológicas.

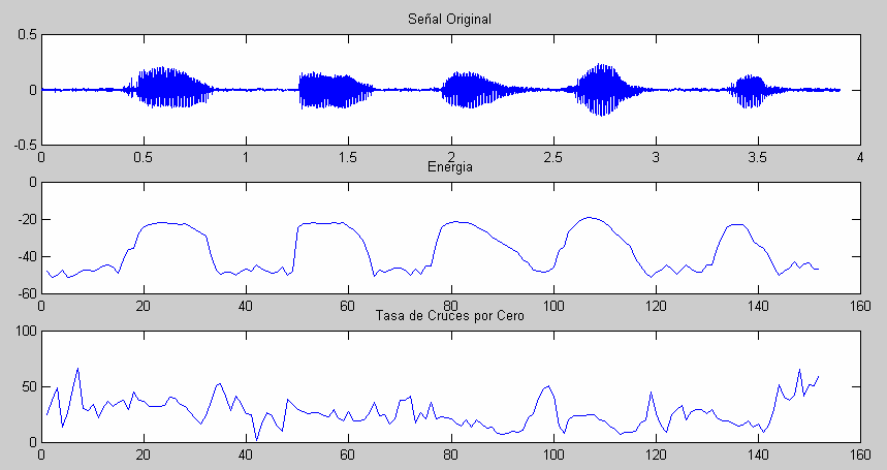
Las etapas de metodología desarrollada que se ejecutan para cada una de las 20 voces de esta sección de la aplicación principal son:

Normalización de niveles: Para efectuar esto a la señal de voz se le resta su media y se divide por su desviación estándar.

Segmentación: Como se dijo anteriormente la segmentación se realiza en dos etapas, una etapa corresponde a la extracción de parámetros que en este trabajo se realizó empleando la energía y la tasa de cruces por cero, y la otra etapa corresponde la clasificación, para lo cual empleó el método estadístico denominado clasificador bayesiano.

El cálculo de la energía y la tasa de cruce por cero a una señal de voz se muestran a continuación

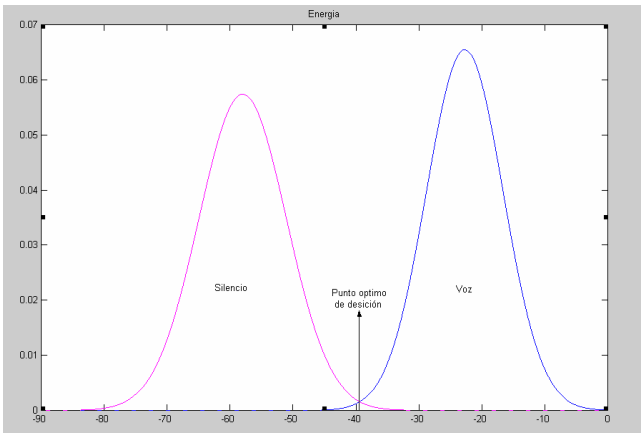
Figura 21. Energía y tasa de cruces por cero.



Como se dijo anteriormente la energía de una señal es mayor para los segmentos de voz y la tasa de cruces por cero es mayor para los segmentos de silencio.

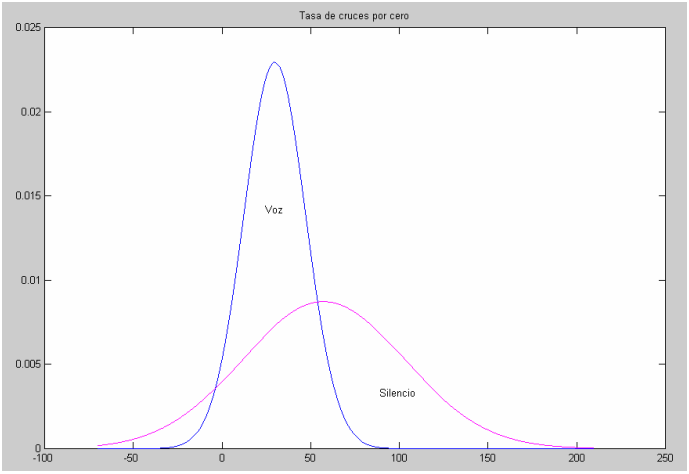
En la implementación del clasificador bayesiano para la segmentación de la señal de voz se tuvo en cuenta las distribuciones normales de los segmentos tanto de voz y de silencio. Los valores representativos de la distribución normal (media y desviación estándar) para los segmentos de voz y de silencio, se encuentran en el archivo llamado **parte1**, que se obtuvo calculando la energía y tasa de cruces por cero de los dos segmentos a clasificar (voz y silencio). Las siguientes figuran ilustran las distribuciones normales para los segmentos de silencio y voz.

Figura 22. Parámetros de segmentación: Energía de la señal



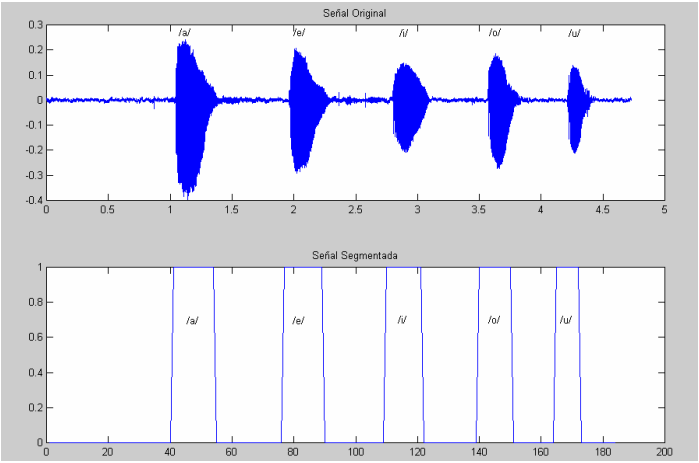
En la figura anterior se muestran las distribuciones normales de la energía tanto para silencio como para voz y de la cual se diferencian claramente los segmentos de sonido y de silencio porque las curvas están claramente separadas y además del punto optimo de decisión que es el punto en el cual se minimiza el error de clasificación garantizando que un segmento pertenezca a su respectiva clase (silencio o voz).

Figura 23. Parámetros de segmentación: Tasa de cruces por cero



Aunque en la figura anterior se nota un traslapamiento en las distribuciones normales de la tasa de cruces por cero de los segmentos de voz y de silencio, al utilizar ambos parámetros de segmentación se obtiene una señal de voz clara mente segmentada, tal como se muestra en la siguiente figura.

Figura 24. Segmentación de voz



De esta manera se obtienen los puntos de inicio y fin de palabra de cada uno de los 5 fonemas vocálicos, los cuales se analizarán cada uno por separado.

Filtro de prénfasis: para la implementación de este filtro digital de primer orden se empleo la siguiente función de transferencia:

$$H(z) = 1 - \alpha z^{-1}$$

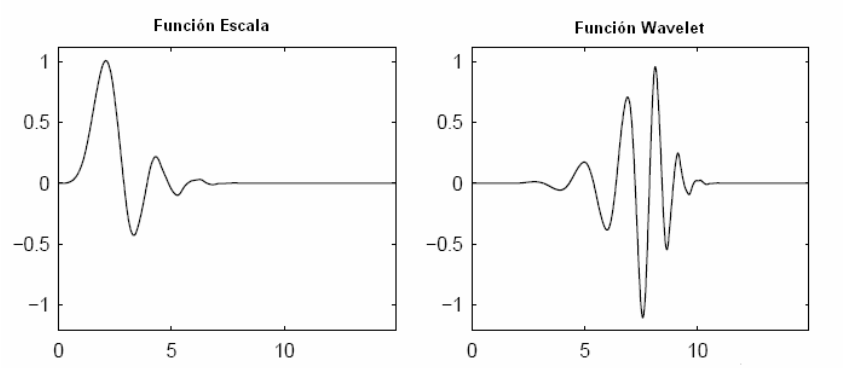
Con un valor para el coeficiente de pre-énfasis α de 0.95.

Extracción de características: El esquema de extracción de características planteado en este trabajo es una modificación del presentado en[23]. Gracias a su propiedad de concentración de energía, la transformada wavelet distribuye características específicas de la señal en diferentes bandas de frecuencia.

Haciendo uso de esta propiedad, el esquema de extracción de características se describe como sigue:

Función madre $\psi(t)$: La función madre utilizada fue la Daubechies de orden 8 (db8), que ha sido empleada con éxito en otros trabajos, dada su similitud con la voz humana y los buenos resultados que ofrece en la etapa de extracción de parámetros.

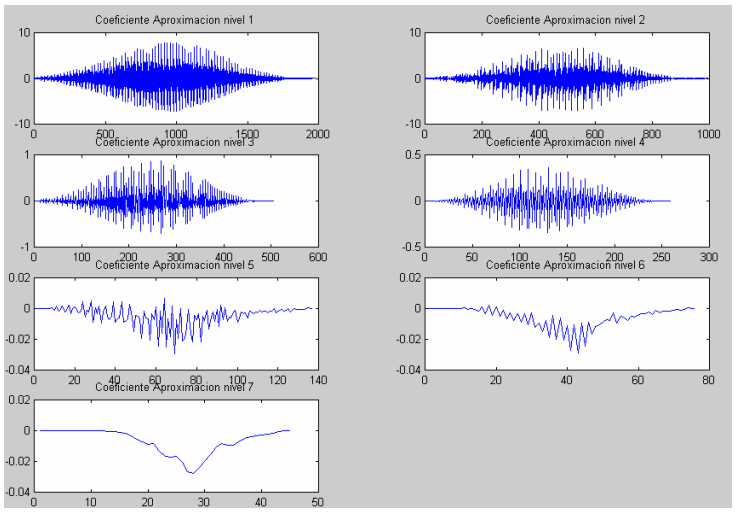
Figura 25. Función de escala y función wavelet para la db8.



Para efectuar el análisis multiresolución empleando el banco de filtros se empleó un número de 6 escalas de aproximación, debido a que después del sexto nivel de aproximación no se retiene información representativa del segmento de voz tal como se ilustra en la siguiente figura.

[23] OJEDA, Op. Cit., p 80.

Figura 26. Siete escalas de aproximación de una vocal /a/.



Como se mencionó anteriormente para el análisis multiresolución a un nivel de aproximación J elegido, la representación del conjunto de coeficientes Wavelet tiene la siguiente estructura $\left[ca_j, cd_j, ..., cd_1\right]$ que representa el conjunto de coeficientes de detalle en todas las resoluciones y los coeficientes de aproximación en la resolución más baja. En este trabajo se calcula la energía de estos coeficientes empleando la siguiente ecuación:

$$Es = 10\log\left(\varepsilon + \frac{1}{N}\sum_{n=1}^{N} s^2(n)\right)$$

En conclusión, la ejecución de este bloque de programa tanto para Voces Normales como para Voces Patológicas darán como resultado 5 matrices de 10x7 donde las filas representan las 10 voces normales o patológicas (según sea el caso) y las columnas representan el vector de energías que se obtiene del calculo de la transformada Wavelet para cada vocal por separado.

6.4.2 Entrenar Red

Inicialmente para el desarrollo de este trabajo se utilizó un espacio de características en la cual se aplicaba la metodología de extracción propuesta a los cinco fonemas vocálicos de una misma señal de voz, de esta forma una señal de voz quedaba representada por 35 características equivalentes a 7 coeficientes wavelet (empleando 6 escalas de aproximación como se dijo anteriormente) por cada una de las 5 vocales, las cuales eran la entrada de una sola red neuronal. Los resultados arrojados no fueron satisfactorios, porque la capacidad de generalización de la red era muy reducida, dado que al analizar patrones de las 5 vocales de una misma voz y a la vez, había una perdida de información importante de los patrones para voces normales y para voces patológicas. Por lo anterior en este trabajo se opto por desarrollar 5 redes neuronales (una para cada vocal), dado que de esta forma si se garantiza el análisis sobre los patrones que determinan si una voz es normal o si es patológica. Cada red desarrollada fue del tipo Percetron multicapa, con 7 neuronas en la capa de entrada, 12 neuronas en la capa de oculta y una neurona en la capa de salida. Como función de activación y de transferencia se seleccionó la tangente hiperbólica.

Como se dijo anteriormente el vector de características X obtenido del cálculo de la energía de cada uno de los coeficientes Wavelet una vez realizado el análisis multiresolución empleando 6 escalas de aproximación, será el vector de entrada a la red neuronal y este tiene la estructura $X = \left[ca_6, cd_6, cd_5, cd_4, cd_3, cd_2, cd_1\right]$. El espacio de salida Y que corresponde a las etiquetas de clases, se definen de forma tal que a las voces normales se le asigna +1 y a las voces patológicas se le asigna -1.

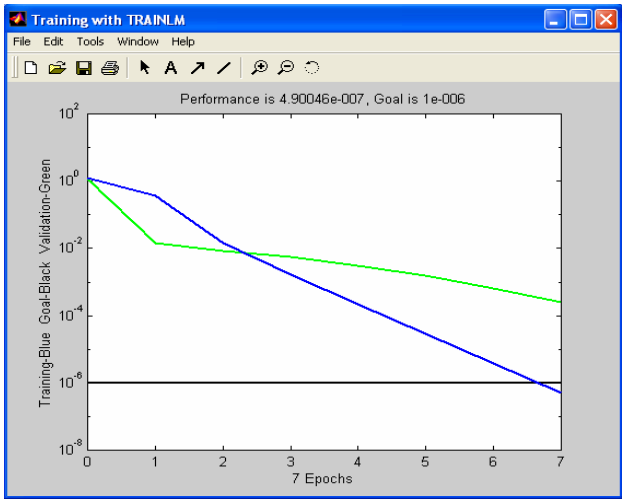
El entrenamiento se efectuó con el criterio de parada anticipada, que emplea un conjunto de datos para entrenar y otro para calcular el error de validación. Tiene la ventaja de que es rápido y mejora la capacidad de generalización de la red. Del total de datos se tomaron el 60% para entrenar, el 20% para validar y el 20% restante como prueba.La forma como se calculan los nuevos valores depende del método de optimización que se emplee para reducir el error. En este trabajo tal y como se mencionó anteriormente, se realizó el entrenamiento con el método de optimización de Levenberg-Marquardt, que es de segundo orden y se caracteriza por ser el más eficiente aunque requiere un consumo elevado de memoria. Como función de desempeño se seleccionó el error medio cuadrático.

Las características de entrenamiento empleadas para la ejecución del algoritmo Levenberg-Marquardt fueron:

red.trainparam.epochs = 1000;	Épocas
red.trainparam.goal = 1e-6;	Error promedio
red.trainparam.lr = 0.115;	Rata de aprendizaje

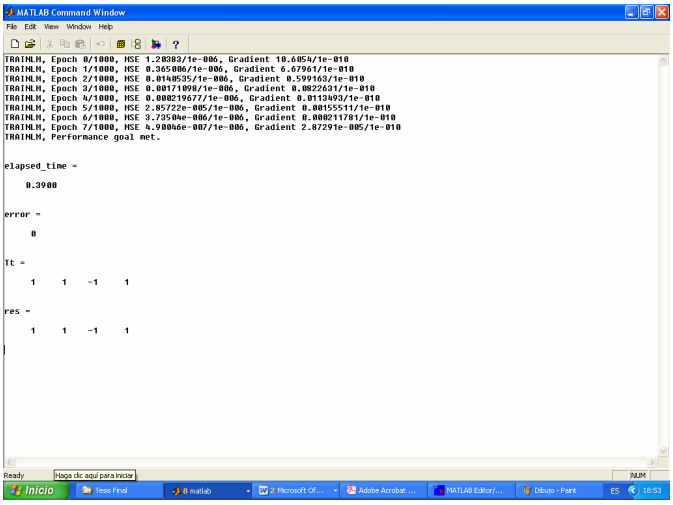
En la ejecución de este bloque de programa para el entrenamiento de cada una de las redes neuronales correspondientes a cada vocal, al pulsar sobre una determinada opción aparece la siguiente figura.

Figura 27. Grafica de avance de error de convergencia



La figura anterior muestra el avance del error medio del set de datos de entrenamiento y de validación a medida que se acercan al error medio cuadrático establecido (goal) respecto al número de épocas. Una vez que aparece la figura anterior se observa en el workspace de MATLAB lo siguiente.

Figura 28. Workspace de MATLAB en el entrenamiento de una red



En el se muestra el tiempo que demora el entrenamiento (elapsed_time) y un error. Este error se calcula haciendo una simulación de la red entrenada con el subconjunto de datos del set de prueba. Se muestra el espacio de salida deseado de los datos de prueba (Tt) y la respuesta de la red al conjunto de datos de prueba. Lo anterior se realiza para tener una idea del efectivo entrenamiento de la red neuronal.

Una vez terminado el entrenamiento se guarda en un archivo con formato redx.mat (donde x es la vocal) el wokspace de matlab.

6.4.3 Clasificación

Esta es la parte final del programa y donde se obtiene el resultado por vocal para la determinación de normalidad o anormalidad de una señal de voz. En esta parte se ejecuta toda la metodología propuesta desde la normalización de niveles hasta la clasificación por medio de la red neuronal.

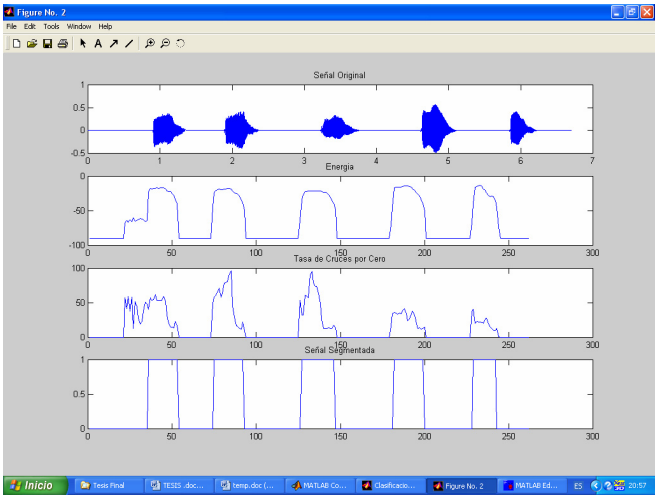
Para seleccionar la señal de voz a analizar, hay que abrir el código fuente correspondiente a la vocal a clasificar cambiando el nombre del archivo y la ubicación de la siguiente línea de código:

```
[y,fs]=wavread('C:\MATLABR11\work\TesisFinal\Voces\Normales\Claudia.wav');
```

Nota: Solo se cambia el nombre (Claudia.wav) y la ubicación (Normales) si se desea elegir una voz patológica por ejemplo.

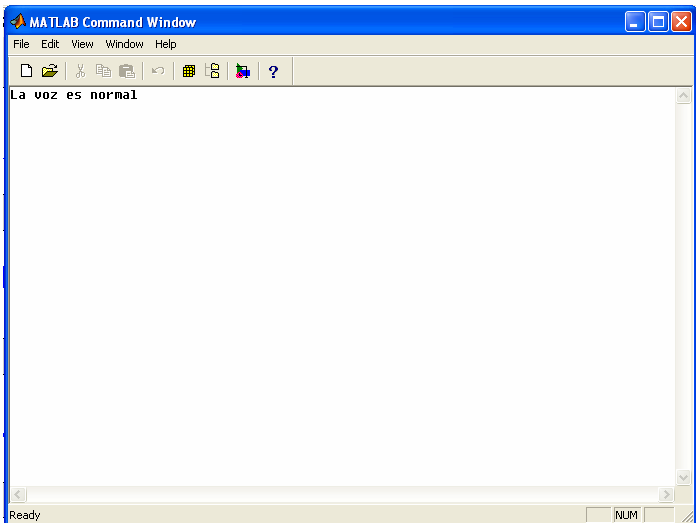
Al ejecutar una sección del programa aparece la siguiente gráfica que ilustra el proceso completo de la segmentación.

Figura 29. Resultado de la segmentación



En el workspace de MATLAB aparece el resultado de la clasificación.

Figura 30. Resultado de la clasificación por vocal



7 RESULTADOS

Como se mencionó con anterioridad en la etapa de extracción de características emplearon 10 voces normales y 10 voces patológicas. El porcentaje restante de muestras de voz pertenecientes a la base de datos se utilizaron para la validación y prueba de clasificador. Se emplearon 5 redes independientes correspondientes a cada fonema vocálico de una muestra de voz. Para la etapa de clasificación se evaluó cada vocal por separado y como criterio final de clasificación entre una voz normal y patológica se empleó el número mayor de aciertos entre las 5 vocales de una misma señal de voz.

En el diseño de las redes neuronales empleadas como clasificadores se resaltan las pruebas realizadas con el objetivo de obtener un mejor desempeño. La siguiente tabla ilustra los resultados obtenidos para cada uno de los 5 fonemas vocálicos empleando redes neuronales con 15 neuronas en la capa oculta.

Tabla 6. Clasificación de vocales separadas con 15 neuronas.

PORCENTAJE DE CLASIFICACIÓN					
	a	**e**	**i**	**o**	**u**
Voces Normales	81,81%	72,72%	36,36%	45,45%	36,36%
Voces Patológicas	83,33%	75,00%	75,00%	58,33%	58,33%

De acuerdo a los resultados anteriores se realizo otra prueba pero esta vez empelando 12 neuronas en la capa oculta con lo que se obtuvieron resultados ilustrados en la siguiente tabla.

Tabla 7. Clasificación de vocales separadas con 12 neuronas.

PORCENTAJE DE CLASIFICACIÓN					
	a	**e**	**i**	**o**	**u**
Voces Normales	90,90%	81,81%	36,36%	54,54%	45,45%
Voces Patológicas	91,66%	83,33%	75,00%	91,66%	66,66%

De lo anterior se resalta el compromiso que se debe tener en cuenta en la elección del número de neuronas en la capa oculta a emplearse, dado que al utilizar un numero de neuronas en la capa oculta igual o incluso mayor de 15 la red se sobre-entrena, con lo cual la generalización es pobre. Por esto el total de neuronas elegido finalmente fue de 12, con el cual se aumentó el porcentaje de aciertos como lo demuestra la tabla de resultados.

De las resultados obtenidos se destacan los resultados de clasificación obtenidos para las vocales /a/ y /e/ tanto para voz normal como para voz patológica que son catalogados como fonemas vocálicos de modo de articulación abierto en los cuales al momento de pronunciarse el aire expelido por los pulmones encuentra pocos obstáculos hasta su salida hacia el exterior. Por otra parte se resaltan los resultados para los obtenidos para las vocales /i/ y /u/ catalogados vocales de abertura mínima de lo cual se concluye que como el aire durante su pronunciación encuentra más

obstáculos que durante la pronunciación de vocales abiertas y el sistema detecta este tipo de vocales con un leve patrón ruidoso que ocasiona su poca generalización al clasificarlo.

En la siguiente tabla se muestran el resultado total de clasificación obtenido del sistema desarrollado, el cual se obtuvo empleando el criterio de mayor número de aciertos de las 5 vocales de una señal de voz.

Tabla 8. Resultado total del clasificador

RESULTADO TOTAL DEL CLASIFICADOR	
Voces Normales	81,81%
Voces Patológicas	100,00%

Los resultados totales que se muestran en la tabla anterior significan que la sensibilidad diagnostica del sistema definido como la probabilidad de clasificar correctamente un individuo enfermo o con alguna alteridad en la voz es del 100% y la especificidad definida como la probabilidad de clasificar correctamente un individuo sano es de 81.81%.

8 CONCLUSIONES

En el presente trabajo se ha detallado la metodología para el desarrollo de un sistema clasificación de voces normales y patológicas que básicamente esta conformado por: una etapa de preprocesamiento de la señal que incluye la normalización de niveles, necesaria para disminuir los efectos externos que ocurren en la captura de señal de voz y también las variaciones de intensidad de la voz de la persona, un proceso de segmentación que evita que se incurra en cálculos innecesarios al delimitar la señal de interés que es la voz a ser procesada y por último un filtrado de preenfasis. Una vez finalizada la etapa de preprocesamiento de la señal de voz se realiza la extracción de características que en este trabajo se realizó utilizando la transformada Wavelet ya que es una herramienta apropiada para obtener una representación parametrizada de señales no estacionarias como es la señal de voz que posee tanto componentes tanto transitorios de alta frecuencia como componentes estacionarios. Por último en la clasificación de voces normales y patológicas se empleó las redes neuronales artificiales.

En la parte de segmentación de voz hay que resaltar los buenos resultados obtenidos en la distinción de voz y silencio para las muestras de voz de la base de datos empleada bajo las condiciones de captura aquí establecidas, que fueron recitos con un ambiente de bajo ruido, una posición erguida del hablante y cercanía del micrófono a una distancia promedio de 3 cm de la boca, porque si estas condiciones de captura varían, como por ejemplo se realizan bajo un ambiente en condiciones de ruido alto, el segmentador presentara errores de clasificación entre voz y silencio dado que su

funcionamiento esta ligado a las muestras de voces bajo las condiciones ya descritas. Para corregir este problema se deben parametrizar nuevamente los segmentos de voz y de silencio de las nuevas señales de voz capturadas aplicando las ecuaciones de energía y tasa de cruces por cero, para que posteriormente el clasificador bayesiano distinga correctamente las dos clases a clasificar.

Los resultados obtenidos sugieren un excelente desempeño en la capacidad de generalización del sistema y hacen prometedora la aplicación de esta alternativa como una herramienta de soporte para el diagnostico de patologías del sistema vocal con la ventaja de no ser de carácter invasivo, esto debido a que el porcentaje de clasificación total resultante es bastante aceptable bajo las condiciones aquí expuestas. Además esta metodología se puede hacer extensiva para cada especialista en el tema que posea su propia base de datos y la emplea para clasificar entre voces normales y patológicas con enfermedades de su interés o cuya incidencia es mayor en su área de investigación que desempeña.

La principal tarea a mediano plazo que hagan posible una extensión y posteriores desarrollos de este trabajo es la ampliación de la base de datos con muestras de voz tanto normales como patológicas para personas de la ciudad de Cartagena de indias de ambos sexos y de cualquier edad, dado que en el presente trabajo se delimitaron las muestras de voz solo a mujeres adultas de la ciudad. Con esto se espera que a medida que se amplíe la base de datos mejore la exactitud del sistema de clasificación dado que se permitirá evaluar la metodología expuesta sobre una muestra más representativa de la población de Cartagena de Indias.

Una extensión interesante de este trabajo sería el estudio y aplicación de la metodología desarrollada cambiando la herramienta de extracción de características de tal forma que se permita hacer uso de habla continua como entrada al sistema de clasificación en lugar de fonemas vocálicos. Así, el paciente presentaría mayor naturalidad para registrar su señal de voz, dado que el habla continua permite variar la frecuencia fundamental de los sonidos emitidos haciendo más flexible el estudio y confrontar los resultados con los métodos tradicionales que emplea un especialista.

Además en la ampliación de la base de datos se puede hacer recopilaciones de voces en donde la persona recite las vocales en desorden, como por ejemplo /a/, /o/,/ u/,/e/ /i/ o recite vocales sostenidas, con el fin de optimizar y mejorar cada una de las etapa de la metodología propuesta.

9 BIBLIOGRAFIA

ACOSTA, Maria Isabel y ZULOAGA, Camilo. Tutorial sobre redes neuronales aplicadas a la ingeniería eléctrica y su implementación en un sitio web. Pereira, 2000.

ACOSTA, Luis y TINOCO, Orlando. Reconocimiento de formas irregulares empleando redes neuronales artificiales. Tesis de grado. 2003.

BALLANTYNE Jhon and GROVES Jhon. Manual de Otorrinolaringología. Barcelona: Salvat Editores, 1982.

BALLENGER, Jhon. Enfermedades de la nariz, garganta y oído. Barcelona: Jims, 1981. p. 624-633.

CONTRERAS, Sonia y CASTELLANOS, Germán. Detección Activa de voz orientada a la clasificación de fonemas aislados. Bucaramanga, 2003. p 45-53. Tesis de Maestría. Universidad Industrial de Santander. Facultad de Ingeniería Físico- Mecánicas.

_____.Segmentación de voz con redes neuronales empleando técnicas de detección bayesianas. VIII Simposio de tratamiento de señales, imágenes y visión artificial.

FAUNDEZ, Pablo y FUENTES, Álvaro. Procesamiento Digital de Señales Acústicas utilizando Wavelets. Instituto de Matemáticas UACH.

FURUI Sadoaki. Digital speech processing, synthesis and recognition. New Cork: Marcel Dekker Inc.1989.

GARCÍA, Léonard. Transformada wavelet aplicada a la extracción de información en señales de voz," Ph.D. dissertation, Departamento de Teoría de Señales y Comunicaciones, Universidad Politécnica de Cataluña, Barcelona, 1998.

HILERA, José y MARTÍNEZ Víctor. Redes Neuronales Artificiales, Fundamentos, Modelos y aplicaciones. Madrid: Alfaomega, 1995.

HUANG, Xuedong and ACERO, Alex. Spoken Language Processing "A guide to Theory, Algorithm, and System Development". New Jersey: Prentice Hall PTR, 2001.

IHYEH, Johnson. "Discrete wavelet transform techniques in speech Processing". CSIRO Division of Radiophysics. Australia, 1996.

JIANG, Jack and LIN Emily. Fisiología de las cuerdas vocales. Clínicas de Norteamérica de Otorrinolaringología. Agosto, 2002.

KRISHNAN, M., NEOPHYTOU, C. and PRESCOTT, G. "Wavelet Transform Speech Recognition Using Vector Quantization, Dynamic Time Warping and Artificial Neural Networks". Preprint. 1994.

LE HUCHE, Fracois, Patología Vocal: Semiología y disfonías disfuncionales. MASON, 1994, vol. 2.

MATHWORKS, Building GUIs with MATLAB 5.3- Version 5, USA: Mathworks Inc.1998.

_____, Neural network toolbox: User's Guide - Version 1, for use with MATLAB 5.3. USA: Mathworks Inc.1998. p. 41-416.

_____, Wavelet Toolbox: User's Guide - Version 1, for use with MATLAB 5.3, USA: Mathworks Inc.1998.

OJEDA, Fabián y CASTELLANOS, Germán. Extracción de Características usando Transformada Wavelet en la Identificación de Voces Patológicas. Manizales, 2003. Trabajo de Grado. Universidad Nacional de Colombia Sede Manizales.

OPPENHEIM, Alan and WILLSKY, Alan. Señales y Sistemas. Segunda ed. México: Prentice Hall, 1998.

STEGMANN Joachim, SCHRÖDER Gerhard and Fischer Kyrill. "Robust Classification of Speech based on the Dyadic Wavelet Transform with application to Celp Coding" IEEE.

TAN, Beng, and LANG, Robert, "Applying wavelet analysis to speech segmentation and classification" IEEE, 1994.

_____, Beng, "The use of wavelet transforms in phoneme recognition" IEEE, 1996.

TOBON, L y CASTELLANOS, Germán. Diseño y desarrollo de un sistema interactivo de análisis acústico de la voz y el habla para la ciudad de Manizales. Manizales

VARGAS, Fabián y CASTELLANOS, Germán. Clasificación automatizada de las características acústicas de la voz normal en la ciudad de Manizales. Manizales, 2001. Trabajo de Grado. Universidad Nacional de Colombia Sede Manizales.

DIRECCIONES EN INTERNET

Descripción de la anatomía de la laringe:
http://escuela.med.puc.cl/paginas/publicaciones/ApuntesOtorrino/AnatomiaLaringea.html

Procedimientos clínicos para la detección de patologías de la voz
http://www.laringeyvoz.com/cirugias.htm

Producción de la señal de voz:
http://www.encolombia.com/medicina/otorrino/otorrinosupl31203-comoseproduce.htm

ANEXOS

ANEXO A. Clasificación de los fonemas del idioma español

En el idioma español estándar de España existen 24 fonemas, mientras que en el de América Latina el sistema se limita a 22 fonemas, al reducirse /z/ y /s/ a uno solo: /s/; y /LL/ e /y/ a: /y/. Los fonemas se dividen en vocales y consonantes.

Fonemas vocálicos

En la producción de estos fonemas, el aire no encuentra obstáculos en su salida desde los pulmones hacia el exterior. Los fonemas vocálicos se definen según los siguientes factores:

El punto de articulación: es la parte de la boca donde se articulan. Pueden ser anteriores (/e/, /i/), medio o central (/a/) o posteriores (/o/, /u/).

El modo de articulación: se refiere a la abertura de la boca al pronunciarlos. Pueden ser de abertura máxima o abierta (/ha/), de abertura media o semiabiertos (/e/, /o/) y de abertura mínima o cerrada (/i/, /u/).

En el cuadro 1.1 se presenta la clasificación de las vocales con el Triangulo de Hellwag.

Clasificación de las vocales

		Localización		
		Anterior	Medio	Posterior
	Mínima	I		U
Abertura	Media	E		O
	Máxima		A	

Fonemas consonánticos

En la articulación de los fonemas consonánticos siempre hay un obstáculo en la salida del aire al exterior. Para clasificarlos pueden tenerse en cuenta varios factores:

- Zona o punto de articulación: es el lugar donde toman contacto los órganos que intervienen en la producción del sonido. Según el punto de articulación, los fonemas pueden clasificarse en:

  - Bilabiales: se articulan uniendo los dos labios
  - Labiodentales: los dientes superiores se apoyan en el labio inferior.
  - Interdentales: la lengua se posiciona entre los dientes.
  - Dentales: la lengua entra en contacto con los dientes superiores.
  - Alveolares: la lengua se ubica sobre la raíz de los dientes superiores.
  - Palatales: se unen la lengua y el paladar.
  - Velares: se unen la lengua y el velo del paladar.

- Modo de articulación: es la postura que adoptan los órganos que producen los sonidos. Según el modo de articulación, los fonemas pueden ser:

  - Oclusivos: se produce un cierre momentáneo que impide el paso del aire.
  - Fricativos: se produce un estrechamiento por el que el aire pasa rozando.
  - Africados: se produce una oclusión seguida de una fricación.
  - Lateral: el aire pasa rozando los lados de la cavidad bucal.
  - Vibrante: el aire hace vibrar la punta de la lengua al pasar.

- Actividad de las cuerdas vocales: permite clasificar los fonemas en sonoros y sordos, teniendo en cuenta si hay vibración de las cuerdas vocales.

- Actividad de la cavidad nasal: si al producir los sonidos parte del aire pasa por la cavidad nasal los sonidos se llaman nasales. Si todo el aire sale por la cavidad bucal se llaman orales.

En el siguiente cuadro se muestra la clasificación de las consonantes teniendo en cuenta los factores anteriormente descritos.

Clasificación de las consonantes

	Bilabial		Labiodental		Interdental		Dental		Alveolar		Palatal		Velar	
	Sonoro	Sordo	Sonoro	Sordo	Sonoro	Sordo	Sonoro	Sordo	Sonoro	Sordo	Sonoro	Sordo	Sonoro	Sordo
Oclusivas	b	p					d	t					g	k
Africados												ch		
Fricativos				f		z				s	y			j
Laterales									l		ll			
Vibrantes									r, rr					
Nasales	m								n		ñ			

ANEXO B. Percepción de la señal de voz

En esta parte se explica cómo se percibe un sonido desde el exterior hasta llegar al cerebro humano, donde cada una de las partes del oído son esenciales para la percepción de este.

Estructura del oído.

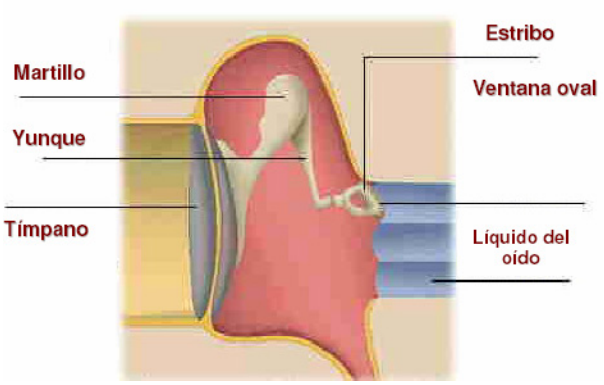
El oído se encuentra dividido en tres zonas: externa, media e interna. La mayor parte del oído interno está rodeada por el hueso temporal.

El oído externo es la parte del aparato auditivo que se encuentra en posición lateral al tímpano o membrana timpánica. Comprende la oreja o pabellón auricular o auditivo (lóbulo externo del oído) y el conducto auditivo externo, que mide tres centímetros de longitud aproximadamente.

El oído medio se encuentra situado en la cavidad timpánica llamada caja del tímpano, cuya cara externa está formada por la membrana timpánica, o tímpano, que lo separa del oído externo. Incluye el mecanismo responsable de la conducción de las ondas sonoras hacia el oído interno. Es un conducto estrecho, o fisura, que se extiende unos 15 mm en un recorrido vertical y otros 15 mm en recorrido horizontal aproximadamente. El oído medio está en comunicación directa con la nariz y la garganta a través de la trompa de Eustaquio, que permite la entrada y la salida de aire del oído medio para equilibrar las diferencias de presión entre éste y el exterior. Hay una cadena formada por tres huesos pequeños y móviles que atraviesa el oído medio.

Estos tres huesos reciben los nombres de martillo, yunque y estribo. Los tres conectan acústicamente el tímpano con el oído interno, que contiene un líquido.

Figura. Estructura del oído.



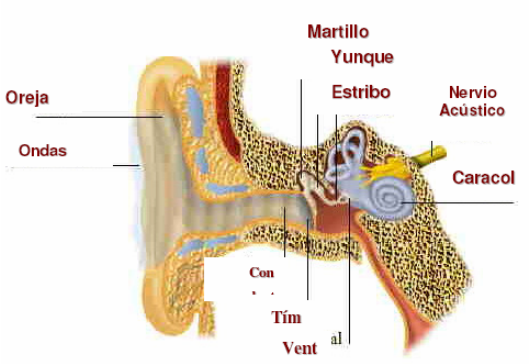
El oído interno, se encuentra en el interior del hueso temporal que contiene los órganos auditivos y del equilibrio, que están inervados por los filamentos del nervio auditivo.

Está separado del oído medio por la fenestra ovalis, o ventana oval. El oído interno consta de una serie de canales membranosos alojados en una parte densa del hueso temporal, y está dividido en: cóclea (en griego, 'caracol óseo'), vestíbulo y tres canales semicirculares. Estos tres canales se comunican entre sí y contienen un fluido gelatinoso denominado endolinfa.

Percepción de las ondas sonoras.

Para poder percibir las ondas sonoras, el cuerpo cuenta con un complejo mecanismo formado por el oído externo, el oído medio y el interno.

Figura. Recorrido de las ondas sonoras a través del oído.



La aurícula recoge las ondas sonoras y las conduce por el conducto auditivo. Las ondas sonoras chocan contra el tímpano, que, como consecuencia, vibra. Las vibraciones se transmiten gracias a una cadena de tres huesecillos: El martillo, el yunque y el estribo. Las vibraciones pasan por la ventana oval y llegan al caracol, ya en el interior del oído interno. Allí las vibraciones se convierten en impulsos nerviosos, estos transcurren por el nervio acústico hasta el cerebro, donde son interpretados como sonidos.

Las ondas sonoras son producidas por cambios en la presión del aire, estas son transmitidas a través del canal auditivo externo hacia el tímpano, en el cual se produce una vibración. Estas vibraciones se comunican al oído medio mediante la cadena de huesecillos (martillo, yunque y estribo) y, a través de la ventana oval, hasta el líquido del oído interno. El movimiento de la endolinfa que se produce al vibrar la cóclea, estimula el movimiento de un grupo de proyecciones finas, similares a cabellos, denominadas células pilosas. El conjunto de células pilosas constituye el órgano de Corti. Las células pilosas transmiten señales directamente al nervio auditivo, el cual lleva la información al cerebro. El patrón de respuesta de las células pilosas a las vibraciones de la cóclea codifica la información sobre el sonido para que pueda ser interpretada por los centros auditivos del cerebro.

El rango máximo de audición en el hombre incluye frecuencias de sonido desde 16 hasta 28.000 ciclos por segundo. El menor cambio de tono que puede ser captado por el oído varía en función del tono y del volumen. Los oídos humanos más sensibles son capaces de detectar cambios en la frecuencia de vibración (tono) que correspondan al 0,03% de la frecuencia original, en el rango comprendido entre 500 y 8.000 vibraciones por segundo. El oído es menos sensible a los cambios de frecuencia si se trata de sonidos de frecuencia o de intensidad bajas.

La sensibilidad del oído a la intensidad del sonido (volumen) también varía con la frecuencia. La sensibilidad a los cambios de volumen es mayor entre los 1.000 y los 3.000 ciclos, de manera que se pueden detectar cambios de un decibelio. Esta sensibilidad es menor cuando se reducen los niveles de intensidad de sonido.

Las diferencias en la sensibilidad del oído a los sonidos fuertes causan varios fenómenos importantes. Los tonos muy altos producen tonos diferentes en el oído, que no están presentes en el tono original. Es probable que estos tonos subjetivos estén producidos por imperfecciones en la función natural del oído medio. Las discordancias de la tonalidad que producen los incrementos grandes de la intensidad de sonido, es consecuencia de los tonos subjetivos que se producen en el oído. Esto ocurre, por ejemplo, cuando el control del volumen de un aparato de radio está ajustado. La intensidad de un tono puro también afecta a su entonación. Los tonos altos pueden incrementar hasta una nota de la escala musical; los tonos bajos tienden a hacerse cada vez más bajos a medida que aumenta la intensidad del sonido. Este efecto sólo se percibe en tonos puros. Puesto que la mayoría de los tonos musicales son complejos, por lo general, la audición no se ve afectada por este fenómeno de un modo apreciable. Cuando se enmascaran sonidos, la producción de armonías de tonos más bajos en el oído puede amortiguar la percepción de los tonos más altos. El enmascaramiento es lo que hace necesario elevar la propia voz para poder ser oído en lugares ruidosos

ANEXO C. Código fuente del programa en MATLAB

INTERFAZ GRÁFICA

```
function fig = Tesis_fig()
% This is the machine-generated representation of a Handle Graphics
object
% and its children.  Note that handle values may change when these
objects
% is re-created. This may cause problems with any callbacks written
to
% depend on the value of the handle at the time the object was
saved.
% This problem is solved by saving the output as a FIG-file.
%
% To reopen this object, just type the name of the M-file at the
MATLAB
% prompt. The M-file and its associated MAT-file must be on your
path.
%
% NOTE: certain newer features in MATLAB may not have been saved in
this
% M-file due to limitations of this format, which has been
superseded by
% FIG-files.  Figures which have been annotated using the plot
editor tools
% are incompatible with the M-file/MAT-file format, and should be
saved as
% FIG-files.

load Tesis_fig

h0 = figure('Color',[0 0.501960784313725 0.501960784313725], ...
   'Colormap',mat0, ...
   'FileName','C:\MATLABR11\work\Tesis Final\Tesis.fig.m', ...
   'HandleVisibility','off', ...
   'Name','Clasificacion de Voces Normales y Patologicas', ...
   'NumberTitle','off', ...
   'PaperPosition',[18 180 576 432], ...
   'PaperUnits','points', ...
   'Position',[1 29 1024 662], ...
   'ResizeFcn','doresize(gcbf)', ...
   'Tag','Clasificacion de Voces Normales y Patologicas', ...
   'ToolBar','figure', ...
   'DefaultaxesCreateFcn','plotedit(gcbf,''promoteoverlay''); ');
h1 = uimenu('Parent',h0, ...
   'HandleVisibility','off', ...
   'Tag','ScribeHGBinObject', ...
   'Visible','off');
```

```
h1 = uimenu('Parent',h0, ...
   'HandleVisibility','off', ...
   'Tag','ScribeFigObjStorage', ...
   'Visible','off');
h1 = uicontrol('Parent',h0, ...
   'Units','points', ...
   'BackgroundColor',[0 0.501960784313725 0.752941176470588], ...
   'Callback','Normales', ...
   'FontName','Tahoma', ...
   'FontSize',10, ...
   'ListboxTop',0, ...
   'Position',[127.5 296.25 94.5 23.25], ...
   'String','Voces Normales', ...
   'Tag','Pushbutton1');
h1 = uicontrol('Parent',h0, ...
   'Units','points', ...
   'BackgroundColor',[0 0.501960784313725 0.752941176470588], ...
   'Callback','Patologicas', ...
   'FontName','Tahoma', ...
   'FontSize',10, ...
   'ListboxTop',0, ...
   'Position',[126.75 249.75 94.5 22.5], ...
   'String','Voces Patologicas', ...
   'Tag','Pushbutton2');
h1 = uicontrol('Parent',h0, ...
   'Units','points', ...
   'BackgroundColor',[0 0.501960784313725 0.501960784313725], ...
   'FontAngle','oblique', ...
   'FontName','Engravers MT', ...
   'FontSize',12, ...
   'ListboxTop',0, ...
   'Position',[232.5 429 348 48.75], ...
   'String','Clasificacion de Voces Normales y Patologicas Empleando
la Transformada Wavelet', ...
   'Style','text', ...
   'Tag','StaticText1');
h1 = uicontrol('Parent',h0, ...
   'Units','points', ...
   'BackgroundColor',[1 1 1], ...
   'FontAngle','italic', ...
   'FontName','Tahoma', ...
   'FontSize',14, ...
   'ListboxTop',0, ...
   'Position',[91.5 352.5 154.5 26.25], ...
   'String','Extraer Características', ...
   'Style','text', ...
   'Tag','StaticText2');
h1 = uicontrol('Parent',h0, ...
   'Units','points', ...
   'BackgroundColor',[1 1 1], ...
   'FontAngle','italic', ...
   'FontName','Tahoma', ...
```

```
    'FontSize',14, ...
    'ListboxTop',0, ...
    'Position',[326.25 353.25 154.5 26.25], ...
    'String','Entrenar Red', ...
    'Style','text', ...
    'Tag','StaticText2');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[1 1 1], ...
    'FontAngle','italic', ...
    'FontName','Tahoma', ...
    'FontSize',14, ...
    'ListboxTop',0, ...
    'Position',[556.5 354 154.5 26.25], ...
    'String','Clasificación', ...
    'Style','text', ...
    'Tag','StaticText2');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[0 0.501960784313725 0.752941176470588], ...
    'Callback','RedA', ...
    'FontName','Tahoma', ...
    'FontSize',10, ...
    'ListboxTop',0, ...
    'Position',[357 297 94.5 23.25], ...
    'String','Red Neuronal A', ...
    'Tag','Pushbutton1');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[0 0.501960784313725 0.752941176470588], ...
    'Callback','RedE', ...
    'FontName','Tahoma', ...
    'FontSize',10, ...
    'ListboxTop',0, ...
    'Position',[358.5 253.5 94.5 23.25], ...
    'String','Red Neuronal E', ...
    'Tag','Pushbutton1');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[0 0.501960784313725 0.752941176470588], ...
    'Callback','RedI', ...
    'FontName','Tahoma', ...
    'FontSize',10, ...
    'ListboxTop',0, ...
    'Position',[359.25 206.25 94.5 23.25], ...
    'String','Red Neuronal I', ...
    'Tag','Pushbutton1');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[0 0.501960784313725 0.752941176470588], ...
    'Callback','RedO', ...
    'FontName','Tahoma', ...
```

```
    'FontSize',10, ...
    'ListboxTop',0, ...
    'Position',[360 165 94.5 23.25], ...
    'String','Red Neuronal O', ...
    'Tag','Pushbutton1');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[0 0.501960784313725 0.752941176470588], ...
    'Callback','RedU', ...
    'FontName','Tahoma', ...
    'FontSize',10, ...
    'ListboxTop',0, ...
    'Position',[360 123 94.5 23.25], ...
    'String','Red Neuronal U', ...
    'Tag','Pushbutton1');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[0 0.501960784313725 0.752941176470588], ...
    'Callback','Clasificacion ClasificacionA;', ...
    'FontName','Tahoma', ...
    'FontSize',10, ...
    'ListboxTop',0, ...
    'Position',[582.75 297.75 105.75 23.25], ...
    'String','Clasificación Vocal A', ...
    'Tag','Pushbutton1');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[0 0.501960784313725 0.752941176470588], ...
    'Callback','Clasificacion ClasificacionE;', ...
    'FontName','Tahoma', ...
    'FontSize',10, ...
    'ListboxTop',0, ...
    'Position',[584.25 255 105.75 23.25], ...
    'String','Clasificación Vocal E', ...
    'Tag','Pushbutton1');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[0 0.501960784313725 0.752941176470588], ...
    'Callback','Clasificacion ClasificacionI;', ...
    'FontName','Tahoma', ...
    'FontSize',10, ...
    'ListboxTop',0, ...
    'Position',[584.25 207.75 105.75 23.25], ...
    'String','Clasificación Vocal I', ...
    'Tag','Pushbutton1');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[0 0.501960784313725 0.752941176470588], ...
    'Callback','Clasificacion ClasificacionO;', ...
    'FontName','Tahoma', ...
    'FontSize',10, ...
    'ListboxTop',0, ...
```

```
    'Position',[584.25 164.25 105.75 23.25], ...
    'String','Clasificación Vocal O', ...
    'Tag','Pushbutton1');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[0 0.501960784313725 0.752941176470588], ...
    'Callback','Clasificacion ClasificacionU;', ...
    'FontName','Tahoma', ...
    'FontSize',10, ...
    'ListboxTop',0, ...
    'Position',[583.5 122.25 105.75 23.25], ...
    'String','Clasificación Vocal U', ...
    'Tag','Pushbutton1');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[0 0.501960784313725 0.501960784313725], ...
    'ListboxTop',0, ...
    'Position',[41.25 108 116.25 18], ...
    'String','Autor:  Reinaldo Rodriguez V.', ...
    'Style','text', ...
    'Tag','StaticText3');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[0 0.501960784313725 0.501960784313725], ...
    'ListboxTop',0, ...
    'Position',[40.5 89.25 116.25 18], ...
    'String','Directora:  Sonia Contreras O.', ...
    'Style','text', ...
    'Tag','StaticText3');
h1 = uicontrol('Parent',h0, ...
    'Units','points', ...
    'BackgroundColor',[0 0.501960784313725 0.501960784313725], ...
    'ListboxTop',0, ...
    'Position',[39.75 70.5 116.25 18], ...
    'String','Trabajo de Grado 2006', ...
    'Style','text', ...
    'Tag','StaticText3');
h1 = axes('Parent',h0, ...
    'CameraUpVector',[0 1 0], ...
    'CameraUpVectorMode','manual', ...
    'Color','none', ...
    'ColorOrder',mat1, ...
    'CreateFcn','', ...
    'HandleVisibility','off', ...
    'HitTest','off', ...
    'Position',[-0.015625 0.06948640483383686 1 1], ...
    'Tag','ScribeOverlayAxesActive', ...
    'Visible','off', ...
    'XColor',[0.8 0.8 0.8], ...
    'XLimMode','manual', ...
    'XTickMode','manual', ...
    'YColor',[0.8 0.8 0.8], ...
```

```
    'YLimMode','manual', ...
    'YTickMode','manual', ...
    'ZColor',[0 0 0]);
h2 = text('Parent',h1, ...
    'Color',[0.8 0.8 0.8], ...
    'HandleVisibility','off', ...
    'HorizontalAlignment','center', ...
    'Position',[0.4995112414467253 -0.01210287443267777
9.160254037844386], ...
    'VerticalAlignment','cap', ...
    'Visible','off');
set(get(h2,'Parent'),'XLabel',h2);
h2 = text('Parent',h1, ...
    'Color',[0.8 0.8 0.8], ...
    'HandleVisibility','off', ...
    'HorizontalAlignment','center', ...
    'Position',[-0.005865102639296186 0.4977307110438729
9.160254037844386], ...
    'Rotation',90, ...
    'VerticalAlignment','baseline', ...
    'Visible','off');
set(get(h2,'Parent'),'YLabel',h2);
h2 = text('Parent',h1, ...
    'Color',[0 0 0], ...
    'HandleVisibility','off', ...
    'HorizontalAlignment','right', ...
    'Position',[0.01466275659824047 0.9288956127080181
9.160254037844386], ...
    'Visible','off');
set(get(h2,'Parent'),'ZLabel',h2);
h2 = text('Parent',h1, ...
    'Color',[0 0 0], ...
    'HandleVisibility','off', ...
    'HorizontalAlignment','center', ...
    'Position',mat2, ...
    'VerticalAlignment','bottom', ...
    'Visible','off');
set(get(h2,'Parent'),'Title',h2);
if nargout > 0, fig = h0; end
```

EXTRAER CARACTERISTICAS

```
% Toma de patrones de entrada

function [Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind);

load parte1 % en este archivo estan los vectores de e y Zr del
silencio y voz solo para voces normales

%**********
y = y-mean(y);
%**************
%sound(y,fs)
t=0:1/fs:length(y)/fs-1/fs;%vector en tiempo real
ts=1/fs;
N=round(0.0256/ts); % longitud de la ventana de 256 muestras
e=floor(length(y)/N);
endpoint=[];
for i=1:e
   v=(i-1)*N+1;
   y1=y(v:v+N-1);

   es(i)=10*log10(eps+(1/N)*sum(y1.^2)); medicion de la energia
   ss = sign(y1);
   vi = find(ss~=0);
   Zr(i) = sum(abs(diff(ss(vi)))/2);% tasa de cruces por cero

   x1=normpdf(es(i),mesil,stdesil);% valores de energia para
silencio
   x2=normpdf(Zr(i),mzrsil,stdzrsil); %valores de cruces por cero
para silencio
   psil=x1*x2;% probabilidad de silencio
    x3=normpdf(es(i),mevoz,stdevoz);% valores de energia para voz
   x4=normpdf(Zr(i),mzrvoz,stdzrvoz); %valores de cruces por cero
para voz
   pvoz=x3*x4;% probabilidad de voz

   if psil>pvoz,
      endpoint(i)=0;
   else
      endpoint(i)=1;
   end

end
```

```
%*********************************************************
% PUNTOS DE INCIO Y FINAL DE LA VOCAL
%*********************************************************

dif = diff(endpoint); % resta el siguiente - el anterior
indc = find(dif==1);  % se encuentra el punto inicial donde comienza
la vocal
indf = find(dif==-1); % se encuentra el punto final

%*********************************************************
% VOCAL A
%*********************************************************

%************
y = y/std(y); %normalización estadística
%***************
%figure
aa = y(indc(1)*N:indf(1)*N);
%k = [1: length(aa)];
%sound(aa,fs)
%plot(k,aa)

W=hamming(length(aa)); % VENTANA DE HAMMING
b=[1 -0.95];
a=1;
saa=filter(b,a,aa); %filtro de preenfasis
saa=saa.*W;% ENVENTANADO
%////////////////////////////////////////////////////////
%DESCOMPOSICION WAVELET
%////////////////////////////////////////////////////////
[Lo_D,Hi_D,Lo_R,Hi_R] = wfilters('db8');
[C,L]=wavedec(saa,6,Lo_D,Hi_D);
%[C,L]=wavedec(saa,6,'db8');  %the sixth-level approximation and the
first three levels of detail) are returned
                                %concatenated into one vector, C.
Vector L gives the lengths of each component.

   cA6 = appcoef(C,L,'db8',6);%To extract the level 6 approximation
coefficients from C
   cD6 = detcoef(C,L,6);      %To extract  detail coefficients from
C,
   cD5 = detcoef(C,L,5);
   cD4 = detcoef(C,L,4);
   cD3 = detcoef(C,L,3);
   cD2 = detcoef(C,L,2);
   cD1 = detcoef(C,L,1);

   Maa(ind,1)=10*log10(eps+mean(cA6.^2));
   Maa(ind,2)=10*log10(eps+mean(cD6.^2));
   Maa(ind,3)=10*log10(eps+mean(cD5.^2));
   Maa(ind,4)=10*log10(eps+mean(cD4.^2));
   Maa(ind,5)=10*log10(eps+mean(cD3.^2));
```

```
   Maa(ind,6)=10*log10(eps+mean(cD2.^2));
   Maa(ind,7)=10*log10(eps+mean(cD1.^2));


%*************************************************************
% VOCAL E
%*************************************************************

ee = y(indc(2)*N:indf(2)*N);
Wee=hamming(length(ee)); % VENTANA DE HAMMING
b=[1 -0.95];
a=1;
see=filter(b,a,ee); %filtro de preenfasis
see=see.*Wee;% ENVENTANADO
%////////////////////////////////////////////////////////////
%DESCOMPOSICION WAVELET
%////////////////////////////////////////////////////////////
[Lo_D,Hi_D,Lo_R,Hi_R] = wfilters('db8');
[C,L]=wavedec(see,6,Lo_D,Hi_D);
%[C,L]=wavedec(see,6,'db8');  %the sixth-level approximation and the
first three levels of detail) are returned
                                %concatenated into one vector, C.
Vector L gives the lengths of each component.

   cA6 = appcoef(C,L,'db8',6);%To extract the level 6 approximation
coefficients from C
   cD6 = detcoef(C,L,6);      %To extract  detail coefficients from
C,
   cD5 = detcoef(C,L,5);
   cD4 = detcoef(C,L,4);
   cD3 = detcoef(C,L,3);
   cD2 = detcoef(C,L,2);
   cD1 = detcoef(C,L,1);

   Mee(ind,1)=10*log10(eps+mean(cA6.^2));
   Mee(ind,2)=10*log10(eps+mean(cD6.^2));
   Mee(ind,3)=10*log10(eps+mean(cD5.^2));
   Mee(ind,4)=10*log10(eps+mean(cD4.^2));
   Mee(ind,5)=10*log10(eps+mean(cD3.^2));
   Mee(ind,6)=10*log10(eps+mean(cD2.^2));
   Mee(ind,7)=10*log10(eps+mean(cD1.^2));

%*************************************************************
% VOCAL I
%*************************************************************

ii = y(indc(3)*N:indf(3)*N);
Wii=hamming(length(ii)); % VENTANA DE HAMMING
b=[1 -0.95];
a=1;
sii=filter(b,a,ii); %filtro de preenfasis
sii=sii.*Wii;% ENVENTANADO
```

```
%////////////////////////////////////////////////////////
%DESCOMPOSICION WAVELET
%////////////////////////////////////////////////////////

[Lo_D,Hi_D,Lo_R,Hi_R] = wfilters('db8');
[C,L]=wavedec(sii,6,Lo_D,Hi_D);
%[C,L]=wavedec(sii,6,'db8');  %the sixth-level approximation and the
first three levels of detail) are returned
                                %concatenated into one vector, C.
Vector L gives the lengths of each component.

   cA6 = appcoef(C,L,'db8',6);%To extract the level 6 approximation
coefficients from C
   cD6 = detcoef(C,L,6);      %To extract  detail coefficients from
C,
   cD5 = detcoef(C,L,5);
   cD4 = detcoef(C,L,4);
   cD3 = detcoef(C,L,3);
   cD2 = detcoef(C,L,2);
   cD1 = detcoef(C,L,1);

   Mii(ind,1)=10*log10(eps+mean(cA6.^2));
   Mii(ind,2)=10*log10(eps+mean(cD6.^2));
   Mii(ind,3)=10*log10(eps+mean(cD5.^2));
   Mii(ind,4)=10*log10(eps+mean(cD4.^2));
   Mii(ind,5)=10*log10(eps+mean(cD3.^2));
   Mii(ind,6)=10*log10(eps+mean(cD2.^2));
   Mii(ind,7)=10*log10(eps+mean(cD1.^2));

%*********************************************************
% VOCAL O
%*********************************************************

oo = y(indc(4)*N:indf(4)*N);
Woo=hamming(length(oo)); % VENTANA DE HAMMING
b=[1 -0.95];
a=1;
soo=filter(b,a,oo); %filtro de preenfasis
soo=soo.*Woo;% ENVENTANADO

%////////////////////////////////////////////////////////
%DESCOMPOSICION WAVELET
%////////////////////////////////////////////////////////

[C,L]=wavedec(soo,6,'db8');  %the sixth-level approximation and the
first three levels of detail) are returned
                                %concatenated into one vector, C.
Vector L gives the lengths of each component.

   cA6 = appcoef(C,L,'db8',6);%To extract the level 6 approximation
coefficients from C
  cD6 = detcoef(C,L,6);    %To extract  detail coefficients from C,
```

```
   cD5 = detcoef(C,L,5);
   cD4 = detcoef(C,L,4);
   cD3 = detcoef(C,L,3);
   cD2 = detcoef(C,L,2);
   cD1 = detcoef(C,L,1);

   Moo(ind,1)=10*log10(eps+mean(cA6.^2));
   Moo(ind,2)=10*log10(eps+mean(cD6.^2));
   Moo(ind,3)=10*log10(eps+mean(cD5.^2));
   Moo(ind,4)=10*log10(eps+mean(cD4.^2));
   Moo(ind,5)=10*log10(eps+mean(cD3.^2));
   Moo(ind,6)=10*log10(eps+mean(cD2.^2));
   Moo(ind,7)=10*log10(eps+mean(cD1.^2));

%*********************************************************
% VOCAL U
%*********************************************************

uu = y(indc(5)*N:indf(5)*N);
Wuu=hamming(length(uu)); % VENTANA DE HAMMING
b=[1 -0.95];
a=1;
suu=filter(b,a,uu); %filtro de preenfasis
suu=suu.*Wuu;% ENVENTANADO

%///////////////////////////////////////////////////////////
%DESCOMPOSICION WAVELET
%///////////////////////////////////////////////////////////

[C,L]=wavedec(suu,6,'db8');  %the sixth-level approximation and the
first three levels of detail) are returned
                                %concatenated into one vector, C.
Vector L gives the lengths of each component.

   cA6 = appcoef(C,L,'db8',6);%To extract the level 6 approximation
coefficients from C
   cD6 = detcoef(C,L,6);      %To extract  detail coefficients from
C,
   cD5 = detcoef(C,L,5);
   cD4 = detcoef(C,L,4);
   cD3 = detcoef(C,L,3);
   cD2 = detcoef(C,L,2);
   cD1 = detcoef(C,L,1);

   Muu(ind,1)=10*log10(eps+mean(cA6.^2));
   Muu(ind,2)=10*log10(eps+mean(cD6.^2));
   Muu(ind,3)=10*log10(eps+mean(cD5.^2));
   Muu(ind,4)=10*log10(eps+mean(cD4.^2));
   Muu(ind,5)=10*log10(eps+mean(cD3.^2));
   Muu(ind,6)=10*log10(eps+mean(cD2.^2));
   Muu(ind,7)=10*log10(eps+mean(cD1.^2));
```

BASE DE DATOS DE ENTRENAMIENTO

Voces normales

```
%voces normales
function Normales;

clc, clear all

[y,fs]=wavread('C:\MATLABR11\work\Tesis Final\Voces\Normales\audio
2_05');
ind=1;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind);


[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Normales\Eliana2.wav');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];


[y,fs]=wavread('C:\MATLABR11\work\Tesis Final\Voces\Normales\audio
1_02');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

[y,fs]=wavread('C:\MATLABR11\work\Tesis Final\Voces\Normales\audio
1_06');
L=Maa;
L1=Mee;
L2=Mii;
```

```
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

[y,fs]=wavread('C:\MATLABR11\work\Tesis Final\Voces\Normales\audio
2_02');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

[y,fs]=wavread('C:\MATLABR11\work\Tesis Final\Voces\Normales\audio
2_07');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

[y,fs]=wavread('C:\MATLABR11\work\Tesis Final\Voces\Normales\audio
2_08');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];
```

```
[y,fs]=wavread('C:\MATLABR11\work\Tesis Final\Voces\Normales\audio
2_11');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Normales\NataliaGiraldo');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Normales\Martha');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

%Mn=[Maa Mee Mii Moo Muu];
%ns=size(Mn,1);
%Mn=[Mn 1*ones(ns,1)];

clear y
save normales
```

Voces Patológicas

```
%voces normales
function Patologicas;
clc, clear all

[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Patologicas\Lourdes.wav');
ind=1;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind);


[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Patologicas\Senit2.wav');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];


[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Patologicas\Jenny.wav');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Patologicas\Mirta2.wav');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
```

```
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Patologicas\Yolanda4.wav');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Patologicas\Amelia2.wav');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Patologicas\CGC1.wav');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Patologicas\CR1.wav');
L=Maa;
L1=Mee;
L2=Mii;
```

```
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Patologicas\AB1.wav');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Patologicas\EA1.wav');
L=Maa;
L1=Mee;
L2=Mii;
L3=Moo;
L4=Muu;
[Maa,Mee,Mii,Moo,Muu]=Principal(y,fs,ind); % CALCULA LA NUEVA MATRIX
Maa=[L;Maa];% LE AGREGA LA ANTERIOR
Mee=[L1;Mee];
Mii=[L2;Mii];
Moo=[L3;Moo];
Muu=[L4;Muu];

%Mp=[Maa Mee Mii Moo Muu];
%ns=size(Mp,1);
%Mp=[Mp -1*ones(ns,1)];



clear y
save Patologicas
```

ENTRENAR RED

```
function Redes(action);


clc
switch (action)

   case 'RedA',

   load normales
   Man=[Maa ones(size(Maa,1),1)];
   load patologicas
   Map=[Maa -1*ones(size(Maa,1),1)];
   M=[Man;Map]';

   case 'RedE',

   load normales
   Men=[Mee ones(size(Mee,1),1)];
   load patologicas
   Mep=[Mee -1*ones(size(Mee,1),1)];
   M=[Men;Mep]';

   case 'RedI',

   load normales
   Min=[Mii ones(size(Mii,1),1)];
   load patologicas
   Mip=[Mii -1*ones(size(Mii,1),1)];
   M=[Min;Mip]';

   case 'RedO',

   load normales
   Mon=[Moo ones(size(Moo,1),1)];
   load patologicas
   Mop=[Moo -1*ones(size(Moo,1),1)];
   M=[Mon;Mop]';

   case 'RedU',

   load normales
   Mun=[Muu ones(size(Muu,1),1)];
   load patologicas
   Mup=[Muu -1*ones(size(Muu,1),1)];
   M=[Mun;Mup]';

 end

nd = size(M,1);
```

```
y = M;

%for k=1:nd-1,
%   [y(k,:), min(k,:), max(k,:)] = premnmx(y(k,:)); % Normalización
de la mariz de entrada en un rango de -1, 1
%end

y = y(:,randperm(size(y,2))); %revolver la matriz
N = size(y,2);

%DIVISION DEL SET DE DATOS: ENTRENAMIENTO, VALIDACIÓN Y PRUEBA

ne = round(0.6*N);
nt = round(0.8*N);
Pe = y(1:nd-1,1:ne); % este vector almacena el 60% de los valores de
las entradas para entrenamiento
Te = y(nd,1:ne);  % este vector almacena el 60% de los valores de
las salidas  para entrenamiento
Pv = y(1:nd-1,ne+1:nt);% este vector almacena el 20% de los valores
de las entradas para validación
Tv = y(nd,ne+1:nt); %este vector almacena el 20% de los valores de
las salidas para validación
Pt = y(1:nd-1,nt+1:end);% este vector almacena el 20% de los valores
de las entradas para prueba
Tt = y(nd,nt+1:end); % este vector almacena el 20% de los valores de
las salidas para prueba

[Pen,minPe,maxPe] = premnmx(Pe); %se normaliza el vector de entrada,
en un rango de -1, 1
[Ten,minTe,maxTe] = premnmx(Te); %se normaliza nuevamente el vector
de salida,

[Pvn] = tramnmx(Pv,minPe,maxPe);% como se esta utilizando datos
normalizados , todas las entradas subsecuentes
[Tvn] = tramnmx(Tv,minTe,maxTe);% deben estar transformadas usando
la misma normalización

[Ptn] = tramnmx(Pt,minPe,maxPe);
[Ttn] = tramnmx(Tt,minTe,maxTe);

V.P = Pvn; % entradas para validación
V.T = Tvn; % salidas para validación

% Red backpropagation feedforward

red = newff (minmax(Pen), [12 1],{'tansig', 'tansig'});

%Algoritmo de Levenberg-Marquardt

%red.trainfcn = 'traingda';            % Algoritmo de entrenamiento
%red.trainparam.lr = 0.115;            % Rata de aprendizaje
```

```
%red.trainparam.lr_inc = 1.9;          % Incremento rata de
aprendizaje
%red.trainparam.lr_dec = 0.1;          % Incremento rata de
aprendizaje
red.trainparam.epochs = 1000;
red.trainparam.goal = 1e-6;            % error promedio
red.trainparam.show = 1;

tic
figure
red = train(red,Pen,Ten,[],[],V);
toc

ff = sim(red,Ptn);

% Ajuste de las salidas a los valores reales
ffr = postmnmx(ff,minTe,maxTe);%post procesa los datos normalizados
ffr=ff

for k=1:length(ffr),
   if ffr(k)>0,
      res(k)=1;
   else
      res(k)=-1;
   end
end

vecerr = find(res~=Tt);
error = length(vecerr)*100/length(ffr)
Tt
res
```

CLASIFICACIÓN

```
function Clasificacion(opcion);

clc %clear all

load parte1 % en este archivo estan los vectores de e y Zr del
silencio y voz solo para voces normales

%[y,fs]=wavread('C:\MATLAB6p5\work\Tesis
Final\Voces\Patologicas\Lesty3.wav'); % Primera prueba //claudiamon
%[y,fs]=wavread('C:\MATLAB6p5\work\Tesis
Final\Voces\Patologicas\LH1.wav');
%[y,fs]=wavread('C:\MATLAB6p5\work\Tesis
Final\Voces\Normales\Mangelica.wav');
```

```
[y,fs]=wavread('C:\MATLABR11\work\Tesis
Final\Voces\Normales\Claudia.wav');

y = y-mean(y);
sound(y,fs)
t=0:1/fs:length(y)/fs-1/fs;%vector en tiempo real
ts=1/fs;
N=round(0.0256/ts); % longitud de la ventana de 256 muestras
e=floor(length(y)/N);
endpoint=[];

for i=1:e
   v=(i-1)*N+1;
   y1=y(v:v+N-1);

   es(i)=10*log10(eps+(1/N)*sum(y1.^2)); %segmentacion: medicion de
la energia

   ss = sign(y1);
   vi = find(ss~=0);
   Zr(i) = sum(abs(diff(ss(vi)))/2);% tasa de cruces por cero

   %Criterio de Bayes

   x1=normpdf(es(i),mesil,stdesil);% valores de energia para
silencio
   x2=normpdf(Zr(i),mzrsil,stdzrsil); %valores de cruces por cero
para silencio
   psil=x1*x2;% probabilidad de silencio

   x3=normpdf(es(i),mevoz,stdevoz);% valores de energia para voz
   x4=normpdf(Zr(i),mzrvoz,stdzrvoz); %valores de cruces por cero
para voz
   pvoz=x3*x4;% probabilidad de voz

   if psil>pvoz,
      endpoint(i)=0;
   else
      endpoint(i)=1;
   end

end
subplot(4,1,1),plot(t,y),title('Señal Original')
subplot(4,1,2),plot(es),title('Energia')
subplot(4,1,3),plot(Zr),title('Tasa de Cruces por Cero')
subplot(4,1,4),plot(endpoint),title('Señal Segmentada')

%*********************************************************
% PUNTOS DE INCIO Y FINAL DE LA VOCAL
%*********************************************************

dif = diff(endpoint); % resta el siguiente - el anterior
```

```
indc = find(dif==1);  % se encuentra el punto inicial donde comienza
la vocal
indf = find(dif==-1); % se encuentra el punto final

y = y/std(y); %normalización estadística

switch (opcion)

   case 'ClasificacionA',

%*************************************************************
% VOCAL A
%*************************************************************

aa = y(indc(1)*N:indf(1)*N);
%k = [1: length(aa)];
%sound(aa,fs)
%plot(k,aa)
W=hamming(length(aa)); % VENTANA DE HAMMING
b=[1 -0.95];
a=1;
saa=filter(b,a,aa); %filtro de preenfasis
saa=saa.*W;% ENVENTANADO

%////////////////////////////////////////////////////////////
%DESCOMPOSICION WAVELET
%////////////////////////////////////////////////////////////

[C,L]=wavedec(saa,6,'db8');  %the sixth-level approximation and the
first three levels of detail) are returned
                                  %concatenated into one vector, C.
Vector L gives the lengths of each component.

   cA6 = appcoef(C,L,'db8',6);%To extract the level 6 approximation
coefficients from C
   cD6 = detcoef(C,L,6);      %To extract  detail coefficients from
C,
   cD5 = detcoef(C,L,5);
   cD4 = detcoef(C,L,4);
   cD3 = detcoef(C,L,3);
   cD2 = detcoef(C,L,2);
   cD1 = detcoef(C,L,1);

      vc(1)=10*log10(eps+mean(cA6.^2));
      vc(2)=10*log10(eps+mean(cD6.^2));
      vc(3)=10*log10(eps+mean(cD5.^2));
      vc(4)=10*log10(eps+mean(cD4.^2));
      vc(5)=10*log10(eps+mean(cD3.^2));
      vc(6)=10*log10(eps+mean(cD2.^2));
      vc(7)=10*log10(eps+mean(cD1.^2));
```

```
% REDES NEURONALES.

load redA

   vcn = tramnmx(vc(1:7)',minPe,maxPe);      % Normalización de los
datos
   rta1 = sim(red,vcn);
   rta = postmnmx(rta1,0,1);


  if rta < 0.5,
        disp('La voz es patologica')
   else
        disp('La voz es normal')
  end

save rta;

case ('ClasificacionE'),

   clc
   %************************************************************
% VOCAL E
%************************************************************

ee = y(indc(2)*N:indf(2)*N);
Wee=hamming(length(ee)); % VENTANA DE HAMMING
b=[1 -0.95];
a=1;
see=filter(b,a,ee); %filtro de preenfasis
see=see.*Wee;% ENVENTANADO

%////////////////////////////////////////////////////////////
%DESCOMPOSICION WAVELET
%////////////////////////////////////////////////////////////

[C,L]=wavedec(see,6,'db8');  %the sixth-level approximation and the
first three levels of detail) are returned
                                %concatenated into one vector, C.
Vector L gives the lengths of each component.

   cA6 = appcoef(C,L,'db8',6);%To extract the level 6 approximation
coefficients from C
   cD6 = detcoef(C,L,6);      %To extract  detail coefficients from
C,
   cD5 = detcoef(C,L,5);
   cD4 = detcoef(C,L,4);
   cD3 = detcoef(C,L,3);
   cD2 = detcoef(C,L,2);
   cD1 = detcoef(C,L,1);

      vc(8)=10*log10(eps+mean(cA6.^2));
```

```
      vc(9)=10*log10(eps+mean(cD6.^2));
      vc(10)=10*log10(eps+mean(cD5.^2));
      vc(11)=10*log10(eps+mean(cD4.^2));
      vc(12)=10*log10(eps+mean(cD3.^2));
      vc(13)=10*log10(eps+mean(cD2.^2));
      vc(14)=10*log10(eps+mean(cD1.^2));

% REDES NEURONALES.

load redE

%      vcn = tramnmx(vc',min,max);     % Normalización de los datos
      vcn = tramnmx(vc(8:14)',minPe,maxPe);     % Normalización de
los datos

      rta1 = sim(red,vcn);

      rta = postmnmx(rta1,0,1);
      if rta < 0.5,
         disp('La voz es patologica')
      else
         disp('La voz es normal')
      end

case ('ClasificacionI'),

   %*********************************************************
% VOCAL I
%*********************************************************

ii = y(indc(3)*N:indf(3)*N);
Wii=hamming(length(ii)); % VENTANA DE HAMMING
b=[1 -0.95];
a=1;
sii=filter(b,a,ii); %filtro de preenfasis
sii=sii.*Wii;% ENVENTANADO

%////////////////////////////////////////////////////////
%DESCOMPOSICION WAVELET
%////////////////////////////////////////////////////////

[C,L]=wavedec(sii,6,'db8');  %the sixth-level approximation and the
first three levels of detail) are returned
                                  %concatenated into one vector, C.
Vector L gives the lengths of each component.

   cA6 = appcoef(C,L,'db8',6);%To extract the level 6 approximation
coefficients from C
   cD6 = detcoef(C,L,6);      %To extract  detail coefficients from
C,
   cD5 = detcoef(C,L,5);
   cD4 = detcoef(C,L,4);
```

```
   cD3 = detcoef(C,L,3);
   cD2 = detcoef(C,L,2);
   cD1 = detcoef(C,L,1);

      vc(15)=10*log10(eps+mean(cA6.^2));
      vc(16)=10*log10(eps+mean(cD6.^2));
      vc(17)=10*log10(eps+mean(cD5.^2));
      vc(18)=10*log10(eps+mean(cD4.^2));
      vc(19)=10*log10(eps+mean(cD3.^2));
      vc(20)=10*log10(eps+mean(cD2.^2));
      vc(21)=10*log10(eps+mean(cD1.^2));


% REDES NEURONALES.

load redI

%      vcn = tramnmx(vc',min,max);     % Normalización de los datos
      vcn = tramnmx(vc(15:21)',minPe,maxPe);    % Normalización de
los datos

      rta1 = sim(red,vcn);

      rta = postmnmx(rta1,0,1);
      if rta < 0.5,
         disp('La voz es patologica')
      else
         disp('La voz es normal')
      end


case ('ClasificacionO'),

   %*************************************************************
% VOCAL O
%*************************************************************
clc
oo = y(indc(4)*N:indf(4)*N);
Woo=hamming(length(oo)); % VENTANA DE HAMMING
b=[1 -0.95];
a=1;
soo=filter(b,a,oo); %filtro de preenfasis
soo=soo.*Woo;% ENVENTANADO

%////////////////////////////////////////////////////////////
%DESCOMPOSICION WAVELET
%////////////////////////////////////////////////////////////

[C,L]=wavedec(soo,6,'db8');  %the sixth-level approximation and the
first three levels of detail) are returned
                                %concatenated into one vector, C.
Vector L gives the lengths of each component.
```

```
   cA6 = appcoef(C,L,'db8',6);%To extract the level 6 approximation
coefficients from C
   cD6 = detcoef(C,L,6);      %To extract  detail coefficients from
C,
   cD5 = detcoef(C,L,5);
   cD4 = detcoef(C,L,4);
   cD3 = detcoef(C,L,3);
   cD2 = detcoef(C,L,2);
   cD1 = detcoef(C,L,1);

      vc(22)=10*log10(eps+mean(cA6.^2));
      vc(23)=10*log10(eps+mean(cD6.^2));
      vc(24)=10*log10(eps+mean(cD5.^2));
      vc(25)=10*log10(eps+mean(cD4.^2));
      vc(26)=10*log10(eps+mean(cD3.^2));
      vc(27)=10*log10(eps+mean(cD2.^2));
      vc(28)=10*log10(eps+mean(cD1.^2));


% REDES NEURONAL.

load redO

      vcn = tramnmx(vc(22:28)',minPe,maxPe);    % Normalización de
los datos

      rta1 = sim(red,vcn);

      rta = postmnmx(rta1,0,1);
      if rta < 0.5,
         disp('La voz es patologica')
      else
         disp('La voz es normal')
      end


case ('ClasificacionU'),

   clc
%*************************************************************
% VOCAL U
%*************************************************************

uu = y(indc(5)*N:indf(5)*N);
Wuu=hamming(length(uu)); % VENTANA DE HAMMING
b=[1 -0.95];
a=1;
suu=filter(b,a,uu); %filtro de preenfasis
suu=suu.*Wuu;% ENVENTANADO

%////////////////////////////////////////////////////////////
```

```
%DESCOMPOSICION WAVELET
%////////////////////////////////////////////////////////

[C,L]=wavedec(suu,6,'db8');  %the sixth-level approximation and the
first three levels of detail) are returned
                                  %concatenated into one vector, C.
Vector L gives the lengths of each component.

   cA6 = appcoef(C,L,'db8',6);%To extract the level 6 approximation
coefficients from C
   cD6 = detcoef(C,L,6);      %To extract  detail coefficients from
C,
   cD5 = detcoef(C,L,5);
   cD4 = detcoef(C,L,4);
   cD3 = detcoef(C,L,3);
   cD2 = detcoef(C,L,2);
   cD1 = detcoef(C,L,1);

      vc(29)=10*log10(eps+mean(cA6.^2));
      vc(30)=10*log10(eps+mean(cD6.^2));
      vc(31)=10*log10(eps+mean(cD5.^2));
      vc(32)=10*log10(eps+mean(cD4.^2));
      vc(33)=10*log10(eps+mean(cD3.^2));
      vc(34)=10*log10(eps+mean(cD2.^2));
      vc(35)=10*log10(eps+mean(cD1.^2));


% REDES NEURONALES.

load redU

     vcn = tramnmx(vc(29:35)',minPe,maxPe);     % Normalización de
los datos

      rta1 = sim(red,vcn);

      rta = postmnmx(rta1,0,1);
      if rta < 0.5,
         disp('La voz es patologica')
      else
         disp('La voz es normal')
      end


end
```