



ESCUELA SUPERIOR POLITÉCNICA DEL EJÉRCITO
ESPE – LATACUNGA
CARRERA DE INGENIERÍA ELECTRÓNICA E
INSTRUMENTACIÓN

PROYECTO DE GRADO PARA LA OBTENCIÓN
DEL TÍTULO EN INGENIERÍA ELECTRÓNICA E
INSTRUMENTACIÓN

**“DISEÑO E IMPLEMENTACION DE UN VOCODER
DE SUBBANDAS, PEDAGOGICO PARA EL IDIOMA
CASTELLANO**

KATYA TORRES V.

Latacunga-Ecuador
2005

CERTIFICACION

**Certificamos que el presente trabajo fue desarrollado por la señora:
Katya Torres Vasquez, bajo nuestra supervisión.**

**Ing. Marcelo Silva
DIRECTOR DE TESIS**

**Ing. Augusto Bourgeat T.
CODIRECTOR DE TESIS**

AGRADECIMIENTO

A la Escuela Politécnica del Ejército

Al Ingeniero Marcelo Silva, por compartir su experiencia y trabajo en la elaboración de esta tesis

Al Ingeniero Augusto Bourgeat codirector de tesis

DEDICATORIA

A mis hijas Katty Gabriela y María Alejandra,
razón y prolongación de mi existencia.

A mi madre y hermanas

INTRODUCCIÓN

La integración de los sistemas de comunicación, junto con el constante crecimiento de la digitalización de las redes que utilizan dichos sistemas, hacen necesario una administración eficiente de los recursos disponibles. Uno de estos recursos, corresponde a la capacidad de memoria de almacenamiento de datos. Más específicamente, se busca que dicha capacidad sea maximizada, a través de la reducción de los paquetes de información almacenados. Es así como surge la compresión de datos como una de las posibles soluciones al problema planteado en la administración eficiente de la capacidad de memoria.

La búsqueda de métodos de compresión de datos está ligada al tipo de datos almacenados, entendiéndose por tipo al origen de dichos datos. Dentro de estos

tipos de datos, tenemos los datos procedentes del proceso de digitalización de la voz humana. A su vez, dichos datos pueden ser nuevamente clasificados según el método de digitalización de voz efectuado.

Dentro de las aplicaciones de la compresión de audio, se distinguen 4 áreas: Broadcasting (difusión), almacenamiento, multimedia y telecomunicaciones. Ejemplos de estos son: almacenamiento en disco (CD audio, video, etc.), televisión por cable y satelital, Internet (audio y video “streaming”), etc.

Este trabajo de investigación tienen como objetivo fundamental la implementación de un sistema de compresión de voz llamado Vocoder (nombre derivado de voice coder-codificador de voz), para el idioma español, utilizando como herramienta de programación el paquete computacional MATLAB. La función principal del Vocoder en Subbandas es la de codificar la voz a niveles de compresión altos comparado con otros sistemas.

El capítulo I, realiza un análisis de los aspectos básicos sobre la generación de la voz. El Capítulo II cubre, los conceptos necesarios sobre las herramientas utilizadas en el procesamiento digital de señales. El Capítulo III, describe las técnicas utilizadas para la codificación de voz. El Capítulo IV explica todos los fundamentos y procesos utilizados para el desarrollo del sistema de compresión. El Capítulo V, proporciona el análisis de pruebas y resultados realizados para la verificación del funcionamiento correcto del sistema vocoder. El Capítulo VI indica las conclusiones y recomendaciones obtenidas en el desarrollo de esta tesis de ingeniería. Se incluyen anexos, que son de gran utilidad en los que se indican: código del programa (anexo A), manual de funcionamiento del sistema (anexo B), encuesta MOST TEST (anexo C) y recopilación de estudios de vocoder(anexo D).

CAPITULO I

ASPECTOS BÁSICOS SOBRE LA GENERACIÓN DE LA VOZ

1.1 INTRODUCCIÓN

Un sistema de comunicación, en esencia, es la transmisión, recepción y procesamiento de información. Pudiendo definir a la información como el conocimiento, la sabiduría o la realidad y puede ser en forma analógica, tal como

la voz humana, información sobre una imagen de video, o música, o en forma digital como números codificados en binario, códigos alfanuméricos, símbolos gráficos, códigos operacionales del microprocesador o información de base de datos. Se propone estudiar un sistema de comunicación específico, el de la comunicación a través de señales de voz, es decir señales acústicas tradicionalmente emitidas y recibidas por seres humanos en forma oral. Algunos de los objetivos son la representación, análisis, modificación y el reconocimiento automático de mensajes vocales pronunciados por seres humanos.

En todo sistema de comunicación hay varios componentes: emisor, receptor, mensaje, código, canal y contexto. Es necesario conocer algunos aspectos de cada uno de ellos para poder integrar sistemas que funcionen de manera eficaz y eficiente. En este caso el emisor es el conjunto integrado por el cerebro que “piensa” el mensaje y el aparato fonatorio que lo “traduce” a una emisión acústica. El receptor es el aparato auditivo que recibe la onda sonora y la transforma en impulsos nerviosos que luego son interpretados por el cerebro. El mensaje es la idea a comunicar. El código es el lenguaje hablado. La combinación del mensaje y el código constituyen la señal. El canal puede ser el medio en el cual se propaga la onda sonora (en general el aire) o un medio de transmisión electrónico que constituye en sí mismo otro subsistema de comunicación cuyas propiedades son bien conocidas y que se aproxima en muchos casos (aunque no siempre) a la idealidad. El contexto puede tener un sin número de componentes, que van desde factores puramente subjetivos o psicológicos, como el interés, la atención, la motivación hasta factores físicos tales como respuesta en frecuencia, interferencias, distorsiones, ruido, etc.

1.1.1 ALGUNOS CONCEPTOS SOBRE LENGUAJE

Lenguaje en términos generales, es la capacidad que tienen los seres vivos de comunicarse y que en los humanos consiste, básicamente, en comunicarse con los demás asociando ciertos significados a determinados sonidos o signos. El lenguaje tiene un carácter universal.

La lengua es el sistema de signos lingüísticos que permiten la comunicación en una comunidad. Es un sistema pues cada uno de sus elementos tiene entidad propia y entidad relativa a su posición o relación con los otros elementos. Es un código de signos, tiene carácter social, ya que es común a una sociedad.

Por otra parte, se emplea la expresión lenguaje natural en oposición a las llamadas lenguas artificiales, creados por el hombre (por ejemplo, el lenguaje de la lógica, el de la matemática, el de los gestos, el de la semiótica)

El habla es el acto concreto de seleccionar los signos de entre los disponibles y organizarlos a través de ciertas reglas. Materializa el código. Es individual, vale decir que cambia de un individuo a otro. Los signos pueden corresponder al lenguaje escrito o al oral.

El lenguaje es un sistema articulado ya que los sonidos y otros componentes se integran entre sí. Está formado por signos lingüísticos, nombre que recibe la señal en el lenguaje. El lenguaje tiene modalidades regionales llamadas dialectos.

Un signo es algo que reemplaza a otra cosa para comunicarla en un mensaje. Los signos lingüísticos se clasifican en dos tipos: significado y significante. La relación entre significado y significante es arbitraria o convencional, aunque no necesariamente discrecional: involucra acuerdos tácitos, explícitos o normativos en una comunidad lingüística.

En el lenguaje escrito, el significante es la grafía escrita, formada por combinaciones de letras, en tanto que en el lenguaje hablado es su realización acústica mediante la palabra hablada.

Las palabras son los elementos libres mínimos del lenguaje. La sintaxis es el conjunto de reglas para la coordinación de las palabras en frases u oraciones. En su versión escrita las palabras están formadas por letras o grafemas, es decir unidades gráficas mínimas, y, en el caso oral, por fonemas.

Los fonemas son la unidad fónica ideal mínima del lenguaje. Se materializan a través de los sonidos, pero de una manera no unívoca. Las variantes de los fonemas se denominan alófonos. Los monemas son unidades mínimas con significado, que puede ser gramatical, dando origen a los morfemas, o léxico, representado por los lexemas. Los morfemas tienen relación con la gramática, o la forma de organizar o dar estructura a las categorías básicas del lenguaje (género, número, tiempo o persona de los verbos, etc.), mientras que los lexemas se refieren a significados externos al lenguaje mismo.

Las palabras constan de al menos un monema, siendo las más comunes bimonemáticas, que incluyen un lexema y un morfema. En la tabla 1.1 se dan dos ejemplos en los que se identifican los componentes de la palabra.

Tabla 1.1 Ejemplos de monemas, grafemas y fonemas

Palabra	Monemas		Grafema	Fonemas
	Lexema	Morfema		
Gato	Gat	o	G,a,t,o	/g/,/a/,/t/,/o/
Amaban	Ama	Ban	A,m,a,b,a,n	/a/,/m/,/a/,/b/,/a/,/n/

1.2 GENERACIÓN DE LA VOZ

La voz humana se produce voluntariamente por medio del aparato fonatorio. Éste está formado por los pulmones como fuente de energía en la forma de un flujo de aire, la laringe, que contiene las cuerdas vocales, la faringe, las cavidades oral (o bucal) y nasal y una serie de elementos articulatorios: los labios, los dientes, el alvéolo, el paladar, el velo del paladar y la lengua (figura 1.1).

Las cuerdas vocales son, en realidad, dos membranas dentro de la laringe orientadas de adelante hacia atrás (figura 1.2). Por adelante se unen en el cartílago tiroides (que puede palparse sobre el cuello, inmediatamente por debajo

de la unión con la cabeza; en los varones suele precarse como una protuberancia conocida como nuez de Adán). Por detrás, cada una está sujeta a uno de los dos cartílagos aritenoides, los cuales pueden separarse voluntariamente por medio de músculos. La abertura entre ambas cuerdas se denomina glotis.

Cuando las cuerdas vocales se encuentran separadas, la glotis adopta una forma triangular. El aire pasa libremente y prácticamente no se produce sonido. Es el caso de la respiración. Cuando la glotis comienza a cerrarse, el aire que la atraviesa proveniente de los pulmones experimenta una turbulencia, emitiéndose un ruido de origen aerodinámico conocido como aspiración (aunque en realidad acompaña a una espiración o exhalación). Esto sucede en los sonidos denominados “aspirados” (como la h inglesa).

Al cerrarse más, las cuerdas vocales comienzan a vibrar a modo de lengüetas, produciéndose un sonido tonal, es decir periódico. La frecuencia de este sonido depende de varios factores, entre otros del tamaño y la masa de las cuerdas vocales, de la tensión que se les aplique y de la velocidad del flujo del aire proveniente de los pulmones. A mayor tamaño, menor frecuencia de vibración, lo cual explica por qué en los varones, cuya glotis es en promedio mayor que la de las mujeres, la voz es en general más grave. A mayor tensión la frecuencia aumenta, siendo los sonidos más agudos. Así, para lograr emitir sonidos en el registro extremo de la voz es necesario un mayor esfuerzo vocal.

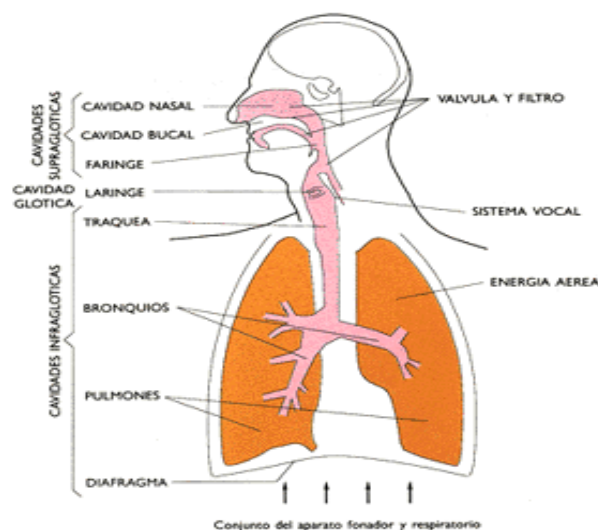


Figura 1.1 Corte esquemático del aparato fonatorio humano

También aumenta la frecuencia (a igualdad de las otras condiciones) al crecer la velocidad del flujo de aire, razón por la cual al aumentar la intensidad de emisión se tiende a elevar espontáneamente el tono de voz.

Finalmente, es posible obturar la glotis completamente. En ese caso no se produce sonido. Sobre la glotis se encuentra la epiglotis, un cartílago en la faringe que permite tapar la glotis durante la deglución para evitar que el alimento ingerido se introduzca en el tracto respiratorio. Durante la respiración y la fonación (emisión de sonido) la epiglotis está separada de la glotis permitiendo la circulación del flujo de aire. Durante la deglución, en cambio, la laringe ejecuta un movimiento ascendente de modo que la glotis apoya sobre la epiglotis.

La porción que incluye las cavidades faríngea, oral y nasal junto con los elementos articulatorios se denomina genéricamente cavidad supraglótica, en tanto que los espacios por debajo de la laringe, es decir la tráquea, los bronquios y los pulmones, se denominan cavidades infraglóticas.

Varios de los elementos de la cavidad supraglótica se controlan a voluntad, permitiendo modificar dentro de márgenes muy amplios los sonidos producidos por las cuerdas vocales o agregar partes distintivas a los mismos, e inclusive producir sonidos propios. Todo esto se efectúa por dos mecanismos principales: el filtrado y la articulación.

El filtrado actúa modificando el espectro del sonido. Tiene lugar en las cuatro cavidades supraglóticas principales: la faringe, la cavidad nasal, la cavidad oral y la cavidad labial. Las mismas constituyen resonadores acústicos que enfatizan determinadas bandas frecuenciales del espectro generado por las cuerdas vocales, conduciendo al concepto de formantes, es decir una serie de picos de resonancia ubicados en frecuencias o bandas de frecuencia que, son bastante específicas para cada tipo de sonido.

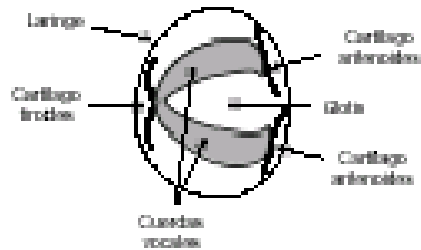


Figura 1.2 Corte esquemático de la laringe según un plano horizontal.

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

1.2.1 CLASIFICACIÓN DE LOS SONIDOS DE LA VOZ

Los sonidos emitidos por el aparato fonatorio pueden clasificarse de acuerdo con diversos criterios que tienen en cuenta los diferentes aspectos del fenómeno de emisión. Estos criterios son:

- a) Según su carácter vocálico o consonántico.
- b) Según su oralidad o nasalidad
- c) Según su carácter tonal (sonoro) o no tonal (sordo)
- d) Según el lugar de articulación
- e) Según el modo de articulación
- f) Según la posición de los órganos articulatorios
- g) Según la duración

1.2.2 VOCALES Y CONSONANTES

Desde un punto de vista mecanoacústico, las vocales son los sonidos emitidos por la sola vibración de las cuerdas vocales sin ningún obstáculo o constricción

entre la laringe y las aberturas oral y nasal. Dicha vibración se genera por el principio del oscilador de relajación, donde interviene una fuente de energía constante en la forma de un flujo de aire proveniente de los pulmones. Son siempre sonidos de carácter tonal (cuasiperiódicos), y por consiguiente de espectro discreto.

Las consonantes, por el contrario, se emiten interponiendo algún obstáculo formado por los elementos articulatorios. Los sonidos correspondientes a las consonantes pueden ser tonales o no dependiendo de si las cuerdas vocales están vibrando o no. Funcionalmente, en el castellano las vocales pueden constituir palabras completas, no así las consonantes.

1.2.3 ORALIDAD Y NASALIDAD

Los fonemas en los que el aire pasa por la cavidad nasal se denominan nasales, en tanto que aquéllos en los que sale por la boca se denominan orales. La diferencia principal está en el tipo de resonador principal por encima de la laringe (cavidad nasal y oral, respectivamente). En castellano son nasales sólo las consonantes “m”, “n”, “ñ”.

1.2.4 TONALIDAD

Los fonemas en los que participa la vibración de las cuerdas vocales se denominan tonales o, también, sonoros. La tonalidad lleva implícito un espectro cuasi periódico.

Como se puntualizó anteriormente, todas las vocales son tonales, pero existen varias consonantes que también lo son: “b”, “d”, “m”, etc. Aquellos fonemas producidos sin vibraciones glotales se denominan sordos. Varios de ellos son el

resultado de la turbulencia causada por el aire pasando a gran velocidad por un espacio reducido, como las consonantes “s”, “z”, “j”, “f”.

1.2.5 LUGAR Y MODO DE ARTICULACIÓN (CONSONANTES)

La articulación es el proceso mediante el cual alguna parte del aparato fonatorio interpone un obstáculo para la circulación del flujo de aire. Las características de la articulación permitirán clasificar las consonantes. Los órganos articulatorios son los labios, los dientes, las diferentes partes del paladar (alvéolo, paladar duro, paladar blando o velo), la lengua y la glotis. Salvo la glotis, que puede articular por sí misma, el resto de los órganos articula por oposición con otro. Según el lugar o punto de articulación se tienen fonemas:

Bilabiales: oposición de ambos labios

Labiodentales: oposición de los dientes superiores con el labio inferior

Linguodentales: oposición de la punta de la lengua con los dientes superiores

Alveolares: oposición de la punta de la lengua con la región alveolar

Palatales: oposición de la lengua con el paladar duro

Velares: oposición de la parte posterior de la lengua con el paladar blando

Glotales: articulación en la propia glotis. A su vez, para cada punto de articulación ésta puede efectuarse de diferentes modos, dando lugar a fonemas:

Oclusivos: la salida del aire se cierra momentáneamente por completo

Fricativos: el aire sale atravesando un espacio estrecho

Africados: oclusión seguida por fricación

Laterales: la lengua obstruye el centro de la boca y el aire sale por los lados

Vibrantes: la lengua vibra cerrando el paso del aire intermitentemente

Aproximantes: La obstrucción muy estrecha que no llega a producir turbulencia

Los fonemas oclusivos (correspondientes a las consonantes “b” inicial, “c”, “k”, “d”, “g”, “p”, “t”) también se denominan a veces explosivos, debido a la liberación repentina de la presión presente inmediatamente antes de su emisión. Pueden ser sordos o sonoros, al igual que los fricativos (“b” intervocálica, “f”, “j”, “h” aspirada, “s”, “y”, “z”). Sólo existe un fonema africado en castellano, correspondiente a la “ch”. Los laterales (“l”, “ll”) a veces se denominan líquidos, y son siempre sonoros.

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

En la tabla 1.2 se indican las consonantes clasificadas según el lugar y el modo de articulación, la sonoridad y la oro-nasalidad. En algunos casos una misma consonante aparece en dos categorías diferentes, correspondiente a las diferencias observadas.

Tabla 1.2 Clasificación de las consonantes de la lengua castellana según el lugar y el modo de articulación y la sonoridad.

Lugar De Articulación	Modo de articulación								
	Oral								Nasal
	Oclusiva		Fricativa		Africada	Lateral	Vibrante	Aproximante	Sonora
	Sorda	Sonora	Sorda	Sonora	Sorda	Sonora	Sorda	Sonora	
Bilabial	p	b,v		b,v				w	m

Labiodental			f						
Linguodenta l			z						
Alveolar	t	d	s	y	ch	l	r,rr		n
Palatal				(y)	(Ch)	ll		i	ñ
Velar	k	g	j						
Glotal			h						

1.2.6 POSICIÓN DE LOS ÓRGANOS ARTICULATORIOS (VOCALES)

En el caso de las vocales, la articulación consiste en la modificación de la acción filtrante de los diversos resonadores, lo cual depende de las posiciones de la lengua (tanto en elevación como en profundidad o avance), de la mandíbula inferior, de los labios y del paladar blando. Estos órganos influyen sobre los formantes, permitiendo su control. Podemos clasificar las vocales según la posición de la lengua como se muestra en la tabla 1.3

Tabla 1.3 Clasificación de las vocales castellanas según la posición de la lengua.

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

Otra cualidad controlable es la labialización, es decir el hecho de que se haga participar activamente los labios. Las vocales labializadas, también definidas como redondeadas, son las que redondean los labios hacia adelante, incrementando la longitud efectiva del tracto vocal. La única vocal labializada en el castellano es la “u”.

En otros idiomas, como el francés, el portugués, el catalán y el polaco, así como en lenguas no europeas como el guaraní o el hindi, existe también el matiz de oralidad o nasalidad. En las vocales orales el velo (paladar blando) sube, obturando la nasofaringe, lo cual impide que el aire fluya parcialmente por la

cavidad nasal. En las vocales nasalizadas (u oronasales) el velo baja, liberando el paso del aire a través de la nasofaringe. Se incorpora así la resonancia nasal.

1.2.7 DURACIÓN

La duración de los sonidos, especialmente de las vocales, no tiene importancia a nivel semántico en el castellano, pero sí en el plano expresivo, a través de la agogía, es decir el énfasis o acentuación a través de la duración. En inglés, en cambio, la duración de una vocal puede cambiar completamente el significado de la palabra que la contiene.

1.2.8 EL ALFABETO FONÉTICO INTERNACIONAL

El castellano es un idioma cuya escritura es eminentemente fonética, ya que salvo pocos casos, hay correspondencia entre grafema y fonema. No todos los idiomas tienen esta característica. El inglés es un caso quizás extremo, a tal punto que George Bernard Shaw ha creado posibles ortografías alternativas para algunas palabras basándose en la forma en que sus fonemas aparecen escritos en otras palabras.

Se ha compilado un extenso conjunto de símbolos fonéticos conocido como el Alfabeto Fonético Internacional (Internacional Phonetic Alphabet, IPA) que contiene una gran cantidad de fonemas de los diversos idiomas, y que permite representar de una manera inequívoca los fonemas independientemente del idioma. El subconjunto correspondiente al idioma castellano se indica en la tabla 1.4.

Tabla 1.4 Los fonemas del alfabeto fonético internacional utilizados en la lengua castellana

Fonemas castellanos					
Sonido	Ejemplo	Sonido	Ejemplo	Sonido	Ejemplo
[p]	P aso	[θ]	Z orzal, lápiz	[ɲ]	Mañana, ñ oño
[b]	B ase, vena	[s]	S olo, cosa	[dʒ]	Y o, Yapeyú

[β]	Labor, lavar	[x]	Giro, jarabe	[j]	Bien, biólogo
[t]	Tres, canto	[tʃ]	Hecho, Chubut	[w]	Hueso, buitre
[d]	Dama, andar	[r]	Arder, jarabe		
[ð]	Cedro, verdad	[rr]	Perro, rojo	[a]	Cama
[k]	Caso, disco	[l]	Loable, fiel	[e]	Espera, ver
[g]	Gula, goma	[λ]	Llanto, calle	[i]	vine, iris
[γ]	Agua, negro	[m]	Mamá, ámbar	[o]	Loro, pos
[f]	Fino, tifón	[n]	Nene, joven	[u]	Burla, huracán

1.3 PERCEPCIÓN DE LOS SONIDOS

1.3.1 AUDICIÓN

El oído percibe las variaciones de presión en forma de sonido cuando su periodicidad está entre las 16 y 16.000 variaciones por segundo (de 20 a 20.000 según otras teorías); es decir, cuando su frecuencia está entre 16 y 16.000 Hz (o 20 a 20.000 Hz).

Esta banda de frecuencias audibles se descompone generalmente en tres regiones: frecuencias graves, medias y agudas.

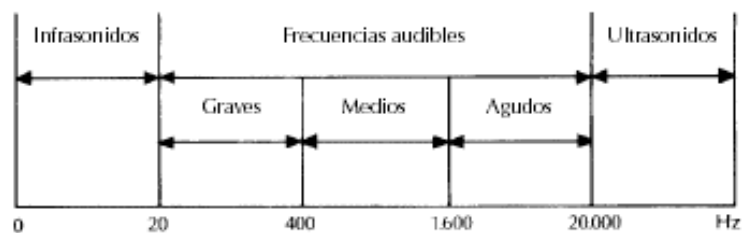


Figura 1.3 Frecuencias Audibles

El modo en que el oído percibe el sonido, es el siguiente (figura 1.4):

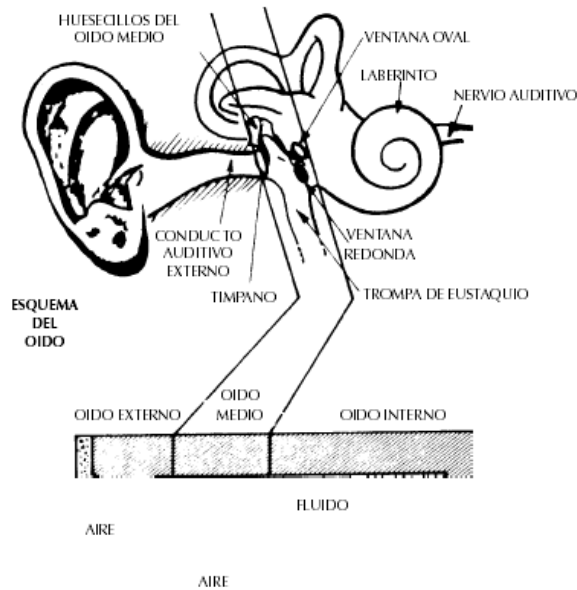


Figura1.4 Esquema del oído Humano

- El oído externo, que fundamentalmente tiene una misión de conducción, pero escasa de percepción.
- El oído medio, que arranca en la membrana del tímpano, que es la que recoge las variaciones de presión.

Estas son transmitidas por un sistema de huesecillos (martillo, yunque y estribo), que actúan como una sucesión de palancas y que constituyen un amplificador (de 55 a 60 veces).

- El oído interno, con apariencia de caracol, está relleno de un líquido (líquido linfático), que es el que transmite finalmente las variaciones de presión al auténtico órgano receptor que es la membrana basal.
- En la membrana basal están las células nerviosas (unas 25.000), son de distinta longitud (similitud con las cuerdas del piano), y según las zonas, recogen los distintos tonos.

También y debido al gran número de células, se hace un análisis de la intensidad, aunque el análisis más fino se realiza, ya, en el cerebro, al que llega esta señal mediante el nervio acústico.

1.3.1.1 INTENSIDAD

Las dos sensaciones fundamentales que nos da el oído, como hemos visto, son el tono y la intensidad.

El tono se puede determinar fácil y objetivamente midiendo la frecuencia.

La intensidad es una magnitud, en parte, subjetiva. Está relacionada con la presión sonora, que es objetivamente medible; sin embargo, dos sonidos de igual presión sonora y de distinta frecuencia no producen la misma sensación de intensidad. Se define como la energía por unidad de superficie y se mide en W/m^2 .

Para que el oído comience a percibir un sonido, la presión acústica debe ser, al menos, de $2 \cdot 10^{-4}$ μ bar. Esto es lo que se denomina Umbral Auditivo.

1.3.1.2 SONORIDAD

Sensibilidad auditiva

Como dijimos en el apartado anterior, el oído humano no es igual de sensible a todas las frecuencias. Fletcher y Munson estudiaron la variación de la sensibilidad del oído con la presión sonora (o, lo que es lo mismo, con el nivel acústico) y resumieron su estudio en unas curvas que dan esta variación de sensibilidad en función de la frecuencia (ver figura 1.5).

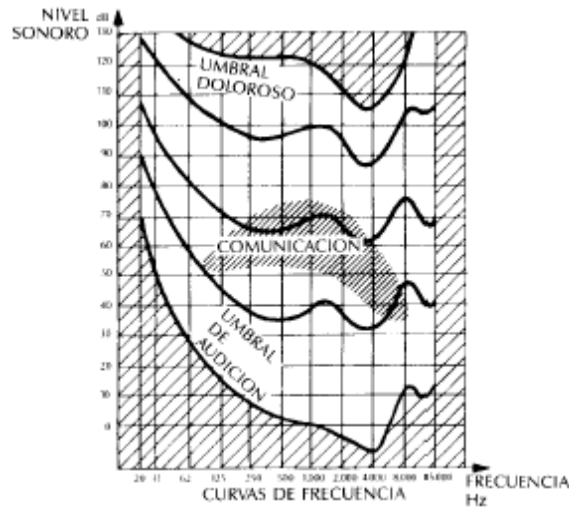


Figura 1.5 Curvas de Frecuencia

Como se ve, la sensibilidad es máxima para 1.000 Hz, es algo menor para frecuencias mayores, y disminuye mucho para bajas frecuencias. Este efecto de sensibilidad depende de las personas y de la edad, la agudeza auditiva disminuye con la edad para frecuencias superiores a los 5.000 Hz.

La sensibilidad del oído humano que hemos visto para tonos puros no es igual en el caso de sonidos y ruidos compuestos de varios tonos. Esto es lo que se conoce como “efecto de enmascaramiento”.

1.3.1.3 MOLESTIA

El ruido, por sus efectos fisiológicos, puede ser una fuente de molestia. La aparición repentina de un ruido inhabitual lleva consigo una modificación de la actividad fisiológica: crecimiento del ritmo cardíaco, modificación del ritmo respiratorio, variación de la presión arterial, etc. Desgraciadamente, la perturbación de un ruido que se debe considerar como molesto no está influenciada solamente por las leyes fisiológicas de la sensibilidad sonora, sino

también por la disposición psicológica, subjetiva y muy variable con el tiempo de cada observador en particular.

El decibelio A

Debido a la subjetividad, es difícil obtener con un solo valor una medida del nivel acústico; es decir, un valor objetivizado que se aproxima lo más posible a la percepción del oído.

Uno de los sistemas empleados para definir con un solo valor el nivel de presión acústica es el decibelio A [dBA]. Esta medida está basada en las curvas antes vistas de Fletcher y Mounson sobre la sensibilidad del oído en función de la frecuencia. Se obtiene mediante la media ponderada entre el espectro del ruido y la curva siguiente, que se conoce como curva de ponderación A.

La medida en dBA se acepta como la valoración simple más aproximada a la sensación producida por música, palabra y ruidos comunitarios más generales, incluidos los de tráfico y electrodomésticos, y siempre que no se trate de ruidos con tonos predominantes.

Esta curva está tomada de la norma UNE 21.314, y se utiliza para compensar las diferencias de sensibilidad que el oído humano tiene para las distintas frecuencias dentro del campo auditivo.

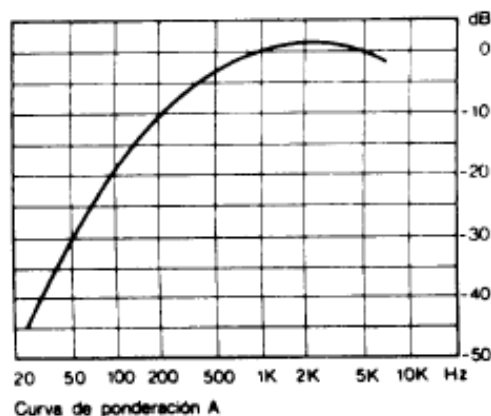


Figura. 1.6 Frecuencias dentro del Campo Auditivo.

En la tabla. 1.5 se especifican los valores que toma la curva de ponderación A para un margen de frecuencias común en la realidad.

Tabla 1.5 Valores de Ponderación A.

Frecuencia en Hz	100	125	160	200	250	315	400	500	630	800
Ponderación en dBA	-19,1	-16,1	-13,4	-10,9	-8,6	-6,6	-4,8	-3,2	-1,0	-0,8
Frecuencia en Hz	1.000	1.250	1.600	2.000	2.500	3.150	4000	5.000		
Ponderación en dBA	0	0,6	1,0	1,2	1,3	1,2	1,0	0,5		

1.3.1.4 VIBRACIONES

Sensación y percepción de vibraciones

Se entiende, en general, por sensación de vibración, la sensación de excitación vibrátil que se produce por contacto directo del cuerpo humano con un cuerpo sólido que vibra. Como no existe un órgano determinado que perciba este tipo de vibración, no es posible una separación clara entre sonido y sensación de vibración, a menos que limitemos la expresión de vibración a las vibraciones por debajo de 16 Hz (o 20 Hz); es decir, los infrasonidos que no se pueden percibir como sonido.

Sin embargo, esta limitación no es razonable ni en sentido físico ni en sentido fisiológico, ya que el oído puede percibir los sonidos que alcanzan y excitan la membrana del tímpano, así como las vibraciones de los huesos del cráneo que excitan directamente al oído interno (audición por conducción ósea, audífonos).

También, las células sensoriales de la piel pueden sentir las vibraciones y, en el caso de ser fuertes, pueden abarcar todo el cuerpo y extender esta sensación a los órganos internos, fundamentalmente a los pulmones y estómago, ya que las bolsas de aire que contienen dichos órganos hacen las veces de amplificador de vibraciones.

1.4 FONÉTICA

En la concepción tradicional, las lenguas se componen de oraciones, las oraciones de palabras y las palabras de sonidos. Pues bien, del estudio de los sonidos de las diferentes lenguas se ocupa la FONÉTICA. Trata por consiguiente de la descripción de los sonidos emitidos por el aparato fonador del hombre, recibidos por los oídos y transmitidos al cerebro, especificando sus rasgos fundamentales y sus relaciones concretas más frecuentes.

La FONÉTICA es hoy una ciencia muy elaborada que ha incorporado determinados dominios de la fisiología y de la física, pero que dispone además de sus propias condiciones de pertinencia, de su propio vocabulario técnico.

A los sonidos se les describe muchas veces de un modo subjetivo y no basado en una técnica, con términos sin sentido alguno: áspero, gutural, suave, plano, oscuro, etc.

Sin embargo, desde un punto de vista objetivo, admiten una descripción a partir de tres perspectivas:

- a) Según las propiedades acústicas de las ondas sonoras que se propagan entre el hablante y el oyente.
- b) Según el modo como son producidos por los órganos de fonación humano.
- c) En función de los efectos físicos que determinan sobre el oído humano y sus mecanismos adyacentes.

Desde esta triple perspectiva, surgen tres divisiones de la FONÉTICA: acústica, articuladora y auditiva.

La fonética acústica: nos dice que el sonido es el resultado de las vibraciones que se transmiten en forma de ondas periódicas cuya frecuencia, amplitud y duración son variables. La transmisión de los sonidos se produce en un medio elástico, como es el aire, pero no lo hace en el vacío. Si se produce una perturbación en un punto del medio elástico, ésta se transmite progresivamente a todos los puntos del mismo, propagándose como una onda. En el aire, a 20° centígrados y 760mm de presión, la velocidad de propagación es de 344 m/s. En el agua, a 17° es de 1430 m/s. Una vibración se corresponde con un movimiento oscilatorio similar al movimiento pendular. La representación de una vibración completa es una onda sonora.

Las cualidades acústicas de los sonidos que se manifiestan en las ondas sonoras son:

1. **Frecuencia (tono)** el número de vibraciones por segundo, es decir, las veces que se repite en un segundo el ciclo completo de la onda. A mayor número de vibraciones, el sonido será más agudo, (chillón); a menos, grave, (ronco). El oído humano es sensible a los sonidos en un margen de frecuencias entre 20 y 20.000 vibraciones por segundo. Por debajo son infrasonidos y por encima ultrasonidos. El campo de audibilidad o zona de frecuencia de los sonidos audibles varía según los sujetos, si bien se admite generalmente que cubre desde las 16 vibraciones hasta las 20 mil. Las ondas mecánicas de frecuencia inferior a las 16 vibraciones se llaman infrasonidos, y las superiores a 20.000, ultrasonidos. La frecuencia de los sonidos de una conversación normal se sitúa entre 512 y 1.624 vibraciones por segundo. En los animales el campo de audibilidad es distinto y lo que para el hombre es ultrasonido para los animales es sonido: el perro alcanza un campo de audibilidad que llega hasta las 30 mil vibraciones; en los murciélagos alcanza las

40 mil y en algunas especies de insectos hasta 100 mil. La frecuencia se mide en hercios: $1\text{Hz} = 1 \text{ ciclo} / 1 \text{ segundo}$. La unidad de frecuencia, el hercio (Hz) recibe este nombre en honor del físico alemán Heinrich Hertz (1857-1894) y se define como el tiempo que tarda una onda en completar una oscilación.

2. **Intensidad**, que es la fuerza con que se emite un sonido dependiendo del esfuerzo muscular o, en general, de la fuerza que haya desencadenado las ondas; se mide en decibelios. Las intensidades acústicas poseen valores muy bajos. Para la máxima intensidad audible el nivel de intensidad es de 120 dB; para la mínima intensidad audible, el nivel sería 0 dB. Un sonido producido por el murmullo de las hojas es de 10 dB; el de una conversación normal 60 dB, el de una calle con tráfico 80 dB; el de un reactor superior a 100 dB. En las proximidades a 120 dB se siente dolor.

3. **Timbre** es la cualidad de un sonido en virtud de la cual lo podemos distinguir de otro de la misma intensidad y frecuencia, pero producido por otra fuente sonora (otro instrumento). El timbre depende del hecho de que, normalmente, una onda sonora va acompañada de otras ondas secundarias o armónicas, que dependen del medio ambiente o de la caja de resonancia.

4. **Tiempo**, esto es, la duración de un sonido. En el ámbito musical, sabemos que, dentro de un mismo compás, (por ejemplo, binario), una blanca es más larga que una negra.

Los rasgos que acabamos de enumerar son los que caracterizan a los sonidos, es decir, ondas periódicas cuyo movimiento vibratorio es regular; pero si el movimiento pendular no es regular ya en su frecuencia, en su amplitud o en su tiempo, entonces tenemos que hablar de ruidos.

El lenguaje humano utiliza ambos, sonidos y ruidos, que se producen en el aparato fonador del hablante, se transmiten por el aire de la respiración y en del ambiente y los capta el oído del oyente o receptor cuando la onda choca contra el tímpano y lo hace vibrar.

La manifestación espectrográfica del nivel fonético acústico de un mensaje articulado constituye lo que se ha dado en llamar lenguaje visible, aunque las investigaciones en laboratorios han permitido la constitución de lenguajes sintéticos.

Como podemos observar, un amplio porvenir aguarda todavía a la Lingüística en este dominio de la fonética acústica.

La fonética fisiológica: también conocida como fonética articulatoria puede considerarse como el método más antiguo y más común en la descripción fonética y se basa en la `articulación' de los sonidos por los órganos del habla. El fonetista entiende por órganos del habla o de fonación órganos fisiológicos que cumplen otras misiones: labios, dientes, lengua, cuerdas vocales, la nariz, los pulmones, etc. Además, teniendo presente que las diferencias individuales en los órganos de fonación que presentan las distintas personas no ejercen ningún efecto notable sobre la clase de sonidos producidos.

La respiración es un fenómeno fisiológico vital. El mismo aire indispensable para nuestra vida, sirve de elemento para transmitir las ondas sonoras producidas por nuestro aparato fonador. En castellano solo se emplea para la fonación el flujo de aire que se expulsa.

En la parte superior de la laringe se encuentran las cuerdas vocales, dos telitas elásticas, unidas cada una a varios músculos que las hacen tomar una posición relajada o tensa, dando la posición oportuna para que vibren al paso del aire. El espacio intercuerdas se denomina glotis, siendo en esta cavidad glótica donde se produce la intensidad acústica de la fonación lingüística.

Además, la faringe, las fosas nasales y la boca, actúan como caja de resonancias. Como el velo del paladar, la lengua y los labios son móviles, se puede hacer variar el volumen y la forma de dicha caja, modificando con ello el timbre.

El aparato fonador del hablante o locutor se complementa con el aparato auditivo del oyente o receptor, (que no vamos a tratar aquí).

El aire, al salir de los pulmones y subir por la laringe, se encuentra con las cuerdas vocales. Si se produce la vibración de éstas entonces la articulación se llama sonora, (lo que ocurre con todas las vocales y con algunas consonantes). Si, en cambio, las cuerdas no vibran, entonces la articulación se denomina sorda.

El aire encuentra después un segundo obstáculo: el velo del paladar, paladar blando, vulgarmente “campanilla”. Si está caído, el aire pasa también por la nariz, produciéndose entonces una articulación nasal (m, n); pero, si el velo del paladar se fija a la pared posterior, todo el aire saldrá por la boca, originándose entonces una articulación oral.

Una vez que el aire ha llegado a la boca encuentra en ella muchos obstáculos, fijos o en movimiento que, además de alterar la caja de resonancia, produce peculiares articulaciones.

Podemos hacer una diferenciación de articulaciones, según el lugar donde se realizan o cómo sale el aire, es decir, el modo.

A. Según el lugar, las articulaciones pueden ser:

- Bilabial: [b],[p],[m]
- Labiodental: [f]
- Linguodental: [t],[d]
- Linguoalveolar: [n],[s],[l],[r],[r]
- Linguopalatal: [ç]
- Linguovelar: [g],[k],[x]

B. Según el modo, la articulación puede ser:

- Oclusiva: cuando el aire sale de repente, en explosión, tras un cierre completo momentáneo. Ejemplo: [p].
- Fricativa: el aire sale de forma continua, pero con cierto roce. Ejemplo: [f].
- Africada: cuando empieza en oclusión y continúa en fricación. Ejemplo: [ç].
- Vibrante: cuando el aire al salir hace temblar la lengua. Ejemplo: [r].
- Lateral: cuando el aire no sale por el centro, sino por los lados de la lengua, que en centro de la boca lo obstaculiza. Ejemplo: [l].

Ya dijimos al referirnos a la fonética acústica que en la onda sonora que emitimos al hablar pueden apreciarse sonidos y ruidos. Según predomine uno u otro se podrá hablar de vocales o consonantes.

Para Bertil Malmberg, " mientras las vocales se caracterizan acústicamente por la carencia de ruido audible y, desde el punto de vista articulatorio, por el libre paso del aire, las consonantes son, o contienen, ruidos y se pronuncian con un cierre o estrechamiento del paso del aire".

También las vocales tienen modo y lugar de articulación:

- [a]: la lengua se sitúa en posición media de la boca y en forma plana. Los labios no actúan.
- [e],[i]: la parte anterior de la lengua se aproxima al paladar, un poco para la [e] y bastante para la [i]. Los labios se estiran en horizontal.
- [o],[u]: la parte posterior de la lengua se acerca hacia el velo del paladar, un poco para la [o] y bastante para la [u]. Los labios se estiran de forma vertical.

Todavía tenemos que distinguir dos articulaciones que en castellano participan de la doble articulación vocal/consonante. Son la llamada yod y el wau, que resultan de la pronunciación de dos o más vocales en la misma sílaba, es decir, en los diptongos y triptongos:

- YOD: es el elemento "i" de un diptongo o triptongo. Es un sonido palatal, más cerrado si cabe que la [i].
- WAU: es el elemento "u" de un diptongo o triptongo. Es un sonido velar, más cerrado si cabe que la [u].

Ambas pueden ser:

- Semivocal: segundo elemento del diptongo, apoyándose en el primero. Ejemplo: yod [j], “aire”; wau [u], “causa”.
- Semiconsonante: primer elemento, apoyándose en el segundo. Ejemplo: yod [j], “pie”; wau [w], “cuento”.

La fonética fisiológica o articulatoria que acabamos de considerar se representa con una serie de signos, que hemos ido detallando, dando lugar al ALFABETO FONÉTICO. Existe un alfabeto fonético internacional referido sólo a las realizaciones normativas que es el AFI, (Alfabeto Fonético Internacional).

Vamos a tratar ahora de la FONOLOGÍA. Entendemos por Fonología la disciplina lingüística que se ocupa del estudio de la función de los sonidos de las lenguas, es decir, estudia los sonidos desde el punto de vista de su funcionamiento en el lenguaje y cómo emplearlos para crear signos lingüísticos.

La FONOLOGÍA nació a partir de 1928 con motivo del I Congreso de Lingüistas celebrado en La Haya. Sus creadores fueron los miembros de la Escuela de Praga, en torno a su gran figura Trubetzkoy. El gran hallazgo, donde radicaba la gran novedad científica, fue el concepto de FONEMA, que venía a aplicar a los sonidos el principio ya formulado por Saussure de que no son los datos físicos sino su organización, lo fundamental en el sistema de una lengua.

Ahora bien, los fonemas no se pronuncian, sino que son las representaciones gráficas de los sonidos. La organización o estructuración de los fonemas de una lengua está basada en un sistema de oposiciones según los rasgos distintivos o pertinentes que se ha establecido. Por ejemplo: [p] y [b] se ha visto que son dos

sonidos, al oponerlos como fonemas, esto es, al realizar su representación gráfica, vemos que ambos elementos de la oposición son bilabiales y oclusivos, pero los diferencia que el primero es sordo y el segundo sonoro, ese es su rasgo pertinente, el rasgo que los diferencia, y que funcionalmente que es como a nosotros nos interesa hablar de fonemas, permite diferenciar, en el nivel léxico, palabras como: pata/bata; pesa/besa; cupo/cubo; tapa/taña.

Estas oposiciones que definen los fonemas sólo raramente aparecen en las lenguas de forma binaria y aisladas, (como el caso de /r/ y /r/ en castellano), por el contrario son correlativas oponiéndose un fonema a dos o más sobre rasgos diferentes: surgen así los haces correlativos:

- /p/,/t/,/k/: común la sordéz y la oclusividad; diferente el punto de articulación, (bilabial, dental, velar), y que se manifiesta en el léxico. Ejemplo: paso, taso, caso; lapa, lata, laca
- /b/,/d/,/g/: común en sonoridad y oclusividad, pero con rasgo pertinente basado en el punto de articulación, (bilabial, dental, velar). Ejemplo: daba, dada, daga; brama, drama, grama.
- /f/,/ /,/x/: común fricatividad y no sonoridad y diferente el lugar de articulación, (labiodental, dental, velar), que produce diferencias en: faca, zaca, jaca.

Estos haces, a su vez, pueden organizarse de otra manera:

- /b/,/p/,/f/: con el rasgo de labialidad común: paja, baja, faja.
- /t/,/d/,/ /: dentalidad: tejo, dejo, cejo.
- /k/,/g/,/x/: velaridad: cota, gota, jota.

Las unidades fonológicas en su interrelación estructural organizan lo que se llama SISTEMA FONOLÓGICO, que en el caso del castellano está compuesto de 24 fonemas: 5 vocálicos y 19 consonánticos. Este sistema constituye un conjunto acabado, cerrado o finito de elementos, es decir que ningún hablante puede modificarlo ni en su número ni en sus relaciones.

Se puede afirmar que, cuantos más haces de correlación existan en la estructura del sistema, tanto más perfecto será en teoría ese sistema. André Martinet ha dicho que: "las correlaciones y los haces contribuyen a reducir el número de articulaciones empleadas con fines distintivos; al ser menos numerosas, se diferenciarán entre sí mucho más perfectamente. Además, al ser más frecuentes en el habla, el locutor tendrá más ocasiones de interpretarlas y reproducirlas".

Estos sistemas que hemos visto están organizados en cuanto que formas aisladas e independientes. Pero en la realidad no es así pues las exteriorizamos por grupos fónicos. Las articulaciones se suceden unas tras otras con rapidez, originándose a menudo asimilaciones entre sonidos contiguos o algo más alejados, aunque dentro del mismo grupo fónico. Por asimilación, siguiendo a M. Grammont "consiste en la extensión de uno o varios movimientos articulatorios más allá de su zona". Es un proceso fonético en el que dos sonidos contiguos o próximos tienden a identificarse o adquirir caracteres similares. Por

ejemplo, la vocal que se encuentra entre sonidos nasales se contagiara de la nasalidad, como en mano,[mano].

A nivel fonológico hay que hablar, por consiguiente de los alófonos que es cada una de las diferentes realizaciones fonéticas de un mismo fonema. Entenderemos esto al hacer la diferenciación entre sonido y fonema. El fonema es único para toda la comunidad lingüística, independientemente del que hable; en cambio el sonido es particular de cada hablante.

No podemos terminar el estudio del sistema fonológico del castellano sin referirnos a dos conceptos que están presentes: la neutralización y el archifonema.

- Se dice que hay neutralización cuando la marca funcional de una oposición deja de funcionar, lo que sólo ocurre en algunas posiciones. Así, por ejemplo, la oposición entre /r/ y /r/ presente en /pero/, /perro/, deja de existir al final de palabra diciéndose entonces que la oposición ha quedado neutralizada: /kantar/- /kantar/.
- El resultado de una neutralización en el archifonema, conjunto de rasgos distintivos comunes a los dos términos de una oposición fonológica cuando se neutraliza el rasgo diferenciador o marca funcional. Se representa en mayúscula: /kantaR/.

Los fonemas pueden también considerarse segmentos de la combinatoria lingüística. Además de los rasgos de pertinencia o diferenciación que les caracteriza hay que añadir una serie de rasgos que se superponen a los segmentos o fonemas y que se denominan suprasegmentos o prosodemas, (por pertenecer a la parte de la gramática que se conoce tradicionalmente con el nombre de prosodia).

La PAUSA: es la interrupción de la cadena fónica, sirviendo de límite a los grupos fónicos. Puede ser lingüísticamente pertinente. Ejemplo: Los niños, que estudiaron, aprobaron / Los niños que estudiaron, aprobaron.

El ACENTO: que corresponde a la mayor intensidad acústica es sin duda el prosodema más conocido. También es pertinente, por ejemplo, en: canto / cantó.

La ENTONACIÓN: esquema melódico que puede ser también pertinente. Ejemplo: ¿Hoy es lunes? / Hoy es lunes.

La CANTIDAD: o duración de un sonido, a tener en cuenta sobre todo en coincidencia de sonidos homólogos: ¡ qué techo!, ¡que te echo!, ¡ qué te he hecho!

CAPITULO II

ANÁLISIS

2.1 MUESTREO

Para procesar señales analógicas por medios digitales es necesario convertirlas a formato digital, esto es, transformarlas en una secuencia de números de precisión finita. Este procedimiento se denomina conversión analógica – digital (A/D) y los dispositivos que realizan este proceso se le denomina “Conversores A/D” (ADC).

Conceptualmente, podemos ver la conversión A/D como un proceso de tres pasos. Este proceso se ilustra en la figura 2.1 y se explica brevemente a continuación:

1. Muestreo. Esta es la conversión de una señal en tiempo continuo a una señal en tiempo discreto obtenida tomando “muestras” de la señal en tiempo continuo en instantes de tiempo discreto. Así, si $x_a(t)$ es la entrada al muestreador, la salida es $x_a(nT) \equiv x(n)$, donde T se denomina el intervalo de muestreo.
2. Cuantificación. Esta es la conversión de una señal en tiempo discreto con valores continuos a una señal en tiempo discreto con valores discretos (señal digital). El valor de cada muestra de la señal se representa mediante un valor seleccionado de un conjunto finito de valores posibles. La diferencia entre la muestra sin cuantificar $x(n)$ y la salida cuantificada $x_q(n)$ se denomina error de cuantificación.
3. Codificación. En el proceso de codificación, cada valor discreto $x_q(n)$, se representa mediante una secuencia binaria de b bits.

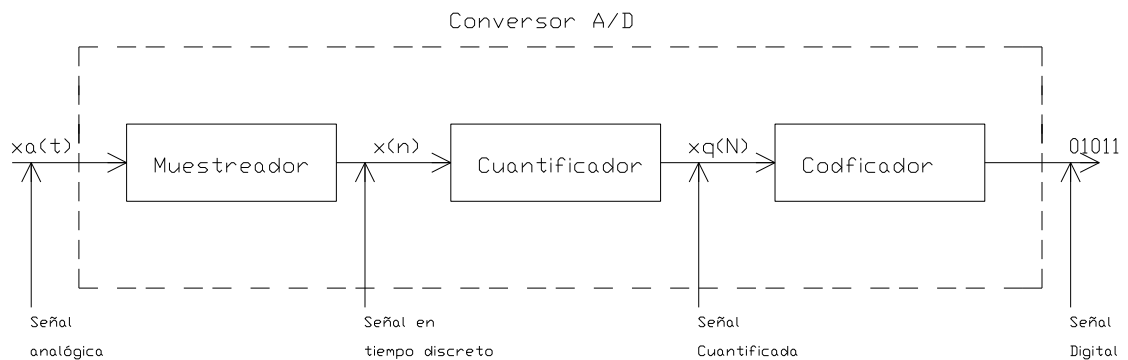


Figura. 2.1 Partes básicas de un convertidor analógico – digital (CAD)

Aunque modelamos el convertidor A/D con un muestreador seguido de un cuantificador, en la práctica la conversión A/D se efectúa en un único dispositivo que toma $x_a(t)$ y produce un número codificado en binario. Las operaciones de muestreo y cuantificación pueden realizarse en cualquier orden, pero, en la práctica, el muestreo siempre tiene lugar antes de la cuantificación.

2.1.1 MUESTREO DE SEÑALES ANALÓGICAS

Existen muchas maneras de muestrear una señal. Aquí presentaremos el muestreo periódico o uniforme, que es el tipo de muestreo usado más a menudo en la práctica. Este se describe mediante la relación

$$X(n) = x_a(nT), \quad -\infty \leq n \leq \infty \quad (2.1)$$

donde $x(n)$ es la señal en tiempo discreto obtenida tomando muestras de la señal analógica $x_a(t)$ cada T segundos. Este proceso se ilustra en la figura 2.2.

El intervalo de tiempo T entre dos muestras sucesivas se denomina **periodo de muestreo** o **intervalo de muestreo**, y su recíproco $1/T = F_s$ se denomina **velocidad de muestreo** (muestras por segundo) o **frecuencia de muestreo** (hertz).

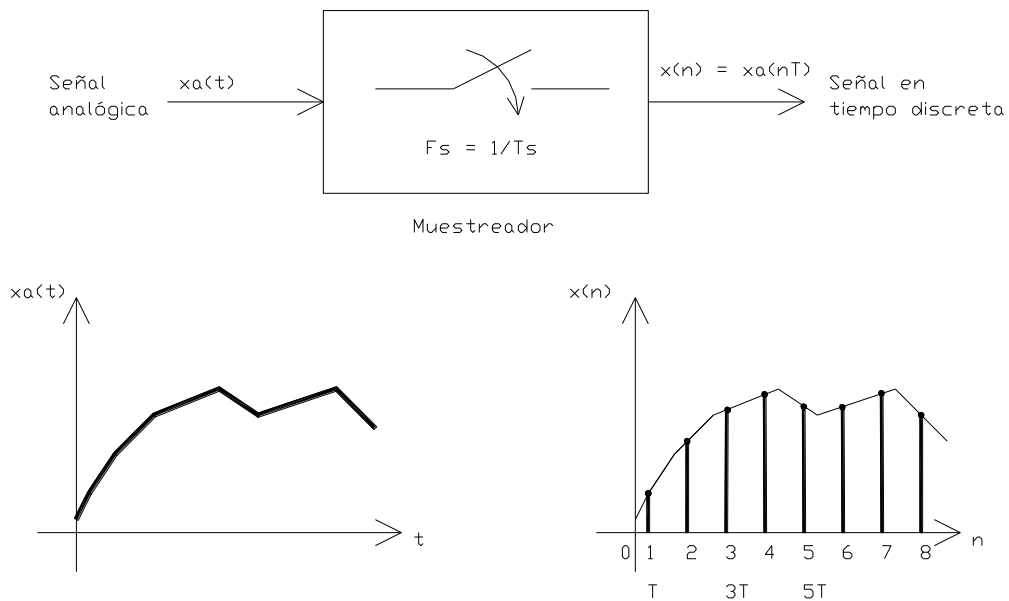


Figura. 2.2 Muestreo periódico de una señal analógica

El muestreo periódico establece una relación entre las variables t y n de tiempo continuo y tiempo discreto, respectivamente. De hecho estas variables se relacionan linealmente a través del periodo de muestreo T o, equivalentemente, a través de la velocidad de muestreo $F_s = 1/T$, como

$$t = nT = \frac{n}{F_s} \tag{2.2}$$

Como consecuencia de (2.2), existe una relación entre la variable de frecuencia F (o Ω) de las señales analógicas y la variable frecuencia f (o ω) de las señales en tiempo discreto. Para establecer dicha relación considere una señal analógica de la forma

$$x_a(t) = A \cos(\pi F t + \Theta) \quad (2.3)$$

que, cuando se muestrea periódicamente a una velocidad de $F_s = 1/T$ muestras por segundo, da lugar a

$$x_a(nT) = x(n) = A \cos(\pi F n T + \Theta) = A \cos\left(\frac{2\pi F n}{F_s} + \Theta\right) \quad (2.4)$$

Si comparamos (2.4) con (2.5)

$$x(n) = A \cos(\pi f n + \Theta), \quad -\infty \leq n \leq \infty \quad (2.5)$$

Observamos que las variables frecuencia F y f están linealmente relacionadas según

$$f = \frac{F}{F_s} \quad (2.6)$$

o, equivalentemente, según

$$\omega = \Omega T \quad (2.7)$$

La relación dada en (2.6), justifica el nombre de frecuencia normalizada relativa, que se usa a veces para describir la variable frecuencia f . Como se ve en (2.6), podemos usar f para determinar la frecuencia F en hertzios solo si la frecuencia de muestreo F_s es conocida.

El rango de la variable frecuencia F o Ω para senoidales en tiempo continuo es:

$$-\infty \leq F \leq \infty \quad -\infty \leq \Omega \leq \infty \quad (2.8)$$

Sin embargo, la situación es diferente para senoidales en tiempo discreto, el cual sus rangos de frecuencias están dados por:

$$\begin{aligned}
 -\frac{1}{2} < f < \frac{1}{2} \\
 -\pi < \omega < \pi
 \end{aligned}
 \tag{2.9}$$

sustituyendo (2.6), (2.7), (2.9) encontramos que la frecuencia de la senoidal en tiempo continuo cuando se muestrea a una velocidad $F_s = 1/T$ debe encontrarse en el rango

$$-\frac{1}{2T} = -\frac{F_s}{2} \leq F \leq \frac{F_s}{2} = \frac{1}{2T}
 \tag{2.10}$$

o equivalente

$$-\frac{\pi}{T} = -\pi F_s \leq \Omega \leq \pi F_s = \frac{\pi}{T}
 \tag{2.11}$$

De estas relaciones se desprende que la diferencia fundamental entre señales en tiempo discreto y señales en tiempo continuo es el rango de valores de las variables frecuencia F y f , ó Ω y ω . El muestreo periódico de una señal en tiempo continuo supone una correspondencia entre un rango de frecuencia infinito correspondiente a la variable F (ó Ω) y un rango de frecuencias finito correspondiente a la variable f ó ω . Dado que la frecuencia máxima de una señal en tiempo discreto es $\omega = \Pi$ o $f = 1/2$, los valores de F y Ω para una velocidad de muestreo F_s son

$$\begin{aligned}
 F_{max} &= \frac{F_s}{2} = \frac{1}{2T} \\
 \Omega_{max} &= \pi F_s = \frac{\pi}{T}
 \end{aligned}
 \tag{2.12}$$

Por lo tanto, el muestreo introduce una ambigüedad; así, la máxima frecuencia de una señal en tiempo continuo que puede determinarse unívocamente cuando dicha señal se muestra a una velocidad $F_s = 1/T$ es $F_{\max} = F_s/2$, ó $\Omega_{\max} = \pi F_s$.

En general, el muestreo de una señal senoidal en tiempo continuo

$$x_a(t) = A \cos(\pi F_0 t + \Theta) \quad (2.13)$$

con una velocidad de muestreo $F_s = 1/T$ produce una señal en tiempo discreto

$$x(n) = A \cos(\pi f_0 n + \Theta) \quad (2.14)$$

donde $f_0 = F_0 / F_s$ es la frecuencia relativa de la senoidal.

Si suponemos que $-F_s / 2 \leq F_0 \leq F_s / 2$, la frecuencia f_0 de $x(n)$ se encuentra en el rango $1/2 \leq f_0 \leq 1/2$, que es el rango de frecuencias para señales en tiempo discreto. En este caso, la relación entre F_0 y f_0 es biunívoca, y por tanto, es posible identificar (o reconstruir) la señal analógica $x_a(t)$ a partir de las muestras $x(n)$.

Por otra parte, si las senoidales

$$x_a(t) = A \cos(\pi F_k t + \Theta) \quad (2.15)$$

Donde

$$f_k = F_0 + k F_s, \quad k = \pm 1, \pm 2, \dots, \quad (2.16)$$

Se muestrea a una velocidad F_s , está claro que la frecuencia F_k se encuentra fuera del rango $-F_s / 2 \leq F \leq F_s / 2$,

La figura 2.3 muestra un ejemplo de aliasing: dos sinusoides de frecuencias $F_0 = 1/8$ Hz y $F_1 = -7/8$ Hz producen muestras idénticas cuando se utiliza una velocidad de muestreo $F_s = 1$ Hz. De (2.16) tenemos que para $k = -1$, $F_0 = F_1 + F_s = (-7/8 + 1)$ Hz = $1/8$ Hz.

Dado que $F_s/2$, que se corresponde con $\omega = \pi$, es la frecuencia más alta que puede ser representada inequívocamente con una velocidad de muestreo F_s , es fácil determinar correspondencia entre cualquier frecuencia alias por encima de $F_s/2$ ($\omega = \pi$) y su frecuencia equivalente por debajo de $F_s/2$. Podemos usar $F_s/2$ o $\omega = \pi$ como el punto de “plegado” y reflejar o doblar la frecuencia alias en el rango $F_s/2$ ($\omega = \pi$). La frecuencia $F_s/2$ ($\omega = \pi$) se denomina “**frecuencia de plegado**”.

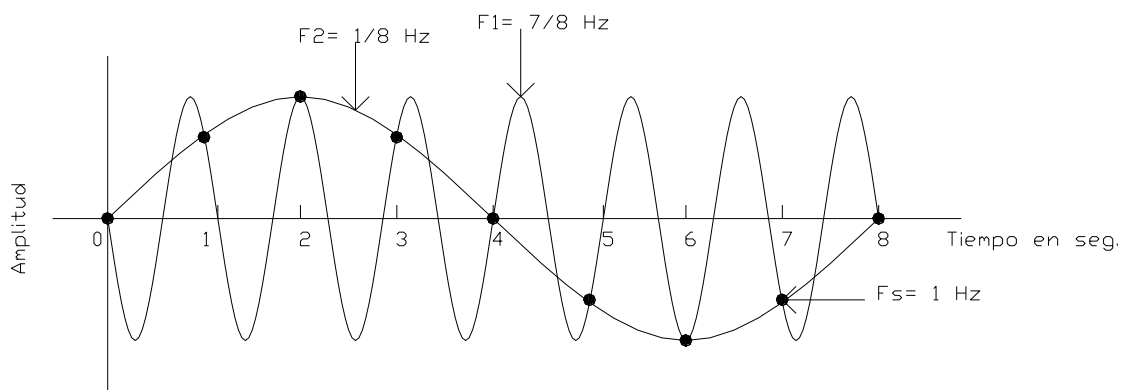


Figura. 2.3 Ilustración del aliasing

2.1.2 Teorema de muestreo

Supongamos que cualquier señal analógica se puede representar como una suma de senoidales de diferentes amplitudes, frecuencias y fases, es decir,

$$x_a = \sum_{i=1}^N A_i \cos(\pi F_i t + \Theta_i) \quad (2.17)$$

donde N indica el número de componentes de frecuencia. Todas las señales, como las de voz o vídeo, se prestan a dicha representación en cualquier intervalo de tiempo pequeño. Supongamos que las frecuencias de una determinada señal no exceden una frecuencia máxima conocida F_{max} .

Dado que la frecuencia máxima puede variar ligeramente, queremos asegurar que F_{max} no sobrepase determinado valor y para ello pasaremos la señal analógica por un filtro que atenúe fuertemente las componentes de frecuencia por encima de F_{max} . Así estaremos seguros de que ninguna señal de la clase que nos interesa tendrá componentes de frecuencia (con amplitud o potencia) significativa por arriba de F_{max} . En la práctica, este filtrado se realiza antes del muestreo. El conocimiento de F_{max} nos permite seleccionar la velocidad de muestreo apropiada. Sabemos que la frecuencia más alta de la señal analógica que puede reconstruirse sin ambigüedad cuando la señal se muestrea a una velocidad $F_s = 1/T$ es $F_s/2$. Para evitar las ambigüedades que resulta del aliasing, se debe seleccionar una velocidad de muestreo lo suficientemente alta, esto es, debemos escoger $F_s/2$ mayor que F_{max} . Por tanto, para evitar el problema del aliasing, se secciona F_s según

$$F_s > 2F_{max} \quad (2.18)$$

donde F_{max} es la frecuencia más alta de la señal analógica. Con la velocidad de muestreo seleccionada de esta manera tenemos que cualquier componente de frecuencia, se corresponde en tiempo discreto con una senoidal de frecuencia.

La condición $F_s > 2F_{max}$ garantiza que todas las componentes senoidales de la señal analógica se correspondan con componentes en frecuencia de tiempo discreto en el intervalo fundamental. Por lo tanto, todas las componentes en frecuencia de la señal analógica están representadas sin ambigüedad en la forma muestreada de la señal y así la señal analógica puede ser reconstruida sin distorsión a partir de las muestras usando un método de interpolación apropiado. La formula de interpolación ideal o apropiada se especifica mediante el teorema de muestreo.

Teorema. Teorema de muestreo. Si la frecuencia mas alta contenida en una señal analógica $x_a(t)$ es $F_{max} = B$ y la señal se muestrea a una velocidad $F_s > 2F_{max} \equiv 2B$, entonces $x_a(t)$ se puede recuperar totalmente a partir de sus muestras mediante la función de interpolación

$$g(t) = \frac{\sin(\pi B t)}{2\pi B t} \quad (2.19)$$

Así, $x_a(t)$ se puede expresar como

$$x_a(t) = \sum_{n=-\infty}^{\infty} x_a\left(\frac{n}{F_s}\right) g\left(t - \frac{n}{F_s}\right) \quad (2.20)$$

donde $x_a(n/F_s) = x_a(nT) \equiv x(n)$ son las muestras de $x_a(t)$. La tasa de muestreo dada por $F_N = 2B = 2F_{max}$, se denomina **tasa de Nyquist**.

2.2 FILTRADO DIGITAL

Los filtros digitales son una clase de sistemas discretos LTI utilizados para extraer características desde el dominio de la frecuencia sobre señales muestreadas. El diseño de los filtros requiere de las siguientes etapas:

- 1) Especificaciones de las propiedades deseadas del sistema.
- 2) Aproximaciones de las especificaciones mediante un sistema causal en tiempo discreto
- 3) La realización del sistema.

El primero es altamente dependiente de la aplicación y el tercero de la tecnología utilizada para la implementación. En términos prácticos, el filtro deseado se realiza utilizando cómputo digital y se emplea para una señal que proviene de tiempo continuo seguido por una conversión analógica/digital.

Cuando se utiliza un filtro digital para realizar el tratamiento de señales en tiempo continuo, empleando una configuración como la indicada en la Figura 2.4, tanto las especificaciones del filtro en tiempo discreto como las que se indica en tiempo continuo se suelen dar en el dominio de la frecuencia. Si se usa un sistema LTI y si la entrada es de banda limitada y la frecuencia de muestreo es lo suficientemente alta para evitar el solapamiento, el sistema completo se comporta como un sistema LTI continuo en el tiempo cuya respuesta en frecuencia es:

$$G_c(\omega) = \begin{cases} G_d(\omega) & |\omega| < \frac{\pi}{T} \\ 0 & |\omega| > \frac{\pi}{T} \end{cases} \quad (2.21)$$



Figura 2.4. Arquitectura del filtro digital

En estos casos es inmediato convertir las especificaciones del filtro efectivo de tiempo continuo, $G_c(\omega)$, en especificaciones del filtro en tiempo discreto, $G_d(\omega)$. La arquitectura presentada parece evidente que es más compleja y costosa que el uso de células Sallen-Key para implementar filtros analógicos. Sin embargo, los filtros digitales exhiben abrumadoras ventajas respecto a los sistemas de tiempo continuo, una enumeración de los beneficios puede enmarcarse dentro de la comparación entre estos dos tipos de filtros:

- Respuesta dinámica: El ancho de banda del filtro digital está limitado por la frecuencia de muestreo, mientras que en los filtros analógicos con componentes activos suelen estar restringidos por los amplificadores operacionales.
- Intervalo dinámico: En los filtros analógicos aparecen derivas que limitan por abajo el rango y se saturan con la alimentación. En cambio en los filtros digitales es fijado por el número de bits que representa la secuencia, y el límite inferior por el ruido de cuantificación y por los errores de redondeo.
- Conmutabilidad: Si los parámetros de un filtro se conservan en registros, los contenidos de dichos registros pueden ser modificados a voluntad. De esta forma, estos filtros se pueden transformar, pudiéndose multiplexar en el tiempo para procesar varias señales de entradas a la vez.
- Adaptabilidad: Un filtro digital puede ser implementado en soporte físico (*hardware*) o mediante un programa de computador (*software*).
- Ausencia de problemas de componentes: Los parámetros de los filtros se representan por medio de números binarios y no derivan con el tiempo. Al no haber componentes, no hay problemas de tolerancia o deriva de componentes, y ningún otro problema asociado con un comportamiento no ideal de resistencias, condensadores, bobinas o amplificadores. Tampoco existen problemas de impedancia de entrada ni salida, ni efectos de adaptación de impedancias entre etapas.
- Complejidad: la potencia de cálculo de los computadores actuales y de los algoritmos desarrollados, permiten implementar prestaciones casi imposibles de diseñar con filtros analógicos.

2.2.1 CLASES DE FILTROS DIGITALES

Una distinción fundamental en los sistemas discretos dinámicos lineales e invariantes, y en particular en los filtros digitales, es la duración de la respuesta ante el impulso. Se habla de sistemas de respuesta de pulso finito o no recursivo (*FIR, Finite Impulse Response*) y de sistemas de respuesta infinita o recursivo (*IIR, Infinite Impulse Response*). Partiendo de la ecuación en diferencias que modela el comportamiento dinámico de estos sistemas:

$$y(k) = a_1y(k-1) + a_2y(k-2) + \dots + a_ny(k-n) + b_0x(k) + b_1x(k-1) + \dots + b_mx(k-m) \quad (2.22)$$

en el caso de tener todos los coeficientes a_i iguales a cero se tendrá un filtro FIR, con lo que quedará la ecuación reducida a:

$$y(k) = b_0x(k) + b_1x(k-1) + \dots + b_mx(k-m) \quad (2.23)$$

siendo m el orden del filtro y tendrá una función de transferencia en z del tipo:

$$G(z) = b_0 + b_1z^{-1} + \dots + b_mz^{-m} \quad (2.24)$$

Obsérvese que en estos tipos de filtros cada valor de la secuencia de salida sólo dependerá de un número finito de valores de la secuencia de entrada. Además también se desprende la carencia de polos en la función de transferencia.

Por lo contrario, las expresiones de los filtros recursivos corresponden a:

$$y(k) = a_1y(k-1) + a_2y(k-2) + \dots + a_ny(k-n) + b_0x(k) + b_1x(k-1) + \dots + b_mx(k-m) \quad (2.24)$$

Y su función de transferencia en z a:

$$G(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_m z^{-m}}{1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_n z^{-n}} \quad (2.25)$$

En estos casos, la secuencia de salida depende tanto de la entrada como de la salida. De estas ecuaciones se deducen las siguientes propiedades.

- Primera, la secuencia de ponderación es infinita para los filtros IIR, aun teniendo un número finito de coeficientes. Mientras la respuesta al impulso de un filtro no recursivo es siempre finita e igual al orden del filtro.
- Segundo, los filtros FIR prácticos son siempre estables, esto es, la secuencia de salida tiene todos sus valores acotados. No es el caso de los filtros recursivos, su estabilidad depende de la función de transferencia, por lo que hay que utilizar alguno de los procedimientos algebraicos, como el de Jury, para analizar su estabilidad.
- Tercera, cualquier filtro recursivo puede ser reemplazado por otro no recursivo con infinitos coeficientes, sus valores vendrán dados por la secuencia de ponderación del IIR. La conclusión inversa no se cumple.

2.2.1.1 FILTROS DE FASE LINEAL O NO DISPERSIVOS

Existen muchas aplicaciones del procesamiento de señales en las cuales las relaciones de fase son importantes y no deben ser perturbadas por el filtrado. Por ello se deben diseñar filtros de fase lineal. Éstos tienen como característica la introducción de un desfase en la respuesta en frecuencia de la secuencia de salida que sea proporcional a la frecuencia. A estos filtros se les llaman de fase lineal o no dispersivos.

Los filtros digitales suelen ser caracterizados en términos de rangos de frecuencia, tanto de la banda pasante como de la supresora. Los cuatro tipos básicos de filtros supresores son ilustrados en la figura 2.5. Al ser éstos sistemas

discretos, sus respuestas frecuenciales son periódicas con la frecuencia de Nyquist, W_N , por lo que sólo se considerará el intervalo $[-W_N, W_N]$.

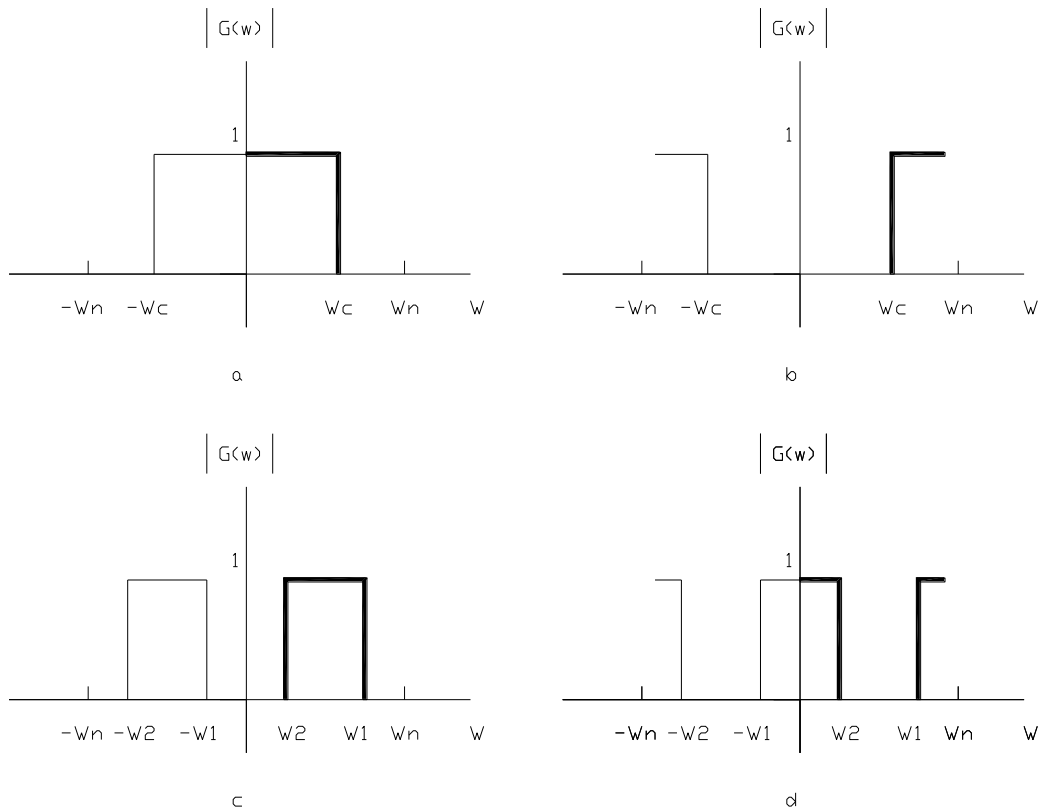


Figura 2. 5 Especificaciones de los filtros digitales a) Paso bajo b) Paso alto c) Pasa banda d) Rechazo de banda

En la figura 2.5 sólo se muestran las respuestas en frecuencias del módulo sin que aparezca el argumento. Sin embargo es deseable que en la banda pasante no se introduzcan desfases o que sean de fase lineal. Con tal propósito se parte de la realización de un filtro con variación lineal del argumento respecto a la frecuencia, esto es, se tendrá que el desfase introducido es del tipo $\varphi = -\lambda\omega T$, donde λ es una constante. El caso particular de desfase nulo será λ igual a cero. Luego si ante una señal de entrada periódica con las condiciones de Fourier, ésta se separa en forma de sumas de sinusoides del tipo $\text{sen}(\omega kT)$, cada una de ellas producirá una respuesta del tipo:

$$|G(w)| \text{sen} [v(kt - \lambda)] \quad (2.26)$$

Por lo tanto, cada armónico de la señal de entrada estará desfasada $-\lambda$ veces, de forma que se obtendrá en la salida una versión no distorsionada de la señal de entrada en el rango de frecuencias dependientes de la banda pasante. A razón de lo expuesto, el diseño de un filtro paso bajo deberá de cumplir con las condiciones mostradas en la figura 2.6.

El módulo será constante en la banda pasante y nulo en la supresora, mientras el desfase tendrá que ser lineal con la frecuencia. Un posible modelo matemático que caracterice este tipo de respuesta correspondería a:

$$G_c(\omega) = \begin{cases} e^{-j\lambda\omega T} & |\omega| \leq \omega_c \\ 0 & \text{En caso contrario} \end{cases} \quad (2.27)$$

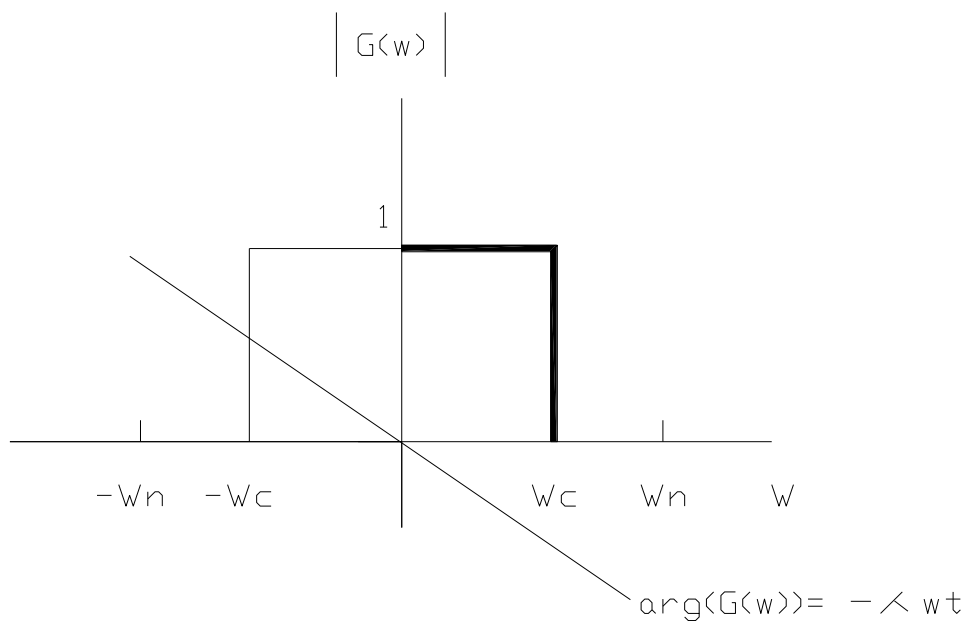


Figura 2. 6 Respuesta en frecuencia de un filtro paso bajo

de manera que si se aplica la transformada inversa de Fourier, la secuencia de ponderación del filtro quedará como:

$$g(k) = \frac{1}{2\pi} \int_{-\omega_N}^{\omega_N} G(\omega) e^{j\omega k T} d\omega = \frac{2\omega_c}{\omega_N} \frac{\text{sen} \left[(k - \lambda) \omega_c T \right]}{(k - \lambda) \omega_c T} \quad k = 0, \pm 1, \pm 2 \dots$$

(2.28)

Desprendiéndose que el filtro paso bajo es no causal y por lo tanto no realizable físicamente.

Dos aspectos destacan.

- Primero, el filtro no es causal, ya que para $k < 0$ los valores de $\{g(k)\}$ no son nulos, por lo que no pueden ser implementados en tiempo real.
- Segundo, la respuesta al impulso no es finita, esto significa que la respuesta del filtro tardaría infinito tiempo en calcularlo, incluso asumiendo que pudiera tener infinitos coeficientes.

De estos aspectos, se concluye que la realización de un filtro ideal es imposible de realizar físicamente; habiendo de aceptar criterios de compromiso para el diseño. Por ejemplo, si se trunca la respuesta al impulso se tendrá un filtro digital realizable físicamente, pero éste sólo será una aproximación al comportamiento ideal.

El truncamiento del orden del filtro hace dejar de comportarse como filtro ideal, apareciendo tanto rizado en la banda pasante como en la supresora. A este efecto se le denomina fenómeno de Gibbs. También se debe intuir el carácter lineal del desfase.

Aunque esta discusión se limita a la realización de un filtro paso bajo, las conclusiones se mantienen en general para todos los filtros selectivos. En resumen, ninguno de los filtros mostrados previamente en la Figura 2.5 son causales y, por tanto, son físicamente irrealizables.

Una pregunta que aflora naturalmente en este punto es la siguiente: ¿ Cuáles son las condiciones suficientes y necesarias que debe satisfacer la respuesta frecuencial para que el filtro sea causal?. La respuesta a esta pregunta viene dada por el teorema de Paley-Wiener.

Del teorema se extraen varias conclusiones:

- Los filtros causales tienen un módulo, $|G(w)|$ que puede ser cero en algunas frecuencias, pero no puede ser cero sobre cualquier banda finita de frecuencias.
- La amplitud de la banda pasante no puede ser constante en un rango finito de frecuencias y la transición de la banda de paso a la de rechazo no puede ser infinitamente abrupta (esto es una consecuencia del fenómeno de Gibbs, causado por el truncamiento de $g(k)$ para lograr causalidad).
- La parte real e imaginaria de $G(W)$ no son independientes y están relacionadas por la transformada de Hilbert discreta. En consecuencia, la magnitud y la fase de $G(W)$ no se pueden elegir arbitrariamente.

Aunque las características de respuesta en frecuencia que poseen los filtros ideales son deseables, no son absolutamente necesarias en la mayoría de las aplicaciones prácticas. Si estas condiciones no son muy estrictas es posible realizar filtros causales que se aproximan a los ideales con tanta precisión como se desee.

En particular, no es necesario insistir en que la ganancia en la banda pasante sea constante. Se puede tolerar un pequeño rizado e igualmente se puede tolerar un valor pequeño distinto de cero en la banda de rechazo. La transición de la banda de paso a la supresora no será nula y existirá una banda de transición. En cualquier problema de diseño de filtros selectivos se podrá especificar:

- 1) El rizado de la banda pasante, δ_1 ,
- 2) El rizado tolerable de la banda de rechazo, δ_2
- 3) La frecuencia de la banda de paso, W_B ,
- 4) La frecuencia de la banda supresora, W_S .

Basándose en estas especificaciones se seleccionarán los coeficientes a_k y b_k de la ecuación en diferencias del sistema LTI, de la manera que mejor se aproxime a la respuesta en frecuencia deseada. El grado con el que $G(W)$ se aproxime a las especificaciones depende del criterio usado en la determinación de los coeficientes así como del orden del polinomio de la FDT del sistema.

2.2.2 DISEÑO DE FILTROS NO RECURSIVOS (FIR)

Los filtros no recursivos tienen ventajas muy interesantes que les hacen ser ampliamente utilizados en múltiples aplicaciones. La característica más destacable es su facilidad de diseño para conseguir una respuesta en frecuencias de fase lineal. Los FIR son por su propia constitución estables, no dan problemas en su diseño o en su fase de implementación. Aunque el diseño de los FIR requiera de una gran cantidad de operaciones de sumas y multiplicaciones, tanto su estructura de programación como su realización en soporte físico resultan fáciles y escalables.

En la práctica, los filtros FIR se emplean en problemas de filtrado donde hay un requisito de fase lineal dentro de la banda de paso del filtro. Si no existe este requisito se pueden emplear tanto filtros FIR como IIR. Sin embargo, como regla general, un filtro IIR tiene menos rizado y el corte es más abrupto que un FIR con el mismo grado de polinomio.

Por esta razón, si se puede tolerar alguna distorsión de fase o ésta no es importante, se prefiere un IIR, principalmente porque su implementación involucra menos parámetros, requiere menos memoria y tiene menor complejidad computacional.

Básicamente hay dos métodos para el diseño de filtros no recursivos. El primero trata de definir la respuesta en frecuencia del filtro para luego determinar los coeficientes del filtro mediante la transformada inversa de Fourier; mientras que la segunda estrategia utiliza métodos de optimización capaz de ir modificando los coeficientes del filtro para aproximarlos a la respuesta en frecuencia deseada. En este trabajo para disminuir la complejidad del análisis se explicará el primer método.

2.2.2.1 FILTROS FIR SIMÉTRICOS Y ANTISIMÉTRICOS

Un filtro FIR tiene fase lineal si su respuesta impulsional satisface la condición de simetría o antisimetría de sus coeficientes. Para su demostración se partirá de la respuesta en frecuencia de un filtro no recursivo que tenga un orden m . Además, para facilitar su comprensión se va a suponer que el filtro sea de orden par, definiéndose que m sea igual a $2N$, por tanto, la respuesta en frecuencia de los filtros de orden par quedará como:

$$G(\omega) = \sum g(k)e^{-j\omega kT}$$

$$G(\omega) = e^{-jN\omega T} \left[g(0)e^{jN\omega T} + g(1)e^{j(N-1)\omega T} + \dots + g(N) + \dots + g(2N)e^{-jN\omega T} \right] \quad (2.29)$$

Si se imponen las condiciones de simetría o antisimetría:

$$g(0) = \pm g(2N)$$

$$\cdot g(1) = \pm g(2N - 1)$$

$$\sim$$

$$\cdot$$

$$\sim$$

$$\cdot$$

$$g(N - 1) = \pm g(N + 1)$$

(2.30)

El signo + indicará simetría en los coeficientes y el – para la antisimetría.

Suponiendo primero que existe simetría y agrupando alrededor de los coeficientes

$e^{\pm j(N-1)\omega T}$ la respuesta en frecuencia del filtro quedará como:

$$G(\omega) = e^{-jN\omega T} \left\{ g(n) + 2 \sum_{i=0}^{N-1} g(i) \cos \left[(N-i)\omega T \right] \right\} \quad (2.31)$$

observándose que el contenido de entre las llaves es real y que el desfase introducido por el filtro es $-N \omega T$, siendo por tanto el argumento lineal con la frecuencia.

De igual manera se actuará con los coeficientes antisimétricos, sin embargo, hay que destacar que si el filtro es de orden par el punto central de la antisimetría será nulo. Es fácil demostrar que la respuesta en frecuencia del filtro FIR de orden par con una respuesta impulsional antisimétrica se puede expresar como:

$$G(\omega) = e^{-jN\omega T + \frac{\pi}{2}} \left\{ 2 \sum_{i=0}^{N-1} g(i) \operatorname{sen} \left[(N-i)\omega T \right] \right\} \quad (2.32)$$

La característica de fase del filtro será $\frac{\pi}{2} - (N-1)\omega T$ Igualmente las expresiones de las respuestas en frecuencias para filtros de orden impar con simetría o antisimetría son sencillas de obtener

$$G(\omega) = e^{-j\frac{m}{2}\omega T} \left\{ 2 \sum_{i=0}^{((m+1)/2)-1} g(i) \cos \left[\frac{m}{2} - i \right] \omega T \right\}$$

$$G(\omega) = e^{-j\frac{m}{2}\omega T + \frac{\pi}{2}} \left\{ 2 \sum_{i=0}^{((m+1)/2)-1} g(i) \operatorname{sen} \left[\frac{m}{2} - i \right] \omega T \right\} \quad (2.33)$$

Estas fórmulas de respuesta en frecuencia generales se pueden usar para diseñar filtros FIR de fase lineal con respuestas impulsionales simétricas y antisimétricas. Nótese que para un filtro simétrico, el número de coeficientes del filtro que especifican la respuesta es $m/2$ cuando m es par o $((m+1)/2)-1$ si m es impar. Por otro lado, si la respuesta impulsional es antisimétrica y el orden es par, el punto central de la antisimetría será nulo y habrá $m/2$ coeficientes que lo definan, en caso de ser orden impar, cada coeficiente tiene un término emparejado de signo opuesto y estará definido por $((m+1)/2)-1$.

La elección de una respuesta impulsional simétrica o antisimétrica depende de la aplicación. Por ejemplo, si el filtro tiene antisimetría, $G(\omega)$ tanto para bajas frecuencias como para frecuencias alrededor de la frecuencia de Nyquist tiene valores próximos a cero, por lo que no es posible utilizarlos ni para filtros paso bajos ni paso alto. Por otro lado, la condición de simetría produce un filtro FIR de fase lineal con una respuesta distinta de cero para bajas frecuencias. En resumen, el problema de diseño de filtros FIR es simplemente el del determinar $m+1$ coeficientes, a partir de las especificaciones en las frecuencias deseadas.

2.2.3 DISEÑO DE FILTROS RECURSIVOS (IIR)

El método tradicional de diseño de filtros IIR en tiempo discreto se basa en la transformación de un filtro analógico en un filtro digital que cumpla las especificaciones preestablecidas. Esta solución es razonable por varios motivos:

- El arte del diseño de filtros IIR analógicos está muy avanzado, y como se pueden obtener resultados útiles, es ventajoso utilizar los procedimientos de diseño que ya se han desarrollado para los filtros en tiempo continuo.
- Muchos métodos útiles de diseño de filtros IIR en tiempo continuo dan como resultado fórmulas de diseño simples en forma cerrada. Por tanto, los métodos de filtros IIR digitales que se basan en esas fórmulas estándar de diseño de filtros IIR continuos son fáciles de realizar.

El hecho de que los diseños de filtros en tiempo continuo se puedan trasladar a diseños de filtros digitales no quiere decir que tengan la misma respuesta frecuencial. Generalmente sucede que el filtro analógico empleado para la aproximación tiene una respuesta en frecuencia diferente de la respuesta en frecuencia efectiva del filtro digital.

Esta circunstancia indica que al diseñar un filtro digital se parte de un conjunto de especificaciones en tiempo discreto; mientras las características del filtro en tiempo continuo se obtienen de la transformación. Al realizar esta conversión se desea que la respuesta en frecuencia del filtro digital preserve las propiedades esenciales del filtro analógico. Esto implica concretamente que se espera que el eje imaginario del plano s se transforme en la circunferencia unidad del plano z .

Una segunda condición es que un filtro estable analógico se debe de transformar en un filtro estable de tiempo discreto. Esto significa que si el filtro continuo tiene los polos en el semiplano negativos de s , el filtro digital tiene que tener los polos dentro del círculo unidad del plano z . Estas restricciones son básicas para las técnicas de diseño de los filtros digitales IIR.

2.2.3.1 DISEÑO DE FILTROS IIR MEDIANTE TRANSFORMADAS BILINEALES

Los filtros recursivos pueden ser diseñados por varios métodos, siendo el más común el basado en las transformaciones bilineales. Este procedimiento requiere del conocimiento de la función de transferencia en el tiempo continuo del filtro a diseñar. Los coeficientes del filtro en el dominio s son transformados a uno equivalente en el dominio z . Los coeficientes de la discretización formarán el filtro IIR.

El origen de este proceder viene dado por la cantidad de experiencia acumulada en el diseño de filtros analógicos. Por tanto, todos los polinomios, tablas, métodos analíticos y gráficos para definir el filtro analógico, empleados anteriormente, serán usados en el diseño de los filtros recursivos.

Si bien hay varios métodos de discretización, la mayoría de ellos tienen problemas de solapamientos en frecuencias, por realizar una relación entre el plano s a z de varias regiones del dominio s a una sola z . Sin embargo, la transformación bilineal consigue una transformación unívoca entre el dominio s a z . Esta transformación se define como:

$$s = \frac{2}{T} \frac{1 - z^{-1}}{1 + z^{-1}} = \frac{2}{T} \frac{z - 1}{z + 1} \quad (2.34)$$

Y su relación inversa es del tipo:

$$z = \frac{1 + (T/2)s}{1 - (T/2)s}$$

(2.35)

y sustituyendo $s = \sigma + j\omega_a$ en la anterior expresión quedará:

$$z = \frac{1 + \sigma(T/2) + j\omega_a(T/2)}{1 - \sigma(T/2) - j\omega_a(T/2)} \quad (2.36)$$

Si $\sigma < 0$, entonces, de la (2.36) se deduce que $|z| < 1$ para cualquier valor de ω_a . Del mismo modo, si $\sigma > 0$, $|z| > 1$ para todo valor de ω_a . Es decir, si los polos del filtro analógico están en el semiplano izquierdo de s , su imagen en el plano z está en el interior de la circunferencia unidad. Por tanto, los filtros en tiempo continuo causales y estables se transforman en filtros en tiempo discreto causales y estables. Seguidamente, para demostrar que el eje $j\omega_a$ se transforma en la circunferencia unidad se procede a sustituir $s = j\omega_a$ en la ecuación (2.36), con lo que se obtiene:

$$z = \frac{1 + j\omega_a(T/2)}{1 - j\omega_a(T/2)} \quad (2.37)$$

con lo que se puede ver que el módulo de z es unitario para cualquier valor de ω_a .

De hecho, para obtener la relación de sus respuestas en frecuencias se sustituye s por $j\omega_a$ y z por $e^{j\omega_d T}$ con lo que queda después de operar que:

$$\omega_a = \frac{2}{T} \tan\left(\frac{\omega_d T}{2}\right) \quad \text{o} \quad \omega_d = 2 \arctan\left(\frac{\omega_a T}{2}\right) \quad (2.38)$$

Observando la ecuación (2. 38) hay que notar cómo el intervalo de la frecuencia digital $0 \leq \omega_d \leq \frac{\pi}{T}$ se transforma en el intervalo de frecuencia analógica $0 \leq \omega_a \leq \infty$.

Las transformaciones bilineales evitan el problema de solapamiento entre s y z , pero el precio que se paga es la compresión no lineal del eje de las frecuencias. Por consiguiente, el diseño de filtros en tiempo discreto mediante transformación bilineal sólo será útil cuando esta compresión se tolera. Así, el uso de la transformación bilineal está limitado al diseño de aproximaciones a filtros con respuesta en amplitud constante a intervalos, como los filtros paso bajo, paso alto o pasa banda. Por otra parte, la distorsión del eje de la frecuencia también se manifiesta en una compresión de la respuesta de fase del filtro, esto es, si el filtro analógico es de fase lineal la transformación bilineal no conserva esta propiedad. Esta afirmación es más cierto sobre todo si la banda pasante se encuentra cerca de la frecuencia de Nyquist.

El método de diseño de los filtros recursivos mediante transformadas bilineales comienza por especificar las características del filtro en el dominio de las frecuencias digitales ($\omega_{d1}, \omega_{d2}, \dots, \omega_{dk}$). Estas especificaciones se pasarán al dominio de las frecuencias analógicas mediante la expresión (2. 38) y que permitirán obtener las frecuencias del filtro analógico:

$$\omega_{ai} = \frac{2}{T} \tan\left(\frac{\omega_{di}T}{2}\right) \quad 1 \leq i \leq k \quad (2.39)$$

A este proceso se le conoce el *prewarping* del filtro analógico. Seguidamente, mediante las soluciones ya tratadas para el caso analógico se obtendrá la función de transferencia del filtro analógico. Sólo restará convertir de s a z empleando (2.34), consiguiendo $G(z)$ del filtro digital. Para facilitar su desarrollo se expone un resumen del procedimiento a seguir:

- 1) Definir las características del filtro digital $Wd1, Wd2, \dots, Wdk$.
- 2) Realizar la operación de *prewarping* de acuerdo con (2.39) obteniendo las frecuencias analógicas $Wa1, Wa2, \dots, Wak$.
- 3) Diseñar el filtro analógico con las frecuencias definidas en el punto 2.
- 4) Reemplazar s en el filtro analógico por la expresión dada en (2.34).

2.2.4 IMPLEMENTACIÓN DE LOS FILTROS DIGITALES

Una vez calculada la FDT (transformada discreta de Fourier) del filtro es posible su construcción. La ecuación en diferencias exige que estén disponibles los valores retrasados de la salida, de la entradas y de las secuencias intermedias. El retardo de los valores de la secuencia implica la necesidad de almacenar los valores pasados de la misma. Además, hay que proporcionar medios para multiplicar por los coeficientes los valores retrasados de la secuencia, así como para sumar los productos resultantes.

Hay dos formas de implementar los filtros digitales: mediante un programa de computador o bien por diseño en soporte físico. Para la ejecución en *software* se elaborará un algoritmo que necesitará almacenar los valores anteriores de la salida y de la entrada y realizar los productos y las sumas correspondientes. En el caso de utilizar lenguajes de alto nivel, se suele usar aritmética de punto flotante, con el objeto de evitar los efectos asociados de la cuantificación con la aritmética de punto fijo, ya que pueden conducir a la inestabilidad y a cambios significativos en la función de transferencia.

Si se usa Electrónica Digital, los elementos básicos necesarios para la realización de un filtro digital son sumadores, multiplicadores y memoria para almacenar los valores retrasados de la secuencia. La interconexión de estos elementos básicos se representan de forma conveniente utilizando diagramas de bloques con los símbolos básicos que se muestran en la figura 2.7.

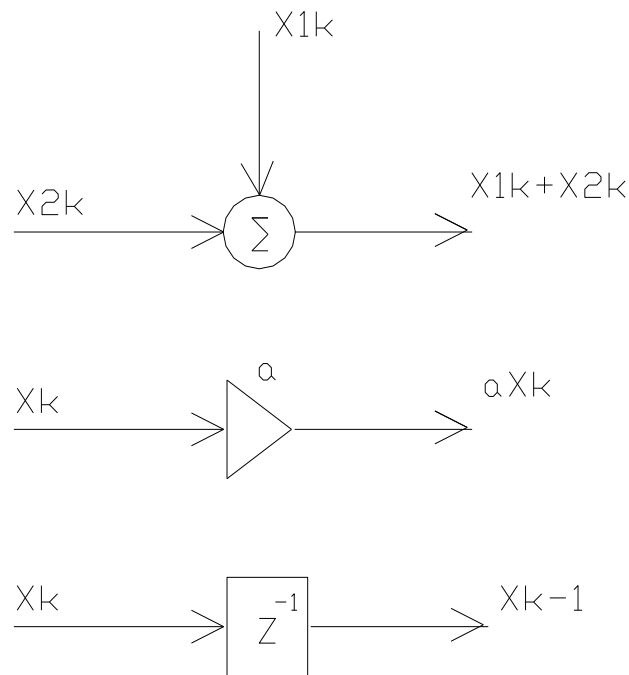


Figura 2. 7. Símbolos del diagrama de bloques: a) suma de dos secuencias, b) multiplicación de una secuencia por una constante, c) Retardo unidad

En la notación general de diagramas de los bloques, un sumador puede tener cualquier número de entradas. Sin embargo, en casi todas las realizaciones prácticas, los sumadores tienen dos entradas. En los diagramas que aquí se exponen esto se indica de forma explícita. En las realizaciones digitales la operación de retardo se puede hacer utilizando un registro de almacenamiento por cada unidad de retardo que se necesite. Si la realización utiliza circuitos integrados, los retardos podrían hacerse empleando un registro de desplazamiento temporizado con la frecuencia de muestreo de la señal de

entrada. En una realización *software*, los retardos en cascada se realizarán con registros de memoria consecutiva.

La realización sistemática en cualquiera de estos dos medios, SW o HW, se ve considerablemente facilitada representando el filtro en diagrama de bloques. Partiendo de las ecuaciones en diferencias del filtro:

$$y(k) = a_1y(k-1) + a_2y(k-2) + \dots + a_ny(k-n) + b_0x(k) + b_1x(k-1) + \dots + b_mx(k-m) \quad (2.40)$$

esta expresión se puede representar por el diagrama de la figura 2.8. Al utilizar sólo sumadores con dos entradas, las sumas se realizan en un orden específico. Es decir, la figura 2.8 indica que deben calcularse los productos a_ny_{k-n} y $a_{n-1}y_{k-(n-1)}$, sumarlos posteriormente y la suma resultante añadirse a $a_{n-2}y_{k-(n-2)}$, y así sucesivamente. Tras calcular y_k , las variables de retardo se deben actualizar introduciendo $y_{k-(n+1)}$ en el registro que tiene y_{k-n} , y así sucesivamente.

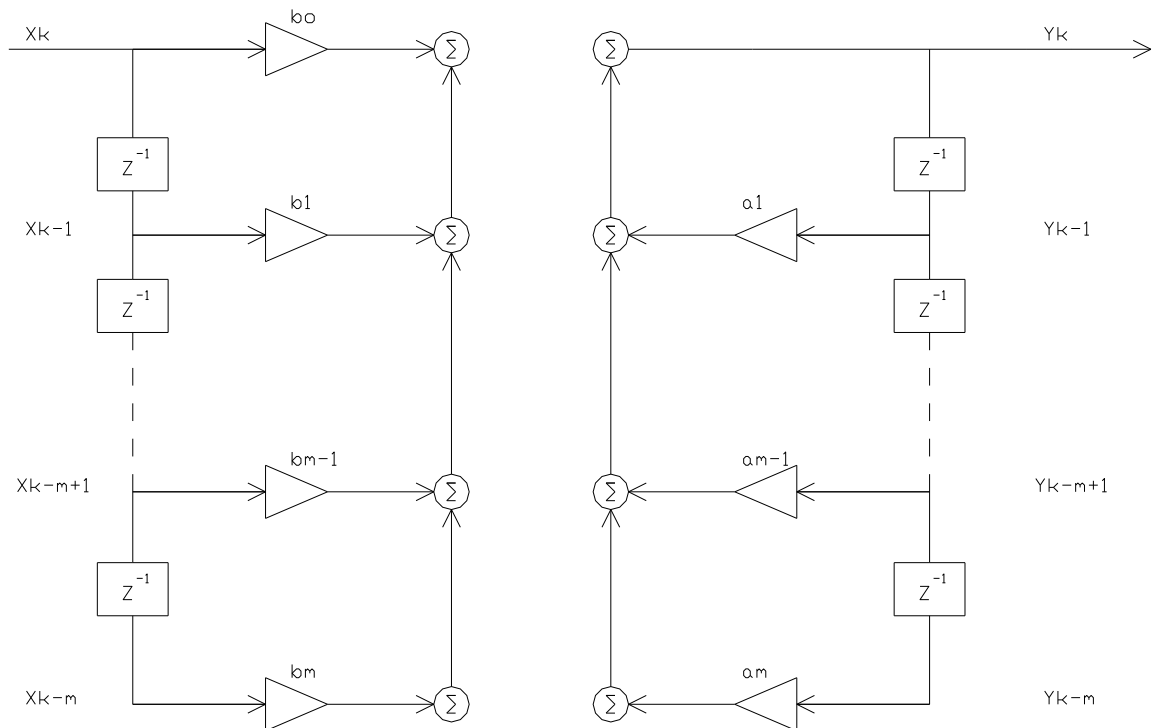


Figura 2. 8 Representación genérica de los filtros a partir de la ecuación en diferencias

Los diagramas de bloques se pueden reestructurar o modificar de diversas maneras sin cambiar la función de transferencia global. Cada reestructuración representa un algoritmo computacional diferente para realizar el mismo sistema. Una forma más adecuada de implementación se basa en la estructura en serie obtenida por la factorización en sistemas de segundo orden de la función de transferencia del filtro:

$$G(z) = k \prod_{i=1}^n \frac{1 + \beta_{1i}z^{-1} + \beta_{2i}z^{-2}}{1 - \alpha_{1i}z^{-1} + \alpha_{2i}z^{-2}} \quad (2.41)$$

La figura 2.9 muestra un diagrama de bloques de un filtro IIR implementado con cascadas de segundo orden. Esta realización tiene cuatro sumas, cuatro multiplicaciones y dos retardos por cada sección de segundo orden. Esta forma de implementación resulta la más efectiva para un rango variado de procesadores de propósito general así como de procesadores específicos DSP.

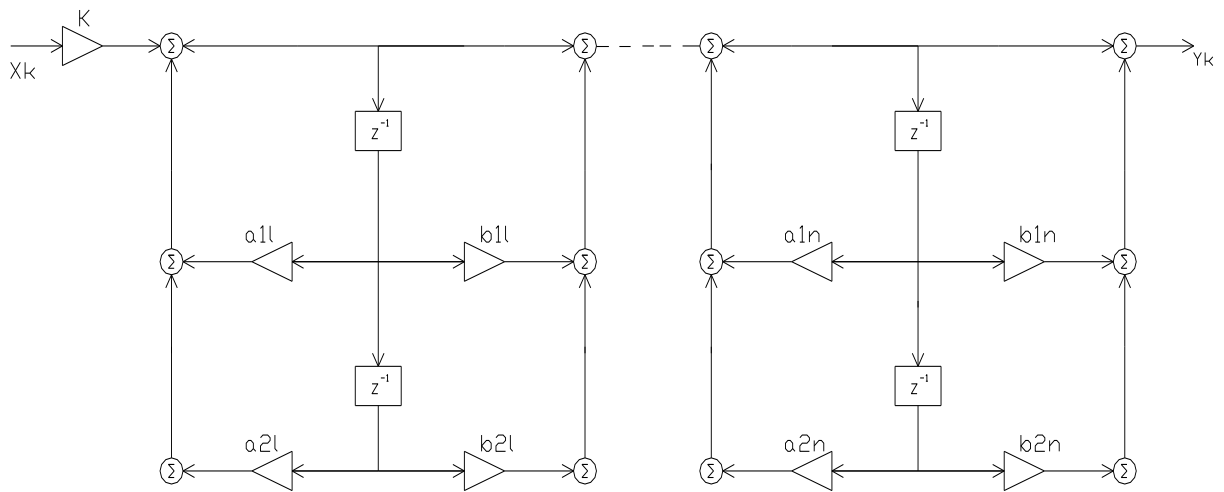


Figura 2. 9 Filtro IIR en cascada

Una realización en cascada se denomina comúnmente implementación en forma canónica. La realización del diagrama de bloques en forma no canónica que muestra la figura 2.8 se denomina implementación en forma directa.

2.2.5 EFECTOS DE LA CUANTIFICACIÓN EN LOS FILTROS

Hasta el momento, en el estudio de los filtros digitales, se ha supuesto que tanto las variables como los coeficientes de los filtros pueden tomar cualquier valor; ahora bien, la realización de estos sistemas se efectúa con dispositivos digitales de cálculo, es decir con computadores, en los que tanto el almacenamiento como la medida de las señales se realiza con números finitos de dígitos. La operación de transformación de los valores de variables y constantes, con un número infinito de cifras, a un número concreto de bits se denomina cuantificación. Este proceso, imprescindible en el tratamiento de información con computador, produce errores que en este último punto se van a analizar y acotar.

En la implantación de los filtros digitales aparecen tres procesos de cuantificación:

- 1) Cuantificación de la señal de entrada
- 2) Cuantificación de los coeficientes de los filtros
- 3) Errores en el redondeo de las operaciones

2.2.6 EFECTOS DE TRUNCAMIENTO Y REDONDEO

En el análisis de error de cuantificación, hay que tener en cuenta la forma de representación de los valores en el computador, según se realice en coma fija o en coma flotante, y, asimismo, también tendrá influencia la forma de realizar el paso de un número cualquiera, equivalente a infinitos dígitos, a un número finito de dígitos, según sea por truncamiento o por redondeo.

En general la representación de los números en computadores se realiza en base 2, es decir en representación binaria, por dos métodos distintos: en coma fija, donde se utiliza un número determinado de dígitos para la parte entera y otra para la parte decimal, o en coma flotante, donde la representación es de la forma:

$$M2^c \quad (2.42)$$

donde M , mantisa, es un número que verifica:

$$\frac{1}{2} \leq M < 1 \quad (2.43)$$

y c , exponente, es un número entero, positivo o negativo; ambos, M y c , se representan por un número fijo de bits.

La forma de representación de los números negativos tiene su importancia según se haga por signo y magnitud, por complemento a dos o por complemento a uno.

Tanto en coma fija como flotante se va a considerar que la parte fraccionaria o mantisa tiene b bits, entonces los números se representan en saltos de 2^{-b} es decir la mínima diferencia entre dos representaciones distintas es de 2^{-b} .

Dado un número de infinitos dígitos, se entiende por truncamiento del mismo la operación de tomar los b primeros bits más significativos, despreciando el resto, mientras que su redondeo es tomar la representación del valor más próximo al mismo.

En la figura 2.10 a) se representan los valores de la señal cuantificada $Q(x)$ en función de los posibles valores de x para el caso de representación por redondeo y coma fija. Como se puede observar en la misma, el error en dicha cuantificación es:

$$\varepsilon = Q(x) - x \tag{2.44}$$

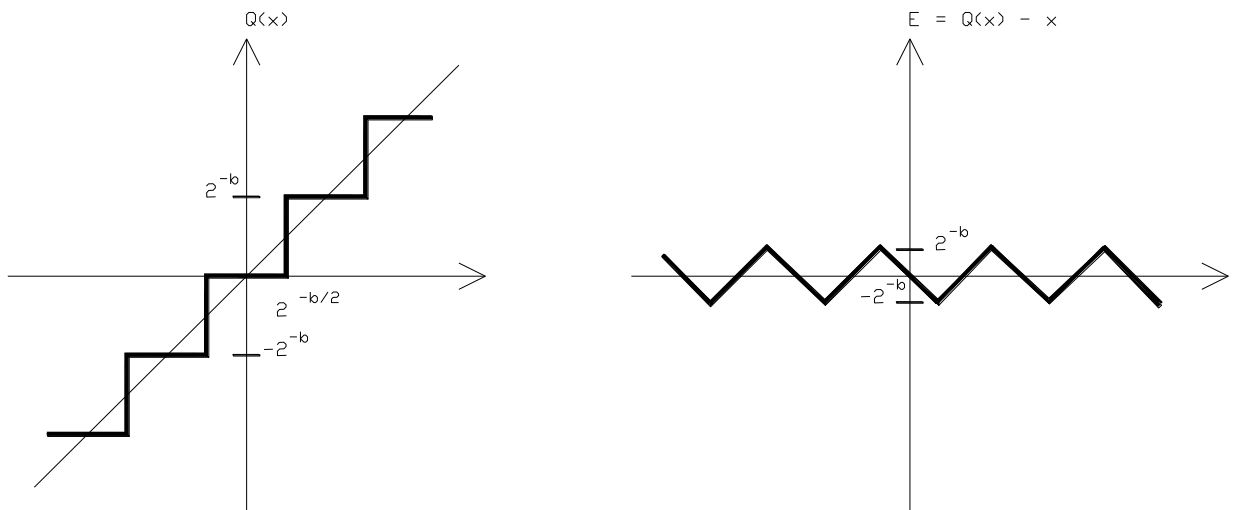


Figura 2. 10 Técnicas de redondeo a) Cuantificación b) Error

En valor absoluto, el error será, como máximo, de medio salto, es decir:

$$\frac{-1}{2}2^{-b} < Q(x) - x < \frac{1}{2}2^{-b}$$

(2.45)

El caso de truncamiento y coma fija, utilizando el complemento a 2 para los negativos, viene representado en la figura 2.11, donde el error está acotado entre los valores:

$$-2^{-b} < Q(x) - x < 0 \quad (2.46)$$

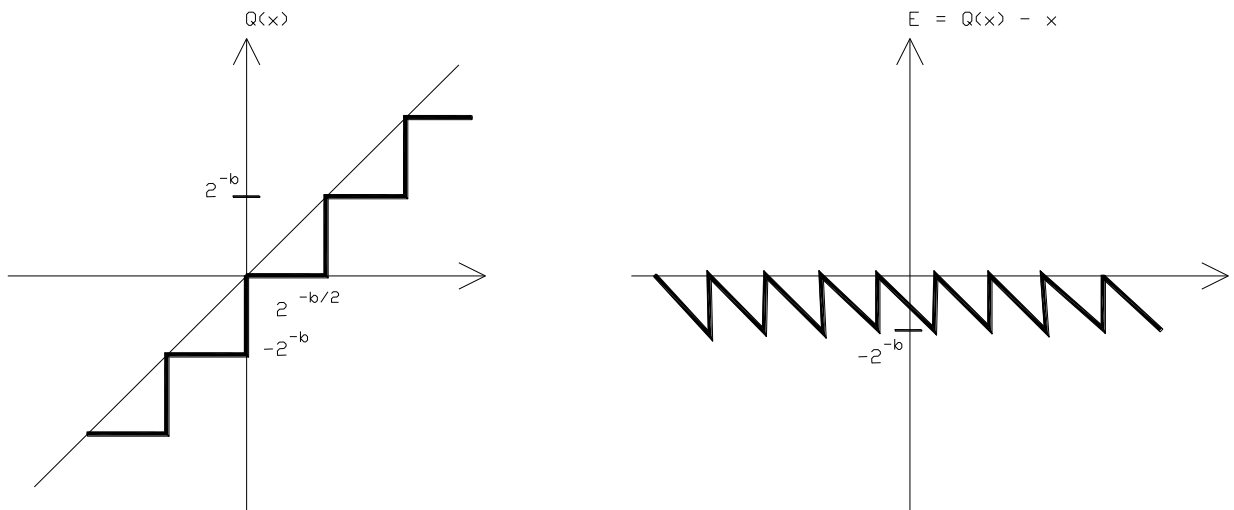


Figura 2. 11 Técnicas de truncamiento a) Cuantificación b) Error

2.2.7 CUANTIFICACIÓN DE SEÑALES

En la conversión analógica/digital de una señal para su tratamiento con computadora además del proceso de muestreo hay que tener en cuenta el de cuantificación, efecto éste debido a la limitación del número de cifras de los registros en los que se almacenan sus valores.

Para el estudio del efecto que esta cuantificación de los elementos de la secuencia de entrada produce sobre los de la secuencia de salida de un filtro

digital, se va a suponer, en primer lugar, como la forma más usual de trabajo de los convertidores A/D, la coma fija con $b+1$ bits, b la parte fraccionaria y 1 de signo, utilizando técnicas de redondeo. Se considera igualmente la señal normalizada a 1, es decir:

$$-1 + \frac{1}{2}2^{-b} < x_k < 1 - \frac{1}{2}2^{-b} \quad (2.47)$$

Una representación equivalente al proceso de cuantificación es la que se obtiene al considerar, según se indica en la figura 2.11b), su efecto como una secuencia de valores aleatorios añadidos a la original.

$$\tilde{x}_k = Q(x_k) = x_k + e_k \quad (2.46)$$

Donde $\{e_k\}$ es una secuencia denominada frecuentemente *ruido* cuyos elementos verifican:

$$-\frac{1}{2}2^{-b} < e_k < \frac{1}{2}2^{-b} \quad (2.47)$$

El proceso de redondeo se caracteriza por una distribución estadística, tal que la secuencia $\{e_k\}$ se define como una variable aleatoria con una función de densidad uniforme, $P(\xi)$ entre $-(2^{-b}/2)$ y $(2^{-b}/2)$ que se representa en la figura 2.12.

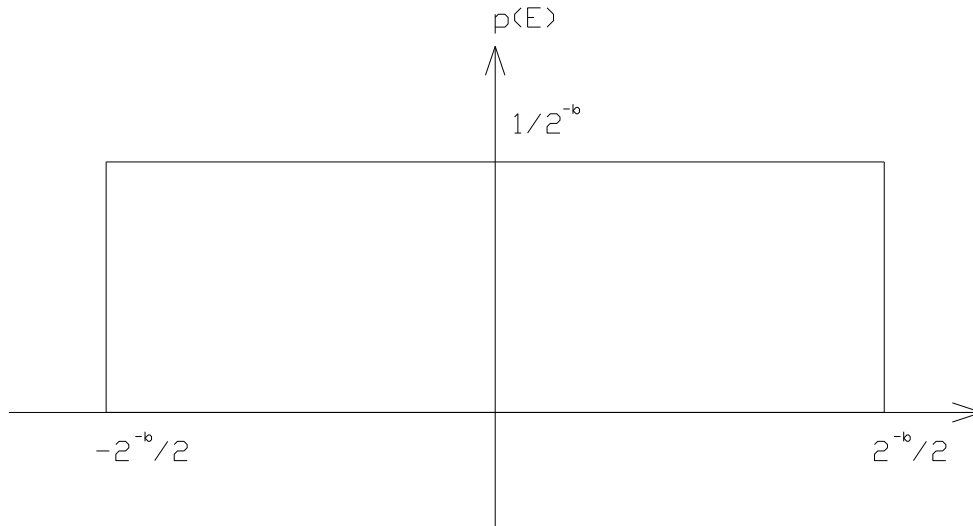


Figura 2. 12 Función de densidad del error por redondeo

Los parámetros estadísticos de la media y la varianza de la misma son:

$$\mu_e = 0 \quad , \quad \sigma_e^2 = \frac{2^{-2b}}{12}$$

(2.48)

El valor que da una idea de la magnitud del error es su varianza. Ésta se suele referir a la amplitud de los valores de la secuencia a cuantificar x_k . Con el fin de comparar parámetros equivalentes se define sobre la secuencia $\{x_k\}$ su varianza:

$$\sigma_x^2 = \frac{1}{N} \sum_{k=0}^N x(n)^2 \tag{2.49}$$

con N igual al número de elementos cuantificados en el proceso en curso de estudio. La medida de error de cuantificación, representada por SNR (*Signal to Noise Ratio*), se define como el cociente entre ambas desviaciones típicas:

$$SNR = \frac{\sigma_x^2}{\sigma_e^2} = \frac{\sigma_x^2}{\frac{2^{-2b}}{12}} = 12(2^{-2b})\sigma_x^2 \tag{2.50}$$

Como se ve en la expresión el índice SNR aumenta con cada incremento en el número de bits que se use para la representación de la señal.

2.2.8 CUANTIFICACIÓN DE LOS COEFICIENTES

Igual que ocurre con los valores de las muestras de la señal, en el proceso de realización de un filtro, los parámetros que lo definen sufren una cuantificación que puede afectar, en algunas ocasiones considerablemente, a las prestaciones del sistema.

Para abordar este problema es necesario centrarse en la ecuación característica del filtro y estudiar como varían sus raíces ante las variaciones en sus coeficientes. Por ejemplo, si un sistema tiene una ecuación en diferencias estimadas:

$$\tilde{y}_k = (a + \Delta a)\tilde{y}_{k-1} + \mu_k \quad (2.51)$$

donde a es el coeficiente que sufre un error de cuantificación, Δa , ésta tendrá por ecuación característica:

$$z - (a + \Delta a) = 0 \quad (2.52)$$

en la que es claro que, si se quiere situar un polo en $z = 0.995$, será necesario almacenar el valor de a con tres cifras decimales ya que la variación máxima de este parámetro, Δa , admisible para garantizar la estabilidad es 0.005.

En general este problema se podrá resolver con un análisis de sensibilidad de las raíces de la ecuación característica a los parámetros de la misma.

Las consecuencias prácticas de la cuantificación de los coeficientes son:

- Los filtros con poco ancho de banda son muy sensibles a la cuantificación. Este efecto se deriva por tener polos muy cerca de $z = 1$.

- El sobremuestreo (*oversampling*) incrementa la sensibilidad a la cuantificación.
- La estructura física del filtro influye mucho.

2.2.9 CUANTIFICACIÓN DE RESULTADOS DE OPERACIONES

A efectos de cuantificación, la operación de mayor interés es la multiplicación ya que el resultado de ésta ocupa el doble de números de cifras que de los factores, por lo que para almacenarlo o realizar otras operaciones habrá que reducir el número de sus cifras.

Este error de cuantificación aparece en cada iteración de la ecuación en diferencias con lo que su efecto se puede estudiar como una secuencia de números aleatorios independientes sumados a la secuencia de salida de los bloques en los que se realizan multiplicaciones. Por ello, la importancia de este error dependerá fuertemente de la estructura elegida para la implementación del filtro.

Utilizando el esquema típico de la construcción de un filtro, según se indicó en la figura 2.9, y siguiendo las propiedades estadísticas enunciadas sobre el error de cuantificación, los ruidos son aditivos, pudiéndose reemplazar por unos cuyos elementos representativos valgan:

$$e = \sum_{i=0}^M e_{bi} + \sum_{i=0}^m e_{ai}$$

(2.53)

Si cada uno de los ruidos considerados tiene iguales propiedades estadísticas, dadas en (2. 54), las del total, suma de los referidos, serán:

$$\mu_e = 0 \quad y \quad \sigma_e^2 = (n + m) \frac{2^{-2b}}{12}$$

(2.54)

indicando que a medida de que se aumente el orden del filtro mayor será el incremento del error cometido por redondeo en las multiplicaciones.

2.3 TIPOS DE ANÁLISIS.

Existen un gran número de métodos de parametrización, todos ellos tienen la característica común de que proporcionan parámetros que varían con el tiempo, es decir, éstos se obtienen a través de una ventana temporal de longitud y tipo adecuado al método de parametrización, que se va desplazando a lo largo de la señal acústica.

Entre todos ellos podemos distinguir los que a continuación se exponen.

2.3.1 EN EL DOMINIO DEL TIEMPO.

Este grupo reúne las técnicas de análisis en el dominio del tiempo, como son: la energía, la amplitud media, la densidad de cruces por cero de la señal y su derivada, etc.

1) *Energía y amplitud media.*

Dada la expresión de la energía de una señal $s(t)$

$$E = \int_{-\infty}^{\infty} s^2(t) dt \quad (2.55)$$

se puede definir la energía dependiente del tiempo con ventana $v(t)$, para señales muestreadas la definición discreta equivalente es como:

$$E(n) = \sum_{m=-\infty}^{\infty} s^2(m)v(m-n) \quad (2.56)$$

en reconocimiento de la palabra, la energía proporciona una primera aproximación para distinguir segmentos vocálicos (alta energía) de segmentos consonánticos (baja energía) y, en el caso de la señal de buena calidad (alta relación señal/ruido), se la puede utilizar para distinguir la voz del silencio (detección de bordes).

El uso de la energía como parámetro puede presentar el inconveniente de su gran sensibilidad a la amplitud de la señal (ésta aparece elevada al cuadrado), que implica la necesidad de un gran margen dinámico o el uso subsiguiente de una transformación logarítmica. Otro inconveniente lo constituye la complejidad de cálculo que supone la elevación de cuadrados (disminución de la velocidad de cálculo, y por consiguiente de la velocidad del reconocimiento).

Para evitar estas dificultades se puede utilizar la amplitud media dependiente del tiempo y su versión discreta es:

$$A(n) = \sum_{m=-\infty}^{\infty} |s(m)v(m-n)| \quad (2.57)$$

2) **Densidad de cruces por cero.**

Se dice que una señal $s(t)$ produce un “cruce por cero”, cuando cambia de signo. Para dicha señal se expresa entonces la “densidad media de cruces por cero” como:

$$Z = \lim_{T \rightarrow \infty} \left(\frac{1}{T} \int_{-\infty}^{\infty} \left| \frac{d}{dt} \text{sig}(s(t)) \right| dt \right)$$

$$\text{Sig}(x) = 1 \text{ para } x > 0 \text{ y } \text{sig}(x) = 0 \text{ para } x \leq 0 \quad (2.58)$$

pudiéndose a partir de esta expresión, definir la “densidad de cruces por cero dependiente del tiempo” y su formulación discreta es:

$$Z(n) = \frac{1}{L} \sum_{m=-\infty}^{\infty} |\text{sig}(s(m)) - \text{sig}(s(m-1))| v(m-n) \quad (2.59)$$

donde L representa la duración efectiva de la función ventana $v(t)$ utilizada.

La *densidad de cruces por cero* ha sido objeto de numerosos estudios teóricos y prácticos. Su utilidad en el reconocimiento del habla radica en que proporciona una estimación aproximada del *contenido frecuencial* de una señal, basada en la idea de que una senoide pura cruza el eje de abscisas dos veces por período. Es un parámetro de muy poca complejidad de cálculo, y se utiliza para detectar segmentos fricativos (señal de pequeña energía y elevada densidad de cruces por cero) y en ocasiones, en combinación con la densidad de cruces por cero de la primera derivada de la señal (preénfasis), para localizar formantes y/o para realizar una clasificación tosca de las vocales. También se suele utilizar para refinar la detección de bordes obtenida a partir de los parámetros de amplitud o energía. Su mayor inconveniente radica en su extremada sensibilidad a las posibles componentes de continua de la señal y al ruido en general.

Realizando las sucesivas derivadas de la señal, se obtienen diferentes parámetros que informan sobre otras características adicionales de la forma de la señal. Los más importantes son:

- 1) **Densidad de máximos y mínimos:** Se obtiene calculando el número de cruces por cero de la derivada de la señal.
- 2) **Densidad de puntos de inflexión:** Se calcula mediante la integral del cruce por cero de la segunda derivada de $s(t)$.

2.3.2 EN EL DOMINIO DE LA FRECUENCIA.

Los métodos de parametrización más conocidos son los espectrales, basados en la *transformada discreta de Fourier*, y a partir de ella pueden utilizarse otros

métodos como el del *análisis homomórfico*, mediante el cual se puede eliminar la componente periódica de la señal acústica.

A partir de la noción de Serie de Fourier (suma trigonométrica de una serie ponderada de senos y cosenos cuyas frecuencias son múltiplos de la frecuencia fundamental) se desarrolló la Transformada de Fourier, una importante transformación matemática que relaciona las representaciones temporal y frecuencial de una función.

La Transformada de Fourier necesita una adecuación que permita trabajar con señales muestreadas en lugar de hacerlo, según su formulación original, con señales continuas. Esta adecuación se conoce como “Transformada Discreta de Fourier”. Desde que se formuló la DFT, se ha realizado una constante labor de investigación para obtener nuevos algoritmos que lleven a cabo el cálculo con un menor tiempo de ejecución. En este sentido, Cooley y Tukey descubrieron en 1965 un algoritmo denominado “Transformada Rápida de Fourier”, que se muestra como el algoritmo de uso más general y efectivo.

En esencia, la transformación de Fourier de una forma de onda o señal $x(t)$ consiste en descomponer o separar dicha señal en una suma de señales sinusoidales de diferentes frecuencias. Dada una señal arbitraria, si se determina la existencia de tal suma entonces se habrá obtenido su Transformada de Fourier. Su representación gráfica es un diagrama que representa la amplitudes y fases de cada señal sinusoidal componente frente a la frecuencia.

Así la Transformada de Fourier de una señal $x(t)$ (t -tiempo) identifica las diferentes señales sinusoidales, representadas a través de una función $X(f)$ (f -frecuencia), cuya combinación forma la señal original $x(t)$. Normalmente, este análisis se asocia con señales periódicas en el tiempo, sin embargo este es un caso particular de su aplicación. Si la señal $x(t)$ no es periódica, entonces la Transformada de Fourier es una función cuyo rango de variación de frecuencias es continuo, al contrario de lo que sucede con las señales periódicas en las que su transformada sólo incluye múltiplos enteros de la frecuencia fundamental. Como conclusión podemos decir que la Transformada de Fourier es una

representación en el dominio de la frecuencia que contiene exactamente la misma información que la señal original. Este cambio de representación permite examinar las señales desde otro punto de vista, quizás con el objetivo de extraer informaciones sobre las señales que no se manifiestan desde la propia evolución temporal.

Mediante un hardware digital sólo puede calcularse la Transformada de Fourier de manera aproximada. En este sentido se debe considerar señales discretas tanto en el tiempo como en la frecuencia, así como aplicarla a un número limitado de muestras de la señal. Si $X(f)$ representa la amplitud y fase para cada frecuencia componente de la función temporal $x(t)$ y se tiene en cuenta los aspectos anteriores se obtiene una aproximación de la Transformada de Fourier que se denomina:

Transformada Discreta de Fourier (DTF):

$$X(k) = \sum_{n=0}^{n=N-1} x(n)W_N^{kn} \quad (2.60)$$

Con , $W_N^{kn} = e^{-j\frac{T}{N}kn}$ siendo $k=0 \dots N-1$.

En realidad $X[n]$ debe interpretarse como $x[nTs]$ y $X[k]$ como $X[kfs/N]$, donde $fs=1/Ts$ representa la frecuencia a la cual se muestrea la señal $x(t)$.

Se han obviado los términos constantes Ts y fs/N por simplicidad, así como el producto por Ts , en el segundo término de la igualdad, que sólo implica un factor de escala en el resultado. La ecuación es la expresión matemática de la DFT y es válida, solamente si la máxima frecuencia contenida en la señal temporal es inferior a la mitad de la frecuencia de muestreo fs (criterio de Nyquist). Por otro lado la aproximación será tanto mejor cuanto más se ajuste el segmento de la señal a procesar a un número entero de periodos de la misma y más muestras N sean tenidas en cuenta.

El coste en tiempo de un algoritmo que realice directamente la DFT es proporcional a N^2 , lo cual puede resultar excesivo si el objetivo es realizar una aplicación en tiempo real. El algoritmo desarrollado por Cooley y Tukey mejora este aspecto considerablemente. La idea principal del algoritmo es de tipo recursivo, y básicamente consiste en dividir el cálculo de una transformada DFT en el cálculo de otras dos DFT con la mitad de puntos, de manera que en una primera división ya tendríamos un costo de ejecución proporcional $N^2/2$. Si $Y[k]$ y $Z[k]$ representan las dos DFT para la mitad de puntos, por ejemplo $Y[k]$ para los puntos pares y $Z[k]$ para los puntos impares y el cálculo se realiza por parejas de puntos separados entre sí en una distancia $N/2$ se obtiene:

El proceso de división, denominado “división en tiempo”, continúa para las funciones $Y[k]$ y $Z[k]$.

$$X(k) = Y(k) + W_N^k Z(k)$$

$$X(k + N/2) = Y(k) - W_N^k Z(k)$$

$$k = 0 \dots\dots\dots N/2-1 \tag{2.61}$$

Cada una de ellas se vuelve a descomponer en otras dos y así sucesivamente hasta que no se puedan realizar más subdivisiones. La pareja de ecuaciones básicas previamente mostradas forman un esquema que se repite sucesivamente y que se conoce con el nombre de “mariposa”.

Este método exige su aplicación sobre un conjunto de muestras $x[n]$ cuyo cardinal N sea una potencia de 2 (orden de la FFT) para que cuadren correctamente las sucesivas subdivisiones. El algoritmo se puede ejecutar contra una única estructura de datos que contiene inicialmente las N muestras $x[n]$ que previamente deberán ordenarse de una forma especial denominada “bit inverso”. Esta ordenación es tal que la expresión del número de orden de la muestra n ésima indicada en bits es la inversa de la posición que ocupa en la lista indicada en bits (ejemplo $x[100$ (en decimal 4)] ocupa la posición 001). El resultado final para las N amplitudes $X[k]$ están ordenadas de forma natural en la misma estructura de datos.

El coste de ejecución del algoritmo FFT es proporcional al número de pasos o subdivisiones por el número de mariposas en cada paso y por el costo de cada mariposa, siendo globalmente $O(N \log^2 N)$, lo que supone una evidente mejora con respecto del cálculo directo de la DFT.

En los últimos años han surgido, y continúan en desarrollo, distintos métodos y técnicas de procesamiento digital, para la detección y evaluación de señales. Entre estos métodos se destacan los basados en el estudio de las distribuciones *tiempo-frecuencia* como la Transformada de Fourier de Corto Tiempo (STFT), o la Transformada Wavelet (WT).

CAPITULO III

CODIFICACIÓN

3.1 CODIFICACIÓN DE LA VOZ

En general, la codificación de voz puede ser considerada como un caso particular en el amplio campo del procesamiento de la voz, el cual incluye análisis y reconocimiento de la misma. El procesamiento de la voz actualmente ha

experimentado un considerable avance debido al estudio de las técnicas y sistemas en el DSP. Recíprocamente, es justo decir que el procesamiento de la voz ha sido la base para el DSP, en el sentido que muchos de los algoritmos ampliamente usados en el procesamiento digital de señales fueron desarrollados o puestos en práctica al inicio, por personas que trabajan en sistemas de procesamiento de la voz.

Estos algoritmos incluyen técnicas de filtrado digital, modulación de código de pulso (PCM), código predictivo lineal (LPC), la transformada de Fourier en tiempo corto (STFT), representaciones generales en tiempo-frecuencia, técnicas de filtrado adaptativo, técnicas de bancos de filtros, modelos de Markow escondidos (HMM) y muchos otros. Por consiguiente, mediante el estudio de la conducta de los sistemas de procesamiento de voz, es posible estudiar la conducta de los más importantes algoritmos en el procesamiento digital de señales.

De todas las áreas en el contexto del procesamiento de voz, la codificación de la voz es la más entendida y la más complicada. Existen tres razones para que esto suceda; la primera resulta del hecho que la voz es una de las señales de más bajo ancho de banda. Las señales telefónicas de calidad tienen un ancho de banda de 3.2 KHz, mientras que una señal de voz de alta calidad puede ser obtenida con un ancho de banda de 5-6KHz. Esto da como resultado que se obtenga rangos bajos de muestreo de 6400-12000 muestras por segundo, lo que provoca que un procesamiento en tiempo real sea efectivamente más caro que para otras áreas de importantes de estudio.

Un segundo factor significativo en el desarrollo de los sistemas de codificación de voz, es su continuo desarrollo debido al revolucionario desarrollo de los circuitos integrados VLSI. La tecnología VLSI tiene un impacto masivo en todas las áreas de la tecnología. Sin embargo hay una especial relación entre los algoritmos de codificación de la voz, la tecnología VLSI, y la industria de las telecomunicaciones.

Los codificadores de la voz, forman la base para un conjunto de algoritmos, tecnologías y aplicaciones.

El tercer factor de importancia es la efectividad de varios algoritmos del DSP para resolver muchos de los problemas fundamentales asociados con los sistemas codificadores de la voz. Las técnicas de procesamiento digital de señales han probado ser muy efectivas en modelar tanto la producción como percepción de la voz. Note que no se quiere decir que el DSP ha resuelto todos los problemas de codificación de la voz.

3.1.1 DEFINICIÓN DE UN CODIFICADOR DE LA VOZ

Todos los elementos primarios de un sistema codificador de la voz están ilustrados en la figura 3.1. En el lado izquierdo de la figura 3.1 están listados los subsistemas los cuales están detallados en los cuadros de la figura 3.1.

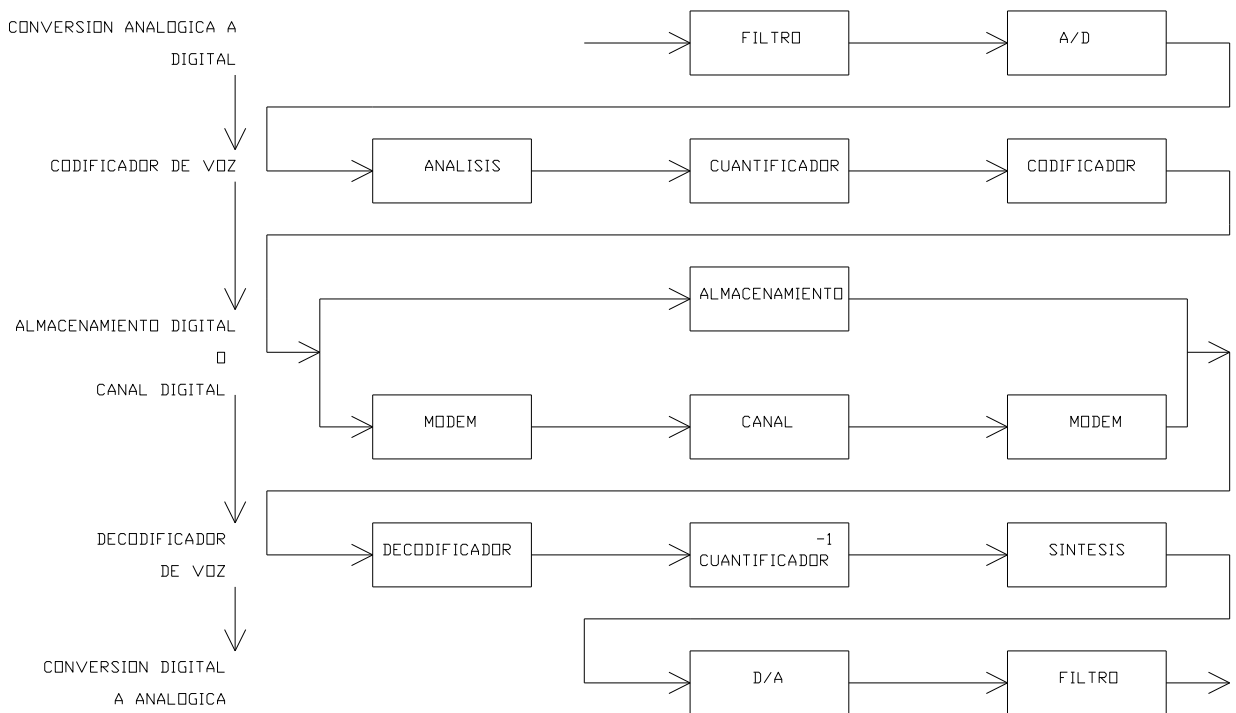


Figura 3.1 Elementos de un sistema de codificación de la voz

La entrada al sistema es una forma de onda de voz continua en el tiempo $s(t)$. Esta señal es filtrada con un filtro pasabajos usando un filtro anti-aliasing y

muestreada por un conversor A/D, entregando una señal de voz digital $s[n]$. Esta señal es la entrada al codificador de voz.

Generalmente el codificador de voz consiste de tres componentes: análisis de la voz, parámetros de cuantificación, y parámetros de codificación. La entrada a la fase de análisis es la señal digital de la voz, mientras que su salida es una nueva representación de la señal de la voz, la cual será cuantificada y codificada.

La salida de la fase de análisis puede variar considerablemente dependiendo del modelo que se haya utilizado en la señal de la voz. Para un sistema PCM, por ejemplo, podría no ser necesario el análisis, debido a que su salida podría simplemente ser la señal digital de la voz. Para otros tipos de codificadores de formas de onda, la salida podría ser una versión procesada de la señal de entrada. Luego del análisis, los parámetros obtenidos no están cuantificados y ellos representan la mejor representación de la señal de la voz que puede ser entregado por el análisis.

Luego del análisis, los parámetros deben ser cuantificados para reducir el número de bits requeridos. En otras palabras, la cuantificación es usada para reducir el rango de información intrínseco de la representación de la voz. La salida del cuantificador puede ser considerada como un ruido de la salida del analizador. La salida del cuantificador se entrega al codificador el cual asigna un código binario único para cada posible representación de la cuantificación. Estos códigos binarios son empaquetados conjuntamente para una eficiente transmisión o almacenamiento.

3.2 TÉCNICAS DE CODIFICACIÓN

Los sistemas modernos de codificación de la voz típicamente toman ventaja de tres clases de características para minimizar la percepción de la distorsión en la codificación: Las características auditivas humanas, las características del tracto vocal, y las características del lenguaje y la individualidad de la persona que habla.

La técnica de codificación de voz más simple, es la modulación de código de pulso lineal (PCM), la cual esencialmente no hace asunciones (excepto para rangos dinámicos) sobre las características de la señal a ser codificada o el eventual uso de la señal codificada. Como resultado, los sistemas PCM lineales requieren de un rango de bits alto para generar una voz de calidad. Los sistemas PCM tienen la ventaja, que estos no tienen preferencia por ninguna señal, por lo tanto pueden usarse para codificar otras clases de señales, tales como datos o música.

Los sistemas de modulación de código de pulso expandido, tales como los sistemas expandidos, 64kbps ley-u, y ley-a usados en las redes de telefonía, hacen uso directo de las características de ruido enmascarado del oído en su forma más básica. En las señales codificadas de la voz, el ruido es correlacionado con la señal, lo cual es bueno, pero el ruido también se distribuye a lo largo del rango de frecuencias sin considerar la presencia de la energía de la señal, lo cual es malo. Los sistemas PCM expandidos son extremadamente simples y su implementación no es costosa, pero el rango al cual ellos deben trabajar para obtener una voz de calidad es todavía relativamente alto.

Sistemas tales como modulación de código de pulso diferencial (DPCM) y modulación delta (DM) hacen uso directo de modelos estadísticos estacionarios de grandes términos para la producción de la voz. Sistemas de modulación de código de pulso diferencial adaptativo (ADPCM) y modulación delta adaptativo (ADM) también hacen uso de la variación lenta de la energía en intervalos cortos de tiempo, provocando que el ruido sea altamente correlacionado con la señal, y produciendo una dramática caída en el canal del ruido.

Varias formas de los codificadores de transformación adaptativa (ATC) y de los codificadores predictivo adaptativo (APC) toman uso de todas las características auditivas y vocales, y estos sistemas son capaces de generar una excelente calidad de la voz a mediano y bajo rango de bits. Lo mismo se puede decir para los más nuevos codificadores paramétricos tales como el codificador predictivo lineal multimpulso- excitado (MPLPC), el codificador predictivo lineal código- excitado (CELP), y el vocoder auto- excitado (SEV). Tales sistemas fueron

originalmente considerados como muy complejos y eran muy sensibles al ruido de fondo y a los errores de transmisión, sin embargo estos problemas han sido solucionados, y muchas implementaciones efectivas existen ahora para esta clase de codificadores de la voz.

3.2.1 CODIFICADORES DE PREDICCIÓN LINEAL (LPC)

El análisis de LPC se utiliza para encontrar los coeficientes que representarán la función de transferencia del filtro que modela el sistema. Si el modelo es capaz de predecir la señal con un error muy bajo, se tiene que el LPC ha sido capaz de almacenar la información necesaria de un trozo de señal como para reproducirla mediante alguna excitación. En analogía con un instrumento musical, el LPC sería un instrumento de viento que al ser soplado emite el sonido con el timbre particular del trozo de voz que representa.

El principio de un LPC es que el valor actual de una muestra de señal de voz, $s(n)$, puede predecirse a partir de un número finito de muestras anteriores: $s(n-1)$, ... , $s(n-p)$, con un error asociado $e(n)$ utilizando un filtro lineal sólo polos:

$$s(n) = e(n) + \sum_{k=1}^p a_k s(n-k) \quad (3.1)$$

El error de predicción (también conocido como señal residual), $e(n)$, es simplemente la diferencia entre el valor actual de la señal, $s(n)$, y el valor que se predijo:

$$e(n) = s(n) - \tilde{s}(n) \quad (3.2)$$

Los factores que otorgan el peso, α_k , son encontrados al minimizar el error cuadrático medio, encontrado en N muestras (E):

$$E = \left(\sum_{i=0}^{N-1} e^2(i) \right) / 2 \quad (3.3)$$

Los coeficientes α_k que minimizan el error de predicción E son calculados igualando el gradiente con respecto a a_i a 0, para $i = 1, \dots, k$. Lo que da como resultado una serie de ecuaciones lineales:

$$\left(\frac{\delta E}{\delta a_k}\right) = 0 \quad \forall k \tag{3.4}$$

Se sabe que $s(n)$ es constante, porque es la señal original, luego al derivar para encontrar el mínimo se tiene que:

$$\left(\frac{\delta E}{\delta a_k}\right) = \sum_i \left(s(i) - \sum_j a_j s(i-j) \right) s(i-k) = 0 \quad j=1 \dots p$$

$$\sum_i \left(s(i) s(i-k) \right) = \sum_j a_j \left(\sum_i s(i-j) s(i-k) \right) \tag{3.5}$$

Si se define $\gamma(i,k) = \sum_n s(n-i) s(n-k)$ se obtiene un sistema de ecuaciones matricial de la forma:

$$\begin{matrix} \gamma(1,1) & \dots & \gamma(1,p) & a_1 & \gamma(1,0) \\ \dots & \dots & \dots & \dots & \dots \\ \gamma(p,1) & \dots & \gamma(p,p) & a_p & \gamma(p,0) \end{matrix} = \dots \tag{3.6}$$

Pero se tiene un problema, para medir en los bordes de la ventana de la señal se necesita salir de la ventana, por lo tanto se pueden hacer 2 suposiciones:

- 1) Medir fuera de la ventana suponiendo que es cíclica.
- 2) Todo fuera de la ventana es 0, lo cual otorga una nueva función g cuyos límites son $-\infty$ y $+\infty$, esto simplifica bastante el problema, llegando al siguiente estado:

$$\gamma(i,k) = r_{|i-k|} = \sum_{n=-\infty}^{\infty} s(n) s(n-(i-k)) \tag{3.7}$$

La ecuación anterior bajo esas condiciones es conocida como método de autocorrelación.

Esto es casi equivalente a suponer que la señal se repite y que la función fuera de la ventana vale 0, por lo tanto la matriz adopta la siguiente forma.

$$\begin{matrix}
 r_0 & \dots & r_{p-1} & \dots & a_1 & & r_1 \\
 \dots & \dots & \dots & \dots & \dots & = & \dots \\
 r_{p-1} & \dots & r_0 & \dots & a_p & & r_p
 \end{matrix} \tag{3.8}$$

Además, $r_k = \sum_n s(n)s(n-k)$ con $n=k, \dots, N-1$ y $k=0, \dots, p$. Es decir, basta calcular $p+1$ valores en la ventana, esquemáticamente se presenta el analizador LPC en la figura 3.2.

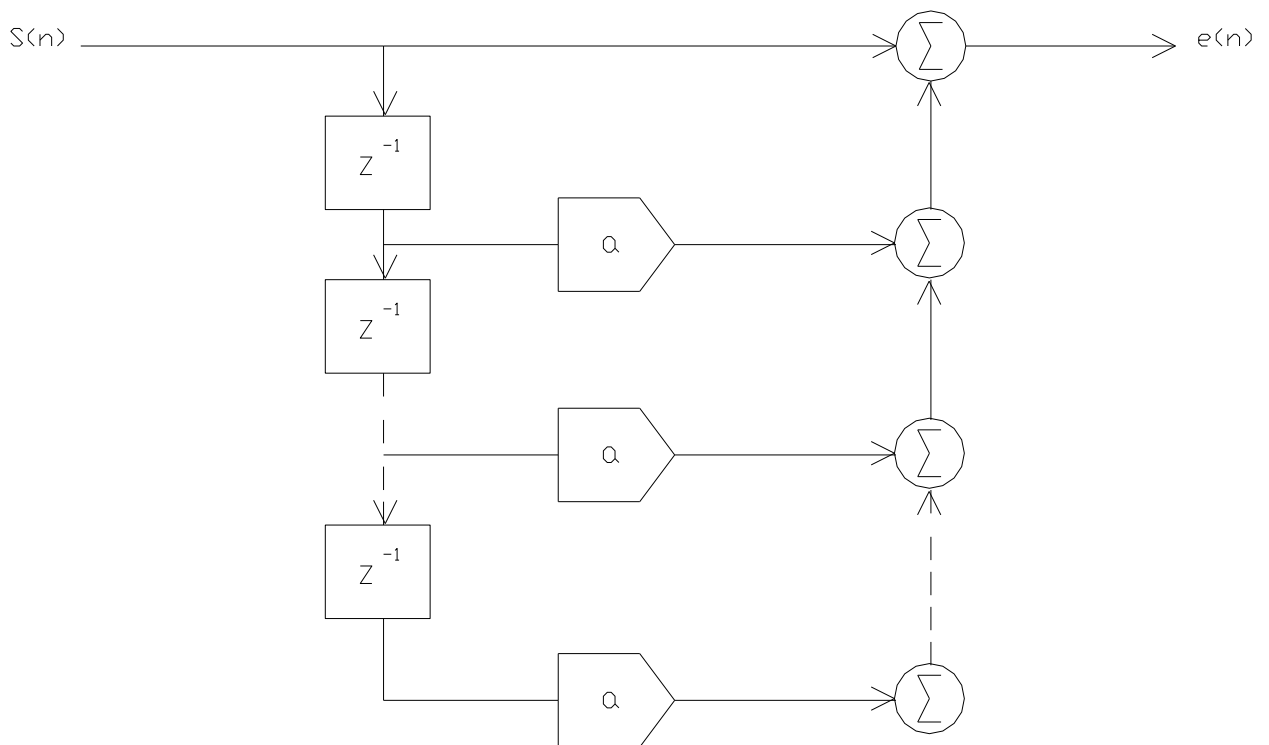


Figura 3.2 Esquema de un analizador LPC

3.3 TÉCNICAS BÁSICAS DE SÍNTESIS DE VOZ.

Muchos sistemas de laboratorio y dispositivos comerciales realizan la conversión automática de un texto a voz y vocoders de voz sintetizadas. El progreso en esta área ha sido posible debido a los avances en la teoría lingüística, en el modelo de caracterización acústica-fonética de los sonidos, en el modelado matemático de generar voz, en la programación estructurada y en el diseño de hardware de los computadores.

3.3.1 SISTEMAS DE SÍNTESIS DE VOZ.

Los métodos para sintetizar voz han cambiado mucho a lo largo de los años, desde los primeros sistemas que utilizaban dispositivos eléctricos y mecánicos resonantes hasta los sintetizadores modernos que emplean computadores o circuitos digitales de diseño específico.

En la actualidad, la clasificación más común de los sistemas de síntesis de voz es atendiendo a las reglas que se siguen para la reconstrucción de la voz, distinguiéndose los siguientes cuatro sistemas:

- 1) Sintetizadores articulatorios. En éstos se realiza una analogía entre parámetros relativos a los órganos articulatorios y sus movimientos con parámetros circuitales.

Pueden proporcionar una calidad altísima, pero es muy difícil obtener y controlar parámetros para un sintetizador de este tipo.

- 2) Sintetizadores por formantes. Son una serie de filtros que modelan el tracto vocal, excitados por fuentes que simulan las cuerdas vocales. Gozan de gran difusión.

- 3) Sintetizadores derivados de las técnicas de predicción lineal (LPC). Son sintetizadores de análisis-síntesis, en los que los parámetros que controlan la función de transferencia del filtro que simula el tracto vocal son parámetros LPC.
- 4) Sintetizadores por concatenación de forma de onda. Concatenan unidades pregrabadas para generar nuevas frases, con lo que intentan aumentar la calidad de la señal generada minimizando el ruido de codificación. La complejidad es alta, pero la calidad obtenida es muy buena.

3.3.1.1 SÍNTESIS POR FORMANTES.

Estos sintetizadores se basan en la teoría acústica de producción de voz, que en su forma más sencilla, dice que es posible ver la voz como el resultado de la excitación de un filtro lineal por una o más fuentes sonoras.

Modelos de la fuente de voz.

Las fuentes primarias del sonido son tonos, producidos por la vibración de las cuerdas vocales y ruido turbulento causado por una diferencia de presión a través de una constricción, es decir, el ruido producido por la fricación.

La fuente de voz usada en los sintetizadores por formantes ha evolucionado desde los trenes de impulsos filtrados o los dientes de sierra de los primeros diseños hasta modelos matemáticos mucho más complejos que permiten tener control sobre parámetros como la frecuencia fundamental, la amplitud, la proporción de tiempo que la glotis está abierta en un periodo, lo abrupto de la forma de onda y diplophonic vibration (los periodos alternos son más similares de los adyacentes). Los modelos de forma de onda glotales más recientes producen espectros que incluyen ceros.

Modelos de la función de transferencia del tracto vocal.

El filtro lineal simula los efectos resonantes del tubo acústico formado por la faringe, la cavidad bucal y los labios. La función de transferencia del este tracto vocal puede modelarse como un conjunto de polos (cada par de polos complejos conjugados produce un pico local en el espectro, conocido como formante). A veces, la representación de la función de transferencia del tracto vocal en términos de polos se ha completado con ceros (antiresonadores) para modelar las propiedades absorbentes del sonido que tienen los tubos secundarios en articulaciones complejas como las nasales, las vocales nasalizadas y las fricativas.

Si se considera que el efecto de los ceros son poco perceptibles acústicamente, se llega a un modelo *de sintetizador de formantes en paralelo* para modelar la función de transferencia del tracto vocal. En este sistema se suman las salidas de un conjunto de resonadores conectados en paralelo y la ganancia de cada formante es controlada por un parámetro independiente.

Parámetros del sintetizador por formantes.

Los parámetros que controlan un sintetizador de formantes son del tipo de “Ancho de banda del primer resonador serie”, “Frecuencia central del cuarto resonador paralelo” o “Ganancia del generador de onda glotal”.

El control de estos parámetros ofrece una subdivisión entre:

- 1) Síntesis por regla: en la que los parámetros se calculan con arreglo a un conjunto de reglas dependientes del contexto.
- 2) Sistemas de análisis-síntesis: en los que los parámetros se obtienen por análisis o parametrización de segmentos de voz natural.

Los sintetizadores de formantes tienen también un conjunto de reglas para predecir las duraciones de los segmentos sintetizados y la curva de tono fundamental.

3.3.1.2 SÍNTESIS POR CONCATENACIÓN.

Otro método o estrategia para sintetizar voz consiste en coger trozos de voz natural pregrabadas como bloques básicos que se concatenan para reconstruir una expresión arbitraria.

Unidades de concatenación.

El tipo de unidad a concatenar es un parámetro crítico para conseguir una buena calidad de la voz sintetizada: hay que llegar a un compromiso entre la calidad intersegmental posible (a mayor longitud de los segmentos, menos puntos de concatenación y por lo tanto mayor calidad) y la cantidad de memoria necesaria para almacenar las unidades pregrabadas.

Los trozos grabados no pueden ser palabras por dos motivos fundamentales. En primer lugar, la pronunciación de una frase es muy diferente a la de una secuencia de palabras recitadas aisladamente, ya que en una frase las palabras tienen una duración más corta que cuando están aisladas y el ritmo, entonación y acentuación, que dependen de factores semánticos y sintácticos, son totalmente antinaturales cuando se concatenan palabras grabadas aisladamente.

Un segundo problema las innumerables palabras existentes en un idioma, si tenemos en cuenta por ejemplo los nombres propios, así como la formación de palabras mediante sufijos, prefijos y conjugaciones.

La sílaba es una unidad interesante lingüísticamente, pero hay un gran número de ellas.

Otra unidad probada fue el fonema, cuyo número es de unos 30, pero el resultado de concatenar fonemas no es satisfactorio debido a efectos coarticulatorios entre fonemas adyacentes que producen cambios de las manifestaciones acústicas de un fonema dependiendo del contexto.

Los efectos coarticulatorios tienden a minimizarse en el centro acústico de un fonema, lo cual llevó a Peterson a proponer el difonema, el trozo de voz que va desde la mitad de un fonema a la mitad del siguiente fonema, como la unidad más satisfactoria para la concatenación.

Además puede ser necesario introducir alófonos diferentes para hacer la distinción entre las vocales acentuadas y átonas o la inclusión de trifenemas, que son una extensión a agrupaciones de tres fonemas cuando los efectos de coarticulación son tan grandes que no es posible la segmentación en difonemas.

Se puede reducir la longitud de la memoria necesitada para el almacenamiento de las unidades:

- 1) Evitando las unidades difonema formadas por consonante + consonante que no se puedan dar en el lenguaje
- 2) Tratando algunos alófonos en determinados contextos como una unidad de tipo fonema, como por ejemplo los fricativos sordos.

3.3.1.3 COMPARACIÓN DE LOS MÉTODOS DE SÍNTESIS.

La principal ventaja del sintetizador de formantes es que permite la manipulación de las características de la fuente de voz. Esto puede ser importante en la síntesis de emociones, puesto que como veremos, la fuente de voz puede cambiar de una emoción a otra, y una de las formas de expresar el enfado consiste en añadir una componente de ruido a la fuente de voz.

Sin embargo, la calidad de la voz normal generada con este método es peor respecto a la generada por el método de concatenación.

Por el contrario, la síntesis mediante concatenación la fuente de la voz es única y corresponde a la grabación de los di fonemas, que generalmente corresponderán a una pronunciación normal.

CAPÍTULO IV

DESARROLLO DEL SISTEMA DE COMPRESIÓN

4.1 TÉCNICAS DE COMPRESIÓN.

Existen varias técnicas de compresión de voz, para otros idiomas tales como el inglés, el alemán, francés, para el caso del idioma español existen muy pocos estudios sobre este tema.

En la actualidad se han realizado estudios de compresión de voz, para el idioma español, utilizando técnicas de decimación e interpolación obteniéndose como resultado un nivel de compresión de 10:1; con este método a demás de la compresión se puede también realizar reconocimiento de voz.

La técnica utilizada en el presente trabajo es vocoder en subbandas, que permite mayores niveles de comprensión, pero no se puede reconocer la voz, ya que esta es reconstruida en forma metálica o robótica.

4.1.1 TÉCNICA DE VOCODER EN SUBBANDAS

Un vocoder está diseñado específicamente para señales de voz, por lo tanto no es aplicable su uso en las redes de telefonía pública, en las cuales otros tipos de señales (como la señal de un MODEM) son transmitidas. Es más, los vocoders típicos producen sonido de voz “artificial” o “poco natural”.

El objetivo principal de un vocoder es codificar solo las características perceptivas importantes de la voz, con la menor cantidad de bits que el común de los codificadores. Debido a esto, los vocoders son utilizados en aplicaciones de limitado ancho de banda, donde otras técnicas no pueden aplicarse.

La voz consiste en una sucesión de “fonemas” (sonidos articulados por el tracto vocal). Cada sonido de voz se caracteriza por su potencia espectral (es decir, el valor cuadrático de la transformada de Fourier de la señal). La envolvente espectral de cada sonido de voz es determinada por el mecanismo humano de generación de voz. A su vez, el sistema de audición humana permite el reconocimiento de dichos fonemas, los cuales forman la voz. Este hecho de analizar el espectro de la voz, es utilizado en el codificador de canal de voz o más conocido por Vocoder.

El vocoder se ha desarrollado, utilizando la naturaleza de la excitación y el promedio de los niveles de potencia. La mayor dificultad que enfrenta el desarrollo de un el vocoder corresponde a la determinación del pitch (Frecuencia fundamental de los pulsos de excitación que excitan las cuerdas vocales) de la señal de voz.

Un vocoder maneja una serie de valores que hacen de su codificación un proceso eficiente en mayor o menor grado, según sea el caso. Por ejemplo, se

puede tomar el ancho de banda de voz, cercana a los 4000 [Hz] y dividirla en 16 subbandas. Tomando muestras de señal (proveniente de la rectificación y filtración en baja frecuencia de la salida de cada filtro pasabanda) cada 20 [ms] y aplicando el teorema del muestreo, se obtiene que cada canal espectral requiere un ancho de banda (para transmitir su información de potencia de la subbanda) de $1/(2 * 20 \text{ [ms]}) = 25 \text{ [Hz]}$. Luego, el ancho de banda total requerido para transmitir la información de señal de voz es de $16 * 25 \text{ [Hz]} = 400 \text{ [Hz]}$.

Esto significa, en pocas palabras, que se ha reducido el ancho de banda en aproximadamente un 90%. No se debe perder de vista que la señal obtenida en el extremo receptor tendrá solo las características más notorias de la voz codificada, por lo que al escuchar un sonido sintético o artificial no debe ser motivo de sorpresa.

Finalmente, se puede observar que la calidad de la señal recibida en el extremo receptor de un vocoder depende del número de filtros pasabanda que tenga el codificador, el orden de dichos filtros (filtros muy abruptos y de bajo ancho de banda implican la utilización de un mayor número de ellos), la duración de cada muestra, y su respectiva generación (excitación) en el decodificador.

.

Todo sistema de compresión de voz, está constituido por dos partes: la compresión y la descompresión de la voz.

El algoritmo utilizado para el proceso de compresión el cual tienen como objetivo la preservación de la amplitud espectral de señales de voz es el indicado a continuación.

- 1.- Conversión del archivo de voz fuente con extensión .WAV en un vector mediante el software Matlab. Lo que implicaría que se ha realizado el proceso de conversión de la señal analógica de voz a una señal discreta.

En el proceso de conversión análogo/ digital intervienen tres etapas:

Muestreo.- El rango de voz es de 0 a 4Khz, y de acuerdo al teorema de muestreo el sistema, trabaja con frecuencias de muestreo mayores a 8 Khz.

Cuantización.- Se utilizan 16 bits para cuantificarlas, obteniéndose una señal con menor error de cuantización, el cual es aceptable para el trabajo con el equipo.

Codificación.- La codificación utilizada es PCM (modulación por codificación de pulso, ya que el software utilizado solo permite emplear esta codificación (MATLAB).

La función empleada para esta etapa es $[Y, Fs, Nbits] = \text{wavread}(\text{archivo .wav})$, la cual convierte un archivo wav en vector de datos, la función devuelve como resultado Y el vector de datos, Fs la frecuencia de muestreo y Nbits el numero de bits de los datos.

- 2.- Filtraje total del vector de datos, dentro de la banda de frecuencia de la voz humana (200 [Hz] a 400[Hz] aproximadamente). Se utilizó un filtro elíptico pasa bajo de cuarto orden, con un rizado de 0.01 dentro de la banda de paso y de 100 fuera de ella y la frecuencia de corte de 4Khz, la función utilizada es **ellip** (orden del filtro, rizado de la banda de paso(dB), rizado en la banda de rechazo(dB), frecuencia de corte).
- 3.- División del archivo de datos (vector) en subventanas de datos. Se separaron los datos en grupos de vectores, según la velocidad de muestreo y el tiempo de la ventana de datos que se va a analizar, mediante sentencias de control de flujo, de lazos no condicionales.
- 4.- Separación de la señal correspondiente a la subventana de datos en distintas subbandas de frecuencia. Se filtró en forma iterativa cada una de las subventanas de datos mediante filtros pasa banda elípticos de cuarto orden, con frecuencias superior e inferior como parámetros del algoritmo.

Utilizándose un filtro elíptico pasa banda de cuarto orden, con un rizado de 0.01 dentro de la banda de paso y de 100 fuera de ella y las frecuencias de corte que van desde 200 a 4000 Hz, linealmente espaciadas, la función utilizada es **ellip** (orden del filtro, rizado de la banda de paso (dB), rizado en la banda de rechazo (dB), frecuencias de corte).

Los filtros diseñados son pasa banda, como lo que nos interesa es el módulo de la respuesta de frecuencia, se optó por utilizar del tipo IIR (Filtros de respuesta al impulso infinito).

Se selecciono el filtro elíptico, por sus características de respuesta, menor banda de rechazo. Los parámetros del filtro como son el orden, el rizado de la banda de paso, el ancho de banda de la banda de rechazo, se obtuvieron a través de pruebas.

- 5.- Rectificación de la señal filtrada. Luego de obtenerse el filtrado de la señal correspondiente a la ventana de datos para cada subbanda, se procede a rectificar dicha señal obtenida, mediante la función valor absoluto **abs**. Con ello se logra adaptar la señal para la próxima etapa.
- 6.- Cálculo de la potencia de la señal en la sub-banda de frecuencia correspondiente. Después de haberse rectificado la señal a la salida de cada filtro pasabanda, la señal obtenida está en condiciones de analizarse y obtener de ella su valor medio, lo cual se realiza con la función **mean**.
- 7.- Creación del vector resultado con los datos comprimidos. Los datos obtenidos en el paso 6 estos se ordenan en un vector, que pasará a ser el vector resultado con los datos comprimidos. Finalizando esta etapa, se procede generación de un archivo (extensión .voc) con los promedios de cada filtro y de todas las ventanas, utilizando sentencias de control de flujo, en este archivo también se deben almacenar, las frecuencia de muestreo, el intervalo de tiempo de la ventana y número de subbandas.

El algoritmo utilizado para el proceso de descompresión es:

- 1.- Abrir el archivo .voc. Se calcula la cantidad de datos que poseerá el archivo descomprimido y se hacen los ajustes y rellenos necesarios utilizando sentencias de control de flujo.
- 2.- Reconstruir el número de muestras de cada ventana, utilizando una función sinusoidal:

$$Y = V \sin (2\pi fn)$$

Donde: V es el promedio de cada filtro

f es la frecuencia intermedia del filtro

n es el número de muestras de la ventana

Para cada ventana, y para cada promedio del filtro se realiza esta operación, y para reconstruir las muestras de la ventana se suman los valores obtenidos para cada filtro. Se asignan todos los valores obtenidos de todas las sumas a un vector resultado correspondiente al archivo descomprimido (recuperado):

- 3.- Regenerar la señal total, ubicando las muestras de cada ventana una a continuación de otra.
- 4.- Conversión de la señal discreta a analógica, con los parámetros utilizados en la conversión analógico/ digital, empleándose la función **soundsc**(archivo .voc, frecuencia de muestreo), la cual automáticamente realiza un escalamiento.

4.2 ANÁLISIS DE SISTEMAS DISPONIBLES.

El estudio de los vocoder se ha concentrado en la mayoría de los casos para el idioma inglés, alemán y francés, en el idioma español se ha utilizado las técnicas de decimación e interpolación, que permite tanto la compresión como el reconocimiento de voz.

La decimación, consiste en eliminar un cierto número de muestras de la señal de voz discreta. La interpolación reconstruye a la señal mediante el uso de filtros en este caso utiliza filtros FIR (Filtros de respuesta al impulso finito). Esta técnica tiene niveles de comprensión relativamente pequeños y un tiempo de respuesta alto comparado con el vocoder de subbandas, el cual posee un nivel de compresión mucho mayor pero no se puede hacer el reconocimiento de voz.

El sistema Mp3, es un sistema de compresión de audio, desarrollado por el Moving Picture Expert Group (MPEG). Esto significa que este formato corresponde a archivos de sonido que de alguna manera ocupan menos espacio que los otros. Para explicar como funciona la compresión de mp3, primero debemos hablar sobre el sonido y sus propiedades. Los sonidos se mueven en una serie de frecuencias, de las cuales el oído humano no puede apreciar más que las comprendidas entre 20Hz y 20Khz. El formato de Mp3 lo que hace para ahorrar espacio es eliminar todos los sonidos que quedan fuera de esas frecuencias y algunas que percibimos con menor intensidad. De esta manera se consiguen **compresiones 12:1** o mayores, lo que significa que una canción en mp3 es 12 veces menor que una canción en formato wav.

Además, se utiliza otra técnica que reduce la cantidad de datos en el archivo: cuando un sonido alcanza la misma intensidad en los dos canales del estéreo, el mp3 lo guarda como uno sólo, pero lo reproduce por los dos.

El formato mp3 utiliza dos parámetros variables: el número de hercios (Hz) y el de bits por segundo (bps). Los hercios miden el número de muestras de audio que se toman por segundo. El número de bits representa la precisión con la que se toman las muestras, que, cuanto mayor sea, más precisa será la conversión. El máximo que puede alcanzarse es de 320 Kbps, dando de esta manera archivos más grandes, pero de una calidad óptima.

Para comprimir al formato mp3, existen diferentes codificadores que son algoritmos, que pueden dar archivos del mismo tamaño, pero de diferentes

calidades. Cuanto más rápido sea un codificador al comprimir la información, menor será la calidad del archivo generado.

En la actualidad para el idioma español, la Escuela Politécnica del Ejército es la pionera a nivel nacional. Estudios similares en Sudamérica, se están realizando únicamente en la república de Chile.

4.3 DESARROLLO DE SOFTWARE.

En el mercado informático, hay muchos paquetes computacionales, para la realización del Vocoder en Subbandas, como C, C++, Visual Basic, Visual C++ y MATLAB, se ha optado por el software de MATLAB.

MATLAB, es un lenguaje de alto nivel, basado en matrices, en continuo crecimiento y muy adaptable a los avances científicos, que resuelve los problemas que presenta la ingeniería en el desarrollo de productos innovadores. En el campo de las comunicaciones, MATLAB permite realizar modelado y diseño de sistemas DSP.

Su entorno básico se complementa en una amplia colección de toolboxes que contiene funciones específicas para determinadas aplicaciones en diferentes ramas de la ciencia y de la ingeniería.

Dentro de los toolboxes, disponibles en MATLAB, los de particular interés para este proyecto son:

- Signal processing Toolbox
- Filter Design Toolbox

MATLAB cuenta con un constructor de interfase de usuario GUI (guide), con objetos gráficos como botones, cuadros de texto, deslizadores, botones de selección, menús, etc..

Debido a las características indicadas anteriormente, la implementación del algoritmo del Vocoder de Subbandas ha sido fácil, flexible, compacto, si se hubiera utilizado otro paquete computacional.

El código del programa se adjunta en anexo A y el manual de operación en el anexo B

4.4 REQUISITOS DEL SISTEMA.

El Vocoder de Subbandas, requiere para su funcionamiento las siguientes especificaciones:

Hardware

- PC- compatible con microprocesador Pentium para trabajar sobre Windows 95, 98, Me, NT, XP o Windows 2000.
- Tarjeta de sonido SoundMAX Digital Audio.
- Tarjeta gráfica VGA y monitor a color.
- 64 Mbytes de memoria RAM.
- Disco duro con un espacio de 1 Gbyte.
- Mouse y unidad de CD-ROM.
- Sistemas de Multimedia.

Software

- MATLAB, versión 6.5.

CAPÍTULO V

ANÁLISIS DE PRUEBAS Y RESULTADOS

5.1 APLICACIÓN DE PRUEBAS

5.1.1 PRUEBA DE FUNCIONAMIENTO DEL VOCODER DE SUBBANDAS

Para determinar el óptimo funcionamiento del Vocoder en Subbandas, se realizaron e implementaron pruebas con personas de los dos sexos y de diferentes edades, de cada uno de ellos se obtuvo un archivo .wav, el cual se procesó en el vocoder con diferentes niveles de compresión.

5.1.2 PRUEBAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS

Con la finalidad de certificar el funcionamiento del vocoder se realizó una encuesta, a varias personas. El instrumento preparado esta apoyado en el Test Internacional de procesamiento de voz, conocido como MOS TEST.

El MOS TEST es una recomendación de ITU P. 800; una prueba MOS realiza comparaciones entre la voz transmitida, que es la referencia de la prueba, con la voz recibida y produce una valoración numérica de 1(malo) a 5 (excelente), esta medida relaciona la calidad del oyente que percibe la voz.

5.1.2.1 PROCEDIMIENTO PARA LA ELABORACIÓN DE UN INSTRUMENTO DE RECOLECCIÓN DE DATOS.

Escuche un mensaje de voz con varios niveles de comprensión que proporciona el software a su disposición. Proceda a calificar la calidad del mensaje de acuerdo a la valoración indicada en la tabla 5.1.

Tabla 5.1 Encuesta MOS TEST

CALIFICACIÓN	DEFINICIÓN	DESCRIPCIÓN
5	Excelente	Por ejemplo, una perfecta recepción de radio
4	Bueno	Por ejemplo, la calidad de una llamada de larga distancia.
3	Regular	Usted requiere algún esfuerzo del oído, para entender el mensaje.
2	Pobre	Por ejemplo, una comunicación entre cortada (difícil de entender el

		mensaje de voz).
1	Malo	No se puede entender el mensaje de voz.

PROCEDIMIENTO

1. Escuche cuidadosamente el mensaje de voz de la prueba (aproximadamente de 10 a 15 segundos cada uno) con un audífono y de su opinión de 5 a 1.
2. Haga su mejor esfuerzo para dar su opinión de la forma más objetiva durante toda la prueba. Su opinión debe reflejar la calidad del mensaje de voz en forma global, o su impresión basada en la claridad y grado de distorsión. Usted puede escuchar muchas veces el mensaje antes de emitir su opinión si lo desea.
3. Repita el procedimiento con los diferentes mensajes de prueba.

5.2 ANÁLISIS DE RESULTADOS

Las pruebas se realizaron en la Escuela Politécnica del Ejército sede Latacunga, ya que cuentan con alumnos principalmente de la región central del país

5.2.1 ANÁLISIS DEL FUNCIONAMIENTO DEL VOCODER.

Se han realizado 30 pruebas para determinar el funcionamiento del Vocoder con personal de la Escuela Politécnica del Ejército sede Latacunga y niños de la ciudad de Latacunga. Distribuida de acuerdo a lo indicado en la tabla 5.2.

Tabla 5.2 Pruebas de funcionamiento del vocoder.

EDAD	PRUEBAS	SEXO	
		HOMBRES	MUJERES
Menores de 12 años	10	5	5
De 12 a 24 años	10	5	5
De 24 a 40 años	10	5	5

TOTAL	30	15	15
--------------	----	----	----

De las pruebas realizadas se comprobó que el vocoder funciona adecuadamente con personas hombres y mujeres comprendidas entre una edad de 12 a 40 años, mientras que para personas menores a 12 años tanto hombres como mujeres funciona no adecuadamente, este fenómeno se produce ya que las personas menores a 12 años el sistema tracto vocal está en pleno desarrollo por lo que la voz de estas personas presentan muchas variaciones en el espectro de frecuencia.

5.2.2 ANÁLISIS DE LOS DATOS RECOLECTADOS.

Cada persona encuestada realizó ocho evaluaciones correspondiendo a los mensajes de voz, con las propiedades indicadas en la tabla 5.3.

Tabla 5.3 Parámetros para las encuestas, para una frecuencia de muestreo de 22050 Hz. y una archivo. Wav de 186 Kbytes..

PRUEBA	DURACION DE LA VENTANA	BANDAS DE FRECUENCIA	NIVEL DE COMPRENSIÓN
1	20 mseg.	5	176:1
2	20 mseg.	10	88:1
3	20 mseg.	15	59:1
4	20 mseg.	20	44:1
5	20 mseg.	40	22:1
6	20 mseg.	60	15:1
7	20 mseg.	80	11:1
8	20 mseg.	100	9:1

La encuesta indicada anteriormente con los diferentes parámetros, se realizaron a 30 personas al azar de la Escuela Politécnica del Ejército sede Latacunga. Los resultados se indican en la tabla 5.4.

Tabla 5.4. Resultados de la prueba MOS TEST.

PRUEBA 1

EVALUACION	NÚMERO DE PERSONAS
5	0
4	0
3	0
2	0
1	30

PRUEBA 2

EVALUACION	NÚMERO DE PERSONAS
5	0
4	0
3	0
2	5
1	25

PRUEBA 3

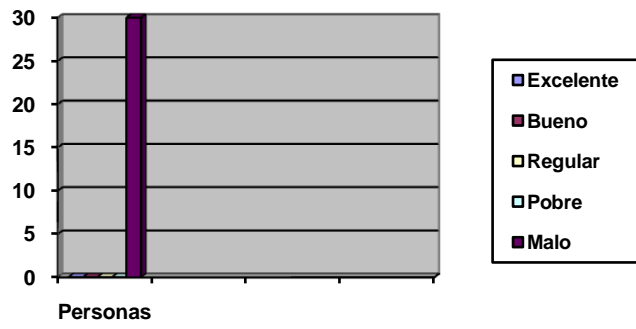
EVALUACION	NÚMERO DE PERSONAS
5	0
4	0
3	3
2	5
1	22

PRUEBA 4

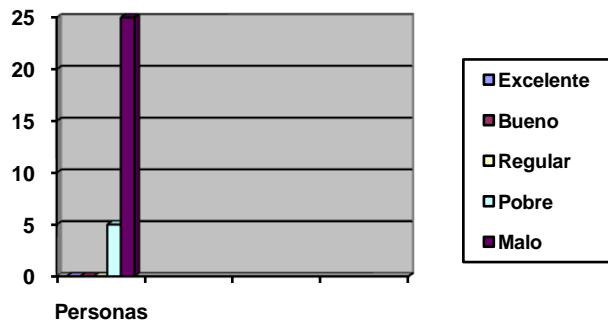
EVALUACION	NÚMERO DE PERSONAS
5	5
4	22
3	2
2	1
1	0

PRUEBA 5	
EVALUACION	NÚMERO DE PERSONAS
5	24
4	5
3	1
2	0
1	0
PRUEBA 6	
EVALUACION	NÚMERO DE PERSONAS
5	30
4	0
3	0
2	0
1	0
PRUEBA 7	
EVALUACION	NÚMERO DE PERSONAS
5	30
4	0
3	0
2	0
1	0
PRUEBA 8	
EVALUACION	NÚMERO DE PERSONAS
5	30
4	0
3	0
2	0
1	0

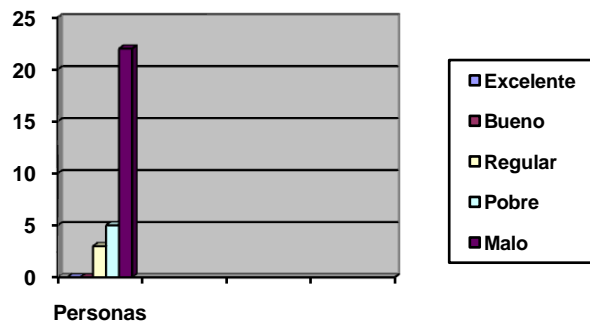
La representación de los resultados en forma gráfica se indican en la figura 5.1.



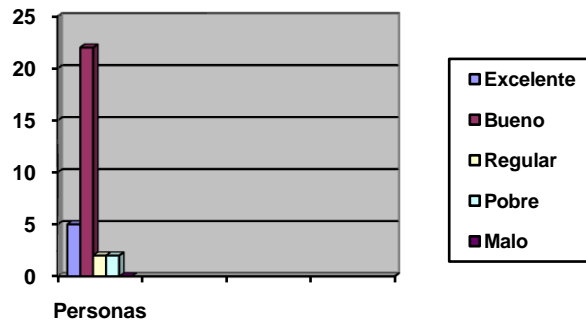
a)



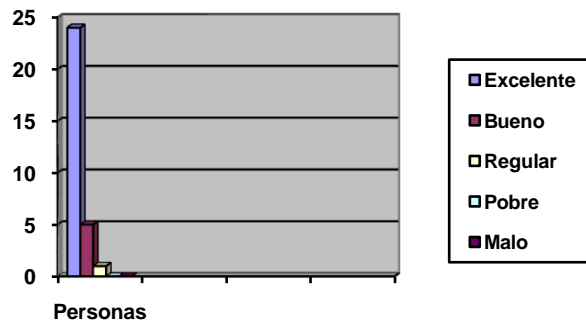
b)



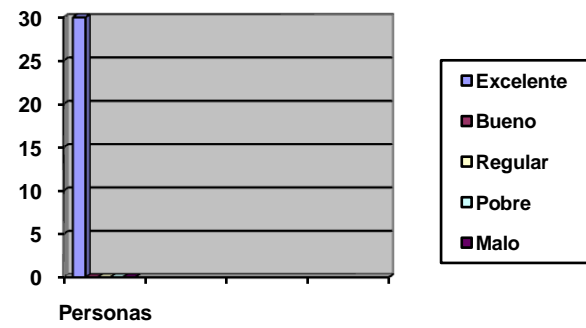
c)



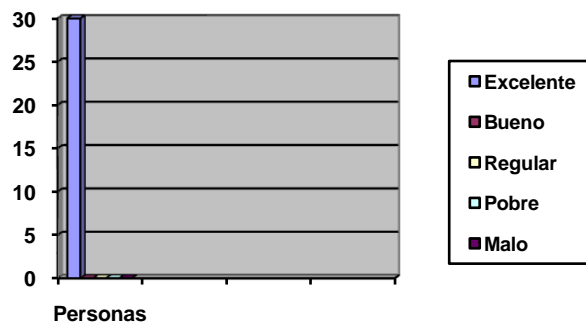
d)



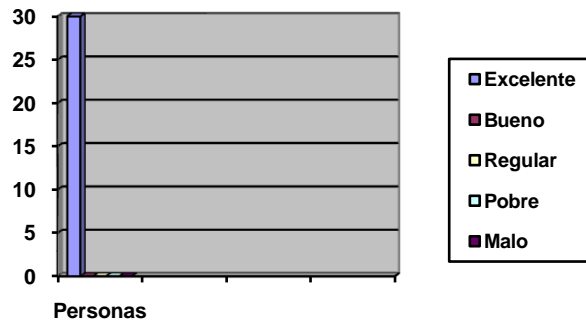
e)



f)



g)



h)

Figura 5.1 Resultado de las pruebas: a) prueba 1, b) prueba 2, c) prueba 3, d) prueba 4, e) prueba 5, f) prueba 6, g) prueba 7, h) prueba 8.

A continuación se describen los resultados obtenidos, mediante el algoritmo de compresión de voz de subbandas de frecuencia :

La prueba 1 tiene una aceptación del 0 % para la evaluación 1 que corresponde a malo.

La prueba 2 tiene una aceptación del 16.6 % para la evaluación 2 que corresponde a pobre y 83.33% para la evaluación 1 correspondiente a malo.

La prueba 3 tiene una aceptación del 10 % para la evaluación 3 que corresponde a regular, para la evaluación de 2 un porcentaje de 16.66 correspondiente a pobre y 73.33 % para la evaluación 1 que equivale a malo.

La prueba 4 tiene una aceptación del 16.66 % para la evaluación 5 que corresponde a excelente, para la evaluación 4 un porcentaje de 73.33% que es bueno, para 3 un porcentaje de 6,66% equivalente a regular y para 2 un porcentaje de 2,82 % correspondiente a pobre.

La prueba 5 tiene una aceptación del 80 % para la evaluación igual a 5 que corresponde a excelente, 16,66% para 4 correspondiente a bueno y 3.33% para la evaluación 3 correspondiente a regular.

La prueba 6 tiene una aceptación del 100 % para la evaluación 5 que corresponde a excelente.

La prueba 7 tiene una aceptación del 100 % para la evaluación 5 que corresponde a excelente.

La prueba 8 tiene una aceptación del 100 % para la evaluación 5 que corresponde a excelente.

Para hacer más didáctico el análisis, este se realizará desde el punto de vista gráfico.

En la figura 5.1, el gráfico superior corresponde a la forma de onda en el tiempo del vector de datos original, este vector fue muestreado a 22050 Hz, con 16 bits de cuantización y codificación PCM, el gráfico inferior es el vector de datos descomprimido, con un nivel de compresión de 44:1.

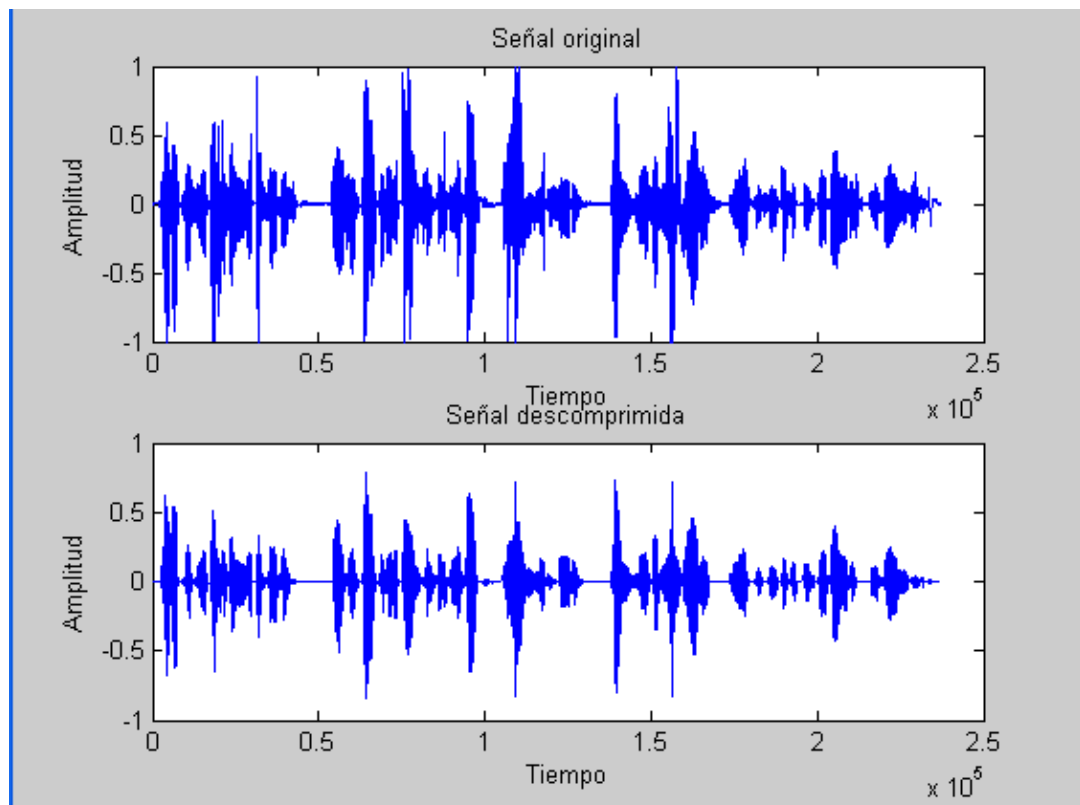


Figura 5.1. Análisis gráfico

Para analizar los resultados obtenidos, se realizaron pruebas de compresión de voz con el programa NERO, el cual permite compresión al nivel de Mp3 y su nueva versión Mp4.

Las pruebas se ejecutaron bajo las mismas condiciones, se obtuvieron niveles de compresión similares, la calidad de la reproducción es mejor con Mp4, debido a que utiliza un conjunto de algoritmos mucho más elaborados.

CAPÍTULO VI

CONCLUSIONES Y RECOMENDACIONES

En este capítulo se presentan las conclusiones y recomendaciones obtenidas del trabajo realizado, cabe indicar que éstas son resultado del análisis realizado de las pruebas con el Vocoder de Subbandas.

6.1 CONCLUSIONES

- Se implemento un vocoder para el idioma catellano.

- Una vez que se implementó el vocoder en Subbandas, tomando como base el software MATLAB versión 6.5, se realizaron diferentes pruebas para analizar la influencia de este en la comprensión y descomprensión de voz para el idioma castellano.
- El Vocoder en Subbandas opera en una forma eficiente para personas adultas de ambos sexos, con un nivel de comprensión de 44:1(para un archivo .wav de 186 Kbytes de tamaño, muestreado a una frecuencia de 22050 Hz.), obteniéndose como resultado una voz generada de tipo metalizada o de robot.
- El Vocoder en Subbandas, tiene un nivel de compresión alto con respecto a otros sistemas similares, pero no permite realizar el reconocimiento de voz.
- El Vocoder en Subbandas para personas menores de 12 años de ambos sexos, da resultados deficientes pobres e incluso en la mayoría de los casos no se puede entender el mensaje de voz, considerando que las pruebas realizadas fueron igual a las ejecutadas para personas adultas, debido a que su sistema vocal esta en crecimiento y su pitch, no esta definido.
- Se ha diseñado e implementado, un Vocoder en Subbandas usando el software de MATLAB, con una interfase de usuario basada en los programas para Windows, lo que hace que el programa sea muy amigable para el usuario y por lo tanto pueda usarse sin dificultades en aspectos pedagógicos.
- Para la implantación del algoritmo correspondiente al Vocoder se hizo un estudio del software de MATLAB, utilizando funciones básicas, funciones de las herramientas de procesamiento de señales y diseño de filtros, lo que facilito la implementación. Para la interfase gráfica se aplico la

programación grafica disponible en este paquete. La programación grafica es más compleja que las que se puede realizar con otros paquetes computacionales de programación visual. Por todo lo indicado se justifica planamente el uso del MATLAB para la implementación del algoritmo del Vocoder.

- La frecuencia de muestreo tiene un efecto sobre el nivel de comprensión y la calidad de voz, son directamente proporcionales, siendo la frecuencia de muestreo optima 22050 Hz, como lo demuestra los resultados obtenidos de las pruebas realizadas.
- En la implantación del vocoder se aplicó las técnicas básicas de procesamiento digital de señales, como son: conversión analoga-digital algoritmos de compresión y descompresión, filtrado digital y conversión digital-analógica, demostrando que el marco teórico, constituye la base fundamental e imprescindible para el desarrollo de este trabajo de investigación.
- El algoritmo propuesto para el desarrollo del vocoder cumplió con los objetivos propuestos.
- En la estación terrena de ANDINATEL para el sistema DOMSAT, requieren de la transmisión de voz vía IP, para disminuir los tiempo de transmisión y el ancho de banda, si además de esto se realizaría una comprensión de voz estos parámetros se minimizarían aún más. El Vocoder en Subbandas puede cubrir este requerimiento.

6.2 RECOMENDACIONES:

- Realizar trabajos de investigación relacionados con el PROCESAMIENTO Digital de Señales, porque en la actualidad está en pleno desarrollo.

- Se recomienda crear variantes del software desarrollado, a fin de que pueda ser utilizado en diferentes plataformas tecnológicas.
- Utilizar el software en aplicaciones en las que no sea imprescindible el reconocimiento de voz.
- Adquirir bibliografía sobre la temática de procesamiento digital, por cuanto no existe en la mayoría de Universidades y Politécnicas del Ecuador.
- Difundir el presente trabajo a fin de que se continúen realizando investigaciones similares.
- Utilizar el software en sistemas de comunicación para optimizar el ancho de banda y minimizar los tiempos de transmisión.

ANEXO A

MANUAL DE OPERACIÓN

INSTALACION DEL SOFWARE

Grabe la carpeta Vocoder en el directorio work de MATLAB.

Antes de correr el programa Vocoder, de la opción file de MATLAB, seleccione la opción **Set path** y añada la carpeta vocoder al directorio de MATLAB.

INSTALACIÓN DEL HARWARE

Conecte en la entrada de audio de su PC un micrófono y en el de salida los parlantes.

INSTRUCCIONES DE USO.

Para ejecutar el programa vocoder en la ventana de comandos de MATLAB, digite vocoder presione **Enter**, lo que hará que aparezca la ventana indicada en la figura 1.

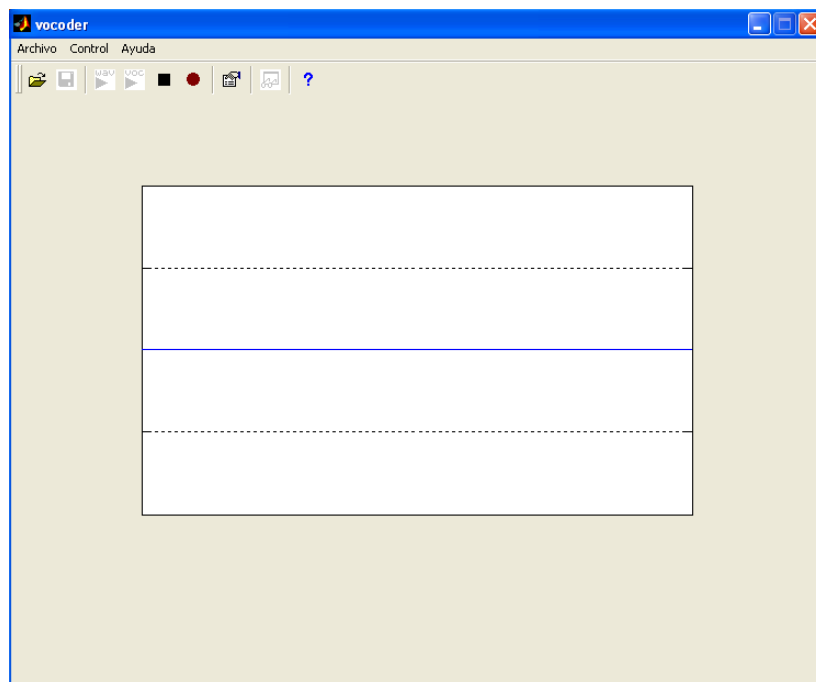


Figura 1. Ventana principal

En la parte superior de la ventana principal, se presenta el menú de la aplicación, las opciones son: Archivo, Control y Ayuda, las cuales a su vez contienen submenús.

La aplicación presenta una barra de herramientas, en la que se encuentran los botones que ejecutan las acciones más comúnmente usadas en una forma rápida.

Como se observa en la figura 1 las únicas opciones disponibles son: **Abrir** un archivo ya sea en el formato .wav o .voc, y grabar un archivo en formato **.wav**, el cual se procesará. En las figuras 2-10 siguientes se indican las diferentes opciones de cada menú:

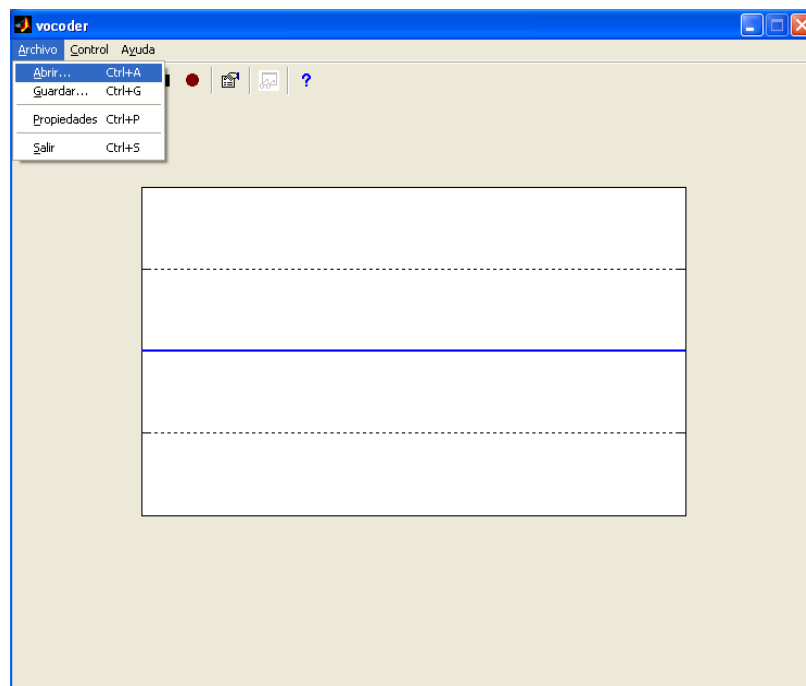


Figura 2. Opciones del menú Archivo.

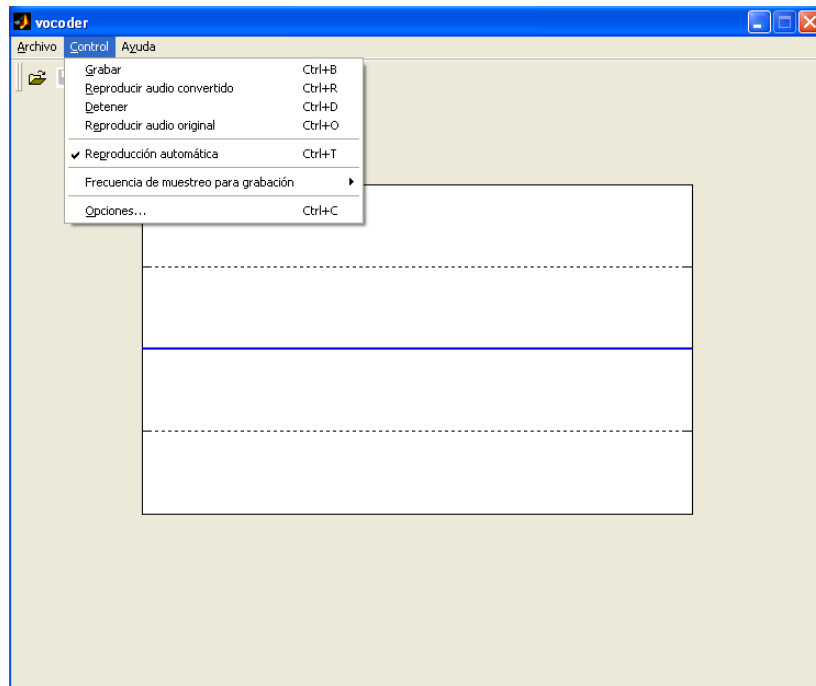


Figura 3. Opciones del menú Control

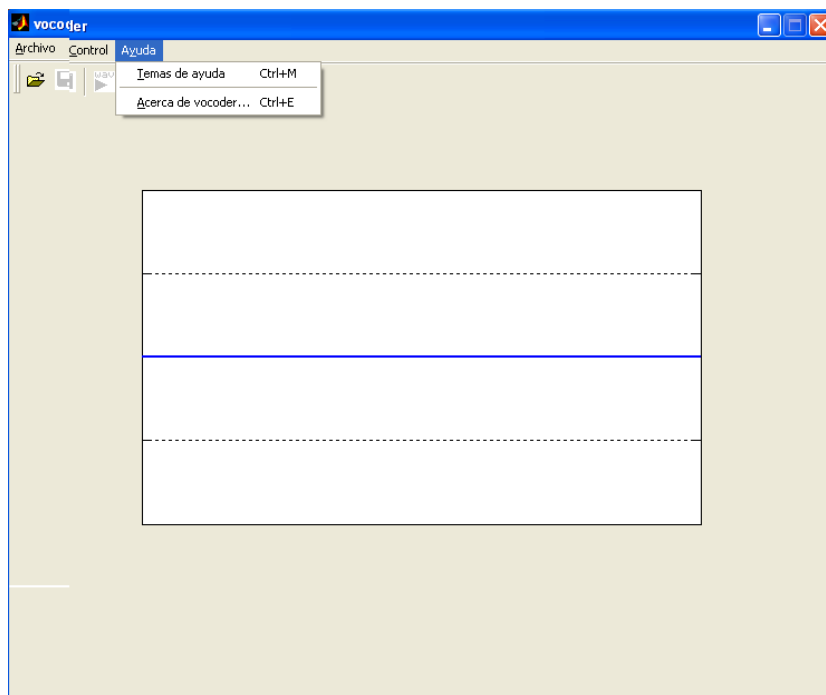


Figura 4. Opciones del menú Ayuda.

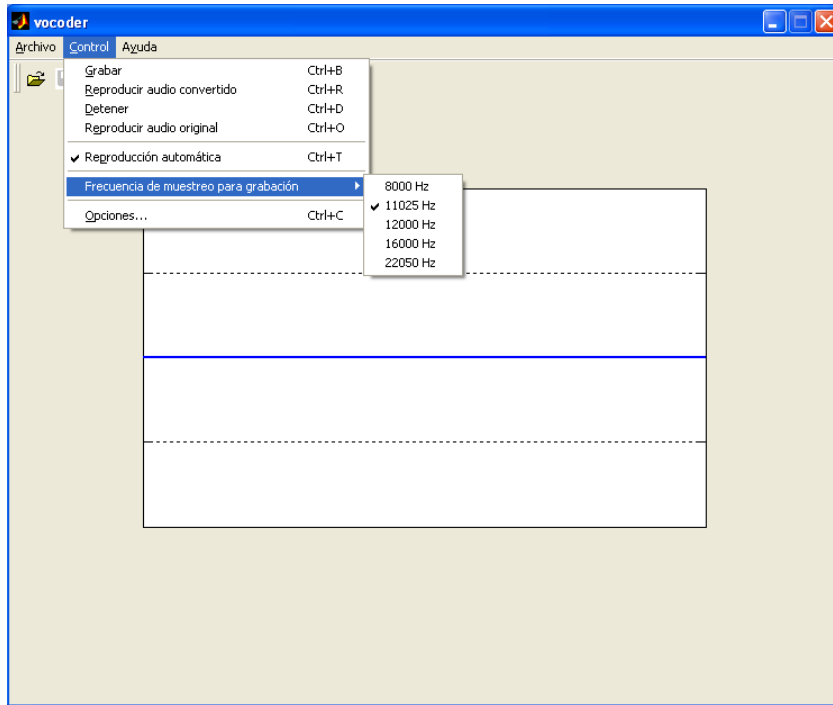


Figura 5. Frecuencias de muestreo para grabación.

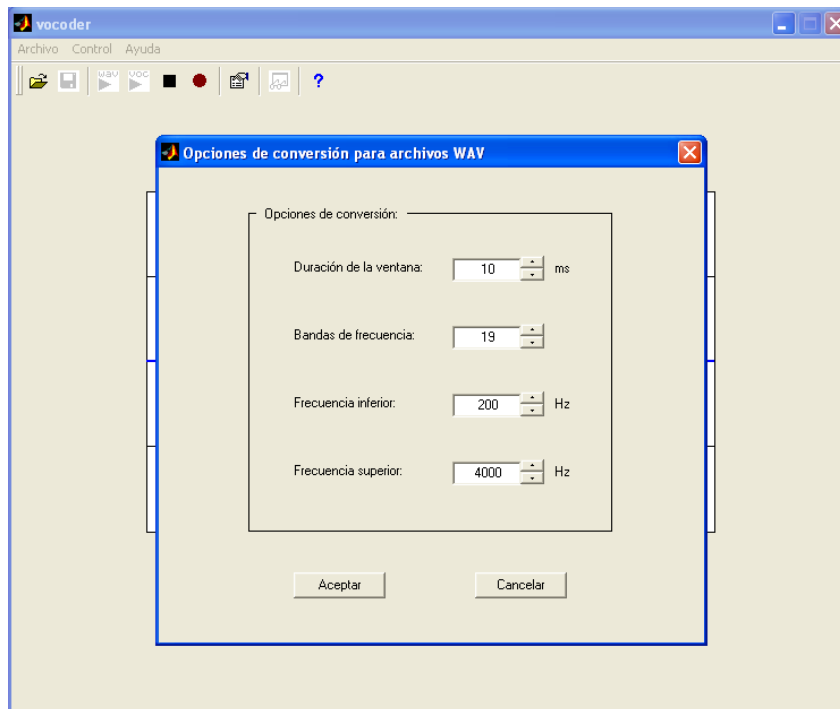


Figura 6. Opciones de conversión para archivos WAV.

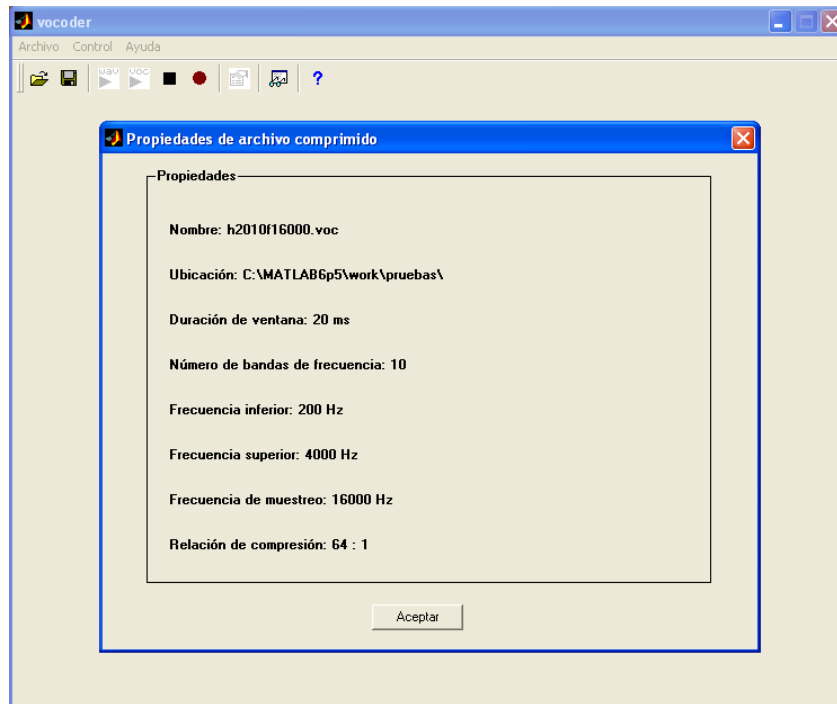


Figura 7. Propiedades del archivo comprimido.

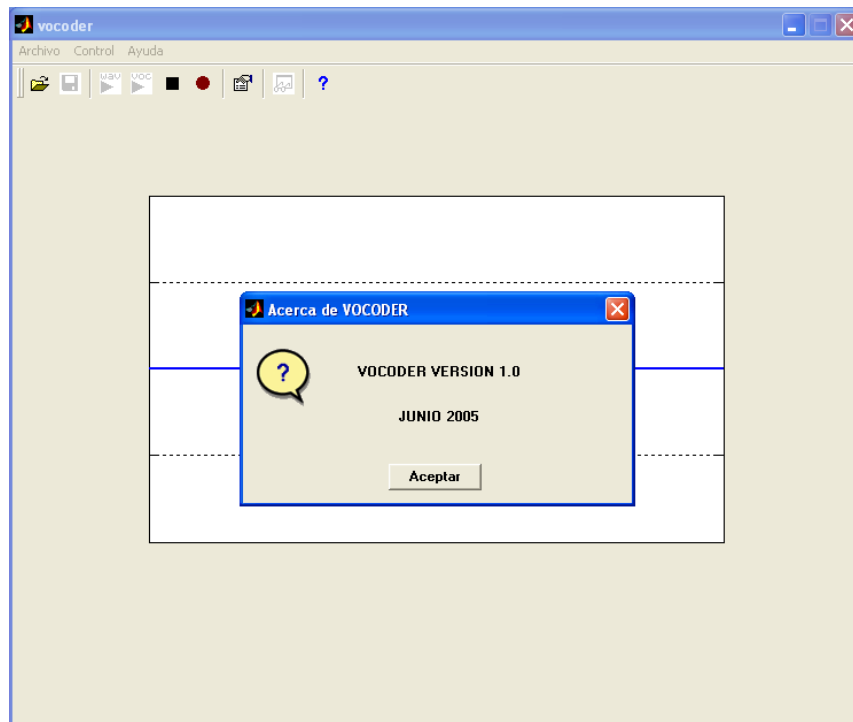


Figura 8. Información del programa.

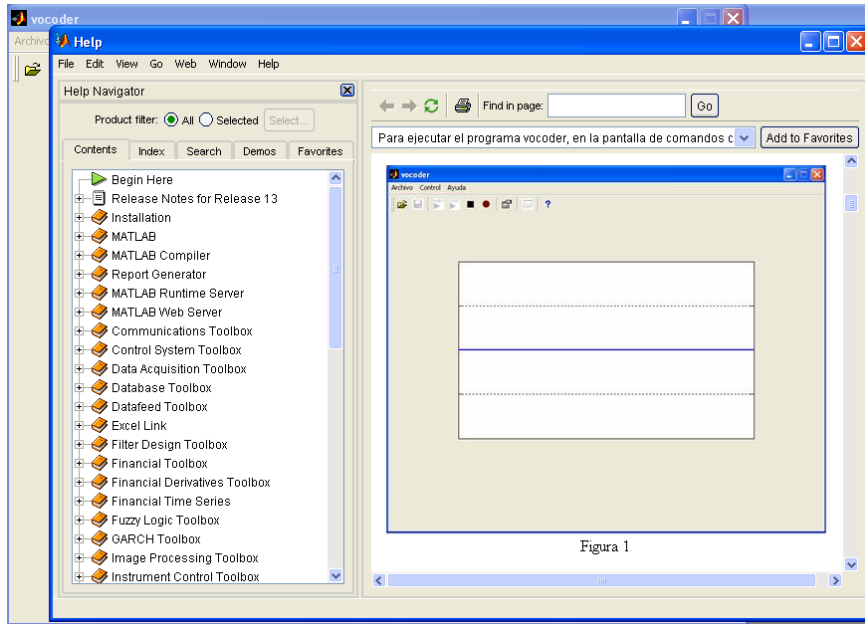


Figura 9. Ayuda.

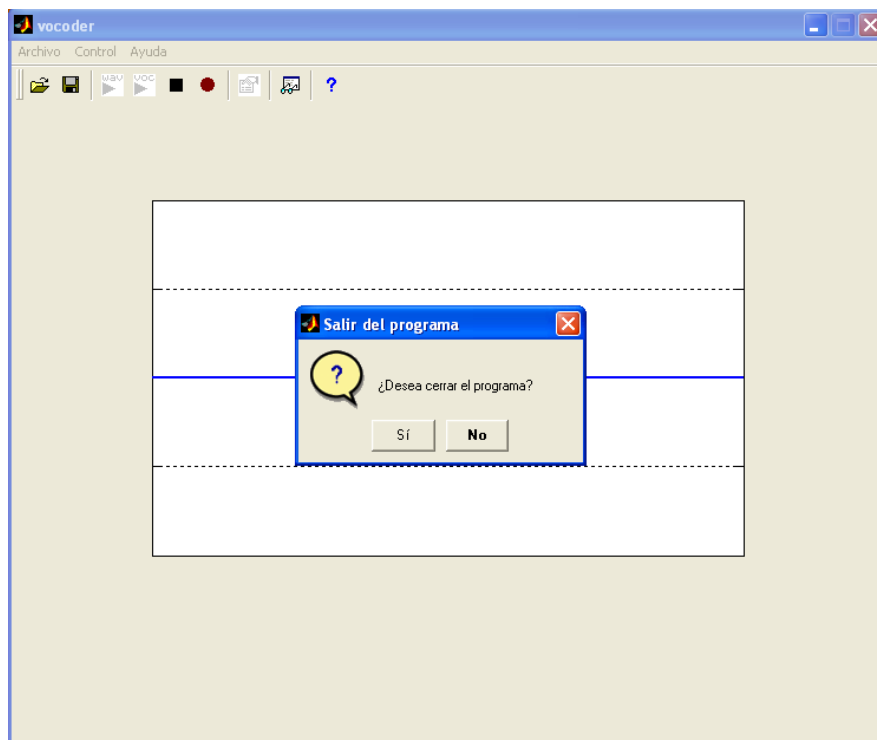


Figura 10. Salir del programa

Las ventanas indicadas pueden ser accedidas a través de los menús, desde los botones de la barra de herramienta, tecleando **ALT** y la letra subrayada o los aceleradores presionando **Control** y la letra asignada para cada caso.

APLICACIÓN CON UN ARCHIVO FORMATO WAV. PREVIAMENTE ALMACENADO

1. Del menú Archivo seleccione **abrir**, como se indica en la figura 11.

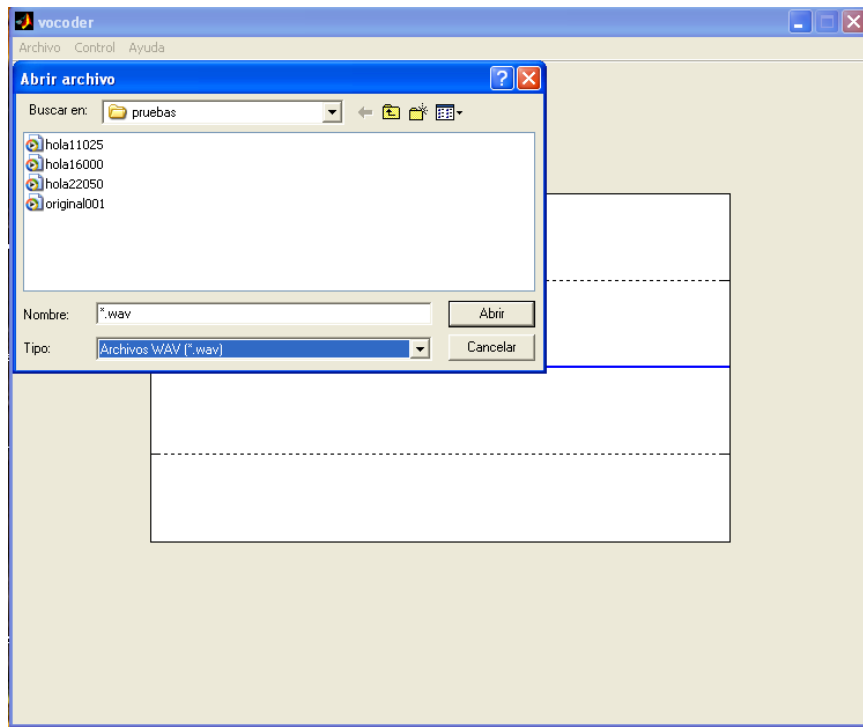


Figura 11. Abrir archivo

2. Automáticamente el archivo se procesa, con las propiedades previamente almacenados en **opciones**.

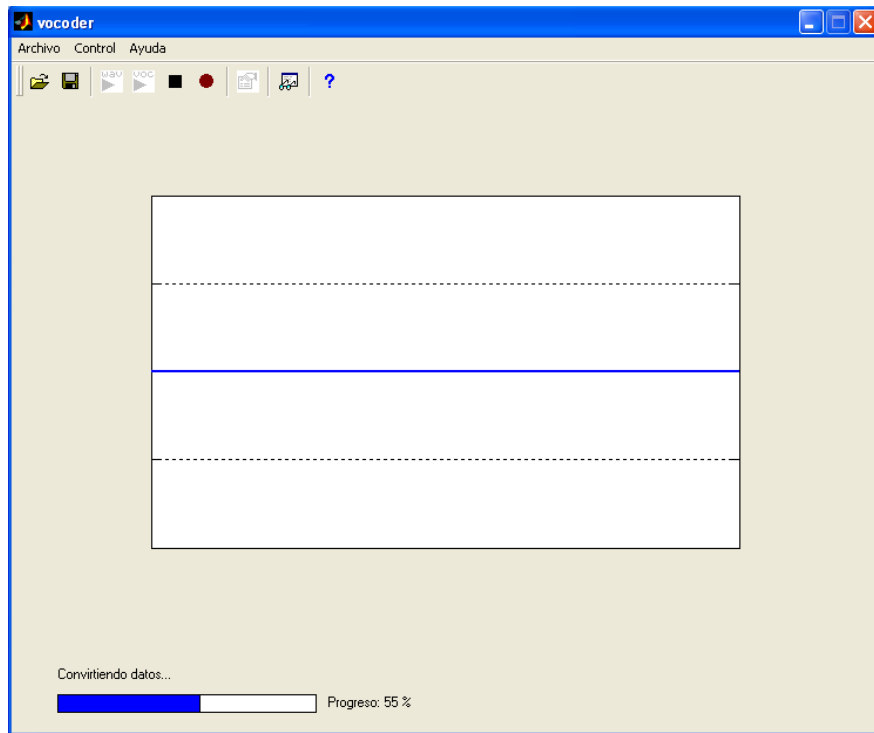


Figura 12. Compresión del archivo

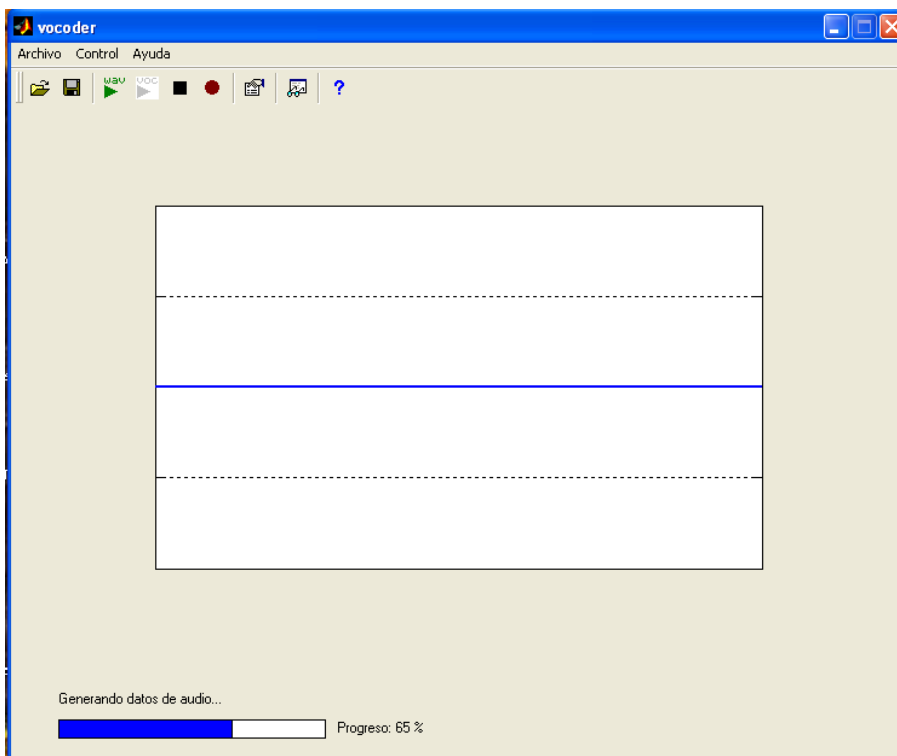


Figura 13. Descompresión del archivo

Con la comprensión se crea el archivo **.voc** el cual queda presente en la memoria RAM que el PC asigna para esta aplicación con el archivo **wav**.

Si desea cambiar las propiedades del archivo, mediante Opciones cambie estos. El archivo se procesará automáticamente con esas nuevas propiedades como se indicó en las figuras 12 y 13.

3. Si desea ver las propiedades del archivo procesado seleccione la opción propiedades y aparecerá una ventana indicándole la duración de la ventana, el número de subbandas, el nivel de compresión, etc, como se indicó en la figura 7.
4. Ud. puede escuchar, el archivo wav o el archivo voc seleccionando la opción respectiva como se indica en las figuras 14 y 15.

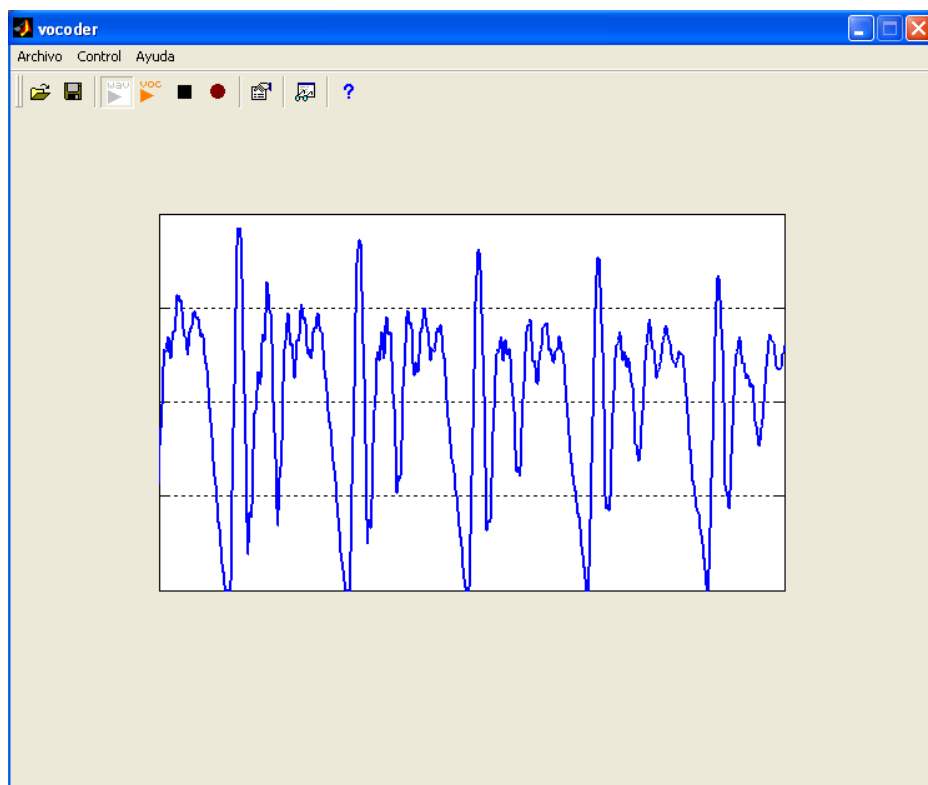


Figura 14. Reproducción del archivo WAV

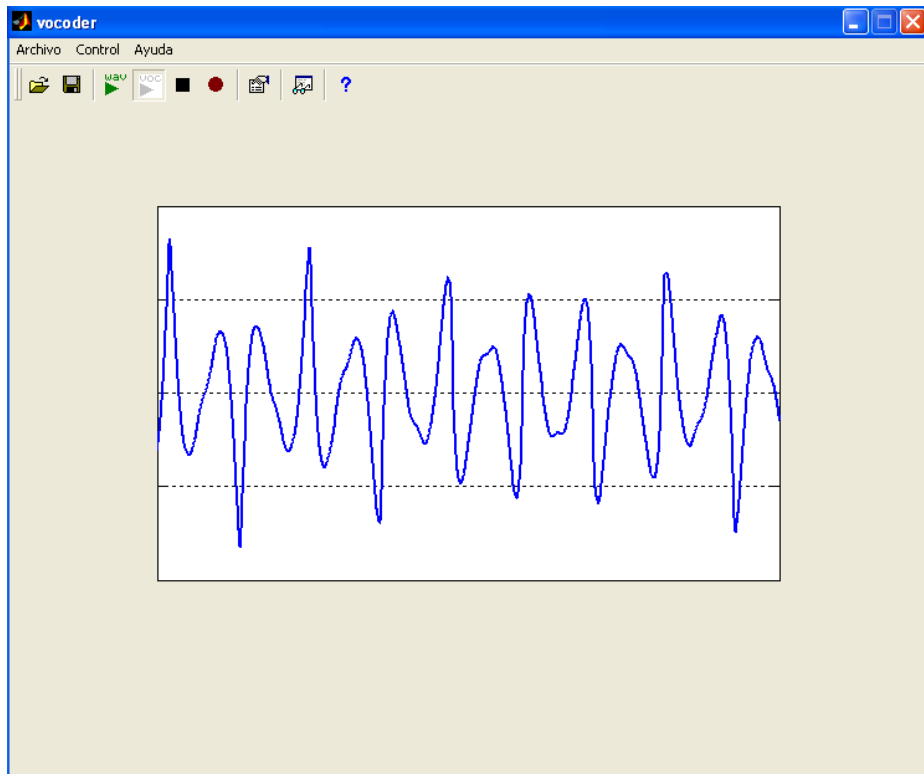


Figura 15.Reproducción del archivo Voc.

5. Usted puede repetir los procesos 3, 4 y 5 hasta que obtenga el resultado que desee.
6. Si desea grabar, los archivo en formato wav y vov seleccione las opciones respectivas como se indica en las figuras 16 y 17.

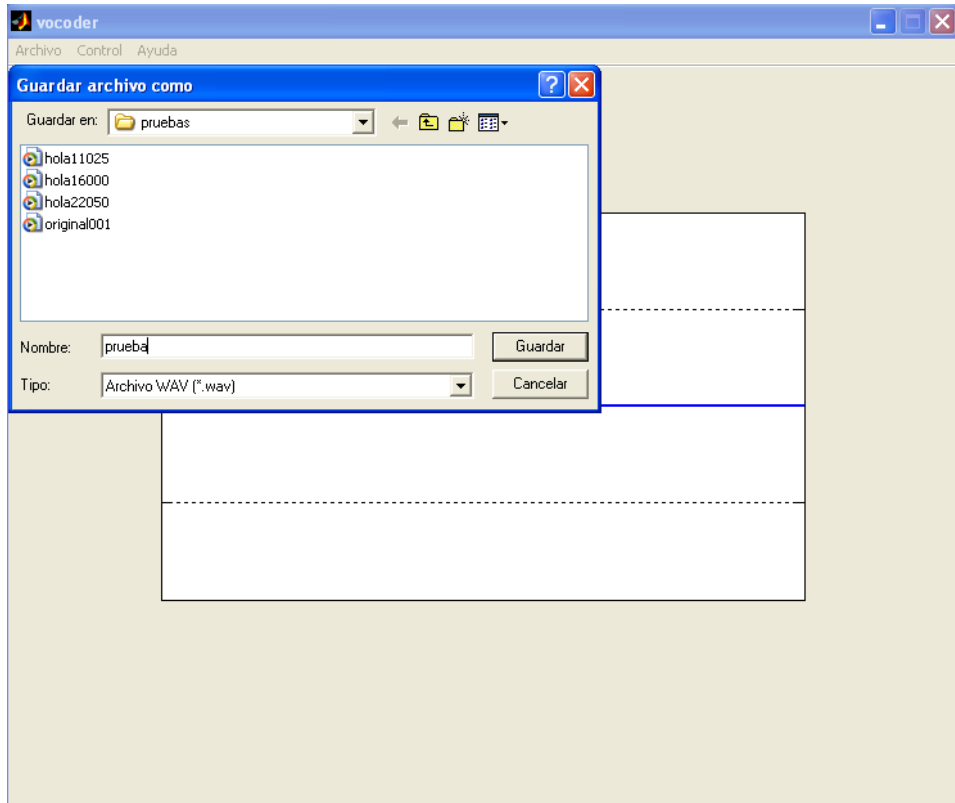


Figura 16. Guardar archivo Wav

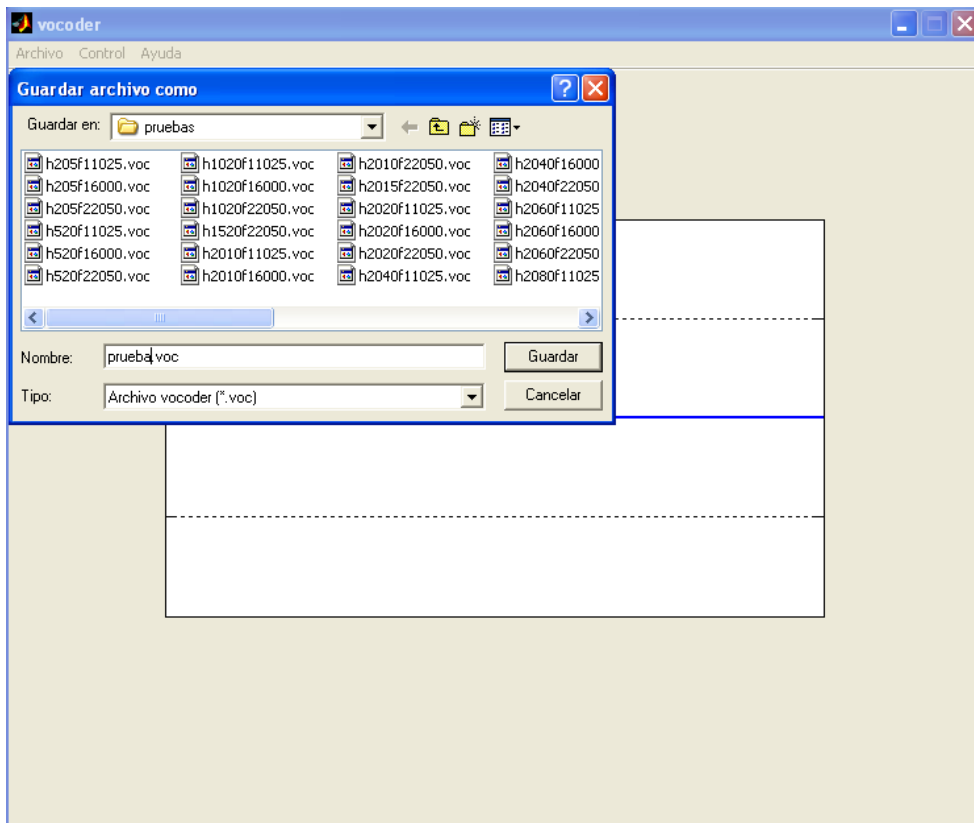


Figura 17. Guardar archivo Voc.

7. Si desea procesar otros archivos proceda como se indico en los pasos anteriores, caso contrario seleccione la opción salir.

En forma predefinida la opción de reproducción automática está activada la cual permite escuchar el archivo procesado una vez que éste ha sido tratado. Si desea hacerlo manualmente desactive esta opción.

APLICACIÓN CUANDO SE GRABA UN ARCHIVO PARA SU TRATAMIENTO.

1. Seleccione la frecuencia de muestreo como se indico en la figura 5.
2. Active la opción grabar y empiece su mensaje de voz utilizando el micrófono, cuando quiera finalizar la grabación active la opción detener.
3. Repita los procedimientos anteriores desde el numeral 2 de la aplicación anterior.

APLICACIÓN CUANDO SE ABRE UN ARCHIVO VOC.

1. Utilize la opción abrir, para recuperar un archivo voc como se indica en la figura 18.

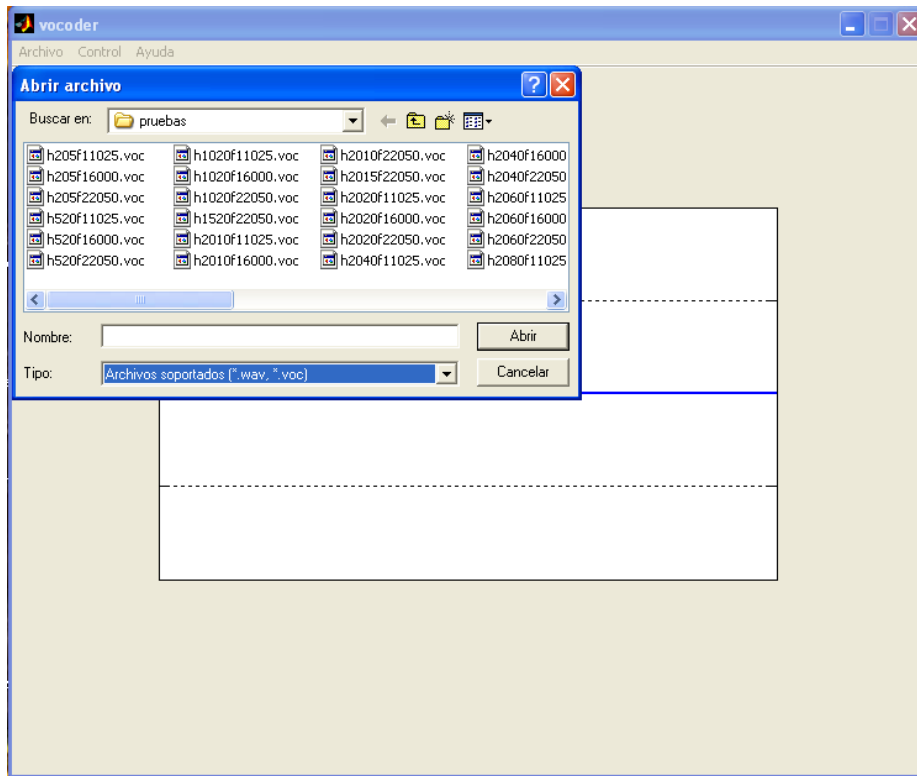


Figura 18. Abrir archivo voc.

2. Con este archivo las únicas posibilidades disponibles son reproducirlo, verificar las propiedades y guardar el archivo.

ANEXO B

CÓDIGO DEL PROGRAMA DEL VOCODER EN SUBBANDAS

```
function varargout = vocoder(varargin)
% VOCODER M-file for vocoder.fig
%   VOCODER, by itself, creates a new VOCODER or raises the existing
%   singleton*.
%
%   H = VOCODER returns the handle to a new VOCODER or the handle to
%   the existing singleton*.
%
%   VOCODER('CALLBACK',hObject,eventData,handles,...) calls the local
%   function named CALLBACK in VOCODER.M with the given input arguments.
%
%   VOCODER('Property','Value',...) creates a new VOCODER or raises the
%   existing singleton*. Starting from the left, property value pairs are
%   applied to the GUI before vocoder_OpeningFunction gets called. An
%   unrecognized property name or invalid value makes property application
%   stop. All inputs are passed to vocoder_OpeningFcn via varargin.
%
%   *See GUI Options on GUIDE's Tools menu. Choose "GUI allows only one
%   instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help vocoder

% Last Modified by GUIDE v2.5 20-Jun-2005 09:55:13

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',    mfilename, ...
                  'gui_Singleton', gui_Singleton, ...
                  'gui_OpeningFcn', @vocoder_OpeningFcn, ...
                  'gui_OutputFcn', @vocoder_OutputFcn, ...
                  'gui_LayoutFcn', [] , ...
                  'gui_Callback', []);
if nargin & isstr(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end

if nargout
    [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT

% --- Executes just before vocoder is made visible.
function vocoder_OpeningFcn(hObject, eventdata, handles, varargin)
```

```

% This function has no output args, see OutputFcn.
% hObject   handle to figure
% eventdata reserved - to be defined in a future version of MATLAB
% handles   structure with handles and user data (see GUIDATA)
% varargin  command line arguments to vocoder (see VARARGIN)

% Determine the position of the dialog - centered on the callback figure
% if available, else, centered on the screen
FigPos=get(0,'DefaultFigurePosition');
OldUnits = get(hObject, 'Units');
set(hObject, 'Units', 'pixels');
OldPos = get(hObject,'Position');
FigWidth = OldPos(3);
FigHeight = OldPos(4);
if isempty(gcf)
    ScreenUnits=get(0,'Units');
    set(0,'Units','pixels');
    ScreenSize=get(0,'ScreenSize');
    set(0,'Units',ScreenUnits);

    FigPos(1)=1/2*(ScreenSize(3)-FigWidth);
    FigPos(2)=2/3*(ScreenSize(4)-FigHeight);
else
    GCBFOldUnits = get(gcf,'Units');
    set(gcf,'Units','pixels');
    GCBFPos = get(gcf,'Position');
    set(gcf,'Units',GCBFOldUnits);
    FigPos(1:2) = [(GCBFPos(1) + GCBFPos(3) / 2) - FigWidth / 2, ...
                  (GCBFPos(2) + GCBFPos(4) / 2) - FigHeight / 2];
end
FigPos(3:4)=[FigWidth FigHeight];
set(hObject, 'Position', FigPos);
set(hObject, 'Units', OldUnits);

inicializar1(handles)

handles = guihandles(hObject);
% Choose default command line output for vocoder
handles.output = hObject;

% Update handles structure
guidata(hObject, handles);

% UIWAIT makes vocoder wait for user response (see UIRESUME)
% uiwait(handles.figure1);

inicializar2(handles)

% --- Outputs from this function are returned to the command line.
function varargout = vocoder_OutputFcn(hObject, eventdata, handles)

```

```
% varargout cell array for returning output args (see VARARGOUT);
% hObject handle to figure
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)
```

```
% Get default command line output from handles structure
varargout{1} = handles.output;
```

```
% -----
function archivo_Callback(hObject, eventdata, handles)
% hObject handle to archivo (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)
```

```
% -----
function abrir_Callback(hObject, eventdata, handles)
% hObject handle to abrir (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)
global WF
global WFo
global INFO
global INFOR
global C
global FNA
global PNA
```

```
cd(PNA)
```

```
[fn,pn]=uigetfile({...
    '*.wav;*.voc','Archivos soportados (*.wav, *.voc)';...
    '*.wav','Archivos WAV (*.wav)';...
    '*.voc','Archivos vocoder (*.voc)'}),...
```

```
'Abrir archivo');
```

```
pause(0.1)
```

```
pf = [pn fn];
```

```
if fn == 0
```

```
    return
```

```
end
```

```
FNA = fn;
```

```
PNA = pn;
```

```
tipo_archivo = fn(findstr('.',fn):end); % Determinar la extensión del archivo
```

```
switch tipo_archivo
```

```
    case '.wav'
```

```
        [WFO Fs] = wavread(pf);
```

```
        [a,b]=size(WFO);
```

```

if b==2
    WFo = (WFo(:,1)+WFo(:,2))/2;
end

INFOR = INFO;
INFOR.Fs = Fs;
axes(handles.progreso)
C = comp(WFo,INFOR,handles);
WF = descomp(C,INFOR,handles);
set(handles.rep_orig,'Enable','on')
set(handles.btn_reproducir1,'Enable','on')
set(handles.btn_opciones,'Enable','on')
case '.voc'
fid = fopen(pf);
datos = fread(fid,'uint8');
fclose(fid);
[C,INFOR]=informacion(datos);
axes(handles.progreso)
WF = descomp(C,INFOR,handles);
set(handles.rep_orig,'Enable','off')
set(handles.btn_reproducir1,'Enable','off')
set(handles.btn_opciones,'Enable','off')
otherwise
    msgbox('Extensión de archivo no soportada','Error de
archivo','warn','modal')
return
end

ra = get(handles.rep_autom,'Checked');
if strcmp(ra,'on')
    set(handles.btn_reproducir2,'State','on','Enable','off')
    reproducir_Callback([], [], handles)
end
set(handles.btn_guardar,'enable','on')
set(handles.btn_propiedades,'enable','on')
% -----
function guardar_Callback(hObject, eventdata, handles)
% hObject    handle to guardar (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
global WF
global WFo
global INFO
global INFOR
global C
global FNA
global PNA

cd(PNA)

```

```
tipo_archivo = FNA(findstr('.',FNA):end); % Determinar la extensión del archivo
```

```
if strcmp(tipo_archivo, '.wav')  
    [fn,pn,fi]=uiputfile({...  
        '*.wav','Archivo WAV (*.wav)';...  
        '*.voc','Archivo vocoder (*.voc)'},...  
        'Guardar archivo como');
```

```
    if fn==0  
        return  
    end  
    nombre = fn(1:findstr('.',fn));
```

```
    if ~isempty(nombre)  
        fn = nombre(1:end-1);  
    end
```

```
    if fi == 1  
        pf = [pn fn '.wav'];  
        wavwrite(WFo,INFOR.Fs,pf)  
    end
```

```
    if fi == 2  
        pf = [pn fn '.voc']  
        datos = convertir(C,INFOR);  
        fid = fopen(pf,'w');  
        fwrite(fid,datos,'uint8');  
        fclose(fid);  
    end
```

```
else  
    [fn,pn,fi]=uiputfile({...  
        '*.voc','Archivo vocoder (*.voc)'},...  
        'Guardar archivo como');
```

```
    if fn==0  
        return  
    end  
    if ~isempty(nombre)  
        fn = nombre(1:end-1);  
    end
```

```
    pf = [pn fn '.voc']  
    datos = convertir(C,INFOR);  
    fid = fopen(pf,'w');  
    fwrite(fid,datos,'uint8');  
    fclose(fid);
```

```
end
```

```

% -----
function propiedades_Callback(hObject, eventdata, handles)
% hObject handle to propiedades (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)
global INFOR
global FNA
global PNA
global C
global WF

REL_COMP = numel(WF)/numel(C)*2;
propiedades({PNA,FNA,INFOR,REL_COMP})

% -----
function salir_Callback(hObject, eventdata, handles)
% hObject handle to salir (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)
close(handles.figure1)

% -----
function control_Callback(hObject, eventdata, handles)
% hObject handle to control (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)

% -----
function grabar_Callback(hObject, eventdata, handles)
% hObject handle to grabar (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)
global AI
global INFOR
global INFO
global FSG
global FNA

INFOR = INFO;
INFOR.Fs = FSG;
set(AI,'SampleRate',FSG);
start(AI)
FNA='.wav';
set(handles.rep_orig,'Enable','on')
set(handles.btn_reproducir1,'Enable','on')

% -----
function reproducir_Callback(hObject, eventdata, handles)
% hObject handle to reproducir (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)
global AO

```

```

global WF
global INFOR

AO.SampleRate = INFOR.Fs;

putdata(AO,WF);
start(AO);
set(handles.btn_reproducir2,'State','on','Enable','off')
% -----
function detener_Callback(hObject, eventdata, handles)
% hObject   handle to detener (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles   structure with handles and user data (see GUIDATA)
global AI
global AO
global WF
global WFo
global C
global INFOR

if strcmp(AI.Running,'On')
    stop(AI)
    WFo = getdata(AI,AI.SamplesAvailable);
    axes(handles.progreso)
    C = comp(WFo,INFOR,handles);
    WF = descomp(C,INFOR,handles);

    ra = get(handles.rep_autom,'Checked');
    if strcmp(ra,'on')
        set(handles.btn_reproducir2,'State','on','Enable','off')
        reproducir_Callback([], [], handles)
    end
    set(handles.btn_guardar,'enable','on')
    set(handles.btn_propiedades,'enable','on')
    return
end

if strcmp(AO.Running,'On')
    stop(AO)
end

% -----
function rep_autom_Callback(hObject, eventdata, handles)
% hObject   handle to rep_autom (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles   structure with handles and user data (see GUIDATA)
umtoggle(hObject);

% -----

```

```

function opciones_Callback(hObject, eventdata, handles)
% hObject   handle to opciones (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles   structure with handles and user data (see GUIDATA)

global INFO
global INFOR
global C
global WF
global WFo
global FNA
global PNA

out=opciones(INFO);

if isstruct(out)
    INFO = out;
    Fs_tmp = INFOR.Fs;
    INFOR = INFO;
    INFOR.Fs = Fs_tmp;

    tipo_archivo = FNA(findstr('.',FNA):end); % Determinar la extensión del archivo
    if strcmp(tipo_archivo, '.wav')
        axes(handles.progreso)
        C = comp(WFo,INFOR,handles);
        WF = descomp(C,INFOR,handles);

        ra = get(handles.rep_autom,'Checked');

        ra = get(handles.rep_autom,'Checked');
        if strcmp(ra,'on')
            set(handles.btn_reproducir2,'State','on','Enable','off')
            reproducir_Callback([], [], handles)
        end
    end
end
end

```

```

% -----
function ayuda_Callback(hObject, eventdata, handles)
% hObject   handle to ayuda (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles   structure with handles and user data (see GUIDATA)

```

```

% -----
function temas_Callback(hObject, eventdata, handles)
% hObject   handle to temas (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles   structure with handles and user data (see GUIDATA)

```


web c:\matlab6p5\work\vocoderfinal\vocoder>manual\vocoder.htm

```
% -----  
function acerca_Callback(hObject, eventdata, handles)  
% hObject handle to acerca (see GCBO)  
% eventdata reserved - to be defined in a future version of MATLAB  
% handles structure with handles and user data (see GUIDATA)  
ayuda
```

```
function inicializar2(handles)  
% Crear y Configurar Objeto Analogico de Entrada  
global AI  
global AO  
AI = analoginput('winsound');  
set(AI,...  
    'Userdata', handles,...  
    'SamplesPerTrigger',inf,...  
    'SamplesAcquiredFcn', @samples_AI,...  
    'StopFcn', @stop_AI)
```

```
addchannel(AI,1);
```

```
% Crear y Configurar Objeto Analogico de Salida  
AO = analogoutput('winsound');  
set(AO,...  
    'Userdata', handles,...  
    'SamplesOutputFcn', @samples_AO,...  
    'StopFcn', @stop_AO)
```

```
addchannel(AO,1);
```

```
global INFO  
global INFOR  
global FSG  
global PNA  
global FNA
```

```
axes(handles.ejes)
```

```
plot([0 1024],[0 0],'b','LineWidth',2);  
set(handles.ejes,'XTick',[],'YTick',[-0.5 0 0.5],...  
    'YTickLabel','', 'YGrid','On',...  
    'YLim',[-1 1], 'XLim',[0 1023])
```

```
try  
    load config  
catch  
    INFO.tv = 0.01;
```

```

INFO.nbf = 19;
INFO.Fmin = 200;
INFO.Fmax = 4000;
INFO.Fs = 11025;
set(handles.fs11025,'Checked','on')
PNA = 'C:\Matlab6p5\work';
ra = 'on';
INFOR = INFO;
FSG = 11025;
end
FNA="";

```

```

switch FSG
case 8000
    set(handles.fs8000,'Checked','on')
case 11025
    set(handles.fs11025,'Checked','on')
case 12000
    set(handles.fs12000,'Checked','on')
case 16000
    set(handles.fs16000,'Checked','on')
case 22050
    set(handles.fs22050,'Checked','on')
end

```

```

set(handles.rep_autom,'Checked',ra)

```

```

INFOR.Fs = FSG;

```

```

set(handles.btn_abrir,'ClickedCallback',@abrir_ccb,'UserData',handles)
set(handles.btn_guardar,'ClickedCallback',@guardar_ccb,'UserData',handles)
set(handles.btn_reproducir1,'ClickedCallback',@reproducir1_ccb,'UserData',handles)
set(handles.btn_reproducir2,'ClickedCallback',@reproducir2_ccb,'UserData',handles)
set(handles.btn_opciones,'ClickedCallback',@opciones_ccb,'UserData',handles)
set(handles.btn_detener,'ClickedCallback',@detener_ccb,'UserData',handles)
set(handles.btn_grabar,'ClickedCallback',@grabar_ccb,'UserData',handles)
set(handles.btn_propiedades,'ClickedCallback',@propiedades_ccb,'UserData',handles)
set(handles.btn_ayuda,'ClickedCallback',@ayuda_ccb,'UserData',handles)

```

```

function inicializar1(handles)
%% Teclas para aceleradores
set(handles.abrir,'accelerator','A')
set(handles.guardar,'accelerator','G')
set(handles.propiedades,'accelerator','P')
set(handles.salir,'accelerator','S')
set(handles.grabar,'accelerator','B')

```

```
set(handles.reproducir,'accelerator','R')
set(handles.detener,'accelerator','D')
set(handles.rep_orig,'accelerator','O')
set(handles.rep_autom,'accelerator','T')
set(handles.frec_muest,'accelerator','F')
set(handles.opciones,'accelerator','C')
set(handles.temas,'accelerator','M')
set(handles.acerca,'accelerator','E')
```

```
uipushtool(...
    'Tag',      'btn_abrir',...
    'CData',    icono('abrir'),...
    'TooltipString','Abrir archivo')
```

```
uipushtool(...
    'Tag',      'btn_guardar',...
    'CData',    icono('guardar'),...
    'TooltipString','Guardar archivo',...
    'enable',   'off')
```

```
uitoggletool(...
    'Tag',      'btn_reproducir1',...
    'CData',    icono('reproducir1'),...
    'Separator', 'on',...
    'TooltipString','Reproducir audio original',...
    'enable',   'off')
```

```
uitoggletool(...
    'Tag',      'btn_reproducir2',...
    'CData',    icono('reproducir2'),...
    'TooltipString','Reproducir audio convertido',...
    'enable',   'off')
```

```
uipushtool(...
    'Tag',      'btn_detener',...
    'CData',    icono('detener'),...
    'TooltipString','Detener reproducción')
```

```
uitoggletool(...
    'Tag',      'btn_grabar',...
    'CData',    icono('grabar'),...
    'TooltipString','Grabar sonido')
```

```
uipushtool(...
    'Tag',      'btn_opciones',...
    'CData',    icono('opciones'),...
    'Separator', 'on',...
    'TooltipString','Opciones de conversión')
```

```
uipushtool(...
```

```
'Tag',      'btn_propiedades',...
'CData',    icono('propiedades'),...
'Separator', 'on',...
'TooltipString','Propiedades',...
'enable',   'off')
```

```
uipushtool(...
  'Tag',      'btn_ayuda',...
  'CData',    icono('ayuda'),...
  'Separator', 'on',...
  'TooltipString','Ayuda')
```

```
function stop_AI(obj,event)
handles = get(obj,'UserData');
set(handles.btn_grabar,'Enable','on','State','off')
```

```
function samples_AI(obj,event)
handles = obj.UserData;
p = get(handles.ejes,'Children');
try
  WF = peekdata(obj,1024);
  set(p,'YData',WF,'XData',0:1023)
end
```

```
function stop_AO(obj,event)
global FNA
```

```
handles = get(obj,'UserData');
```

```
tipo_archivo = FNA(findstr('.',FNA):end); % Determinar la extensión del archivo
if strcmp(tipo_archivo,'.wav')
  set(handles.btn_reproducir1,'Enable','on','State','off')
end
set(handles.btn_reproducir2,'Enable','on','State','off')
```

```
function samples_AO(obj,event)
global WF
global WFO
global FNA
handles = obj.UserData;
so = obj.SamplesOutput;
p = get(handles.ejes,'Children');
```

```
v = get(handles.btn_reproducir1,'State');
if strcmp(v,'on')
  try
```

```

        set(p,'YData',WFO(so-1023:so),'XData',0:1023)
    end
end

v = get(handles.btn_reproducir2,'State');
if strcmp(v,'on')
    try
        set(p,'YData',WF(so-1023:so),'XData',0:1023)
    end
end
end
% --- Executes when user attempts to close figure1.
function figure1_CloseRequestFcn(hObject, eventdata, handles)
% hObject    handle to figure1 (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: delete(hObject) closes the figure
global AI
global AO
global FSG
global PNA
global INFO

btn = questdlg('¿Desea cerrar el programa?','Salir del programa','Sí','No','No');
if strcmp(btn,'Sí')

    ra = get(handles.rep_autom,'Checked');

    save config INFO FSG PNA ra

    delete(hObject);
    stop(AI)
    stop(AO)
    delete(AI)
    delete(AO)
    clear global
end

% -----
function frec_muest_Callback(hObject, eventdata, handles)
% hObject    handle to frec_muest (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% -----
function fs8000_Callback(hObject, eventdata, handles)
% hObject    handle to fs8000 (see GCBO)

```

```
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)
global FSG
```

```
set(handles.fs8000,'Checked','on')
set(handles.fs11025,'Checked','off')
set(handles.fs12000,'Checked','off')
set(handles.fs16000,'Checked','off')
set(handles.fs22050,'Checked','off')
```

```
FSG = 8000;
```

```
% -----
function fs11025_Callback(hObject, eventdata, handles)
% hObject handle to fs11025 (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)
global FSG
```

```
set(handles.fs8000,'Checked','off')
set(handles.fs11025,'Checked','on')
set(handles.fs12000,'Checked','off')
set(handles.fs16000,'Checked','off')
set(handles.fs22050,'Checked','off')
```

```
FSG = 11025;
```

```
% -----
function fs12000_Callback(hObject, eventdata, handles)
% hObject handle to fs12000 (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)
global FSG
```

```
set(handles.fs8000,'Checked','off')
set(handles.fs11025,'Checked','off')
set(handles.fs12000,'Checked','on')
set(handles.fs16000,'Checked','off')
set(handles.fs22050,'Checked','off')
```

```
FSG = 12000;
```

```
% -----
function fs16000_Callback(hObject, eventdata, handles)
% hObject handle to fs16000 (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)
global FSG
```

```
set(handles.fs8000,'Checked','off')
set(handles.fs11025,'Checked','off')
```

```
set(handles.fs12000,'Checked','off')
set(handles.fs16000,'Checked','on')
set(handles.fs22050,'Checked','off')
```

```
FSG = 16000;
```

```
% -----
function fs22050_Callback(hObject, eventdata, handles)
% hObject   handle to fs22050 (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles   structure with handles and user data (see GUIDATA)
global FSG
```

```
set(handles.fs8000,'Checked','off')
set(handles.fs11025,'Checked','off')
set(handles.fs12000,'Checked','off')
set(handles.fs16000,'Checked','off')
set(handles.fs22050,'Checked','on')
```

```
FSG = 22050;
```

```
% -----
function rep_orig_Callback(hObject, eventdata, handles)
% hObject   handle to rep_orig (see GCBO)
% eventdata reserved - to be defined in a future version of MATLAB
% handles   structure with handles and user data (see GUIDATA)
global AO
global WFo
global INFOR
```

```
AO.SampleRate = INFOR.Fs;
```

```
putdata(AO,WFo);
start(AO);
set(handles.btn_reproducir1,'State','on','Enable','off')
```

```
function abrir_ccb(obj,event)
handles = get(obj,'UserData');
abrir_Callback([], [], handles)
```

```
function guardar_ccb(obj,event)
handles = get(obj,'UserData');
guardar_Callback([], [], handles)
```

```
function reproducir1_ccb(obj,event)
handles = get(obj,'UserData');
set(obj,'Enable','off')
rep_orig_Callback([], [], handles)
```

```
function reproducir2_ccb(obj,event)
handles = get(obj,'UserData');
set(obj,'Enable','off')
reproducir_Callback([], [], handles)
```

```
function opciones_ccb(obj,event)
handles = get(obj,'UserData');
opciones_Callback([], [], handles)
```

```
function detener_ccb(obj,event)
handles = get(obj,'UserData');
detener_Callback([], [], handles)
```

```
function grabar_ccb(obj,event)
handles = get(obj,'UserData');
set(handles.btn_grabar,'Enable','off')
grabar_Callback([], [], handles)
```

```
function propiedades_ccb(obj,event)
handles = get(obj,'UserData');
propiedades_Callback([], [], handles)
```

```
function ayuda_ccb(obj,event)
handles = get(obj,'UserData');
temas_Callback([], [], handles)
```


ANEXO C

ESCUELA POLITÉCNICA DEL EJÉRCITO SEDE LATACUNGA

CARRERA DE ELECTRÓNICA E INSTRUMENTACIÓN

ENCUESTA MOST TEST PARA EL VOCODER EN SUBBANDAS

LATACUNGA, JULIO 2005

PRUEBA 1		PRUEBA 2		PRUEBA 3		PRUEBA 4		PRUEBA 5		PRUEBA 6		PRUEBA 7		PRUEBA 8		PRUEBA 9		PRUEBA 10	
5		5		5		5		5		5		5		5		5		5	
4		4		4		4		4		4		4		4		4		4	
3		3		3		3		3		3		3		3		3		3	
2		2		2		2		2		2		2		2		2		2	
1		1		1		1		1		1		1		1		1		1	

GRACIAS POR SU COLABORACIÓN

