Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ciencias

RECONSTRUCCIÓN DE SEÑALES DE AUDIO

Pedro Fernando Morales Almazán

Asesorado por el Phd. Ing. Enrique Edmundo Ruiz Carballo

Guatemala, octubre de 2007

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

RECONSTRUCCIÓN
DE SEÑALES DE AUDIO

TRABAJO DE GRADUACIÓN
PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

PEDRO FERNANDO MORALES ALMAZÁN
ASESORADO POR EL PHD. ING. ENRIQUE EDMUNDO RUIZ CARBALLO

AL CONFERÍRSELE EL TÍTULO DE
INGENIERO ELECTRÓNICO

GUATEMALA, OCTUBRE DE 2007

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA

NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Murphy Olympo Paiz Recinos
VOCAL I	Inga. Glenda Patricia García Soria
VOCAL II	Lic. Amahán Sánchez Álvarez
VOCAL III	Ing Julio David Galicia Celada
VOCAL IV	Br. Kenneth Issur Estrada Ruiz
VOCAL V	Br. Elisa Yazminda Vides Leiva
SECRETARIA	Inga. Marcia Ivonne Véliz Vargas

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Murphy Olympo Paiz Recinos
EXAMINADOR	Ing. Romeo Lø’opez
EXAMINADOR	Ing. Luis Solares
EXAMINADOR	Ing. Luis Durán
SECRETARIA	Inga. Marcia Ivonne Véliz Vargas

HONORABLE TRIBUNAL EXAMINADOR

Cumpliendo con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

Reconstrucción de
Señales de Audio,

tema que me fuera asignado por la Coordinación de la Carrera de Ingeniería Electrónica, el 8 de mayo de 2007.

Pedro Fernando Morales Almazán

AGRADECIMIENTOS A:

Dios	Por haberme permitido concluir con esta meta.
Mi Mamá	Por haberme apoyado en todo lo necesario para llegar hasta aqui.
Mi familia	Por apoyarme en todo lo que hago.
Mi asesor	Por toda la ayuda y consejo brindado para realizar este trabajo de graduación.
Mis amigos	Por haber pasado juntos tantos momentos buenos durante los proyectos y los desvelos.
La Universidad de San Carlos	Por haberme brindado la oportunidad de ser parte de esta casa de estudios.

A mis Amigos

"Todos somos muy ignorantes. Lo que ocurre es que no todos ignoramos las mismas cosas."

Albert Einstein

ÍNDICE GENERAL

LISTA DE ILUSTRACIONES

LISTA DE TABLAS

LISTA DE SÍMBOLOS

Símbolo	Significado
$\neq$	Diferente de
$\|\|\|\|$	Norma de una función
$\boldsymbol{\partial}$	Derivada parcial
$\mathbb{R}$	Campo de los reales
$\in$	Pertenece
$\boldsymbol{\Delta}$	Diferencial
j	Raiz cuadrada de -1
$\mathbf{log}$	Logaritmo base 10
$\approx$	Aproximadamente
$\mathbf{:}$	Proporción
$\boldsymbol{sgn(x)}$	Función signo
$\boldsymbol{<,>}$	Producto interno
$\mathbb{Z}$	Conjunto de Enteros
$\mid$	Tal que
$\equiv$	Identicamante
$\subset$	Subconjunto
$\boldsymbol{\{\}_{n=0}^{\infty}}$	Sucesión
$\boldsymbol{rect(\omega, a)}$	Rectángulo de ancho a

RESUMEN

En el mundo de hoy, la música, la comunicación celular y telefónica, y en general, cuanquier tipo de comunicación que involucre sonido, es parte de nuestra vida cotidiana y cada vez más se convierte en algo no puramente recreativo, sino más bien, necesario.

Por la creciente demanda de este tipo de comunicaciones, el aprovechamiento máximo de los recursos de transmisión, cables, fibra ótica, ancho de banda, etc., es de vital importancia en las telecomunicaciones de hoy en día. Por esto, la compresión de archivos resulta ser un tema muy importante en la tecnología actual, sin embargo no hay que dejar a un lado que la calidad del sonido sea lo mejor posible. Por esto la reconstrucción de señales de audio resulta ser de suma importancia para poder obtener buenos resultados a la hora de realizar una comunicación.

OBJETIVOS

General

- Poder reconstruir una señal de audio original con la mayor fidelidad a partir de un archivo con baja calidad.

Específicos

1. Obtener un algorítmo que permita mejorar la calidad de archivos de auido no comprimidos.
2. Obtener un método para poder mejorar la calidad en comunicaciones de audio en tiempo real.
3. Determinar características suficientes que permitan realizar una compresión en archivos de audio.
4. Analizar métodos aritméticos y de tratamiento de señales digitales que permitan realizar de forma sencilla una compresión sin pérdidas.

INTRODUCCIÓN

La comunicación audible es la primer forma de comunicación que utilizó el hombre, desde cuando se comenzó a agrupar en comunidades. Se comenzó a desarrollar el lenguaje hablado y a perfeccionar este tipo de comunicación entre personas, es por esto que es uno de los tipos de comunicación mas utilizados en la actualidad.

Con la exploración del mundo y el crecimiento poblacional, se convirtió en imperante un tipo de comunicación a distancia. En un principio se recurrió a la comunicación escrita, sin embargo contaba con la desventaja de ser demasiado lento, y además solo es de una via, no podía establecerse un intercambio rapido de información. Sin embargo, luego se implementó la comunicación por voz en tiempo real a grandes distancias, el teléfono. Además se inició la grabaci'on de musica y conferencias para ser posteriormente reproducidas.

Con esto se inició el manejo de señales audibles, voz y musica. En un principio se realizó todo el tratamiento de forma analógica, pero con el tiempo se comenzó a utilizar más la tecnología digital por diversas razones, la reducción del ruido, el mejoramiento de la calidad, el aparecimiento de las computadoras, etc.

Con la cuantización de las señales de sonido aparece una distorción inevitable a la hora de reproducir los sonidos y es el ruido de cuantización. El objetivo es cuantizar de tal forma que la distorción producida este en un rango no audible, asi de esta forma el oido no notará la diferencia entre la señal original y la reconstruida.

1. PSICOACÚSTICA

1.1. El Sonido

El sonido ha sido el medio por el cual el hombre comenzó a comunicarse y a intercambiar información. Al principio imitando sonidos naturales y luego creando sonidos a los cuales se les comenzó a dotar de significado, con esto creando un código para establecer una comunicación.

El sonido es la sensación producida en el oído por las variaciones de presión generadas por un movimiento vibratorio transmitidas a través de un medio elástico. Es decir, para que halla sonido deben de haber dos factores, una fuente mecánica de vibración y un medio elástico por el cual se propaguen las ondas sonoras.

Cualquier tipo de vibración mecánica producirá sonido, puesto que esta creará cambios de presión dentro del medio que se encuentre, lo cual generá una propagación de cambios de presión dentro del medio. Generalmente el medio al cual estamos acostumbrados es el aire, puesto que en él estamos inmersos, sin embargo, al percibir sonidos en otros medios como metales, objetos rígidos, agua, etc. nos damos cuenta que el sonido se propaga en estos a diferentes velocidades.

El primer acercamiento sobre la velocidad del sonido en los diversos medios los podemos observar cuando miramos algún acontecimiento a gran distancia, por ejemplo, si se observa a una persona martillando en el techo de una casa, se verá como golpea y momentos después se oirá el sonido, así como también en una noche de fuegos artificiales se verá primero la explosión y luego se oirá esta.

Esto nos indica que la luz viaja mas rápido que el sonido, de echo mucho mas rápido. Este echo puede ser utilizado de varias maneras, por ejemplo para determinar la distancia a la cual se produjo un relámpago, contanto la diferencia de tiempo en que se vio el rayo y en que se escuchó.

1.1.1. Velocidad del Sonido

La velocidad de propagación del sonido depende, como se vió anteriormente, del medio en el cual se mueva. Esto es debido a la velocidad con que el medio puede llevar los cambios de presión.

En esto influyen muchos factores como la temperatura, humedad, viscosidad, densidad, compresibilidad, etc. del medio.

La forma en que se progaga el sonido es a nivel molecular, siendo creado por vibraciones mecánicas, desplaza moléculas del medio, las cuales a su vez desplazan a las moléculas adyacentes, generando una reacción en cadena a lo que se le conoce como ondas sonoras.

La propagación de las ondas sonoras es similar a la propagación de las ondas electromagnéticas. Además se puede considerar a la fuente de sonido equivalente a una antena transmisora, y dependiendo del medio, isotrópica y onmidirecciónal, esto es, las ondas sonoras se propagan en todas direcciones, formando ondas esféricas, y con la misma intensidad en todas las direcciones, dependiendo del medio.

Como tienen características de onda, a las señales de sonido se les asigna una frecuencia y una longitud de onda. Estas cantidades se relacionan junto con la velocidad de propagación del sonido en el medio a través de

$$\nu = f\lambda \tag{1.1}$$

donde λ es la longitud de onda, f la frecuencia y ν la velocidad del sonido en el medio. Esta ecuación es equivalente a la utilizada en teoría electromagnética que relaciona a la longitud de onda con su frecuencia y la velocidad de propagación del la onda en el medio.

Se puede modelar la propagación de las ondas sonoras u ondas de presión en un medio en el cual no hay ninguna otra fuente de presión como

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial p(r,t)}{\partial r}\right) - \frac{1}{\nu^2}\frac{\partial^2 p(r,t)}{\partial t^2} = 0 \tag{1.2}$$

en donde $p(r,t)$ es la presión sobre la superficie de una esfera de radio r en el tiempo t, localizando la fuente de del sonido en el origen.

Se tiene entonces que la solución a la ecuación para $r >> \lambda$ se puede representar como

$$p(r,t) = p_0 + \frac{\Delta p}{r}\sin\left(2\pi f t - 2\pi\frac{r}{\lambda} + \phi_0\right) \tag{1.3}$$

con p la presión inicial sobre la esfera y Δp el cambio máximo de presión.

Con esto se logra demostrar que las ondas sonoras basicas son ondas sinusoidales, además por la linealidad de la ecuación diferencial, se tiene que los medios de propagación son lineales, es decir, que en ellos se cumple el principio de superposición.

Ahora la interrogante radica en la obtención de la velocidad de propagación de las ondas sonoras en un medio, sabiendo sus propiedades principales. Esto se logra establecer por medio de la siguiente relación

$$\nu = \sqrt{\frac{B + \frac{4}{3}S}{\rho}} \tag{1.4}$$

con B el módulo de volumen del medio, S el módulo de corte y ρ la densidad.

En el caso de gases, se tiene que

$$S = 0 \tag{1.5}$$

$$B = \gamma p \tag{1.6}$$

con γ la constante adiabática, por lo que 1.4 queda como

$$\nu = \sqrt{\frac{\gamma p}{\rho}} \tag{1.7}$$

sin embargo, de acuerdo con la ecuación de los gases ideales

$$pM = \rho RT \tag{1.8}$$

con R la constante de los gases, T la temperatura del gas y M la masa molecular del gas, se tiene que 1.7 puede escribirse como

$$\nu = \sqrt{\frac{\gamma RT}{M}} \tag{1.9}$$

así por ejemplo, en el aire, a temperatura ambiente,

$$
\begin{aligned}
T =& 298 \quad \mathrm{K} \\
M =& 29{,}0 \times 10^{-3} \quad \mathrm{Kg/mol} \\
R =& 8{,}31 \quad \mathrm{J/mol \cdot K} \\
\gamma =& 1{,}4
\end{aligned}
$$

se tiene que

$$
\nu = \sqrt{\frac{(1{,}4)(8{,}31\mathrm{J/mol \cdot K})(298\mathrm{K})}{29{,}0 \times 10^{-3}\mathrm{Kg/mol}}} = 345{,}8 \quad \mathrm{m/s}
$$

1.1.2. Parámetros del Sonido

Básicamente se pueden agrupar los parámetros que determinan el tipo de sonido en 3 carácteristicas principales

- Intensidad
- Tono
- Timbre

La intensidad tiene que ver principalmente con la amplitud de la onda sonora.

Formalmente la intensidad está definida como la cantidad de energía que pasa por un lugar durante cierto tiempo. En lo que respecta a las ondas sonoras, al ser producidas, el movimiento vibratorio que las propaga realiza un desplazamiento de energía, la energía que está moviendo a las partículas del medio para provocar los cambios de presión causantes del sonido.

Se puede entonces definir la *Intensidad* como

$$I = \frac{P}{A} \tag{1.10}$$

en donde P es la potencia de la señal y A es un área sobre la cual se esta midiendo la intensidad.

La potencia, es decir, la energía transmitida de la onda, está en funsión de la amplitud de la onda.

Por la naturaleza esférica de las ondas sonoras, se puede pensar en la intensidad como la *cantidad* de sonido que pasa a traés de una esfera de un cierto radio, que encierra a la fuente del sonido.

Para poder calcular la intensidad de una onda sonora, se debe de encontrar el comportamiento de la amplitud de la onda sonora a través del medio, y por 1.3 se vio que esta disminuye con forme la distancia aumenta, esto es, la Intensidad del sonido disminuye con forme las ondas sonoras se alejan de la fuente.

La intensidad es medida en el SI por [W/m^2], aunque muchas veces es más común expresar la medida de intensidad como una ganancia en una escala logarítmica, expresada en decibeles,

$$[\mathrm{dB}] = \log \frac{I}{I_0} \tag{1.11}$$

con I_0 una intensidad de referencia. Generalmente, como se verá más adelante, para audición, la referencia I_0 es el umbral auditivo que es

$$I_0 = 1 \times 10^{-2}\mathrm{W/m^2}$$

Por otra parte, el Tono de un sonido se refiere a la frecuencia a la cual la fuente sonora esta vibrando, esto es, la frecuencia fundamental del sonido.

La otra característica importante del sonido la constituye el Timbre.

El timbre se refiere a la propiedad que distingue particularmente a los sonidos. Por ejemplo, el La central de un piano, un violín y una flauta tienen el mismo Tono, 440Hz, sin embargo es posible distinguir que sonido pertenece a cada instrumento. Esto es debido a que las formas de onda de las ondas sonoras producidas por cada uno de los tres instrumentos son diferentes, a pesar que poseen el mismo período.

La diferencia de timbre se produce ya que, a pesar que las tres poseen la misma *frecuencia fundamental*, cada una tiene diferentes *armónicas* que cambian la forma de la onda, y producen un *tipo* de sonido diferente.

Con esto se termina la descripción general del sonido y las ondas sonoras que serán de utilidad para el estudio que se realizará.

1.2. El Oído Humano

Para poder establecer los límites del tratamiento de las ondas sonoras por medios electrónicos, específicamente digitales, es necesario comprender el funcionamiento y limitaciones del transductor natural que utiliza el ser humano, el oído.

El oído cumple la función de concentrar el sonido, conducirlo y transformar las ondas sonoras en impulsos eléctricos hasta el nervio auditivo, para así poder ser transmitidos al cerebro y ser percibidos.

Esto se cumple a través del *Oído Externo*, *Medio* e *Interno*.

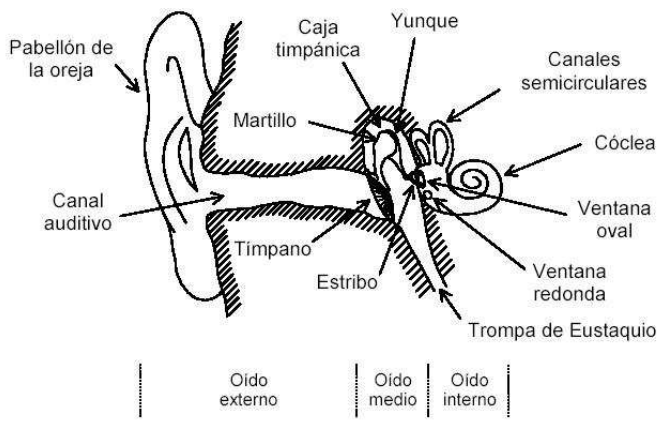

Figura 1: Oído Humano

1.2.1. Oído Externo

La función principal del oído externo es la de captar y concentrar las ondas sonoras.

Está constituido por el pabellón auricular u oreja y el canal auditivo externo.

El pabellón auditivo sirve como un receptor que recoge las ondas sonoras del ambiente por medio de difracción y reflexión de las mismas, para así conducirlo a través del canal auditivo externo del oído.

Es decir, este resulta ser una especie de *antena* para las ondas sonoras, que concentra toda la información para que se conduzca a través del túnel que llevará la información al oído medio.

El oído externo humano es bastante directivo y puede recibir muy bien las ondas sonoras incidentes, sin embargo es menos directivo que el de otros seres vivos como los oídos de un perro o un gato, los cuales a su vez tienen movimientos musculares voluntarios para mejorar la directividad.

Otra característica que logra el oído externo es la determinación del lugar de proveniencia de los sonidos, *el sonido 3D*.

Solamente con la intensidad de una onda sonora se puede tener una idea de la distancia de la fuente, sin embargo no es posible tener un dato acerca de la dirección de la proveniencia del sonido, sin embargo por la disposición de las orejas, es posible encontrar un desfase en la incidencia de las ondas sonoras en ambos oídos y así se puede determinar la dirección de proveniencia del sonido.

1.2.2. Oído Medio

La función principal es acoplar al oído externo con el oído interno. Se puede ver al oído medio como un acople de impedancias acústicas entre la alta impedancia del medio, y la baja impedancia del oido interno.

Esta conformado por el tímpano, los huesecillos y la trompa de Eustaquio.

La función de estos es básicamente convertir las ondas sonoras captadas por el oído externo en vibraciones que serán captadas por el oído interno. Además, no solo realizará esta conversión, sino que también amplificará las ondas sonoras que llegan al sonido, pues cuando llegan al oído no llegan con la potencia necesaria para poder ser reconocidas.

La división entre el oído externo y el oído medio está constituida por la membrana del tímpano. Esta membrana vibra con forme a las ondas sonoras incidentes, y transforma estas en vibraciones que pasan a la *caja timpánica*, donde se encuentran los huesecillos.

El tímpano se encuentra adherido al primero de los huesecillos, el *martillo*. Luego este se conecta al *yunque* y por ultimo al *estribo*.

Debido a que el tímpano es una membrana muy delicada, y para no afectar la vibración que las ondas sonoras le producen, es necesario que de ambos lados de él se encuentre la misma presión, la presión atmosférica.

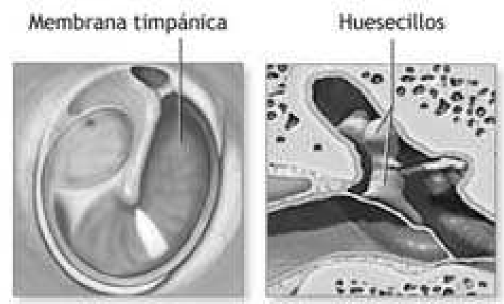

Figura 2: Membrana del Tímpano

La trompa de Eustaquio es la encargada de cumplir con esta misión. La trompa de Eustaquio conecta el oído medio con la garganta. Permanece normalmente abierta, sin embargo cuando se bosteza o se traga, la trompa de Eustaquio se abre, y regula la presión interna del odio medio. Al hacer esto, la presión que tiene la membrana del tímpano se encuentra equilibrada de ambos lados, y con esto puede funcionar adecuadamente.

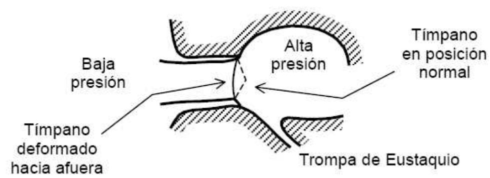

Figura 3: Presión en el Tímpano

Al estar equilibradas las presiones en ambos lados del tímpano, este puede vibrar adecuadamente con forme a las ondas sonoras incidentes, puesto que al estar iguales las presiones, cualquier perturbación, en este caso las ondas sonoras, hará que el tímpano trate de restaurar su estado normal.

Al pasar la membrana del timpano, las ondas sonoras se transfoman en vibraciones que manejan los huesecillos de la caja timpánica. Estos se encuentran unidos entre sí por medio de ligamentos y músculos.

La finalidad de esta caja timpánica es de aumentar la potencia de las señales que llegan al oido. Básicamente se encargan de convertir señales de alta amplitud y baja presión, en señales de baja amplitud y gran presión, con el objeto de poder pasar al oído interno y así poder estimular al líquido que se encuentra allí.

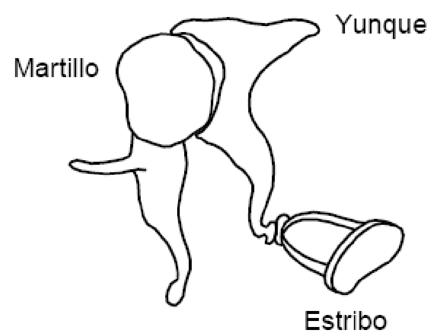

Figura 4: Martillo, Yunque y Estribo

Los tres huesecillos actúan como un amplificador mecánico, con ganancia de 1.3, es decir que la fuerza que ejerce el estribo es 1.3 veces más grande que la ejercida sobre el martillo por el tímpano. Además, el área de contacto de la entrada, el martillo, y la salida, el estribo, son diferentes y esto causa la ganancia de presión.

El área de contacto que el martillo posee con el tímpano es de $0{,}6\text{cm}^2$ en promedio, y el área entre el estribo y *la ventana oval*, la parte que conecta al oído medio y el interno, es de $0{,}04\text{cm}^2$ en promedio, así que se tiene que la ganancia de presión obtenida a través de dicho sistema mecánico es de

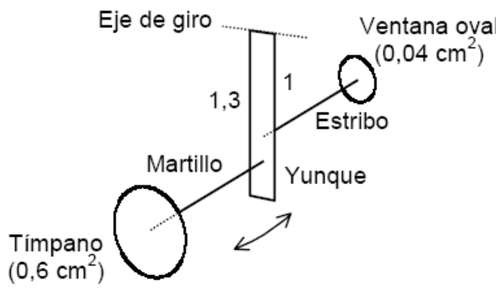

Figura 5: Modelo Mecánico de los Huesecillos

$$G_p = \frac{p_o}{p_i} = \frac{F_o A_i}{A_o F_i} = 1{,}3\frac{0{,}6}{0{,}04} \approx 20 \tag{1.12}$$

1.2.3. Oído Interno

El oído interno esta constituido principalmente por 3 partes, el *laberinto*, el *vesíbulo* y el *caracol*.

El laberinto es la cámara ósea que contiene a los canales semicirculares. Estos son 3 conductos que son los responsables del equilibrio en el cuerpo. Están llenos del líquido *endolinfático*. Estos tres conductos se encuentran dispuestos ortogonales entre sí, es decir, uno se encuentra paralelo al suelo, otro a una pared lateral de la cabeza, y el otro a la parte frontal. Están recubiertos internamente por vellosidades que detectan el movimiento del líquido endolinfático, para así poder determinar la posición de la cabeza y guardar el equilibrio. Cuando la cabeza se mueve, el líquido presiona las vellosidades y estas mandan impulsos eléctricos al cerebro para determinar la inclinación.

El vestíbulo sirve de enlace entre el caracol y el laberinto y además establece la comunicación con la caja timpánica por medio de la *ventana oval* y la *ventana circular*.

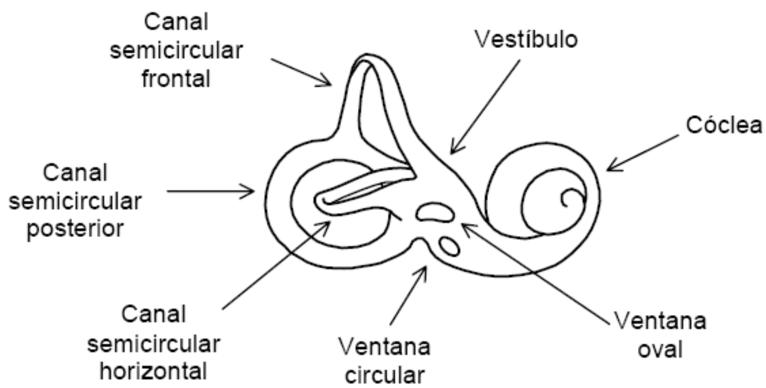

Figura 6: Oído Interno

La parte fundamental del oído interno, y en general, de la audición, es la *cóclea.* La cóclea se encuentra en el caracol del oído y este es un tubo enrollado en espiral de aproximadamente 2 vueltas y media.

Este está dividido en tres secciones, la sección baja es la *rampa timpánica*, la superior es la *rampa vestibular* y la última es la *rampa coclear.*

Tanto la rama timpánica como la rama vestibular contienen líquido *perilinfático* el cual es rico en sodio. Estas se unen cerca del vértice del caracol, a través de un pequeño orificio llamado *helicotrema.* La rama coclear contiene el líquido endolinfático, el cual es rico en potasio.

Las primeras dos rampas se comunican con el oído medio a través de la ventana oval y la ventana circular respectivamente. La rampa coclear contiene a la membrana *basiliar* y la membrana *tectorial.* Entre estas membranas se encuentra el *órgano de Corti*, en el cual se produce la transducción entre las vibraciones y los impulsos eléctricos.

En el órgano de Corti se encuentran las células *ciliadas* o *pilosas*, las cuales son las responsables de la conversión. Estas células, aproximadamente 24,000, están

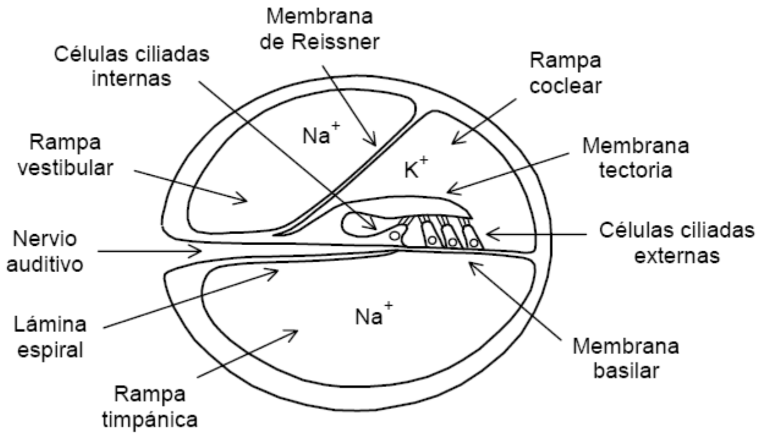

Figura 7: Estructura interna de la Cóclea

dispuestas en 4 filas a lo largo del órgano de Corti, y responden de acuerdo a las vibraciones de la membrana basiliar.

La membrana basiliar responde de diferente forma a cada frecuencia aplicada, es por esto que cada punto responde diferente a las vibraciones, de acuerdo con la frecuencia de resonancia de cada punto.

Al ser estimuladas las células pilosas, estas generan un químico que reacciona generando pulsos eléctricos de aproximadamente unos 90 mV, para luego ser transmitidos por el nervio auditivo, que se encuentra sobre ellas, hacia el cerebro.

La membrana basiliar mide aproximadamente unos 35mm de longitud y 0,004mm de ancho en el extremo *basal* y unos 0,005mm en el extremo *apical*.

Esta membrana tiene una densidad no uniforme, siendo más densa en la parte más angosta para poder así funcionar como un filtro discriminador de frecuencias.

Cuando una onda sonora llega al oído interno, esta estimula el líquido perilinfático de la rama, produciendo así una diferencia de presión entre la parte superior

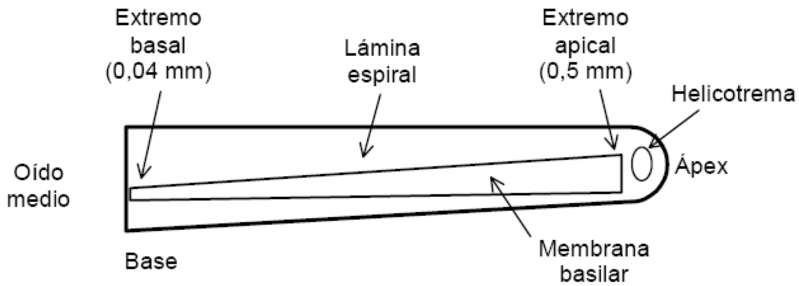

Figura 8: Vista Superior de la Membrana Basiliar

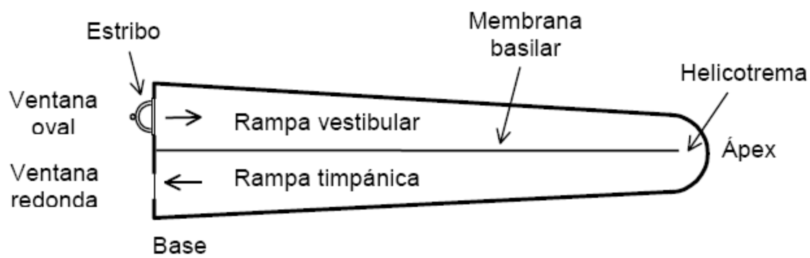

Figura 9: Vista Lateral de la Membrana Basiliar

de la membrana basiliar y la parte inferior en donde se encuentra el líquido endolinfático, esta onda de presión se propaga a través del liquido, a lo largo de la membrana, la cual, por su densidad, amplifica la senñal a medida que se acerca al *apex*.

Al llegar al final, la onda es reflejada creando así una onda estacionaria sobre la membrana. Al suceder esto, se asentuará una frecuencia, que será la frecuencia fundamental de la onda sonora, a la cual responderá un punto específico de la membrana, y este será el que mande la información al cerebro.

Lo que genera los pulsos eléctricos es la mezcla de los líquidos que se encuentran arriba y por debajo de la membrana, ya que las células pilosas actúan como pequeñas compuertas que se abren cuando son estimuladas por la frecuencia adecuada, dejando

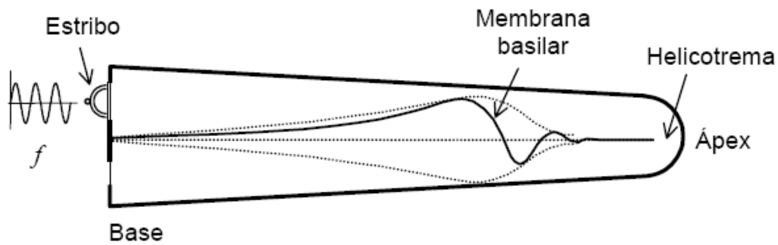

Figura 10: Onda Estacionaria

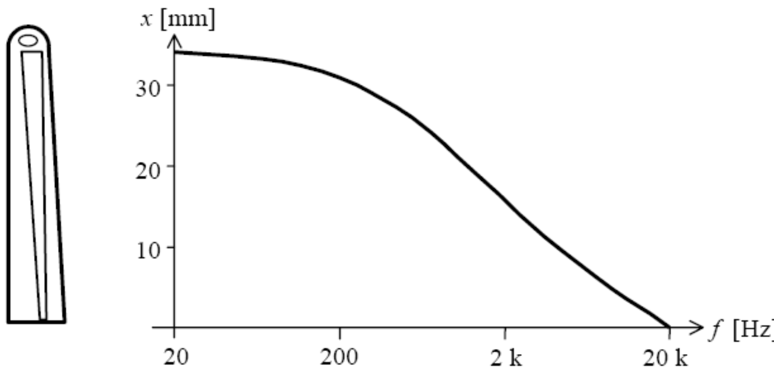

Figura 11: Relación entre la frecuencia de la onda y la distancia de la membrana

así que halla una reacción química entre ambos líquidos y generando así el potencial eléctrico requerido para la percepción del sonido.

El impulso eléctrico dura al rededor de unos 2ms y es transmitido al cerebro por medio de las neuronas. La velocidad con que las neuronas transmiten el impulso puede variar desde $1m/s$ hasta los $100m/s$. Esto depende de la cantidad de *mielina* que recubre la neurona. La mielina es un compuesto de un alcohol llamado esfingol, una cadena de ácido graso, fosfato, y colina, el cual mejora la velocidad de transmición de impulsos en los *axones* de las neuronas.

Luego de los 2ms que dura el impulso, existe un tiempo en el cual no se puede generar otro pulso, por lo que no puede detectarse otro sonido de la misma frecuencia.

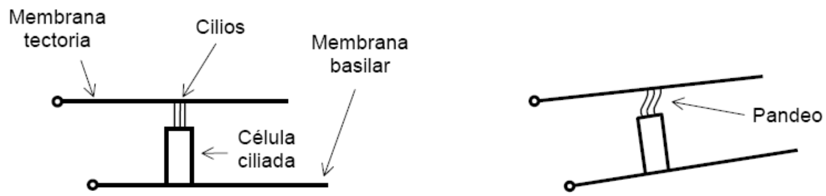

Figura 12: Movimiento de las células ciliadas

A este tiempo se le conoce como *período refractario*.

Cuando se aplica un potencial constante en las *dendritas* de una neurona, el potencial de la neurona tiende a acercarse, con cierta constante de tiempo, al potencial aplicado. Si el potencial aplicado es suficientemente alto, en algún momento se supera el umbral y se produce el disparo, volviendo la neurona a su estado inicial. Después de completarse el período refractario, el ciclo vuelve a empezar, lo cual lleva a que se genere un tren de potenciales de acción. La frecuencia de este tren de pulsos aumenta al aumentar el potencial constante aplicado. Resulta, así, que la neurona se comporta en forma similar a un modulador de frecuencia, codificando las señales recibidas a través de la frecuencia de los potenciales de acción.

1.3. Percepción del Sonido

El oído humano presenta un cierto rango en el cual es capar de *percibir* o detectar los sonidos que inciden en él.

Se ha comprobado que el oído humano es capaz de detectar sonidos que van desde los $20\mu Pa$ hasta los $20Pa$ de presión. Con este rango se puede ver que el oído es capaz de percibir una gran variedad de sonidos, sin embargo, para poder realizar un mejor estudio del comportamiento del oído humano, es razonable cambiar la escala de medición de los sonidos audibles, para hacerlos mas manejables.

En primer lugar, se puede estudiar, en lugar de las presiones de los sonidos, las intensidades de los mismos. La intensidad puede ser expresada como

$$I = \frac{P^2}{\rho\nu} \tag{1.13}$$

con ρ la densidad del medio de propagación y ν la velocidad de propagación del sonido en el medio.

Al estudiar la intensidad de los sonidos en lugar de la presión se ha realizado un cambio de medición cuadrático, lo cual, a un principio, aumentará el rango de valores permisibles a ser estudiados, sin embargo generalmente para determinar si un sonido es fuerte o no es necesario compararlo con otro, es decir, es imperante establecer una *intensidad de referencia* para poder establecer una comparación en cuanto a intensidades en este caso.

Sea I_0 esta intensidad de referencia, que generalmente se toma como el umbral auditivo o el sonido mádébilil que el oído puede percibir, entonces es conveniente estudiar el comportamiento de

$$\frac{I}{I_0} \tag{1.14}$$

lo cual dará una medición de la intensidad tomando como referencia el valor de I_0, sin embargo el rango de los valores aún es grande y un pocdifícilil de poder manipular, por lo que se acostumbra a trabajar en una escala logarítmica, definiendo así el *nivel de presión sonora* como

$$L_P = 10 \log \left(\frac{I}{I_0} \right) \tag{1.15}$$

o equivalentemente en su expresión utilizando presiones,

Sonido	Presión [Pa]	Nivel de Presión Sonora [dB]
Umbral de dolor	20	120
Discoteca a todo volumen	6.3	110
Martillo neumático a $2m$	3.6	105
Ambiente industrial ruidoso	0.63	90
Piano a $1m$ con fuerza media	0.20	80
Automóvil silencioso a $2m$	0.063	70
Conversación normal	0.020	60
Ruido urbano de noche	0.0063	50
Habitación interior (día)	0.0020	40
Habitación interior (noche)	0.00063	30
Estudio de grabación	0.00020	20
Cámara sonoamortiguada	0.000063	10
Umbral de audicin a $1kHz$	0.000020	0

Cuadro I: Valores de presión y nivel de presión para diversas situaciones

$$L_P = 20 \log \left(\frac{P}{P_0} \right) \tag{1.16}$$

Debido a la definición de este parámetro, el nivel de presión sonora es medido en *decibeles*.

Algunos de los valores de presiones y niveles dpresiónon sonora más comunes se muestran en la siguiente tabla:

Como se observa en la tabla, los límites auditivos para las presiones van desde, como se mencionó anteriormente, los $20\mu Pa$ hasta los $20Pa$, teniendo un rango de $1:10^6$, mientras que en los niveles de presión sonora se tiene un rango desde $0dB$ hasta $120dB$, siendconsiderablementete más pequeño que el anterior.

2. AUDIO DIGITAL

La grabación de sonidos ha sido una de las actividades que el hombre ha venido desarrollando desde mucho tiempo atrás. Esto surge por la necesidad de poder reproducir y exhibir arte. A diferencia de la pintura y escultura, la música es un tipo de arte que necesita del tiempo para poder ser interpretado, al igual que el teatro y la danza. Debido a esto, el hombre trata de *guardar* de alguna manera la música, para poder reproducirla después.

Un primer acercamiento, fue la utilización de aparatos mecánicos, como las *Cajas de Música*. Las cajas de música son aparatos mecánicos que ejecutan música grabada en un rodillo dentado. Estas se desarrollaron en el siglo XVIII, aunque su popularidad se dió durante el siglo XIX. Se puede decir que el primer tipo de grabación sonora realizado fue de este tipo.

Con este tipo de *grabaciones* era posiblúnicamentete reproducir música monofónica, ni pensar en voz humana, sin embargo en 1877 se logra un acercamiento a una mejor grabación de sonidos. A cargo de Thomas Alba Edison, se inventa el *Fonógrafo*.

El fonógrafo utiliza un sistema de grabación mecánica analógica en el cual las ondas sonoras son transformadas en vibraciones mecánicas mediante un transductor acústico-mecánico. Estas vibraciones mueven una aguja que labra un surco helicoidal sobre un cilindro de fonógrafo, en el cual queda grabado el sonido. Para reproducir el sonido se invierte el proceso.

Al principio se utilizaron varios tipos de cilindros para realizar las grabaciones, sin embargo el más duradero y de mayor calidad resulto ser el cilindro hecho de cera sólida.

Figura 13: Fonógrafo

Luego vino el *Gramófono* inventado por Emile Berliner, el cual era parecido al fonógrafo, sin embargo utilizaba discos en lugar de cilindros para la grabación y reproducción de sonidos.

El gramófono de Berliner, al igual que los tocadiscos desarrollados posteriormente, consta de un plato giratorio, un brazo, una aguja y un amplificador. Un motor eléctrico hace girar el plato a una velocidad constante de 33, 45 o 78 revoluciones por minuto.

El gramófono ganó popularidad y desplazó al fonógrafo rápidamentete, y este sistema de grabación y reproducción prevaleció por mucho tiempo como una buena forma de grabación analógica durante el siglo XIX y principio del siglo XX.

Con el descubrimiento del *electromagnetismo* a principios del siglo XX se comenzó a desarrollar una nueva manera de grabar y reproducir sonidos, la *Cinta Magnética*.

La grabación en cinta magnética utiliza principios electromagnéticos para su funcionamiento. Se puede decir que es la primer forma de grabación que realmente utiliza medios eléctricos y electrónicos para funcionar, pues anteriormente solo se

Figura 14: Gramófono

hacían uso de sistemas mecánicos para dicho efecto.

El proceso de grabación magnética consiste en convertir la señal sonora en una señal de voltaje a través de un transductor, un *Micrófono*. Esta señal de voltaje estimula un electroimán, el cual genera un campo magnético que incide sobre una cinta recubierta de material que posee alta *remanencia magnética*.

Así, la señal de audio original queda grabada como una señal magnética en la cinta. Para reproducir el sonido se realiza el proceso inverso, esto es, el campo magnético remanente en la cinta estimula el electroimán, generando este una corriente eléctrica proporcional, la cual alimenta las bocinas, que funcionan bajo el mismo principio que el micrófono, y estas generan el sonido.

Luego a principios de los años 80, surge una nueva forma de grabar datos, y en particular sonidos, el *Audio Digital*. Este proceso consistía en tomar la misma señal de voltaje producida por el transductor de sonido-voltaje como un micrófono, pero luego se realizaba un proceso denominado *conversión analógica-digital*, en la cual se tomaban muestras de la señal de voltaje cada cierto tiempo, y a la amplitud de la señal de voltaje se le asignaba un número de *bits*. De esta manera se guardaba el sonido en forma de *0's y 1's*.

Los primeros medios de almacenaje digital fueron también cintas magnéticas. Este cambio fue útilutil debido a que ahora era necesario distinguir entre dos estados guardados en la cinta, 0 ó 1, y no como antes que era un conjunto continuo de valores posibles.

Luego a mediados de los 80 surge el *Disco Compacto*, el cual guardaba también el sonido de manera digital, solo que ahora de forma óptica.

Con esto se puede ver de qué manera ha venido evolucionando la manera de guardar sonidos para su posterior reproducción.

2.1. Digitalización del Sonido

2.1.1. Muestreo

Para obtener una señal digital, primero debenrecogerseerse muestras en el tiempo de la señal original, puesto que la digitalización consiste en discretizar el tiempo en el cual se tiene la señal.

Para esto, se tiene un circuosciladorndor a una *frecuencia* f_s de muestreo, el cual conmuta la señal de origen con una salida, teniendo así muestras de la señal a una frecuencia f_s, es decir que cada muestra dura $t_s = \frac{1}{f_s}$ segundos.

Al realizar esto, se tienen ventanas de tiempo en las cuales se puede seguir a

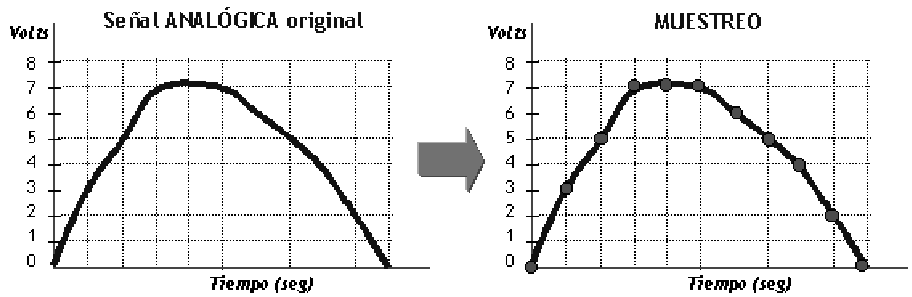

Figura 15: Muestreo de una Señal en el Tiempo

la señal original, sin embargo se tiene el problema de saber a que frecuencia debe de muestrearse para tener una buena representación de la señal.

Para poder determinar esto, es necesario determinarfrecuenciancia necesaria para poder reconstruir la señal original a partir de la señal muestreada, esto se logra por medio del teorema de Nyquist-Shannon.

Este teorema establece que:

2.1 Teorema (Nyquist-Shannon)**.** si se tiene una señal $s(t)$ que tiene componentes en frecuencia menores a f_m, puede ser muestreada y reconstruida sin pérdida de información si la frecuencia de muestreo es de al menos $2f_m$.

Con esto es posible tener una buena opción para la velocidad de muestreo mínima necesaria para poder reconstruir la señal, esto es, para que la conversión digital de la señal no afecte la calidad de la grabación, y mucho menos la reproducción del audio.

El problema luego consiste en que, por lógica se pensaría en que, a mayor frecuencia de muestreo, mejor es la calidad de la señal discretizada, pero esto no ocuasí asi.

Según el teorema de Nyquist-Shannon, la señal puede ser reconstruida totalmente si la frecuencia de muestreo es mayor que el doble de la frecuencia más alta de la señal original, por lo tanto esto provee una solución exacta y no aproximada a la señal fuente, por lo que da lo mismo, matemáticamente hablando, hacerlo a una frecuencia exageradamente alta, que a una levemente mayor que la frecuencia de Nyquist.

Al muestrear a una frecuencia muy elevada,úniconico que se obtiene es redundar en la información obtenida, puesto que con la muestra ala frecuencia de Nysquist es suficiente para poder realizar un buen modelado de la señal.

Así pues,comúnmenteente usada la frecuencia de Nyquist como la frecuencia de muestreo en las señales para así poder reducir el espacio que ocupara la señal al ser guardada digitalmente.

Si se realiza un muestreo por debajo de la frecuencia de Nyquist, se produce un efecto conocido como *Aliasing*. El Aliasing es el efecto que se produce al digitalizar una señal y no ser capaz de poreconstruirla delade una única manera.

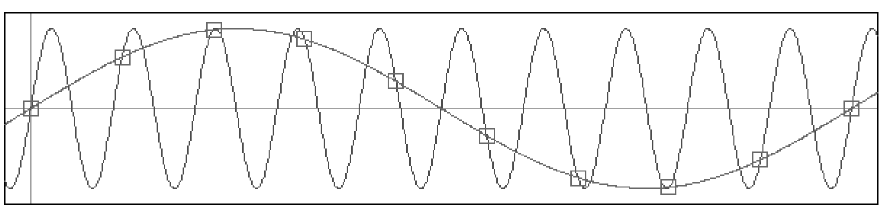

Figura 16: Ejemplo de Aliasing

En la imagen se puede observar que de la misma muestra se pueden reconstruir dos señales diferentes, sin embargo solo una cumple con el criterio de Nyquist, por lo tanto, sino se muestrea con este criterio presente, se produce una situación en donde es imposible de saber cuál es la imagen original.

Luego de ser muestreada la señal se procede a *cuantizarla*.

2.1.2. Cuantización

La cuantización de una señal consiste en asignarle un conjunto finito de valores a la señal. Así, una señal $s(t)$ se cuantifica en $n > 2$ estados $a_1, a_2, \cdot, a_n$ de acuerdo con un numero $n-1$ de valores de la señal, $v_1, v_2, \cdots, v_{n-1}$ de acuerdo con

- Si $s(t_s) \leq v_1$, se asigna a_1
- Si $v_{i-1} < s(t_s) \leq v_i$, para $1 < i < n$, se asigna a_i
- Si $v_{n-1} < s(t_s)$, se asigna a_n

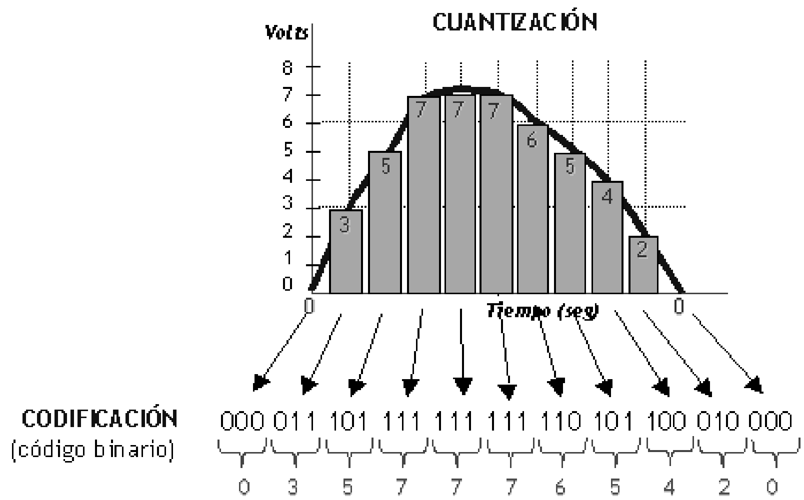

Figura 17: Proceso de Cuantización

Con esto, la señal muestreada se convierte en una señal de pulsos rectos, con la cual ya se vuelve más manejable para tratarla en un circuito lógico o una computadora.

Al realizar esta conversión se introduce *ruido* a la señal llamado *Ruido de Cuantización* debido a la pérdida al realizar la asignación de los estados finitos.

Este es un tipo de ruido que no se puede obviar, puesto que,importandoanto que tan precisa sea la conversión y el numero de niveles de cuantización, siempre no se tendrá el valor exacto de la señal originalmente muestreada.

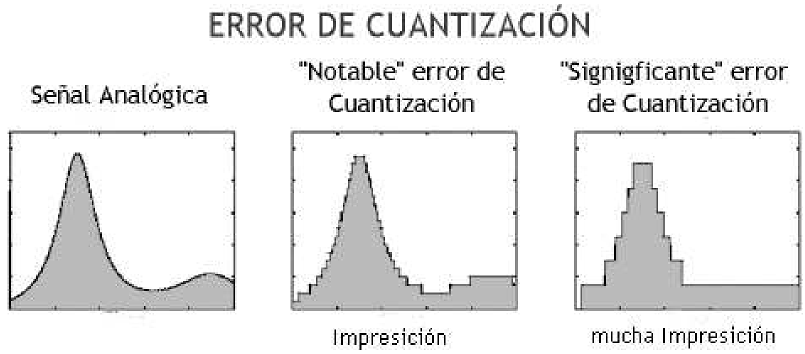

Figura 18: Error de Cuantización

2.1.3. Codificación

Luego de muestrear y cuantizar la señal, esta se *codifica*, esto es, se le asigna una secuencia de *bits* o números binarios a cada estado a_k con lo cual se maneja más fácilmenteente la señal en una computadora o sistema electrónico.

Dependiendo de n, el número de estados de cuatización, será el número de bits que serán utilizados para la codificación de la señal.

La codificación *física* de la señal puede codificarse de varias maneras, y esto sirve para poder evitar errores de transmisión o de *decodificación*.

Dentro de las codificaciones *Polares* o de dos estados se tienen

- NRZ
- RZ

- Bifase

2.1.3.1. Códigos NRZ

Este sistema de codificación, *Non Retrun to Zero*, consiste en que los bits se codifican con voltajes positivos y negativos, esto es, un valor positivo de voltaje se toma como un nivel alto y un negativo como un nivel bajo ó viceversa, dependiendo el tipo que se utilice.

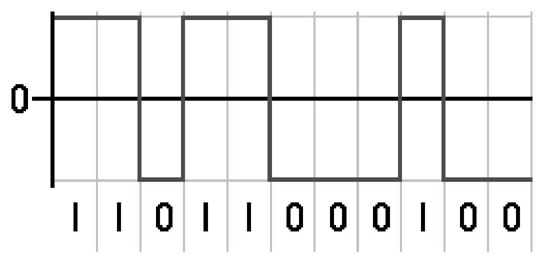

Figura 19: Codigo NRZ

Se utilizan dos tipos de NRZ, el NRZ-L y el NRZ-I

Codificación NRZ-L En esta codificación se interpreta un valor positivo de la señal como un nivel bajo y un negativo como un nivel alto.

Codificación NRZ-I Este es un tipo de codificación *diferencial*. Al recibirse un valor positivo de la señal quiere decir que no se produjo un cambio y al recibir un valor negativo indica que se produjo un cambio de 1 a 0 ó viceversa.

Este código pierde la señal de reloj, y su principal ventaja es que al emplear pulsos de larga duración requiere menor ancho de banda que otros sistemas de codificación que emplean pulsos más cortos.

2.1.3.2. Código RZ

En este tipo de código, la señal toma valores positivos o negativos durante la primera mitad del tiempo del pulso, dependiendo el estado lógico de la señal, y luego, durante la segunda mitad del tiempo del pulso, se va a cero.

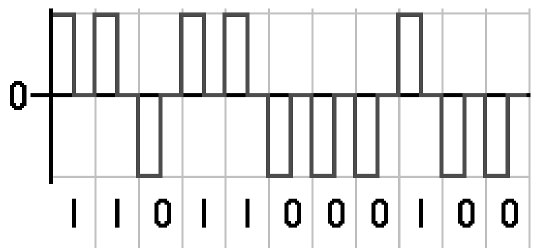

Figura 20: Codigo RZ

Este código utiliza el doble del ancho de banda que el anterior, sin embargo, si es posible obtener la señal de reloj debido a que en reloj del receptor queda sincronizado por la cadencia de los pulsos que llegan del transmisor puesto que todos los bits tienen una transición, esto permite identificar a cada bit en una larga cadena de unos o ceros.

2.1.3.3. Códigos Bifase

En esta codificación se tiene que la señal se mantiene en un valor la primera mitad del tiempo del pulso y cambia de polaridad la segunda mitad, por lo tanto, la señal toma valores positivos ó negativos, no cero.

Los dos tipos principales tipos de esta codificación son la codificación *Manchester* y *Manchester Diferencial*.

Codificación Manchester En esta codificación, si la señal tiene un estado alto, la primera mitad del tiempo del pulso es positiva y la segunda es negativa, y cuando la señal tiene un estado bajo, la primera mitad del tiempo del pulso es negativa y la segunda es positiva.

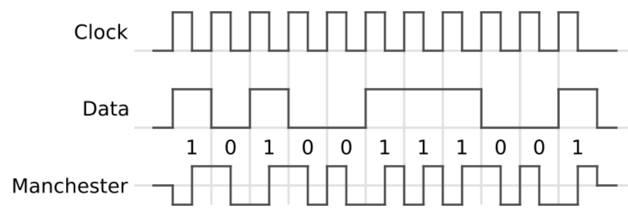

Figura 21: Código Manchester

La codificación Manchester provee una forma simple de codificar secuencias de bits, incluso cuando hay largas secuencias de periodos sin transiciones de nivel que puedan significar la perdida de sincronización, o incluso errores en las secuencias de bits.

En esta codificación se mandan la información de datos así como de la señal de reloj.

Codificación Manchester Diferencial Es similar a la anterior, solo que esta codifica en lugar de los estados de la señal, las transiciones de esta.

La codificación de un 0 se representa por la presencia de una transición al principio del intervalo del bit, y un 1 1a se representa mediante la ausencia de transición.

El código Manchester diferencial tiene las mismas ventajas de los códigos Manchester con la adición de las ventajas derivadas de la utilizacin de una aproximacin diferencial.

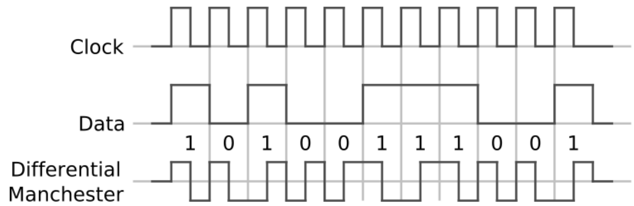

Figura 22: Codigo Manchester Diferencial

Para mejorar el manejo de las señales eléctricas, se desarrollaron sistemas de codificación *Bipolares*, los cuales utilizan 3 estados para representar los niveles lógicos de la señal.

Se usan tres valores, positivo, negativo y cero. El nivel de voltaje cero se utiliza para representar un bit 0. Los bits 1 se codifican como valores positivo y negativo en forma alterna. Si el primer 1 se representa por una amplitud positiva, el segundo se representa por una amplitud negativa, el tercero positiva, así de forma alternada. Siempre se produce una alternancia entre los valores de amplitud para representar los bits 1, aunque estos bits no sean consecutivos.

Los 3 tipos más comunes de codigos bipolares son

- AMI
- B8ZS
- HDB3

2.1.3.4. AMI

En la codificación AMI, *Alternate Mark Inversion*, los bits 1 se codifican con las polaridad inversa a la del anterior 1 codificado, el bit 0 con cero voltios. Esta es la manera sobre las cuales se basan las demás codificaciones bipolares.

2.1.3.5. B8ZS

La codificación *Bipolar 8-Zero Substitution*, consiste en que cuando aparecen 8 bits 0 consecutivos, B8ZS introduce cambios artificiales (violaciones y transiciones de polaridad) en el patrón, basados en la polaridad del último bit 1 codificado de la siguiente manera:

- V: Violacin, mantiene la polaridad anterior en la secuencia.
- B: Transicin, invierte la polaridad anterior en la secuencia.

Los ocho ceros se sustituyen por la secuencia: 000V B0VB

Esta es utilizada principalmente Estados Unidos.

2.1.3.6. HDB3

Esta es parecida a la anterior.En la *High Density Bipolar 3*, cuando aparecen cuatro bits 0 consecutivos, estos se sustituyen por una de las dos siguientes secuencias:

- Si el nmero de bits 1 es impar desde la última sustitución o se trata de la primera sustituci'on realizada, HDB3 los sustituye por la secuencia: 000V
- Si el número de bits 1 es par desde la última sustitución, HDB3 los sustituye por la secuencia: B00V

Esta codificación es usada en Europa y Japón.

2.1.4. Modulacion Digital

También como parte del proceso de digitalización de la señal, se procede a realizar modulación digital. Esto es, la manera en que se interpretan los datos binarios para producir la señal de audio de salida.

Las principales formas de realizarlo son

- PAM
- PCM
- DPCM
- ADPCM

2.1.4.1. PAM

La modulacin de amplitud de pulso, *Pulse Amplitud Modulation*, es la forma más simple de modulación de pulso. Esta técnica transmite datos variando las amplitudes del voltaje o de la energía de pulsos individuales en una secuencia sincronizada de pulsos electromágneticos. Es decir los datos que se transmitirán se codifican en la amplitud de una serie de pulsos de la señal.

Si se ve esto de un punto de vista puramente teórico, las amplitudes de pulso posibles en la modulación de amplitud de pulso pueden ser infinitas. Éste es el caso con la modulación de amplitud análoga de pulso. Una modulación de amplitud de pulso discreto tendrá un número finito de estados que puede tomar la señal eléctrica.

Su mayor importancia radica en que favorese el multiplexado de señales dentro de un mismo canal.

2.1.5. PCM

Es un proceso digital de modulación para convertir una señal analógica en un código digital. La señal analógica se muestrea, es decir, se mide periódicamente. En un convertidor analógico/digital, los valores medidos se cuantifican, se convierten en un número binario y se descodifican en un tren de impulsos. Este tren de impulsos es una señal de alta frecuencia portadora de la señal analógica original.

La modulación PCM, *Pulse Code Mudulation*, modifica los pulsos creados por PAM para crear una señal completamente binaria. Para hacerlo, PCM, en primer lugar, cuantifica los pulsos de PAM de acuerdo como en 2.1.2.

De esta manera, a cada amplitud de PAM se le hace corresponder un *Byte*, en donde 7 bits representan la amplitud de la señal y 1 bit indica el signo, positivo o negativo.

2.1.6. DPCM

La modulación DPCM, *Differential Pulso Code Modulation*, está basada en la modulación PCM. Esta no codifica la señal en sí, sino que toma la diferencia del valor anterior y el presente de la señal, convirtiendo así la diferencia de igual manera, asignándole un byte que tiene la información de la amplitud y el signo de la *diferencia* de la señal.

Cuando DPCM se codifica utilizando únicamente 1 bit se tiene una modulación especial llamada *Modulación Delta*.

2.1.7. ADPCM

Este sistema, *Adaptive Differencial Pulso Code Modulation*, se comporta de igual manera que el DPCM, solo que el tamaño del escalón del convertidor digital analógico varia automáticamente, dependiendo de las características de amplitud de la señal de entrada analógica. Cuando el resultado del transmisor es una cadena de unos o ceros consecutivos , indica que el resultado de la pendiente de la señal del conversor digital analógico es menor que la pendiente de la señal analógica en la dirección positiva o negativa.

El convertidor ha perdido noción exacta de donde están las muestras analógicas. Con un ADPCM, después de un número predeterminado de unos o ceros consecutivos , el tamaño del escalón se incrementa automáticamente. Si la amplitud de la salida del conversor aún esta debajo de la amplitud de la muestra , el siguiente escalón se incrementa aun más , hasta que al fin el conversor alcanza a la señal analógica. Cuando esta ocurriendo una secuencia alternada de unos y ceros, esto indica que la posibilidad de que ocurra ruido es alto ,pero el conversor se revertirá automáticamente a su tamaño mínimo de escalón y , por lo tanto, reducirá la magnitud del error de ruido.

2.2. Calidad del Audio Digital

Ya dentro del ambiente computacional, existen varias formas de tratar las senñales digitales que continen la información de audio, esto con el objetivo principal de reducir el espacio que ocupan los archivos y de aumentar la calidad con la que estos pueden reproducir los sonidos.

Luego de tener establecida la manera en que se comportan las señales digitales a nivel fśico o eléctrico, viene el tratamiento ya en sí de las representaciones binarias de los datos, dentro de una computadora.

En el ámbito computacional y de circuitos lógicos, estas señales son representadas simplemente por cadenas de 0's y 1's, y dependiento de las *convenciones* utilizadas o las formas de *codificación* utilizadas, es posible alcanzar los dos objetivos antes mencionados.

2.2.1. Parámetros del Audio Digital

A la hora de reproducir un archivo de audio digital, es necesario conocer ciertas propiedades de como fue guardada la señal originalmente.

Las tres propiedades básicas que se deben considerar son

- Número de Canales
- Tasa de Muestreo
- Número de Bits por Muestra

2.2.1.1. Número de Canales

Este parámetro indica cuantas señales de audio se reproducirán simultaneamente al abrir el archivo. Al principio, las primeras grabaciones realizadas de manera analógica fueron *Monofónicas* , es decir, solamente tenian una señal de audio, luego se incorporaron las grabaciones *Estereofónicas*, en donde se tenian dos señales a reproducir, una para la bocina Izquiera y otra para la Derecha.

El sonido estereofónico fue muy difundido por mucho tiempo, y en la actualidad es muy utilizado, en la música de los CD, en las transmiciones de las estaciones de radio, programas de TV, etc.

Otros sistemas fueron desarrollandose poco a poco, siendo uno de los primeros el *Cuadrafónico*, siendo este la base de los sistemas actuales 5.1 en adelante.

En la acutalidad los sistemas más usados son

- 1 Canal
- 2 Canales
- 2.1 Canales
- 4 Canales
- 5.1 Canales
- 6.1 Canales
- 7.1 Canales

Los ultimos 3 sistemas de mas de 5 canales son conocidos como sistemas de sonido envolventes. y son utilizados mayormente en salas de cine y salas de exposiciones.

Cuando se tiene solamente amplificada una frecuencia de una señal de audio y se transmite por otro canal, se dice que este en sí no constituye un canal completamente, puedo que solamente reproduce sonidos en cierto rango reducido de frecuencias, y por sí solo no aporta información inteligible, por lo que se denota como 0.1 canales. En los sistemas 2.1, 5.1, 6.1 y 7.1 se tiene pues, que se poseen 2, 5, 6 y 7 canales respectivamente, de señales de audio completas, y se tiene otro destinado unicamente a realzar los sonidos bajos.

2.2.1.2. Tasa de Muestreo

Como se vió en 2.1.1, para convertir una señal analógica en digital, el primer paso consiste en realizar un muestreo de ésta, o lo que es igual, tomar diferentes muestras de tensiones o voltajes en diferentes puntos de la señal. La frecuencia a la que se realiza el muestreo se denomina razón, tasa o también frecuencia de muestreo y se mide en kilohertz $[kHz]$. En el caso de una grabación digital de audio, a mayor cantidad de muestras tomadas, mayor calidad y fidelidad tendrá la señal digital resultante.

Durante el proceso de muestreo se asignan valores numéricos equivalentes a la tensión o voltaje existente en diferentes puntos de la señal, con la finalidad de realizar a continuación el proceso de cuantización.

Las tasas o frecuencias de muestreo más utilizadas para audio digital son las siguientes:

- 24 000 muestras por segundo (24 kHz)
- 30 000 muestras por segundo (30 kHz)
- 44 100 muestras por segundo (44.1 kHz) (Calidad de CD)
- 48 000 muestras por segundo (48 kHz)

Por tanto, una señal cuyo muestreo se realice a 24 kHz, tendrá menos calidad y fidelidad que otra realizada a 48 kHz. Sin embargo, mientras mayor sea el número de muestras tomadas, mayor será también el ancho de banda necesario para transmitir una señal digital, requiriendo también un espacio mucho mayor para almacenarla en un CD o un DVD.

En la grabación de CDs de música, los estudios de sonido utilizan un estándar de muestreo de 44.1 kHz a 16 bits. Esos son los dos par'ametros requeridos para que una grabación digital cualquiera posea lo que se conoce como calidad de CD.

2.2.1.3. Número de Bits por Muestra

Este parámetro indica la *resolución* con que se tiene el sonido. Viene dado por el número de niveles de cuantización utilizados para digitalizar el sonido, como está en 2.1.2.

Entre mayor la cantidad de bits utilizados, mayor será la cantidad de estados de cuantización, y por lo tanto, menor el *ruido de cuantización* de la señal obtenida.

Así, se tiene que la relación entre el número de bits n y el número de estados es 2^n.

Por ejemplo:

Bits	Estados
1 bit	2 estados
8 bits	256 estados
16 bits	65,536 estados
32 bits	4,294,967,296 estados

2.2.2. Compansión

La compasión es una técnica utilizada con el objetivo de mejorar la calidad del sonido a la hora de realizar la cuantización. Además ayuda a disminuir el ruido a la hora de realizar una transmición de datos.

Basicamente consiste en dos etapas, una etapa de *Compresión* y la otra de

Expensión.

La idea es utilizar más estados de cuantización para describir los niveles bajos de la señal, y menos para los niveles más altos. Esto debido a que los sonidos con poca amplitud son más probables que los sonidos con una amplitud mayor, así, con esta técnica es posible disminuir el ruido debido a la cuantización que sufre la señal.

A la hora de realizar la cuantización de una señal como en 2.1.2, habitualmente se toman todos los estados de igual amplitud, i.e $v_{i+1} - v_i = c$ para $0 \leq i \leq n$, pero en la compansión se utiliza un rango *dinámico* de cuantización, en el cual los estados más bajos son más pequeños que los altos, por lo tanto se tiene que los sonidos bajos tendrán mejor resolución en compansión que en cuantización lineal.

En general, la compansión consiste en mejorar la resolución de un rango de la señal que es más probable de ocurrir y a cambio se pierde calidad en los rangos menos probables.

Por lo tanto la característica de la cuantización por compansión vendrá dada por el tipo de señales a tratar. Por ejemplo, para señales de voz es necesario un rendimiento cuadrático medio relativamente constante, lo que significa que la distorsión debe ser proporcional a la amplitud de la señal para cualquier nivel de señal de entrada. Esto requiere una razón de compresión logarítmica. Existen dos métodos de compresión analógicos que se aproximan a una función logarítmica, y son conocidos como Ley μ y Ley A.

2.2.2.1. Ley μ

Este es un sistema de cuantificación logarítmica de una señal de audio. Es utilizado principalmente para audio de voz humana dado que explota las características de ésta. Su aplicacin cubre el campo de comunicaciones telefónicas actualmente. Este sistema de codificación es usado en EEUU y el Japón.

Las señales de voz están formadas en gran parte por amplitudes pequeñas, ya que son las más importantes para la percepción del habla, por lo tanto éstas son las más probables. En cambio, las amplitudes grandes no aparecen tanto, por lo tanto tiene una probabilidad de aparición muy baja.

En el caso de que una señal de audio tuviera una probabilidad de aparición de todos los niveles de amplitud por igual, la cuantificación ideal sería la uniforme, pero en el caso de la voz humana esto no ocurre, estadísticamente aparecen con mucha más frecuencia niveles bajos de amplitud. El algoritmo Ley μ explota el factor de que los altos niveles de amplitud no necesitan tanta resolución como los bajos. Por lo tanto, si damos más niveles de cuantificación a las bajas amplitudes y menos a las altas conseguiremos más resolución, un error de cuantificación inferior y por lo tanto una relación *Señal a Ruido* superior que si efectuáramos directamente una cuantificación uniforme para todos los niveles de la señal.

Cuando una señal pasa a través de un compresor, el intervalo de las amplitudes pequeñas de entrada es representado en un intervalo más largo en la salida, y el intervalo de las amplitudes más elevadas pasa a ser representado en un intervalo más pequeño en la salida.

Digitalmente, todo este esquema es equivalente a aplicar una cuantificación no uniforme (logarítmica) a la señal original, donde se tendrá pequeños pasos de cuantificación para los valores pequeños de amplitud y pasos de cuantificación grandes para los valores grandes de amplitud. Para recuperar la señal en el destino se tendrá que aplicar la función inversa.

Por lo tanto, la implementación del sistema consiste en aplicar a la señal de entrada una función logarítmica y una vez procesada realizar una cuantificación uniforme. Es lo mismo que decir que el paso de cuantificación sigue una funcin del tipo logarítmico

$$F(x) = sgn(x)\frac{\log(1+\mu|x|)}{\log(1+\mu)} \tag{2.1}$$

en donde la x representa la señal de entrada, la cual está entre -1 y 1, y la μ indica el factor de compresión utilizado.

2.2.2.2. Ley A

La Ley A de compansión es muy parecida a la Ley μ, y se basa en los mismos hechos que esa. Esta Ley es más utilizada en Europa y es la contrapartida de la Ley μ.

La función que define a esta Ley está dada por

$$F(x) = \begin{cases} sgn(x)\frac{A|x|}{1+\log A} & |x| < \frac{1}{A} \\ sgn(x)\frac{1+\log A|x|}{1+\log A} & \frac{1}{A} < |x| \leq 1 \end{cases} \tag{2.2}$$

donde A es el parámetro de compresión. En Europa $A = 87{,}7$. También se usa el valor 87,6

3. FORMATOS DE AUDIO

Luego de haber convertido la señal de audio en una forma de audio digital, es decir, a una cadena de 0's y 1's, el manejo de las señales de audio se tranforma en el análisis y tratamiento de estas cadenas de bits.

El manejo y manipulación de estas cadenas de bits se realiza para obtener principalmente dos resultados

- Mejorar Calidad
- Reducir Espacio

Por el constante intercambio de información que se realiza actualmente es necesario optimizar el espacio ocupado por los archivos, sin embargo, también es de suma importancia lograr este objetivo tenien siempre una buena calidad en los archivos de audio, para tener la menor pérdida de calidad posible a la hora de reproducirlos.

A la manera de *transformar* las cadenas de bits se les conoce como *Formatos de Audio*. Estos simplemente la manera de guardar la información de audio en un archivo para posteriormente ser reproducidos.

Prinipalmente se puede hablar de formatos con compresión y sin compresión.

Los formatos sin compresión guardan la información tal y como se obtiene como resultado de la conversión análoga digital. Estos no procesan de ninguna manera los datos, y por lo tanto no tienen ninguna propiedad específica sobre manejo y calidad de los archivos de audio.

Para guardar las información en archivos, de una manera comprimida, se realiza una *codificación* de la información, la cual tiene como objetivos

- Evitar errores a la hora de transmitir archivos
- Comprimir el espacio utilizado
- Evitar errores en la reproducción

3.1. Formatos no Comprimidos

Estos tipos de formatos se basan principalmente en la codificación PCM de la señal al cuantizarla.

Los tres tipos de formatos de audio no comprimido más conodios son

- WAV (Windows)
- AIFF (Mac)
- AU (Unix)

Este tipo de formatos usan el mismo espacio para guardar cualquier tipo de sonido, puesto que cada muestra se guarda exactamente como fue registrada, ya sea una muestra de silencio o de ruidos aleatorios.

3.1.1. WAV

WAVEform audio format o WAV, es un formato de audio digital sin compresión de datos desarrollado y propiedad de Microsoft y de IBM que se utiliza para almacenar sonidos en el PC, admite archivos mono y estéreo a diversas resoluciones y velocidades de muestreo, su extensin es .wav.

Es una variante del formato RIFF (Resource Interchange File Format, formato de fichero para intercambio de recursos), método para almacenamiento en ”paquetes”, y relativamente parecido al IFF y al formato AIFF usado por Macintosh. El formato toma en cuenta algunas peculiaridades de la CPU Intel, y es el formato principal usado por Windows.

A pesar de que el formato WAV puede soportar casi cualquier codec de audio, se utiliza principalmente con el formato PCM (no comprimido) y al no tener pérdida de calidad puede ser usado por profesionales. Por ejemplo, para tener calidad disco compacto se necesita que el sonido se grabe a 44100 Hz y a 16 bits, por cada minuto de grabación de sonido se consumen unos 5 megabytes de disco duro. Una de sus grandes limitaciones es que solo se puede grabar un archivo de hasta 4 gigabytes, que equivale aproximadamente a 6,6 horas en calidad disco compacto. Es una limitación propia del formato, independientemente de que el sistema operativo donde se utilice sea MS Windows u otro distinto, y se debe a que en la cabecera del fichero se indica la longitud del mismo con un número entero de 32 bit, lo que limita el tamao del fichero a 4 GB.

3.1.2. AIFF

El formato AIFF, *Audio Interchange File Format* está muy extendido en plataformas Apple. Se fundamenta en el formato IFF de Electronic Arts, que permite almacenar la informacin en segmentos o *chunks*.

Al tratarse de un fichero de formato IFF, debe contener al comienzo una cabecera con un primer campo de 4 bytes que contiene la palabra ”FORM”, un segundo campo de 4 bytes que indica la longitud del resto del fichero. Por último para identificar el fichero IFF como contenedor de audio AIFF, los 4 bytes siguientes a la cabecera deben contener la palabra .AIFF"

A partir de la cabecera y la identificación el resto del fichero se compone de una secuencia de segmentos. Cada segmento se compone a su vez de una cabecera de segmento compuesta por 4 bytes de identificación y 4 bytes de longitud del campo de datos. Esta longitud no incluye ni la cabecera ni el posible byte que pueda haber para hacer que su longitud total sea par. El orden de estos segmentos es irrelevante.

Solo existe un segmento obligatorio denominado Segmento Común (ÇOMM") y en el caso de que la forma de onda tenga longitud mayor que cero, también es obligatoria la existencia del segmento Datos de Sonido ("SSND"). El resto de segmentos son opcionales y los programas de reproducción podrán ignorarlos selectivamente. Sin embargo a la hora de copiar el fichero se deben copiar la totalidad de los segmentos incluidos los que son ignorados en la reproducción.

La ordenación de los bytes en formato AIFF es de tipo*big-endian* como en el microprocesador 68000 de Motorola. Las muestras de la señal se almacenan en el menor número entero bytes, rellenando los bits sobrantes con ceros. En cada muestra los bits de información se sitúan en las posiciones de mayor peso, quedando el relleno de 0's en las posiciones menos significativas.

Las reproducciones multicanal se organizan de la siguiente forma: los muestras se agrupan en tramas de muestra, que son un conjunto de muestras, cada una de las cuales corresponde a un canal distinto. Está definido el siguiente orden para las siguientes situaciones:

- Estéreo: Izquierdo Derecho
- Tres canales: Izquierdo Derecho Central
- Cuadrafónico: Delantero Izquierdo Delantero Derecho Trasero Izquierdo Trasero Derecho

- Cuatro canales: Izquierdo Central Derecho Entorno
- Seis canales: Central Izquierdo Izquierdo Central Central Derecho Derecho Entorno

Las muestras pertenecientes a una trama de muestra se empaquetan una tras otra, sin rellenos, al igual que las tramas de muestra entre sí.

3.1.3. AU

Es un formato de fichero muy asociado a máquinas Sun y Next. Su estructura es muy sencilla, las razones de compresión que puede llegar a ofrecer son pequeñas y destaca sobre todo su soporte de longitudes de muestras muy altas comparadas con otros formatos (32 y 64 bits). Se compone de tres secciones:

- Una cabecera, en la que describe la codificación de audio utilizada
- Un campo de longitud variable para almacenar otro tipo de informacin como texto en formato ASCII
- El conjunto de los datos de audio

Tras la cabecera AU se puede colocar un campo de información de propósito y formato libre. La longitud de este campo está determinado por el campo offset de la cabecera. La cabecera tiene longitud fija, de 24 bytes, por lo que este campo tiene una longitud de offset 24 bytes. Sus usos fundamentales son la inclusión de información de copyright en el fichero y la descripción del mismo utilizando caracteres ASCII imprimibles.

Luego viene el campo de datos. Este comienza en la posición indicada por offset y puede tener longitud no definida. En configuraciones de formato multicanal las tramas se agrupan en tramas de muestra, tal que el campo de datos es una sucesión de tramas de muestra. Una trama de muestra contiene tantas muestras como número de canales, y cada una de las muestras corresponde a un canal diferente. Para la reproduccin del fichero será necesario que se obtenga del mismo una tasa de samplerate tramas de muestra por segundo.

3.2. Codecs

A la forma de tratar las señales para mejorar su calidad y minimizar su espacio en disco se le conoce como *Codec*, que proviene de *Codificación-Decodificación*. Un codec basicamente provee las reglas para *empaquetar* la información y luego desempaquetarla.

Estos se dividen en dos categorías

- Sin pérdida
- Con pérdida

Los codecs sin pérdida tienen como objetivo principal el presenvar la calidad de las señales de audio y guardar la información de tal manera que se pueden evitar y corregir errores a la hora de reproducir o transmitir los archivos, mientras que los codecs con pérdida tienen como fin primordial el comrimir el tamaño de los archivos, teniendo un poco de pérdida en la calidad del sonido.

Los codecs de audio se caracterizan por los siguientes parámetros

- Número de Canales
- Frecuencia de Muestreo

- Número de Bits por Muestra

- Pérdida

3.2.0.1. Número de Canales

El Número de canales se refiere a que un flujo de datos codificado puede contener una o más señales de audio simultáneamente. De manera que puede tratarse de señales como en 2.2.1.1. Los codec de audio multicanal se suelen utilizar en sistemas de entretenimiento de *Teatro en Casa* o salas de cine.

3.2.0.2. Frecuencia de Muestreo

La Frecuencia de muestreo se tiene que, de acuerdo con el teorema de Nyquist, determina la calidad percibida a través de la máxima frecuencia que es capaz de codificar, que es precisamente la mitad de la frecuencia de muestreo. Por tanto, cuanto mayor sea la frecuencia de muestreo, mayor será la fidelidad del sonido obtenido respecto a la señal de audio original, puesto que, a pesar que como en 2.1 se dice que se tiene el mismo resultado al muestrear a frecuencias muy altas que a la frecuencia de Nyquist, esto se daba porque se tenía un filtro pasa bajo ideal que reconstruía la senñal que hacía compensación de la información, sin embargo en la práctica esto no es posible, por lo tanto a mayor frecuencia se tiene una mejor calidad en la señal. Por ejemplo, para codificar sonido con calidad CD nunca se usan frecuencias de muestreo superiores a 44,1 Khz, ya que el oído humano no es capaz de escuchar frecuencias superiores a 22 kHz.

3.2.0.3. Número de Bits

Como se vió en 2.2.1.3, el Nmero de bits por muestra determina la precisión con la que se reproduce la señal original y el rango dinámico de la misma. Se suelen utilizar 8 (para un rango dinámico de hasta 45 dB), 16 (para un rango dinámico de hasta 90 dB como el formato CD) o 24 bits por muestra (para 109 a 120 dB de rango dinámico). El más común es 16 bits.

3.2.0.4. Pérdida

La Pérdida se refiere a que algunos codecs pueden eliminar frecuencias de la señal original que, teóricamente, son inaudibles para el ser humano. De esta manera se puede reducir la frecuencia de muestreo. En este caso se dice que es un codec con pérdida o *Lossy* codec. En caso contrario se dice que es un codec sin pérdida o *Lossless* codec.

El parámetro tasa de bits o bit-rate es el número de bits de información que se procesan por unidad de tiempo, teniendo en cuenta la frecuencia de muestreo resultante, la profundidad de la muestra en bits y el número de canales. A causa de la posibilidad de utilizar compresión (con o sin pérdidas), la tasa de bits no puede deducirse directamente de los parámetros anteriores

3.2.1. Codecs sin Pérdida

Estos codifican los datos de manera que no se pierda la información que contienen. Los codecs sin pérdidas que se han desarrollado son capaces de completar su misión reduciendo también el tamaño de los archivos, es decir, realizando una compresión de ellos, aunque en menor escala que los formatos con pérdidas.

Los principales codecs sin pérdida que se utilizan actualmente son

- Apple Lossless (ALAC).
- Direct Stream Transfer (DST).
- FLAC (Free Lossless Audio Codec).
- Lossless Audio (LA).
- LOSSLESS AUDIO COMPRESSION WITH Ltac
- LPAC (Lossless Predictive Audio Codec).
- Monkey's Audio (APE).
- OptimFROG.
- RealAudio Loseless.
- RKAU.
- Shorten (SHN).
- True Audio (TTA).
- WavPack.
- Meridial Lossless Packing (MLP).

3.2.2. Codes on Pérdida

Estos proveen una mayor tasa de compresión por medio de la eliminación de cierta información que puede ser impresindible.

Se basan en los rangos de frecuencias perceptibles al oído humano, y eliminan las frecuencias que teóricamente no pueden ser escuchadas por el oído, con esto se reduce la información a ser guardada y por lo tanto el espacio necesario para su almacenamiento se disminuye.

Con estos codecs con pérdidas, es imposible lograr reconstruir integramente la señal original, puesto que se pierde información de esta.

Los principales codecs con pérdida que se utilizan actualmente son

- MP1 (MPEG audio layer-1).
- MP2 (MPEG audio layer-2).
- MP3 (MPEG audio layer-3).
- Advanced Audio Coding (AAC).
- Ogg Vorbis
- WMA (Windows Media Audio).
- Musepack
- AC3 (Dolby Digital A/52).
- DTS (Digital Theather Systems).
- ADPCM.
- ADX (usado en videojuegos).
- ATRAC (Adaptive TRansform Acoustic Coding).
- Perceptual Audio Coding
- TwinVQ

4. RECONSTRUCCIÓN DE SEÑALES DE AUDIO

Gran parte del análisis de las señales de audio se realizan en el *dominio del tiempo* como se ha visto hasta ahora, sin embargo, muchas propiedades muy importantes de estas señales se analizan de una mejor manera estudiandolas en el *domino de la frecuencia*, esto es, analizar las características de estas ondas en las frecuencias que las componen.

Para realizar este análisis en el dominio de la frecuencia se utiliza el *Análisis de Fourier*.

4.1. Análisis de Fourier

Joseph Fourier fué un matemático francés que realizó varios aportes en el área del análisis real. Uno de sus aportes más utilizados es el *Análisis de Fourier*. En un principio tuvo más que ver con termodinámica, sin embargo, actualmente es ampliamente utilizado en análisis de señales y en Procesamiento Digital de Señales.

4.1.1. Series de Fourier

El análisis de Fourier se centra en el estudio de las funciones continuas definidas en un intervalo cerrado $[a, b]$, denotadas por $C_{[a,b]}$, y denotando por T a $b - a$.

4.1 Definición (Producto Punto)**.** Si se tienen dos funciones, $f, g \in C_{[a,b]}$, se define el *Pruducto Punto* o *Producto Interno* como

$$< f, g >= \int_a^b f(t)g(t)dt \tag{4.1}$$

El producto punto así definido cumple con ser *lineal*, esto es:

- $< f, g >=< g, f >$
- $< f, g + h >=< f, g > + < f, h >$
- $< f, cg >= c < f, g >$ con c una constante

4.2 Definición (Ortogonalidad)**.** Si dos funciones $f, g \in C_{[a,b]}$ cumplen con que

$$< f, g >= 0 \tag{4.2}$$

se dicen que son *ortogonales entre sí*.

Un ejemplo muy útil de funciones ortogonales está dado por el siguiente teorema

4.1 Teorema. Sea

$$\mathcal{B} = \{\cos\left(\frac{2\pi nt}{T}\right), \sin\left(\frac{2\pi nt}{T}\right) |n \in \mathbb{Z}, t \in [a, b]\} \subset C_{[a,b]}$$

el conjunto de funciones senoidales y cosenoidales y sus múltiplos. Entonces si $f, g \in \mathcal{B}$ y $f \neq g$, se tiene que f y g son ortogonales.

Demostración. La demostración se realiza tomando la definición de producto punto. $\square$

Este hecho resulta de suma importancia, puesto que con esta propiedad, el conjunto $\mathcal{B}$ tiene características importantes que lo relacionan con los elementos de $C_{[a,b]}$

4.3 Definición (Dependencia Lineal)**.** Un conjunto $\{f_1, f_2, \cdots, f_n\} \subset C_{[a,b]}$ se dice *Linealmente Dependiente* si existen $\{\gamma_1, \gamma_2, \cdots, \gamma_n\} \in \mathbb{R}$ no todos cero, tales que

$$\gamma_1 f_1 + \gamma_2 f_2 + \cdots + \gamma_1 f_n \equiv 0 \tag{4.3}$$

Si el conjunto no es linealmente dependiente, se dice que es *Linealmente Independiente.*

Con esto se puede definir un objeto que es la base fundamental del manejo de funciones por medio del análisis de Fourier.

4.4 Definición. Un conjunto $\mathbb{B} = \{v_i\}_{i=0}^{\infty} \subset C_{[a,b]}$ se llama una *Base Ortogonal* de $C_{[a,b]}$ si cumple con que

- Los v_i sean linealmente independientes

- Si $v_i, v_j \in \mathbb{B}$, $i \neq j$, entonces $< v_i, v_j >= 0$

- Para toda $f \in C_{[a,b]}$, existen $\{\alpha_i\}_{i=0}^{\infty} \in \mathbb{R}$ tales que

$$f = \sum_{i=0}^{\infty} \alpha_i v_i \tag{4.4}$$

Se tiene que precisamente $\mathcal{B}$ cumple con estas propiedades, por lo que se tiene el siguiente teorema

4.2 Teorema. $\mathcal{B}$ es una base ortogonal de $C_{[a,b]}$

Esto quiere decir que si se tiene una función continua f definida en un cierto intervalo $[a, b]$, se pueden encontrar constantes $\{\alpha_i, \beta_i\}_{i=0}^{\infty} \in \mathbb{R}$ de tal manera que

$$f(t) = \frac{1}{2}\alpha_0 + \sum_{i=1}^{\infty} \alpha_i \sin\left(\frac{2\pi it}{T}\right) + \beta_i \cos\left(\frac{2\pi it}{T}\right) \tag{4.5}$$

Gracias a la propiedad de $\mathcal{B}$ de ser ortogonal, pueden encontrarse los α_i, β_i realizando productos punto. Así se tendrá que

$$
\begin{aligned}
< f, \sin\left(\frac{2\pi nt}{T}\right) > &=< \alpha_0 + \sum_{i=1}^{\infty} \alpha_i \sin\left(\frac{2\pi it}{T}\right) + \beta_i \cos\left(\frac{2\pi it}{T}\right), \sin\left(\frac{2\pi nt}{T}\right) > \\
&=< \alpha_0, \sin\left(\frac{2\pi nt}{T}\right) > + \\
&+ \sum_{i=1}^{\infty} < \alpha_i \sin\left(\frac{2\pi it}{T}\right) + \beta_i \cos\left(\frac{2\pi it}{T}\right), \sin\left(\frac{2\pi t}{T}\right) > \\
&= \sum_{i=1}^{\infty} < \alpha_i \sin\left(\frac{2\pi it}{T}\right), \sin\left(\frac{2\pi nt}{T}\right) > + \\
&+ \sum_{i=1}^{\infty} < \beta_i \cos\left(\frac{2\pi it}{T}\right), \sin\left(\frac{2\pi nt}{T}\right) > \\
&= 2T\alpha_n
\end{aligned}
\tag{4.6}
$$

de donde se tiene que

$$
\alpha_n = \frac{2}{T}\int_a^b f(t)\sin\left(\frac{2\pi nt}{T}\right) dt \tag{4.7}
$$

similarmente se puede obtener que

$$
\beta_n = \frac{2}{T}\int_a^b f(t)\cos\left(\frac{2\pi nt}{T}\right) dt \tag{4.8}
$$

Así, 4.5 es llamada, la *representación en serie trigonométrica de Fourier de* $f(t)$. Por conveniencia es frecuente llamar $\omega_n = \frac{2\pi n}{T}$, con esto se puede escribir 4.5 como

$$
f(t) = \frac{1}{2}\alpha_0 + \sum_{i=1}^{\infty} \alpha_i \sin(\omega_i t) + \beta_i \cos(\omega_i t) \tag{4.9}
$$

A menudo es conveniente el uso de números complejos para representar amplitud y fase de señales eléctricas, por lo que combinando 4.5 con la *identidad de Euler*

$$
e^{i\theta} = \cos\theta + i\sin\theta \tag{4.10}
$$

se obtiene la *representación en serie exponencial de Fourier de* $f(t)$, dada por

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{i\omega_n t} \tag{4.11}$$

De manera similar, se tiene que el conjunto de funciones $\{e^{i\omega_n t}\}_{n=-\infty}^{\infty}$ constituyen una base ortogonal de $C_{[a,b]}$, por lo que se tiene que

$$< e^{i\omega_n t}, e^{i\omega_m t} >= 0 \tag{4.12}$$

si $n \neq m$, y de igual manera puede utilizarse este hecho para obtener los valores de c_n a partir de $f(t)$

$$c_n = \frac{1}{T} \int_a^b f(t) e^{-i\omega_n t} dt \tag{4.13}$$

aca c_n toma valores sobre los números complejos.

El motivo de llamarle ω al argumento de la exponencial es por el hecho que $e^{-i\omega_n t}$ representa un vector *unitario* que gira al rededor del origen con una velocidad angular de ω_n. Al tomar el producto punto con $f(t)$, o lo que es lo mismo, hacer 4.13, se obtiene el contenido o componente de $f(t)$ que tiene a la frecuencia angular ω_n, por este motivo, los coeficientes c_n representan la *magnitud* y *fase* de la componente de $f(t)$ en la frecuencia $\nu_n = \frac{\omega_n}{2\pi}$.

Así, para analizar el comportamiento de $f(t)$ en frecuencia se realiza el estudio de su expanción en serie de Fourier, ya sea trigonométrica o exponencial.

4.1.2. Transformada de Fourier

Hasta aca se tiene que la función analizada solo existe en el intervalo $[a, b]$, por lo que se puede interpretar que esta es una función *periódica* con período T, sin embargo si se trata de una función no periódica, se puede extender este análisis, tomando el límite cuando T se vuelve infinito, así, una función no periódica puede interpretarse como una periódica con período infinito. Para esto, de 4.13 y 4.11 se tiene que una función periódica puede representarse por

$$f(t) = \sum_{n=-\infty}^{\infty} \left(\frac{1}{T} \int_a^b f(t) e^{-i\omega_n t} dt \right) e^{i\omega_n t} \tag{4.14}$$

de donde, si se supone a f como una función no periódica y se toma el límite, esto se convierte en

$$\begin{aligned} f(t) &= \lim_{T\to\infty} \sum_{n=-\infty}^{\infty} \left(\frac{1}{T} \int_a^b f(t) e^{-i\omega_n t} dt \right) e^{i\omega_n t} \\ &= \sum_{n=-\infty}^{\infty} \left(\lim_{T\to\infty} \frac{1}{T} \int_a^b f(t) e^{-i\omega_n t} dt \right) e^{i\omega_n t} \\ &= \sum_{n=-\infty}^{\infty} \left(\lim_{T\to\infty} \frac{2\pi}{2\pi T} \int_a^b f(t) e^{-i\omega_n t} dt \right) e^{i\omega_n t} \end{aligned} \tag{4.15}$$

de $\omega = \frac{2\pi}{T}$ se tiene que si $T \to \infty$, $\frac{2\pi}{T} \to d\omega$, y $\omega_n \to \omega$ por lo que

$$\begin{aligned} f(t) &= \sum_{n=-\infty}^{\infty} \left(\lim_{T\to\infty} \frac{d\omega}{2\pi} \int_a^b f(t) e^{-i\omega t} dt \right) e^{i\omega t} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \right) e^{i\omega t} d\omega \end{aligned} \tag{4.16}$$

Se denota como

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \tag{4.17}$$

y a esto se le conoce como la *Transformada de Fourier de* $f(t)$, y 4.16 se convierte en

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega \tag{4.18}$$

que se conoce como la *Transformada Inversa de* $F(\omega)$.

Cuando se realiza la transformada de una función periódica se obtiene la expanción en serie de dicha función, así pues $F(\omega)$ representa el contendio *espectral* de la frecuencia ω en $f(t)$, de la misma manera que c_n lo representaba en la serie de Fourier, además, $F(\omega)$, al igual que c_n, es generalmente un número complejo que representa la magnitud y fase de la componente en frecuencia.

Por esta razón a $|F(\omega)|$ se le conoce como la *densidad espectral de potencia*, puesto que da el tamaño de la componente de la función en la frecuencia ω.

A la transformada de Fourier de una función f suele denotarsele como $\mathcal{F}\{f(t)\}$ y a la transformada inversa por $\mathcal{F}^{-1}\{F(\omega)\}$, y por su definición, estas son *lineales* en t y ω respectivamente.

4.1.3. Convolución

La *Convolución* es una operación muy utilizada en el tratamiento y análisis de señales digitales, puesto que posee muchas interpretaciones.

4.5 Definición (Convolución)**.** Se define la convolución de dos funciones f y g como

$$(f \otimes g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau \tag{4.19}$$

La importancia de la convolución se puede ver cuando se realiza la transformada de esta, puesto que por la definición de transformada pueden obtenerse los siguientes resultados

4.3 Teorema. Si f y g son dos funciones cuyas transformadas de Fourier existen, entonces

$$\mathcal{F}\{f \otimes g\} = \mathcal{F}\{f\}\mathcal{F}\{g\} \tag{4.20}$$

4.4 Teorema. Si F y G son dos funciones cuyas transformadas inversas de Fourier existen, entonces

$$\mathcal{F}^{-1}\{F \otimes G\} = 2\pi\mathcal{F}^{-1}\{F\}\mathcal{F}^{-1}\{G\} \tag{4.21}$$

Estos resultados son de suma importancia, puesto que, como se se vió anteriormente, la transformada de una función representa su comportamiento en el dominio de la frecuencia, así, si se tiene que una de las funciones representa una señal de entrada y la otra la *función de transferencia* de un sistema, esto es, su comportamiento en frecuencia, se tendra que para obtener la salida del sistema basta con convolucionar la entrada con la función de transferencia.

Así mismo, un sistema puede representarse por medio de su función de transferencia, denotada por $H(\omega)$ en el caso de la frecuencia, y por $h(t)$ en el cso del tiempo. Esta función de transferencia tiene una relación estrecha con el comportamiento del sistema cuando es exitado por un pulso en la entrada. Esto puede verse por medio del resultado obtenido en 4.21.

Un pulso puede representarse por medio de la *función Delta de Dirac*, $\delta(x)$. Esta función es la derivada de la función escalón unitario, por lo que se tiene que

$$\int_{-\infty}^{\infty} \delta(x)dx = 1 \tag{4.22}$$

más aún, se tiene que si f es una función cualquiera, se tiene que

$$\int_{-\infty}^{\infty} f(x)\delta(x-a)dx = f(a) \tag{4.23}$$

Así que, si la función de entrada de un sistema es $f = \delta$ y la función de transferencia del sistema es H, se tiene que la salida del sistema en frecuecuia será

$$\mathcal{F}\{f\}H = \mathcal{F}\{\delta\}H = H \tag{4.24}$$

puesto que $\mathcal{F}\{\delta\} = 1$, así, se tiene que H representa la respuesta en frecuencia a un impulso en la entrada, por lo tanto se tiene que la salida será $h(t)$.

Con esto, se tiene que si f representa una señal de entrada y h la respuesta al impuslo del sistema, la convolución $f \otimes h$ es la salida del sistema.

Por lo tanto, si se tiene un filtro pasa bajo ideal con frecuencia de corte B_c, se tiene que su función de transferencia es

$$H(\omega) = \begin{cases} 1 & \text{si } |\omega| \leq 2\pi B_c \\ 0 & \text{en otro caso} \end{cases} \tag{4.25}$$

si se le llama ω_c a $2\pi B_c$, la respuesta al impulso del filtro será

$$\begin{aligned} \mathcal{F}^{-1}\{H(\omega)\} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} H(\omega)e^{-i\omega t}d\omega = \frac{1}{2\pi}\int_{-\omega_c}^{\omega_c} e^{-i\omega t}d\omega \\ &= \frac{1}{2\pi}\left.\frac{e^{-i\omega t}}{-it}\right|_{-\omega_c}^{\omega_c} = \frac{\sin(\omega_c t)}{\pi t} \\ &= \frac{1}{\pi}\frac{\sin(\omega_c t)}{\omega_c t} \end{aligned} \tag{4.26}$$

la última expresión se conoce como la función *sinc*, definida por

$$\operatorname{sinc}(x) = \frac{\sin(x)}{x} \tag{4.27}$$

por lo tanto, la respuesta al impulso de un filtro pasa bajo ideal es

$$h(t) = \frac{1}{\pi}\operatorname{sinc}(\omega_c t) \tag{4.28}$$

y recíprocamente, se tiene que

$$\mathcal{F}^{-1}\{\operatorname{sinc}(at)\} = \frac{\pi}{a} \cdot \operatorname{rect}(t, a) \tag{4.29}$$

donde

$$\operatorname{rect}(\omega, a) = \begin{cases} 1 & \text{si } |\omega| < a \\ 0 & \text{en otro caso} \end{cases} \tag{4.30}$$

Este resultado es de gran importancia para poder realizar la demostración del teorema del muestreo y para resultados posteriores.

4.2. Demostración del Teorema del Muestreo

El teorema del muestreo es de suma importancia en la adquisición de audio, puesto que da un parámetro para poder realizar una buena digitalización de la señal, de manera que a la hora de reproducirla, se tenga una alta fidelidad con la señal original.

El teorema del muestreo, como en 2.1 dice que

4.5 Teorema (Nyquist-Shannon)**.** Si se tiene una señal $s(t)$ que tiene componentes en frecuencia menores a f_m, puede ser muestreada y reconstruida sin pérdida de información si la frecuencia de muestreo es de al menos $2f_m$.

Demostración. Como se supone que la señal no posee componentes espectrales mayores a f_m, se puede decir que $S(\omega) = 0$ si $|\omega| \geq \omega_m = 2\pi f_m$, y por lo tanto si se pasa la señal a través de un filtro pasa bajo ideal como en 4.26, se tendrá la misma señal.

Para representar la señal muestreada se puede definir una *función muestreadora* a una frecuencia de muestreo $f_s = \frac{1}{T_s}$ como

$$\mathrm{sa}(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \tag{4.31}$$

que es un tren de pulsos con frecuencia f_s.

Con esto la señal muestreada se puede escribir como

$$\begin{aligned} s_a(t) = s(t)\mathrm{sa}(t) = s(t) &\sum_{n=-\infty}^{\infty} \delta(t - nT_s) \\ = &\sum_{n=-\infty}^{\infty} s(t)\delta(t - nT_s) \\ = &\sum_{n=-\infty}^{\infty} s(nT_s)\delta(t - nT_s) \end{aligned} \tag{4.32}$$

por lo tanto, por la linealidad del filtro, se puede obtener la salida como una superposición de cada uno de los términos $s(nT_s)\delta(t - nT_s)$.

Así, por 4.28, la salida será

$$\begin{aligned}&\sum_{n=-\infty}^{\infty} s(nT_s)\left(\frac{1}{\pi}\text{sinc}\left(\omega_m(t-nT_s)\right)\right)\\ &=\sum_{n=-\infty}^{\infty} \frac{s(nT_s)}{\pi}\text{sinc}\left(\omega_m(t-nT_s)\right)\end{aligned} \tag{4.33}$$

Ahora, se procederá a encontrar la densidad espectral de potencia de la salida del filtro, que será

$$\begin{aligned}&\mathcal{F}\left\{\sum_{n=-\infty}^{\infty} \frac{s(nT_s)}{\pi}\text{sinc}\left(\omega_m(t-nT_s)\right)\right\}\\ &=\sum_{n=-\infty}^{\infty} \frac{s(nT_s)}{\pi}\mathcal{F}\left\{\text{sinc}\left(\omega_m(t-nT_s)\right)\right\}\\ &=\sum_{n=-\infty}^{\infty} \frac{s(nT_s)}{\pi}e^{-i\omega nT_s}\mathcal{F}\left\{\text{sinc}\left(\omega_m t\right)\right\}\\ &=\sum_{n=-\infty}^{\infty} \frac{s(nT_s)}{\pi}e^{-i\omega nT_s}\frac{\pi}{\omega_m}\ \text{rect}(\omega,\omega_m)\end{aligned} \tag{4.34}$$

por lo que para $|\omega| < \omega_m$ la salida será

$$\sum_{n=-\infty}^{\infty} \frac{s(nT_s)}{\omega_m}e^{-i\omega nT_s} = \sum_{n=-\infty}^{\infty} \frac{s(-nT_s)}{\omega_m}e^{i2\pi n\omega/\omega_s} \tag{4.35}$$

y cero en otro caso.

Por otro lado, de 4.32 y 4.4 se tiene que

$$S_a(\omega) = \frac{1}{2\pi}\mathcal{F}\{s(t)\} \otimes \mathcal{F}\{\text{sa}(t)\} \tag{4.36}$$

$\mathcal{F}\{s(t)\} = S(\omega)$, la densidad espectral de potencia de la señal original. Para encontrar $\mathcal{F}\{\text{sa}(t)\}$, se tiene que esta es una función periódica con período T_s, por lo tanto

$$\begin{aligned} c_n &= \frac{1}{T_s}\int_{-T_s/2}^{T_s/2} \delta(t)e^{-i\omega_n t}dt \\ &= \frac{1}{T_s} \end{aligned} \tag{4.37}$$

y

$$\mathrm{sa}(t) = \sum_{n=-\infty}^{\infty} c_n e^{-i\omega_n t} = \frac{1}{T_s}\sum_{n=-\infty}^{\infty} e^{-i\omega_n t} \tag{4.38}$$

por lo que la transformada de $\mathrm{sa}(t)$ estará dada por

$$\begin{aligned} \mathcal{F}\{\mathrm{sa}(t)\} &= \mathcal{F}\left\{\frac{1}{T_s}\sum_{n=-\infty}^{\infty} e^{-i\omega_n t}\right\} \\ &= \frac{1}{T_s}\sum_{n=-\infty}^{\infty} \mathcal{F}\left\{e^{-i\omega_n t}\right\} \\ &= \frac{1}{T_s}\sum_{n=-\infty}^{\infty} 2\pi\delta(\omega-\omega_n) \\ &= \frac{2\pi}{T_s}\sum_{n=-\infty}^{\infty} \delta(\omega-\omega_n) \\ &= \omega_s\sum_{n=-\infty}^{\infty} \delta(\omega-\omega_n) \end{aligned} \tag{4.39}$$

Por lo que

$$\begin{aligned} S_a(\omega) &= \frac{1}{2\pi}S(\omega)\otimes\omega_s\sum_{n=-\infty}^{\infty}\delta(\omega-\omega_n) \\ &= \frac{\omega_s}{2\pi}\sum_{n=-\infty}^{\infty} S(\omega)\otimes\delta(\omega-\omega_n) \end{aligned} \tag{4.40}$$

El término $S(\omega)\otimes\delta(\omega-\omega_n)$ lo que hace es centrar el espectro en ω_n, así se tiene que $S_a(\omega)$ es periódica con período ω_s. Si se realiza la expansión en serie de Fourier de $S_a(\omega)$ se tendrá que

$$S_a(\omega) = \sum_{n=-\infty}^{\infty} c_n e^{i2\pi n\omega/\omega_s} \tag{4.41}$$

donde, si se supone que $\omega_s/2 > \omega_m$ se tiene que

$$c_n = \frac{1}{\omega_s}\int_{-\omega_s/2}^{\omega_s/2} S_a(\omega)e^{i2\pi n\omega/\omega_s}d\omega = \frac{1}{\omega_s}\int_{-\omega_m}^{\omega_m}\frac{\omega_s}{2\pi}S(\omega)e^{i2\pi n\omega/\omega_s}d\omega \tag{4.42}$$

y

$$c_n = \frac{1}{2\pi}\int_{-\omega_m}^{\omega_m} S(\omega)e^{i2\pi n\omega/\omega_s}d\omega \tag{4.43}$$

Además, por 4.18 se tiene que

$$s(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{i\omega t}d\omega = \frac{1}{2\pi}\int_{-\omega_m}^{\omega_m} S(\omega)e^{i\omega t}d\omega \tag{4.44}$$

así que si $t = -nT_S = -2\pi n/\omega_s$, se tiene que

$$s(-nT_s) = \frac{1}{2\pi}\int_{-\omega_m}^{\omega_m} S(\omega)e^{i\omega 2\pi n/\omega_s}d\omega \tag{4.45}$$

y por 4.43 y 4.45 se tiene que

$$c_n = s(-nT_s) \tag{4.46}$$

por lo tanto 4.35 no es más que la representación en serie de Fourier de la densidad epectral de potencia de la señal original, por lo que 4.35 es igual $S(\omega)$ y la salida del filtro es igual a la señal orginal como se requería.

Sin la condición de $\omega_s/2 > \omega_m$ no se podrían determinar los c_n, por lo que no sería posible reconstruir la señal original. $\square$

4.3. Mejoramiento de Calidad de Archivos de no Comprimidos

La demostración del Teorema del Muestreo en la sección precedente, además de mostrar un resultado importante para la adquisición de señales, tambien da una forma de poder reconstruir la señal original de una manera muy certera, a partir de muestras que simplemente cumplan la condición de Nyquist.

Así, la ecuación 4.33 provee una manera de *interpolar* los datos muestreados para reconstruir la señal original.

Por lo tanto, para una frecuencia de muestreo f_s y una frecuencia de corte f_m, se tiene la identidad

$$f(t) = \sum_{n=-\infty}^{\infty} f(nT_s)\mathrm{sinc}\,(\omega_m(t - nT_s)) \tag{4.47}$$

si y solo si $f_s > 2f_m$.

Idealmente, este efecto paso-bajo ideal debería de darse al reproducir un archivo de sonido en una computadora, puesto que las bocinas actúan como un filtro pasa bajo ya que el electroimán que se encarga de producir los sonidos en la bocina se puede tomar como una inductancia en serie con la resistencia de carga de la bocina, lo cual es un filtro pasa bajo, sin embargo, al no ser ideal su respuesta, no se obtiene una respuesta al impulso como en 4.28, y por lo tanto la ecuación de interpolación 4.47 no funciona.

Si se tuvieran bocinas con respuestas ideales, al reproducir un archivo que cumpla con las condiciones de Nyquist, este sería escuchado de forma fiel con la señal original, sin embargo al no poder realizarse esto de manera física, la calidad del audio reproducido depende de la frecuencia de muestreo que se tome, y entre más alta la tasa de muestreo, se tendrá una mejor calidad de sonido.

Así, una manera de poder mejorar la calidad de los archivos de sonido con tasas de muestreo bajas, siempre que cumplan Nyquist, es simular el paso de la señal a través de un filtro pasa bajo, esto por medio de la ecuación de interpolación 4.47.

Con esto, se puede mejorar la calidad, aunque nunca podrá reconstruirse en su enteridad a la señal original, puesto que para tener igualdad en 4.47, se tiene que realizar una suma infinita, cosa que no es posible realmente por algún medio electrónico.

Además, al ser reproducido por una computadora o sistema digital, tampoco es posible obtener una salida analógica como la función sinc $(\omega_m t)$, solamente valores discretos de esta, por lo que, por este lado tampoco es posible reproducir a cabalidad la señal original, lo más que puede hacerse es tener una aproximación tan fiel como se desee, sin embargo, entre mejor calidad, más complejidad computacional y más procesos necesarios se tendrán que realizar.

La idea es realizar la simulación del filtro digital de tal manera que esta se complemente con la respuesta del filtro intríseco que representan las bocinas. Además, al esuchar, el oído humano también actúa como un filtro pasa bajo, por lo tanto junto con lo anterior, se tendría una aproximación de la respuesta de un filtro digital y dos filtros analógicos intrísecos, lo cual ayuda a mejorar la selectividad del *filtro* total, puesto que se aumentarían los *polos* del filtro equivalente analógico, y así poder lograr un mejor efecto.

4.3.1. Error de Truncamiento

Debido a que no es posible realizar la suma infinita en 4.47, es necesario determinar que tanto error se estará comentiendo al realizar solamente una aproximación finita de la interpolación.

Para esto se puede definir la *aproximación por suma finita* como

$$f_N(t) = \sum_{n=-N}^{N} f(nT_s)\text{sinc}\left(\omega_m(t - nT_s)\right) \tag{4.48}$$

y con esto se puede definir el *error cuadrático medio* por

$$\epsilon_N = \int_{-\infty}^{\infty} |f(t) - f_N(t)|^2 \, dt \tag{4.49}$$

por lo tanto, como $f(t) - f_N(t)$ está limitada en frecuencia y de *energía finita*, ϵ_N, se tiene que existe una constante M tal que

$$|f(t) - f_N(t)| \leq M \tag{4.50}$$

por lo que se tiene que

$$|f(t) - f_N(t)|^2 = \left|\frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega) - F_N(\omega)d\omega\right|^2 = \left|\frac{1}{2\pi}\int_{-\omega_m}^{\omega_m} F(\omega) - F_N(\omega)d\omega\right|^2$$
$$\leq \frac{1}{2\pi}\int_{-\omega_m}^{\omega_m} |F(\omega) - F_N(\omega)|^2 \, d\omega \frac{1}{2\pi}\int_{-\omega_m}^{\omega_m} d\omega = \frac{\epsilon_N \omega_m}{\pi} \quad (4.51)$$

y con esto se puede encontrar una cota al error que tendrá la señal

$$|f(t) - f_N(t)| \leq \sqrt{\frac{\epsilon_N \omega_m}{\pi}} \quad (4.52)$$

Supóngase que se tiene una señal de audio muestreada en un archivo de sonido $A = \{a_n\}_{n=0}^{N}$, donde cada a_n sepresenta un byte. De esta manera, se tiene que

$$a_n = f(nT_s) \quad (4.53)$$

Por lo tanto, puede suponerse que $f(t) = 0$ para el tiempo que está fuera del tiempo en que se realizó la muestra, por lo que el error de aproximación finita de f_N sería cero, y habría error solamente si se toma un f_M para un $M < N$.

Para mejorar la calidad de A, se construirá $B = \{b_k\}_{k=0}^{K}$ de tal manera que A sea una *subsucesión* de B, y que la nueva tasa de muestreo $\hat{f}_s$ sea un múltiplo de f_s, por lo tanto, se tiene que K es múltiplo de N.

Para esto, se definirá

$$\begin{aligned} b_k &= \sum_{n=0}^{N} a_n \mathrm{sinc}\left(\omega_m(k\hat{T}_s - nT_s)\right) \\ &= \sum_{n=0}^{N} a_n \mathrm{sinc}\left(\omega_m T_s\left(k\frac{K}{N} - n\right)\right) \end{aligned} \quad (4.54)$$

Así, B será un mejoramiento de la calidad de sonido del archivo A y 4.54 provee el algorítmo para realizar esta mejora.

4.3.2. Mejora en Tiempo Real

El algortimo descrito anteriormente necesita tener el archivo completo para poder realizar la mejora, puesto que trabaja con una respuesta *no causal* de la simulación del filtro paso bajo, y los valores actuales de la señal son afectados por los valores anteriores y posteriores, sin embargo, para realizar un tratamiento de archivos en tiempo real, como sería el caso de un *Streaming* o *VoIP*, se puede considerar el comportamineto *causal* del filtro, simplemente imponiendo la condición que exista respuesta al impulso hasta que este halla llegado al filtro, esto es, cambiando la respuesta por

$$\mathrm{sgn}(t)\mathrm{sinc}\left(\omega_m t\right) \tag{4.55}$$

donde $\mathrm{sgn}(t)$ es la función signo, la cual es 0 si $t < 0$ y 1 en otro caso.

Por lo tanto, el algoritmo de mejora de 4.54 se convierte en

$$b_k = \sum_{n=0}^{N} a_n \mathrm{sgn}\left(k\frac{K}{N} - n\right) \mathrm{sinc}\left(\omega_m T_s \left(k\frac{K}{N} - n\right)\right) \tag{4.56}$$

4.3.3. Implementación del Algoritmo

A la hora de realizar la implementación por medio de una computadora o circuito digital, se tiene que realizar el cálculo dado por 4.54 y 4.56 resulta muy tardado, puesto que para cada nuevo valor se tienen que calcular $N + 1$ operaciones, las cuales, para un valor promedio de tasa de muestreo, andan en alrededor de $22,000$ datos por segundo, lo que hace que un archivo de un minuto tenga un $N = 60 * 22000 = 1320000$, que resultaría un proceso muy lento, ya que tendría que realizarse esto un total de $K + 1$ veces.

Para esto, se puede hacerse una aproximación de 4.54 y 4.56 como

$$b_k = \sum_{n=k\frac{K}{N}-i}^{k\frac{K}{N}+i} a_n \mathrm{sinc}\left(\omega_m T_s \left(k\frac{K}{N} - n\right)\right) \tag{4.57}$$

en donde para cada interpolación, unicamente se utilizan $2i+1$ valores, los que estas más proximos al valor que se está calculando.

4.4. Compresión sin Pérdidas

Al trabajar con archivos de audio, uno de los parmámetros más importantes a considerar, a parte de la calidad del sonido que se tiene, lo constituye el tamaño del mismo, puesto que, aunque las comunicaciones actuales son más rápidas y se dispone de mayor ancho de banda para la transmisión, siempre es algo de suma importancia el poder aprovechar al máximo los recursos disponibles.

Con el fin de poder realizar esto, se ha venido utilizando el manejo de archivos comprimidos, y de una manera un tanto más popular, se usan compresiones con pérdidas, puesto que estas logran alcanzar un mayor nivel de compresión que las compresiones sin pérdidas, ya que que la mayoría se basan en desechar parte de la información que en promedio en oído humano no es capaz de detectar, por lo que el contenido neto que es guardado se reduce.

Sin embargo, para realizar una compresión sin pérdidas, solamente se debe de realizar una modificación en los datos en la manera en que son guardados, puesto que el objetivo principal de esto es no perder nada de la información que se tiene.

Este objetivo se alcanza realizando una *codificación* de los datos que serán guardados.

Al hacer esto se dispone de más información en el *transmisor* y el *receptor* de un sistema de comunicaciones sobre la señal que será transmitida, y a cambio es posible mandar esta senãl utilizando menos recuros en el canal, es decir, se puede interpretar los datos que son transmitidos y no solamente recibir los datos.

4.4.1. Codificación

La codificación consiste en establecer una correspondencia entre cada uno de los símbolos de un alfabeto fuente y una secuencia de símbolos de una alfabeto destino. Al alfabeto destino se le denomina alfabeto código y a cada una de las secuencias de símbolos de este alfabeto que se corresponda con un símbolo del alfabeto fuente se denomina palabra de código.

El alfabeto fuente contiene los símbolos originales que se quieren codificar. El alfabeto código contiene las palabras de código equivalentes en que se codificarán los símbolos originales. Estas palabras de código son aptas para ser transmitidas por un sistema de comunicaciones.

Existen principalmente dos tipos de códigos

- Códigos de Bloque
- Códigos Compactos

Los códigos de bloque son aquellos en los que todas las palabras de código correspondientes a cada símbolo del alfabeto fuente tienen la misma longitud.

Estos a su vez se dividen en dos tipos

- Singulares
- No Singulares

Los singulares son aquellos que a cada símbolo del alfabeto fuente le corresponde una única palabra de código, y los No Singulares son los que a cada símbolo del alfabeto fuente le corresponde dos o más palabras de código.

Por otra parte, en los Códigos Compactos se busca que a cada símbolo del alfabeto fuente le corresponda una palabra de código de longitud mínima según algún criterio de minimización dado.

El objetivo de la codificación es obtener una representación eficiente de los símbolos del alfabeto fuente. Para que la codificación sea eficiente es necesario tener un conocimiento de las probabilidades de cada uno de los símbolos del alfabeto fuente. El dispositivo que realiza esta tarea es el codificador de la fuente. Este codificador debe cumplir el requisito de que cada palabra de código debe decodificarse de forma única, de forma que la secuencia original sea reconstruida perfectamente a partir de la secuencia codificada.

Se denomina compresión de datos al conjunto de técnicas que permiten que un conjunto de datos de una determinada longitud pueda ser reducido en su tamaño, sin alterar el significado de la información que contiene.

Hay dos tipos de compresión

- Lógica: se trata de reducir los datos desde el momento del diseño.
- Física: proceso de reducción de la cantidad de datos antes de poner los datos en el medio de transmisión y deshacer el proceso en el receptor. Tiene en cuenta la frecuencia de ocurrencia de los caracteres.

La compresión modifica la velocidad de transferencia de información y además reduce la probabilidad de que se produzcan errores durante la transmisión a través de un canal con ruido.

Por lo tanto, al realizar una codificación con el objetivo de comprimir datos se está también obteniendo una posible manera de detectar y corregir errores. Por errores se puede entender cuando se producen diferencias entre las secuencias de datos enviadas a través de un canal y las secuencias de datos recibidas debidas a la existencia de ruido en el canal.

4.4.2. Compresión por Diferenciación

En 2.1.6 se habó sobre DPCM como una manera de registrar la señal de audio de entrada. Se vió como una forma de representar la señal cuando era muestreada y cuantizada, es decir, un proceso que se realizaba al inicio, cuando se obtenía la señal, tomando las diferencias de esta y manejando esta información.

En otras palabras, DPCM es un proceso meramente circuital por medio del cual se puede representar una señal, sin embargo, puede utilizarse la idea de tomar las diferencias de la señal para realizar un tipo de codificación de esta.

Por ejemplo, si se tiene un archivo de audio como en 4.53, se puede definir la *Primer Compresión por Diferencias* como un nuevo archivo por definido por

$$a_n^1 = a_n - a_{n-1} \tag{4.58}$$

Esta es la misma idea de DPCM, solo que ahora visto desde el punto de vista de *Procesamiento Digital de Señales*. La idea de esto es, primero, guardar a_0 y posteriormente guardar las diferencias entre los siguientes a_i, es decir, el nuevo archivo queda determinado por

$$a_n^1 = \begin{cases} a_0 & n = 0 \\ a_n - a_{n-1} & n > 0 \end{cases} \tag{4.59}$$

Este nuevo archivo así definido aprovecha la propiedad de ser continuas las señales de audio, por lo que las diferencias guardadas serán pequeñas, y más pequeñas entre más grande la tasa de muestreo utilizada. Esto resulta de gran utilidad, puesto que se tendrá que si los a_n están cuantizados con m_0 bits, los a_n^1, $n > 0$, estarán cuantizados con $m_1 \leq m_0$ bits.

Esto puede interpretarse analizando la señal de entrada como una función $f(t)$, y al tomar las diferencias a_n^1 lo que se está realizando es una aproximación de la derivada $\frac{f(t)}{dt}$ de la función.

Por ser $f(t)$ una función continua y proveniente de una respuesta a un fenómeno ondulatorio, 1.2, se tiene que $\left\| \frac{df(t)}{dt} \right\| \leq \|f(t)\|$, donde $\|f(t)\|$ se define como

$$\|f(t)\| = \max_{t \in \mathbb{R}} f(t) \tag{4.60}$$

Por esto, si se vuelve a obtener las diferencias, pero ahora de $A_1 = \{a_n^1\}_{n=0}^N$, se tendrá que el nuevo archivo $A_2 = \{a_n^2\}_{n=0}^N$ estará cuantizado con $m_2 \leq m_1 \leq m_0$ bits.

Esto resulta como tomar la segunda derivada de $f(t)$ y guardar esta información en lugar de guardar la información de $f(t)$.

Por lo tanto, para recuperar la señal original, lo único que se tiene que hacer es *integrar* la información guardada, es decir, para pasar de $\{a_n^1\}_{n=0}^N$ a $\{a_n\}_{n=0}^N$, se obtiene la integral de A_1 como

$$a_n = a_n^1 + a_{n+1} \tag{4.61}$$

esto es equivalente a realizar la integración de la derivada de la función, es decir

$$f(t) = \int \frac{df(t)}{dt} dt + C \tag{4.62}$$

aca, la constante de integración viene a representar el punto de inicio, es decir a_0^1.

De esta manera, pueden definirse la *i-ésima Compresión por Diferencias* como

$$a_n^i = \begin{cases} a_n^{i-1} & n < i \\ a_n^{i-1} - a_{n-1}^{i-1} & n \geq i \end{cases} \tag{4.63}$$

y $A_i = \{a_n^i\}_{n=0}^{N}$. Para $i = 0$ se tiene que A_0 es el archivo original.

Con esto se tiene que las cantidades de bits necesarios para representar los A_i conforman una sucesión no creciente de números naturales, $m_0 \geq m_1 \geq m_2 \geq \cdots \geq m_n$.

Debido a que se posee un número finito de muestras, solamente es posible obtener n compresiones por diferencias, puesto que de alli en adelante ya no es posible obtener diferencias entre los datos, por lo que este sistema de compresión resulta ser un algoritmo, es decir, es finito.

Al realizar una aplicación real de esto se tiene que los valores de n resultan ser muy grandes como para obtener A_n, sin embargo se tiene que en los primeros pasos del algoritmo se logra obtener una reducción considerable del tamaño total del archivo guardado. Esta reducción y el número de etapas necesarias para alcanzarla dependen de la *forma* y del tipo de señal al que le es aplicado este algoritmo, en algunos se va tener que se puede obtener una buena reducción en el tamaño, mientras que en otros puede no haber reducción alguna, tal es el caso de $f(t) = \cos(t)$, puesto que al derivarlo varias veces se tiene que

$$\left\| \frac{d^i f(t)}{dt^i} \right\| = 1 \tag{4.64}$$

para todo valor natural de i.

4.4.3. Interpretación por medio de Filtrado

Si se encuentra la respuesta en frecuencia del sistema diferencial definido por 4.58 se tiene que este es

$$H(z) = 1 - z^{-1} = \frac{z-1}{z} \tag{4.65}$$

la cual, tiende a 1 cuando $|z| \to \infty$ y tiende a $-\infty$ cuando $|z| \to 0$, por lo que esto representa un filtro pasa altos con cero en $z = 1$ y polo en $z = 0$.

Para recuperar la señal se aplicaba el proceso inverso, una integración, 4.61, la cual tiene una función de transferencia dada por

$$H(z) = 1 + z^{-1} = \frac{z+1}{z} \tag{4.66}$$

que corresponde a un filtro pasa bajos, con frecuencia de corte en $z = -1$ y polo en $z = 0$

Por lo tanto, al realizar la Compresión por Diferencias, lo que se está haciendo en realidad es pasar varias veces la señal por medio de un filtro pasa altos, esto es equivalente a pasar la señal a través de un filtro pasa alto más selectivo, o de mayor orden.

Así, se puede decir que una manera de obtener una forma de comprimir archivos de audio es realizando filtrado con un filtro pasa altos en la codificación, y un filtro pasa bajos en la decodificación.

De esta manera, se tiene que una manera de poder realizar esta compresión sería utilizando un filtro paso alto ideal, cuya función de transferencia está dada por

$$H(\omega) = \begin{cases} 1 & |\omega| > \omega_c \\ 0 & \text{en otro lado} \end{cases} \tag{4.67}$$

sin embargo, para estos propositos, es necesario calcular la transformada inversa de Fourier de la respuesta en frecuencia, ya que con esto se tiene la respuesta al impulso del sistema con lo que se puede modelar el sistema. En este caso, la función de transferencia no posee transformada inversa de Fourier, así que, como se está trabajando con funciones limitadas en banda, el efecto del filtro pasa altos ideal puede obtenerse con un filtro pasa banda ideal.

Para este caso, la respuesta del filtro está dada por

$$H(\omega) = \begin{cases} 1 & \omega_2 > |\omega| > \omega_1 \\ 0 & \text{en otro lado} \end{cases} \tag{4.68}$$

por lo tanto se obtiene que la respuesta al impulso de esta función esta dada por la transformada inversa de Fourier que es

$$f(t) = 2\omega_2 \text{sinc}\,(\omega_2 t) - 2\omega_1 \text{sinc}\,(\omega_1 t) \tag{4.69}$$

y debido a la relación que guardan la Transformada de Fourier y la Transformada Z, se tiene que $z = e^{j\omega}$ y $\omega = -j \log z$.

Con esto, al tener un archivo de entrada $A = \{a_n\}_{n=0}^{N}$ puede calcularse la salida del filto ideal de manera nuérica, esto es, haciendo $Y(z) = H(z)X(z)$ y obteniendo la expansión en Serie de *McLaurin* de la respuesta.

Para lograr *decodificar* el archivo se puede pasar ahora el archivo comprimido por medio de un filtro pasa bajos ideal, dado por

$$H(\omega) = \begin{cases} 1 & |\omega| > \omega_c \\ 0 & \text{en otro lado} \end{cases} \tag{4.70}$$

y cuya respuesta al impulso ya se calculó anteriormente.

De la misma manera puede determinarse la salida de manera numérica, ya que de forma algebraica no es posible obtener una expresión que modele su comportamiento en el caso general.

La relación que deben guardar ambos filtros, pasa bajo y pasa alto (aca pasa banda), para realizar la bien la codificación y decodificación del archivo, es que tengan los polos y los ceros cambiados de signo, así que, en lo que respecta a los ceros, esta condición se transforma en que $\omega_c = \omega_1$ y la condición de polos no importa puesto que ambos filtros no poseen polos. Además, debe de tenerse que $\omega_2 \gg f_m$ con f_m la frecuencia máxima de la señal.

CONCLUSIONES

1. El análisis de los archivos de sonido es un tema con mucha aplicación en la actualidad, debido a la creciente necesidad actual de tener mejores y más rápidas telecomunicaciones.12pt

2. La simulación de sistemas ideales y sistemas analógicos por medios digitales, logra permitir una manera de mejorar el desempeño de sistemas cuyas respuestas no son naturalmente anaógicos, tal es el caso de las computadoras y los circuitos combinacionales.

3. El Teorema del Muestreo provee un medio por el cual se puede evaluar la posibilidad de reconstruir posteriormente una señal de audio y de saber que tan fiel es dicha reconstrucción.

4. Por medio de la demostración del Teorema del Muestreo es posible obtener un proceso que permite realizar una mejora en la calidad de un archivo de audio no comprimido..

5. Es posible realizar compresión de archivos de audio mediante la simulación de un sistema de filtros paso alto, con esto se reduce el tamaño del archivo y no se pierde la calidad del mismo.

6. Para poder realizar la decodificación de una compresión por medio de filtrado, se realiza el proceso inverso, esto es, la simulación de un sistema de filtros paso bajo.

RECOMENDACIONES

1. Los métodos descritos en este trabajo proporcionan formas de manejar archivos de sonido, las cuales, de manera combinada, pueden producir muy buenos resultados a la hora de trabajar en telecomunicaciones y transmición de datos, pues proveen una forma de disminuir el tamaño de un archivo sin disminuir su calidad, y además es posible de mejorar dicha calidad a la hora de reproducirlo.

BIBLIOGRAFÍA

1. Distefano, JosephJ. **Teroía y Problemas de Retroalimentación y Sistemas de Control** Estados Unidos: McGraw-Hill, 1967.

2. Krishna, Hari. **Computational Number Theory and Digital Signal Proccessing.** Estados Unidos: CRC Press, 1994.

3. Rudin, Walter. **Principles of Mathematical Analysis.** Estados Unidos: McGraw-Hill, 1964.

4. Walker, Stuart. **Fourier Analysis.** Estados Unidos: Oxford University Press, 1988.