



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

La Universidad Católica de Loja

MODALIDAD CLASICA

ESCUELA DE ELECTRÓNICA Y TELECOMUNICACIONES

**Desarrollo de un Códec Basado en Ogg-Vorbis e Implementación de un
Codificador G722 Aplicados a Sonidos Estetoscópicos**

Trabajo de fin de carrera previa a la obtención
del título de Ingeniería en Electrónica y
Telecomunicaciones

AUTORES

Castro Hurtado Hugo Alberto
Ordóñez Mendieta Ángel José

DIRECTOR

Ing. Mgs. Aguirre Reyes Daniel Fernando

LOJA-ECUADOR

2010

CERTIFICACIÓN: ACEPTACIÓN PROYECTO DE FIN DE CARRERA

Loja, octubre de 2010

Ing. Mgs.

Daniel F. Aguirre R.

Escuela de Electrónica y Telecomunicaciones

Dejo constancia de haber revisado y estar de acuerdo con el proyecto de fin de carrera, titulado: "Desarrollo de un códec basado en Ogg-Vorbis e implementación de un codificador G722 aplicados a Sonidos Estetoscópicos".

Presentado por:

Hugo Alberto Castro Hurtado

Ángel José Ordóñez Mendieta

Particular que comunico para los fines legales pertinentes.

Ing. Mgs. Daniel Aguirre

Visto Bueno Dirección de Escuela

F).....

Ing. Susana Arias

DIRECTORA DE LA ESCUELA DE ELECTRÓNICA Y
TELECOMUNICACIONES

Octubre de 2010

CESIÓN DE DERECHOS

Hugo Alberto Castro Hurtado y Ángel José Ordóñez Mendieta, declaramos ser autores del presente trabajo y eximamos expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales.

Adicionalmente declaramos conocer y aceptar la disposición del Art. 67 del Estatuto Orgánico de la Universidad Técnica Particular de Loja que en su parte pertinente textualmente dice: "Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado que se realicen a través o con el apoyo financiero, académico o institucional (operativo) de la Universidad".

Los Autores

.....
Hugo A. Castro

.....
Ángel J. Ordoñez

AUTORÍA

Las ideas, opiniones, conclusiones, recomendaciones y más contenidos expuestos en el presente informe de tesis son de absoluta responsabilidad de los autores.

Hugo Alberto Castro Hurtado
Ángel José Ordóñez Mendieta

INTRODUCCIÓN

Los avances realizados durante la última década en las ramas de electrónica, telecomunicaciones y procesamiento de audio, han dado lugar al progreso de la mayoría de dispositivos y mecanismos médicos. El tema propuesto funciona en base al estetoscopio, uno de los dispositivos más útiles en el área médica, que hoy en día ha evolucionado en electrónico y digital[1]. Por medio de este dispositivo se percibe los diferentes sonidos tanto cardíacos como respiratorios, normales y anormales, lo que permite grabarlos, para luego comprimirlos y transmitirlos, siendo un gran aporte para la telemedicina. Esta rama se encuentra estrechamente relacionada con las telecomunicaciones debido a que se puede enviar información importante que puede producir un caso asistencial, desde los lugares más alejados de nuestro país, para que estos puedan ser revisados por profesionales capacitados que se encuentran en las cabeceras provinciales. Para esto debemos acondicionar la información y optimizar la forma de transmisión. La información que se acondicionará está compuesta por sonidos estetoscópicos comprimidos.

El presente trabajo se enfoca en desarrollar, modificar y adaptar el codificador Ogg-Vorbis y el estándar G.722 a archivos de audio estetoscópicos tanto cardíacos y respiratorios que se encuentran en formato WAV. El programa a desarrollar permitirá comprimir los sonidos a la mayor escala posible y así optimizar el ancho de banda utilizado para transmitirlos hasta su destino. Este programa es aplicable al proyecto de telemedicina instalado en el cantón Yacuambi, Provincia de Zamora Chinchipe, liderado por la UTPL.

Objetivos

Objetivo general.-

Desarrollar un software compresor de sonido utilizando la Transformada Discreta del Coseno Modificada (MDCT) utilizada en el códec Ogg-Vorbis, e implementar el códec (codificador – decodificador) G.722 mediante un programa previamente establecido.

Objetivos específicos.-

- Analizar los sonidos captados por el estetoscopio y estudiar sus valores en tiempo y frecuencia.
- Establecer los conceptos básicos de compresión de audio y estudiar el códec implementando la MDCT y ADPCM (modulación por codificación de impulsos diferencial adaptativa).
- Desarrollar un software del códec MDCT para la compresión tanto de los sonidos cardíacos y respiratorios captados por el estetoscopio, en el programa MatLab.

Dedicatorias

A mis padres

Quienes con sus enseñanzas en valores y sus grandes esfuerzos me han ayudado a conseguir cada una de mis metas propuestas de forma responsable, valiente, y siempre guiadas por Dios.

A los Chuku's

De especial manera a todos mis compañeros quienes me dieron su apoyo en cada momento académico y fuera de él, hasta llegar a la consecución de este logro en mi vida. Gracias.

Hugo Castro

A mis padres

Que gracias a su apoyo y a su confianza lograron que este sueño se vuelva una realidad, un logro que lo hemos alcanzado todos como una familia.

A mis amigos

Los cuales brindándome su apoyo incondicional y su ayuda en momentos difíciles contribuyeron a que este sueño sea una realidad. Y en forma conjunta agradecerle a nuestro director de tesis que gracias a su dedicación contribuyo a la finalización del proyecto de la mejor forma. ¡Gracias a todos!

Ángel José

RESUMEN

En el presente trabajo se ha desarrollado un programa de compresión de sonido, el mismo que está destinado a sonidos estetoscópicos tanto cardiacos como respiratorios.

En el primer capítulo realizamos un breve análisis a la telemedicina, tanto en nuestro país, como los proyectos que realiza la universidad en este campo, e indicando en que parte se tiene previsto la implementación de este proyecto. Además se analizan las ventajas y desventajas de la Telemedicina, además del futuro que le espera.

En lo que respecta al capítulo dos, se describe las clases de sonido que existen, así como también se indica que es la auscultación en general, y las definiciones de auscultación cardiaca y pulmonar. Continuando con el capítulo tres, se explica los tipos de compresión que existen, y en cual a basar el proyecto, Además se analiza los parámetros más importantes que tenemos que tomar en cuenta a la hora de desarrollar un códec de sonido.

En el capítulo cuatro, se desarrolla un análisis a dos tipos de códecs, como son: Ogg Vorbis y G722, en los cuales se va a basar el desarrollo del programa, aquí se explica su funcionamiento, sus ventajas y desventajas. Posterior a esto en el capítulo cinco se describe como se desarrollo el programa, que parámetros se tomaron en cuenta, y las nuevas características que va a tener el programa.

Finalmente en el capítulo seis se muestra el programa desarrollo en su totalidad, se indica su funcionamiento, y se indica los resultados del mismo. Adicional a esto en anexos, tenemos datos importantes como tablas con valores de compresión del programa, pruebas con un especialista en el campo de la medicina, el código en si, entre lo más importante.

TABLA DE CONTENIDO

Cesión de derechos.....	I
Autoría	I
Introducción	III
Objetivos.....	IV
Dedicatorias.....	V
Tabla de contenido	VI
Lista de figuras	IX
Lista de tablas.....	X
CAPÍTULO 1. Telemedicina.....	1
1.1 Introducción	1
1.2 Principales proyectos.....	1
1.3 Funciones y aplicaciones.....	2
1.4 Ventajas y futuras implementaciones	3
CAPÍTULO 2. Adquisición de información	3
2.1 Sonido y clases de sonidos	3
2.1.1 Clasificación de acuerdo a la frecuencia	3
2.1.1.1 Infrasonido.....	4
2.1.1.2 Sonido Audible	4
2.1.1.3 Ultrasonido	4
2.1.2 Clasificación de acuerdo a su grabación y reproducción	5
2.1.2.1 Sonido monofónico.....	5
2.1.2.2 Sonido polifónico	5
2.1.2.3 Sonido estéreo	6
2.2 Auscultación	6
2.2.1 Auscultación Cardíaca	7
2.2.2 Auscultación Pulmonar	7
CAPÍTULO 3. Compresión de audio.....	8
3.1 Tipos de compresión	8
3.1.1 Compresión Sin Pérdida.....	8
3.1.2 Compresión Con Pérdida	9
3.2 Tratamiento del sonido	9

3.2.1 Almacenamiento	10
3.2.2 Compresión	11
3.2.3 Transmisión Digital del Sonido	12
3.3 Modelo Psicoacústico	13
3.4 Bandas Críticas	14
3.4.1 El Bark	15
3.5 Métodos de codificación	16
3.5.1 Codificadores Perceptuales	16
3.5.2 Codificadores Paramétricos	17
3.5.2.1 Vocoders	18
CAPÍTULO 4. Códecs de compresión de audio	20
4.1 OggVorbis (MDCT)	20
4.1.1 Introducción	20
4.1.2 Características y Funcionamiento	21
4.1.3 Detalles técnicos	28
4.1.4 Ventajas y desventajas	29
4.2 G.722 (ADPCM)	30
4.2.1 Introducción	30
4.2.2 Características y Funcionamiento	31
4.2.3 Detalles técnicos	33
4.2.4 Ventajas y desventajas	34
CAPÍTULO 5. Descripción General	34
5.1 Desarrollo de codificador MDCT (Ogg-Vorbis)	34
5.1.1 Entramado y segmentación	35
5.1.2 SMR (Relación señal a máscara) y aplicación FFT	35
5.1.3 Asignación de bits	37
5.1.4 Cuantización y Codificación	37
5.1.5 Almacenamiento	37
5.1.6 Diagramas de Bloques	38
5.2 Implementación del Codificador G.722	40
5.2.1 Codificador ADPCM	40
5.2.2 Inclusión al Programa Principal	41

CAPÍTULO 6. Resultados	41
6.1 Procedimiento general del software utilizado	42
6.2 Comparación de sonidos estetoscópicos utilizando los dos códecs	45
Conclusiones	48
Recomendaciones	49
Bibliografía.....	50
Glosario de términos.....	53
Anexos.....	56

Lista de Figuras.

Figura	2.1	Clasificación	por	la
frecuencia.....				¡Error! Marcador no definido.
Figura				2.2
Estetoscopio.....				¡Error!
Marcador no definido.				
Figura 3.1	Compresión en almacenamiento.....			8
Figura 3.2	Capacidad auditiva.....			14
Figura 3.3	Codificador perceptual.....			16
Figura 3.4	Codificación paramétrica.....			17
Figura 3.5	Esquema de un vocoder.....			18
Figura 3.6	Codificador híbrido.....			19
Figura 4.1	Codificación Ogg Vorbis.....			22
Figura 4.2	Solapamiento de dos ventanas con la misma extensión.....			22
Figura 4.3	Solapamiento de una ventana grande y una pequeña.....			23
Figura 4.4	Proceso de decodificación.....			27
Figura 4.5	Diagrama de bloques Códec G.722.....			31
Figura 4.6	Diagrama de bloques codificador G.722.....			32
Figura 4.7	Diagrama de bloques decodificador G.722.....			32
Figura 5.1	Parte de código utilizado.....			38
Figura 5.2	Procesos de codificación del archivo de audio.....			39
Figura 5.3	Procesos de decodificación del archivo de audio.....			39
Figura 5.4	Configuración del programa.....			40
Figura 5.5	Llamado del programa desde Matlab.....			41
Figura 6.1	Pantalla principal del software.....			42
Figura 6.2	Pantalla códec ADPCM.....			43
Figura 6.3	Pantalla del programa Audio Music Batch Converter.....			43
Figura 6.4	Pantalla códec MDCT.....			44
Figura 6.5	Parte de codificación MDCT.....			44
Figura 6.6	Parte de decodificación MDCT.....			45
Figura 6.7	Espectros de la señal cardíaca anormal estenosismitral.....			45
Figura 6.8	Comparación de las señales.....			46
Figura 6.9	Comparación en fase de las señales.....			47

Lista de Tablas.

Tabla 4.1 Características generales G.722.1 [36] 33

CAPÍTULO 1

Telemedicina

1.1 Introducción

Desde la década de los 90, los sistemas sanitarios buscan nuevas fórmulas de organización y alternativas a los modos tradicionales de proporcionar servicios sanitarios. Las tecnologías de la información y la comunicación se han revelado como elementos facilitadores y herramientas de cambio, haciendo viables modelos organizativos, como la continuidad asistencial o el acercamiento de la atención al entorno del paciente [2].

Telemedicina quiere decir, una posibilidad de tratamiento globalizador de todo el proceso de atención sanitaria (procedimientos administrativos, diagnóstico, tratamiento, gestión de datos clínicos, formación y educación sanitarias, etc.) utilizando redes de comunicación y sistemas interoperables de información [3].

1.2 Principales Proyectos en el Ecuador

Entre los principales proyectos se encuentra el que desarrolla la UTPL, a través de su Departamento de Telemedicina de CITTES de Ciencias Médicas, trabaja desde hace cuatro años en telesalud rural con el Proyecto Tutupaly que beneficia directa e indirectamente a las poblaciones de Yacuambi, al norte de Zamora Chinchipe. Atiende en las especialidades de pediatría, dermatología, ginecología y obstetricia, medicina interna y cirugía. Sus consultorios, equipados con sistemas de videoconferencia se enlazan con el Hospital Universitario UTPL que brinda consultas de especialidad y segunda opinión.

El proyecto de la UTPL estableció un convenio con el Ministerio de Salud Pública enviando a diez médicos recién graduados a dar atención en puestos de salud rural en la provincia de Zamora Chinchipe, el mismo que entro en marcha en noviembre del 2006, en los siguientes sectores [4]:

- Sector Yacuambi: La Esperanza, La Paz, Jembuentza, Tutupali y 28 de Mayo.
- Sector El Pangui: El Pincho, Pachicutza, El Zarza, Tundayme y El Pangui.

La Telemedicina no es nueva en el país, sobre todo para las universidades. Como ejemplo tenemos a la UTE (Universidad Tecnológica Equinoccial) que trabaja en el tema hace cinco años. Un convenio con la Universidad de Nuevo México en Estados Unidos, dio paso al primer Programa de Telemedicina en Nuevo Rocafuerte, frontera con Perú en el bajo Napo y el Yasuní. Tras un año abrieron otro centro en el Hospital Oskar Jandl de la Isla San Cristóbal en Galápagos.

Otro de los proyectos más importantes que se encuentra en desarrollo es el impulsado por la Fundación Ecuatoriana de Telemedicina y Salud; y el Instituto de Telemedicina y Salud de la Universidad San Francisco de Quito. Por medio de una alianza con la Fundación Telesalud Internacional y la Fundación de Mujeres, concretaron en septiembre de este año un proyecto en la Sierra Central –Cotopaxi, Tungurahua, Chimborazo y Bolívar–. En Riobamba y Latacunga hay dos centros con médicos y enfermeras adonde acuden los pacientes de las comunidades. Cuando requieren la ayuda de especialistas toman contacto con el centro de Telemedicina del Instituto en la Universidad San Francisco [38].

1.3 Funciones y aplicaciones.

Los servicios de Telemedicina incluyen aplicaciones asistenciales (Teleconsulta, Telediagnóstico, Teleeducación y Teleepidemiología), relacionadas con la administración y gestión de pacientes, e información y formación a distancia para usuarios y profesionales.

Algunos ejemplos de aplicaciones clínicas ensayadas con éxito incluyen Teleradiología, Telecardiología, Teledermatología, Telepsiquiatría, como algunos ejemplos. Se están empleando servicios de Telemedicina en diversos sistemas sanitarios y en una variedad de escenarios, tales como zonas rurales, áreas urbanas, áreas sanitarias, prisiones, cuidados a domicilio, emergencias, conflictos bélicos, y otros [5].

Muchos aspectos tecnológicos relacionados con Telemedicina están pendientes de ser evaluados definitivamente, como por ejemplo: requerimientos de ancho de banda, estándares en tecnología o protocolos de transmisión, evaluación de la efectividad y eficiencia de las prácticas de telemedicina en comparación con las prácticas tradicionales, aceptación de pacientes y profesionales, y además, todo lo relacionado con el coste de implantación de estos servicios y las fórmulas de pago a los

profesionales, diferentes aspectos legales relacionados con la seguridad, confidencialidad y acreditación [6].

El proyecto se encuentra se encuentra destinado a la implementación de la tele estetoscopia, en la cual con la ayuda un tele-estetoscopio (o tele-fonendo), se adquiere los sonidos que son evaluados normalmente por los especialistas, y se realiza un tratamiento a los mismos ya sea para amplificarlos, filtrarlos o transmitirlos, lo cual ayuda a un mejor diagnostico del paciente.

1.4 Ventajas y futuras implementaciones.

Las ventajas que se obtienen al aplicar la Telemedicina repercuten principalmente en los pacientes, los familiares de estos, el personal médico, los hospitales, los estudiantes y los residentes.

Pacientes: Disponen de diagnóstico y tratamiento especializado más rápido.

Familiares: Están más cerca de los enfermos.

Médicos:

- Tienen la posibilidad de consultar a los especialistas sin tener que desplazarse por este motivo.
- Tienen garantizada la educación continua, así como el intercambio de información y producción científicas.
- La cooperación entre atención primaria y la especialidad mejora considerablemente.

Hospitales:

- Reducción del peligro de pérdidas de imágenes.
- Mayor rapidez y precisión en los tratamientos.
- Utilización más eficaz de los servicios.

Estudiantes y residentes: Podrán asistir, por ejemplo, a intervenciones quirúrgicas realizadas en cualquier parte del mundo e incluso, comunicarse directamente con el cirujano [7].

Las perspectivas de crecimiento de esta área tecnológica de la Medicina son prácticamente ilimitadas y van de la mano del desarrollo de la informática, la robótica y las telecomunicaciones.

La tele-robotización de prácticas quirúrgicas, la formación médica de pre y post grado en áreas remotas, la aparición de nuevas redes globales del estilo de la Internet, pero dedicadas por entero a la Telerradiología y la Telemedicina en general, son emprendimientos que están lo suficientemente avanzados como para poder utilizarlos rutinariamente en los siguientes años.

CAPÍTULO 2

Adquisición de información.

El presente capítulo detalla el tipo de información que va a ser adquirida, en este caso el estudio y clasificación de los sonidos que serán utilizados en el proyecto.

2.1 Sonido y clases de sonidos.

Existen variados y numerosos tipos de sonido, por lo cual se elaboró dos tipos de clasificación, que contienen a los principales sonidos que se encuentran a nuestro alrededor, la primera de acuerdo a la frecuencia y la segunda de acuerdo a su grabación - reproducción.

2.1.1 Clasificación de acuerdo a la frecuencia

Esta clasificación se realizó en base al rango de frecuencia que poseen los sonidos, divididos en tres clases: infrasonido, sonido audible y ultrasonido

2.1.1.1 Infrasonido

Es un tipo de sonido que posee una frecuencia tan baja que no puede ser percibida por el oído humano y que normalmente se aplica desde los 16 Hertz hasta los 0,001 Hertz. Se tiene conocimiento que algunos animales tienen la capacidad de percibir este tipo de sonido y usarlo (por ejemplo los tigres lo utilizan en sus técnicas de cacería) [25]. El infrasonido tiene la característica de poder cubrir grandes distancias y traspasar objetos con poca disipación. Se utiliza en sismográfica para monitorear terremotos y también se lo emplea para la detección de objetos [8].

2.1.1.2 Sonido Audible

El sonido humanamente audible consiste de ondas sonoras compuestas de muchas oscilaciones en base a la presión del aire. Estas ondas mecánicas son convertidas en señales eléctricas por el oído humano e interpretadas por el cerebro. Un oído sano y joven es sensible a las frecuencias comprendidas entre los 20 Hz y los 20 KHz (Fig.2.1). No obstante, este margen varía según cada persona y se altera con la edad (presbiacusia es la pérdida de audición con la edad).

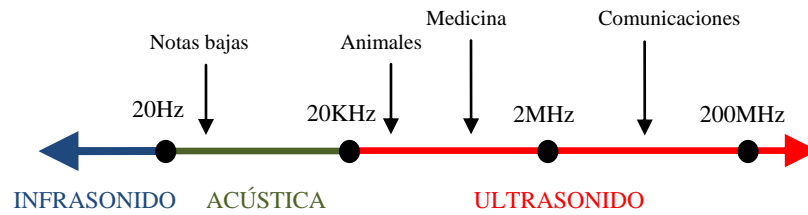


Fig. 2.1. Clasificación por la frecuencia

En el sonido no hay un desplazamiento neto de las moléculas del aire aunque si instantáneo. Lo que se comunica entre las ondas es la vibración a una velocidad aproximada de 340 m/s en el aire en condiciones normales de temperatura y de presión [38].

2.1.1.3 Ultrasonido

Onda acústica que no puede ser percibida por el hombre por estar en una frecuencia más alta de lo que puede captar el oído. Este límite se encuentra en los 20 KHz. Existen animales que tienen la capacidad de oír estas frecuencias como los murciélagos, delfines y perros, que las utilizan como radar, para orientarse y cazar.

El ultrasonido se utiliza en muchos ámbitos de las ciencias y las tecnologías. Por ejemplo, en medicina se emplea para el diagnóstico por ultrasonido (ultrasonografía), fisioterapia, etc. En el campo médico se los llama equipos de ultrasonido a dispositivos tales como el doppler fetal, el cual utiliza ondas ultrasónicas entre 2 y 3 MHz para detectar la frecuencia cardíaca fetal dentro del vientre materno [9].

2.1.2 Clasificación de acuerdo a su grabación y reproducción

De acuerdo a las características de grabación y reproducción de un archivo de audio se realizó la siguiente clasificación: sonido monofónico, polifónico y estereofónico.

2.1.2.1 Sonido monofónico

El sonido monofónico se le conoce al grabado y reproducido de audio por un solo canal. Un ejemplo claro de este sonido es utilizado por las estaciones de radio AM. Este tipo de sonido no es tan utilizado como el sonido estereofónico, ya que es de una calidad más baja, porque no se hace una separación de los sonidos [26]. Por ejemplo, cuando escuchamos una canción en donde se escucha el sonido de la

batería y la guitarra, los sonidos son transmitidos por un solo canal, en cambio, el sonido estereofónico hace una división de canales de audio en donde se puede distinguir mejor el sonido de cada instrumento [10].

En este proyecto se utiliza esta clasificación de sonido debido a que los archivos de audio implementados tienen un tamaño de bits pequeño y además el sonido está compuesto por una sola fuente, por lo cual, es suficiente el uso de un solo canal para su grabación y reproducción.

2.1.2.2 Sonido polifónico

El sonido polifónico es el conjunto o la unión de varios y diversos sonidos en un archivo de audio. Debido a esto, se necesita más de un canal para su grabación y reproducción. Estos archivos poseen una magnitud mucho mayor que los monofónicos. Además se tiene que considerar que la fuente que produce estos sonidos no es la misma, sino que pueden ser varias.

El sonido polifónico está muestreado a 44100Hz (calidad de audio) y está compuesto, aparte de los tres tonos fundamentales, por 20 armónicos cuya amplitud sufre un decaimiento exponencial por cada uno de los tonos fundamentales [11].

2.1.2.3 Sonido estéreo

El sonido estereofónico o estéreo es el nombre que se le da al grabado y reproducción en dos canales. El sonido estereofónico permite crear la sensación de dirección y profundidad tanto en las emisiones radiofónicas como en las grabaciones. Cuando escuchamos en la radio una orquesta, por ejemplo, distinguimos los violines, los instrumentos de viento y los de percusión.

Aunque el sonido estéreo pueda tener dos canales monaurales independientes, habitualmente la señal en un canal está relacionada con la señal del otro canal. Aunque se refiere exclusivamente a sistemas de dos canales, el término se puede aplicar a cualquier sistema de audio que usa más de un canal como es el caso del audio de 5.1 canales y los sistemas de 6.1 que se usan en películas y producciones televisivas [12]. En la actualidad los discos compactos de audio, la mayoría de las estaciones de radio FM, los canales de TV y televisión vía satélite, transmiten señales de audio estéreo. El propósito de grabar en sonido estéreo es recrear una experiencia

más natural al escucharlo, y donde, al menos en parte, se reproducen las direcciones izquierda y derecha de las que proviene cada fuente de sonido grabada [13].

2.2 Auscultación

La auscultación es un método de exploración clínica que consiste en aplicar el sentido del oído para recoger todos aquellos sonidos o ruidos que se produce en los órganos. Estos sonidos pueden ser generados por el propio órgano (contracción cardiaca, borborismos intestinales etc.) o en el órgano mismo pero sin que este tome una participación inmediata en su propia producción (soplos intracardiacos).

Tipos de auscultación

Directa: en el que se acerca la oreja, ejerciendo cierta presión directamente sobre el cuerpo del examinando con interposición de un ligero paño de hilo. Es una técnica fácil que no requiere de instrumental y obtiene los ruidos intensos y puros, evitando ruidos sobre añadidos debidos al roce del fonendoscopio con la piel.

Indirecta: esta se escucha mediante un aparato llamado estetoscopio o fonendoscopio (Fig.2.2). Es el método más utilizado para eludir muchos de los inconvenientes que se puede tener con la auscultación directa [14].



Fig.2.2 Estetoscopio¹

2.2.1 Auscultación cardiaca

Es uno de los métodos clínicamente más valiosos para el examen del corazón, especialmente para el diagnóstico de las afecciones valvulares que pueden ser

¹ Tomado de ALLSERIESLATINEWS
<http://allserieslinamarcela.wordpress.com/2009/03/05/medicos-de-television-que-dan-rating/>, última visita 20 de septiembre 2010.

reconocidas por su intermedio, antes de que se haya determinado modificaciones en la forma y tamaño cardiacos o del electrocardiograma. Sirve también para identificar los estados de claudicación miocárdica y las alteraciones del ritmo cardiaco. Constituye un método de exploración, difícil de reemplazar.

Técnica de Auscultación Cardiaca

1. Debe practicarse en un lugar tranquilo y mejor con luz velada para que nada distraiga nuestra atención y evitar la influencia de otros ruidos exteriores.
2. La posición que adopta el médico puede ser variable, la mejor manera es la que se realiza con el paciente tranquilo, relajado y tumbado sobre la espalda, con la parte superior del cuerpo ligeramente incorporada.
3. Se utiliza el fonendoscopio, el mismo que debe ser colocado en los cinco focos de auscultación valvular, los mismos que no corresponden exactamente a su proyección anatómica, ya que la propagación de las vibraciones acústicas se hace preferentemente en los puntos en donde el corazón y los grandes vasos tienen contacto más íntimo con la pared torácica.

2.2.2 Auscultación pulmonar

Este procedimiento aporta un dato importante sobre el estado del parénquima pulmonar y de las pleuras, consiste en la exploración auditiva de los ruidos normales o no, que se producen en el aparato respiratorio. Estos ruidos son generados por flujo de aire que ingresa a través de las vías aéreas los mismos que causan turbulencias, que originan las vibraciones que percibimos como ruidos.

Estas turbulencias se producen en las zonas donde la velocidad del aire es mayor y en aquellas con condiciones geométricas que dificultan un flujo laminar. Esto ocurre principalmente en la laringe y en las bifurcaciones de los bronquios mayores, lobulares y segmentarios. En las vías aéreas más periféricas el área de sección va aumentando, por lo que la velocidad disminuye rápidamente, siendo muy lenta cerca de los alveolos, en los que el flujo es laminar. Por lo tanto, a nivel distal no hay turbulencias y tampoco se originan ruidos respiratorios. Estas características son de importancia ya que durante la exploración clínica nos permite identificar los hallazgos auscultatorios normales [15].

CAPÍTULO 3

Compresión de audio.

La compresión de audio es un método que realiza una compresión de datos, es decir, busca la reducción del tamaño de los archivos de audio. Para esto se emplea algoritmos de compresión de audio llamados códecs. Existen dos tipos de códecs: basados en algoritmos de compresión sin pérdida, y, basados en algoritmos de compresión con pérdida [18].

Uno de los principales objetivos de éste trabajo es la reducción del tamaño final de los sonidos, si que se altere la calidad de los mismos y así realizar su transmisión (Fig. 3.1). Para esto se realiza el uso de diversas técnicas de compresión, como implementación de la FFT, entramado, asignación de bits, entre otras, esto para conseguir la reducción del tamaño de los sonidos que se utilizan proporcionando una aceptable calidad de audio en una determinada aplicación.

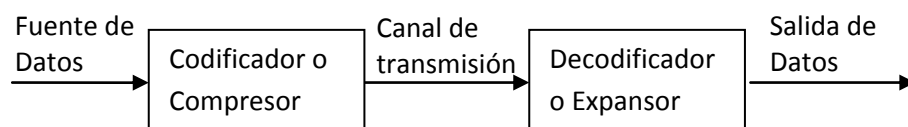


Fig. 3.1 Compresión en almacenamiento

3.1 Tipos de compresión

Las técnicas de compresión de audio normalmente se las clasifica en dos categorías básicas: sin pérdidas y con pérdidas.

3.1.1 Compresión sin pérdida

La compresión sin pérdidas se caracteriza porque los datos del expansor (procesador de dinámica del audio) son idénticos bit por bit a los datos originales de la fuente. Este tipo de compresión también se la conoce con el nombre de compresión reversible, por la razón que no pierde ninguna cantidad de información. La señal de audio comprimido puede ser descomprimida sin omitir ninguna muestra de información. Una de las principales desventajas de este tipo de compresión es que solo permite una mínima reducción de tasa de bits [18].

Se debe considerar que para un codificador sin pérdidas no se puede llegar a definir

un factor de compresión en particular y debido a esto los vínculos de comunicaciones y los procedimientos de almacenamiento que se utilizan deben tener la capacidad de funcionar con una tasa de datos variable.

Se puede realizar el uso de códecs sin pérdida al incluirlos en planes de prueba de la tasa de bits erróneos. También existe la posibilidad de poner en serie o en tándem códecs sin pérdida, sin precauciones especiales [18].

3.1.2 Compresión con pérdida.

La compresión con pérdidas se caracteriza porque los datos desde el expansor no son idénticos bit a bit a los datos de la fuente. Este tipo de compresión con pérdida es irreversible y debido a esto, la señal de audio reconstruida, perderá algunas de sus muestras originales.

La compresión con pérdida tiene la gran ventaja de permitir una mayor reducción de la tasa de bits. De la misma forma, al tener una mayor compresión aumentará la cantidad de distorsión en la señal de audio, y para solucionar este problema los códecs utilizan técnicas cuidadosamente seleccionadas. Con esto la distorsión resultante puede hacerse casi imperceptible al oído humano. Este tipo de códecs con pérdida no son de utilidad con respecto a datos informáticos, sin embargo son comúnmente utilizables para aplicaciones de audio, puesto que las pérdidas que se producen en los archivos de audio pasan desapercibidas. Por esta razones es el tipo de compresión que se va a implementar en el proyecto, debido a que obtenemos un mejor nivel de compresión que en los compresores sin pérdida,

Debido a la gran cantidad de compresores con pérdidas que existen, se debe señalar que códecs con pérdida exitosos son aquellos en los cuales las diferencias se corrigen para que el oído humano los encuentre subjetivamente difíciles de percibir. Para esto los códecs con pérdida deben estar basados en una comprensión de percepción psicoacústica, los mismos que toman el nombre de códigos perceptivos [18].

3.2 Tratamiento del sonido.

El tratamiento del sonido está conformado por varias etapas a las que se somete la señal de audio. Las principales se dan durante el almacenamiento, compresión y

transmisión del sonido.

3.2.1 Almacenamiento

El almacenamiento y reproducción del sonido puede hacerse siguiendo dos criterios que son: el magnetófono y la partitura. El primero pretende almacenar la forma de onda de la forma más fiel posible para reproducirla después. Es el método utilizado en los antiguos dispositivos analógicos como el gramófono o el magnetófono (el propio teléfono). Por otro lado, el sistema de partitura establece las reglas por las que se formará el sonido. Es la información contenida en la partitura la que informa sobre los resultados que debe conseguir.

La informática ha reproducido los criterios anteriores adaptándolos a los dispositivos digitales. De forma resumida el sistema del magnetófono digitaliza la onda por el procedimiento de anotar repetidamente su amplitud en intervalos de tiempo muy pequeños. Este es el denominado sistema de muestreo, del que existen muchas variedades. En todos los casos la onda es transformada en una serie de valores numéricos que se almacenan en un fichero.

Posteriormente el fichero puede ser leído por un programa especial y sus datos ser enviados secuencialmente a un dispositivo que es capaz de regenerar una tensión eléctrica proporcional a los valores (numéricos) recibidos, de forma que semeja la forma de onda original. La tensión producida es entregada a su vez a un altavoz, que la transforma finalmente en "sonido audible".

El sistema de la partitura utiliza un lenguaje especial parecido al de los músicos para almacenar información sobre cada nota. Es desde luego un lenguaje informático como cualquier otro (Algol, C, Fortran, Basic, etc.), aunque adaptado a las necesidades de la representación del sonido. Existe una amplia variedad de estos lenguajes musicales para ordenador cuyo funcionamiento es parecido. La "fuente" es aquí una partitura electrónica que indica cuándo debe sonar una nota, por cuánto tiempo, con qué volumen y cómo debe terminar. Después de un cierto proceso o "compilación", en vez de producir un ejecutable normal, producen un "binario de sonido", es decir, un fichero .wav, .au, .snd, .mid, .mp3, .aiff, etc.

Los ficheros de sonido pueden ser leídos por programas adecuados que los transforman en una serie de valores numéricos, que son enviados secuencialmente a

la tarjeta de sonido del ordenador. Esta los transforma en valores analógicos (tensiones), que finalmente se transforman en sonido en los altavoces.

Una forma especial del sistema de la partitura es el denominado sistema MIDI (Interfaz Digital de Instrumentos Musicales), que se creó para intercambiar información musical entre dispositivos electrónicos digitales. Recientemente han aparecido sistemas híbridos, que están a medio camino entre el sistema de muestreo y el modelo MIDI. Los ficheros resultantes (.mod, .xm, .mtm y .s3m) son menores que los de muestreo puro, pero mayores que los MIDI [19].

3.2.2 Compresión.

Debido a las necesidades de transmisión y almacenamiento de sonido, se recurre a comprimir la información resultante de su digitalización. Este recurso, que se emplea de forma recurrente en informática, se basa en que, cualquier información acaba siendo una sucesión de ceros y unos del tipo: 01101011100..., etc. que conforma un fichero en algún lugar del ordenador.

Basándose en el estudio de determinadas secuencias de repeticiones, los programadores han diseñado algoritmos matemáticos que permiten, mediante un proceso de cálculo (codificación) que realiza el propio ordenador, reducir el tamaño del fichero resultante. Un proceso inverso (decodificación) restituye el contenido original. Hay que reseñar que, dependiendo de la naturaleza del fichero almacenado, algunos de estos algoritmos consiguen compresiones realmente notables. Por ejemplo, el denominado MP3 (MPEG-1 Audio Layer 3) consigue tasas del orden de 1/20 en la compresión de sonido de calidad CD estéreo [20].

En una forma más técnica, definimos la compresión de audio como un proceso que atenúa la señal eléctrica en una determinada cantidad y a partir de un determinado nivel de entrada. El objetivo es conseguir que la excursión dinámica resultante sea inferior a la original, proteger ciertos equipos frente a los posibles picos de señal o si se trata de un sonido saturado intentar disimular el error. En el campo de la música, su uso va desde aplicaciones para grabaciones musicales hasta el sonido directo. En algunos casos estas técnicas de compresión se ven reforzadas por el hecho de que no se pretende una reconstrucción exacta del fichero original, sino que se permite una cierta diferencia, o pérdida de calidad, que en este caso se traduciría en una distorsión del sonido resultante respecto del original. Esta distorsión puede estar

dentro de lo admisible o ser sencillamente inapreciable.

Se ha utilizado tiempo y talento para conseguir los algoritmos de compresión y descompresión más rápidos y eficaces; los resultados son diferentes módulos software llamados códecs, casi todos patentados, que resuelven el problema de comprimir imágenes y sonido. Un ejemplo de algoritmo conservativo es el conocidísimo WinZip; por su parte, el MP3 sería un ejemplo de algoritmo que permite una cierta distorsión (no conservativo). Algunos de estos algoritmos se apoyan en dispositivos de hardware específicos (códecs hardware) para conseguir mayor rapidez de ejecución [13].

3.2.3 Transmisión digital del sonido.

El archivo codificado puede ser transmitido a través de Internet de dos formas: como archivos estáticos de transmisión y como una transmisión en vivo.

3.2.3.1 Archivos estáticos de transmisión.

El término se refiere a la entrega de archivos audiovisuales pre-grabados a través de la internet en tiempo real. Normalmente cuando se trata de archivos audiovisuales en línea estamos hablando de archivos estáticos de transmisión. Un archivo largo de video codificado en varios archivos de transmisión, es un ejemplo de una colección de archivos estáticos de transmisión. Este tipo de contenido es también conocido como 'bajo demanda' o 'contenido archivado'. Este contenido puede ser obtenido mediante procesos de Descarga y Reproducción, Descarga Progresiva y Streaming Verdadero [21].

3.2.3.2 Transmisión en Vivo.

Se refiere a la obtención de señales en directo de audio y/o video a través de internet. Esto permite al usuario disfrutar un evento en el mismo momento en que este sucede. Existen muchos ejemplos de esto como son las estaciones de radio en línea o la posibilidad de ver conciertos en vivo, por ejemplo. Únicamente el Streaming Verdadero soporta la transmisión en vivo.

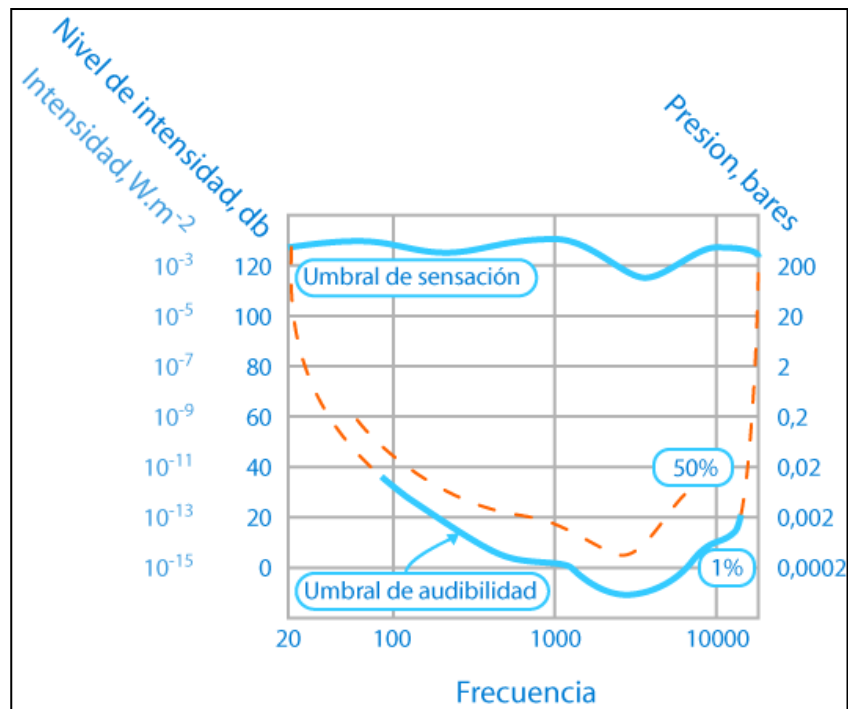
Esta tesis básicamente no está enfocada hacia el Streaming Verdadero tanto de contenidos audiovisuales en vivo como archivados o pre-grabados.

3.3 Modelo Psicoacústico.

La compresión de audio se encuentra basada en realizar una reducción del margen dinámico irrelevante. Esto quiere decir, de forma más específica, en utilizar el margen de la capacidad del sistema auditivo para detectar los errores de cuantificación en condiciones de enmascaramiento. Para esto los códecs realizan la división de la señal en bandas de frecuencia, las mismas que se aproximan a las bandas críticas, que son bandas preestablecidas. Luego de esta división se cuantifica cada subbanda en función del umbral de detección del ruido dentro de esa banda. En todo este proceso se utiliza la implementación del modelo psicoacústico, que utiliza un método denominado predicción polinómica, el cual consiste en analizar la señal de audio y calcular la cantidad de ruido que se puede introducir en función de la frecuencia, es decir, realiza el cálculo del umbral de enmascaramiento en función de la frecuencia [22].

Esta información es usada por los códecs para determinar la mejor forma de utilizar los bits disponibles (cantidad de bits preconfigurados en el codificador por el usuario).

De acuerdo a investigaciones realizadas en base al modelo psicoacústico, la distorsión generada es imperceptible para el oído en condiciones normales desde los 192 kbps, y, en casos especiales como personas con el oído no experimentado, con 128 kbps o 96 kbps [22]. En general, para personas que tienen una capacidad auditiva desarrollada (Fig.3.2), una codificación de 192 o 256 kbps es suficiente para escuchar correctamente, como ejemplo, la mayor parte de la música que circula por Internet se encuentra codificada entre 128 y 192 kbps [23].

Fig.3.2 Capacidad auditiva²

A continuación se muestran los pasos generales para el cálculo psicoacústico de la señal [40]:

1. Alineación en tiempo.
2. Representación espectral.
3. Componentes tonales y no-tonales.
4. Función de dispersión.
5. Umbral de enmascaramiento individual.
6. Umbral de enmascaramiento global
7. Umbral de enmascaramiento mínimo.
8. Relaciones señal a máscara.

3.4 Bandas críticas.

De acuerdo a los estudios de la discriminación en frecuencia del oído, se ha comprobado que en bajas frecuencias tonos con unos cuantos Hertz de separación pueden ser distinguidos. Sin embargo, al hablar de frecuencias altas, el oído necesita que estos tonos se encuentren separados por cientos de Hertz para poder

² Tomado de Educared
http://portales.educared.net/wikiEducared/index.php?title=Intensidad_del_sonido_e_intensidad_de_la_onda, última visita 20 de septiembre 2010.

distinguirlos. De forma general se conoce que el oído responde al estímulo más fuerte que se presente en sus diferentes regiones de frecuencia. A este comportamiento se le da el nombre de bandas críticas [56]. Las bandas críticas son más estrechas en las bajas frecuencias que en las altas. Así, en porcentaje tenemos que el 75% de las bandas críticas están por debajo de los 5 KHz, lo que implica que el oído recibe más información en las bajas que en las altas frecuencias. La distancia de una banda crítica se denomina Bark. La siguiente fórmula(3.1) da un valor aproximado del ancho de las bandas críticas:

$$\text{Ancho de la banda crítica(Hz)} = 24.7 (4.37 * F + 1) \quad (3.1)$$

Donde F es la frecuencia en Hz.

Las bandas críticas son comparables a un analizador de espectro con frecuencia central variable. Hay que tomar en cuenta que las bandas críticas no son fijas, ya que estas son continuamente variables en frecuencia y cada tono puede crear una banda crítica centrada en él [24].

3.4.1 El Bark.

El bark en honor al físico alemán Georg Heinrich Barkhausen, es la unidad de frecuencia perceptual. Un bark mide la tasa de banda crítica, lo que quiere decir que una banda crítica tiene un ancho de un bark. La escala bark relaciona la frecuencia absoluta (en Hz) con las frecuencias medidas perceptualmente (el caso de las bandas críticas). Con la ayuda del bark, se puede realizar la transformación de un sonido en el dominio de la frecuencia al dominio psicoacústico. De esta forma se puede representar en una curva de enmascaramiento psicoacústico un tono puro. El científico Eberhard Zwicker modeló el oído con 24 bandas críticas arbitrarias para frecuencias por debajo de 15 KHz con una banda adicional que ocupa la región entre 15 y 20 KHz, las cuales son utilizadas actualmente por los codificadores que utilizan el modelo psicoacústico en su funcionamiento [25]. El bark puede calcularse con las fórmulas(3.2) y (3.3):

$$1 \text{ bark (Hz)} \cong \frac{f}{100} \quad \text{para } f < 500 \text{ Hz} \quad (3.2)$$

$$1 \text{ bark (Hz)} \cong 9 + 4 \log\left(\frac{f}{100}\right) \quad \text{para } f < 500 \text{ Hz} \quad (3.3)$$

Se presenta en el Anexo 4, la relación entre la escala “Bark” y la escala estándar en Hertz; esta escala “Bark” corresponde con las distancias físicas en la membrana basilar de la cóclea del oído humano.

3.5 Métodos de codificación.

En la actualidad existe una gran y diversa cantidad de codificadores de audio los mismos que a su vez constantemente se están actualizando. Por esta razón en una forma global se pueden clasificar en dos grupos, los cuales son:

3.5.1 Codificadores perceptuales.

Los codificadores perceptuales utilizan las limitaciones en la percepción del sistema auditivo humano para codificar el flujo de datos. De esta forma para la codificación del formato PCM (Modulación por Impulsos Codificados) se realiza una transformada al dominio frecuencial y las tramas se cuantifican y codifican a partir de un conjunto de datos obtenidos en un modelo psicoacústico, como se observa en la Fig.3.3. Esta implementación es lo que determina la calidad final y es aquí donde se realiza el aprovechamiento del enmascaramiento temporal, conociendo que es el umbral de enmascaramiento el que controla la cuantificación.

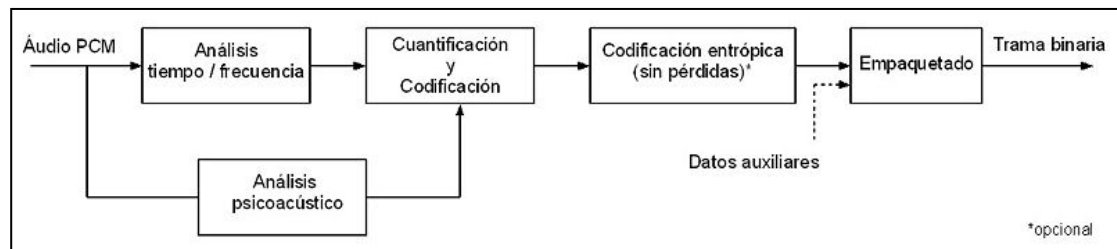


Fig.3.3 Codificador perceptual[28]

Para la cuantificación de audio se emplean varios tipos de codificadores como son: uniforme, no uniforme y diferencial. De estos, el que normalmente se utiliza para el tratamiento de la música es el no uniforme; con respecto a señales de voz se emplea la cuantificación diferencial y para un tratamiento más avanzado de la voz se utiliza la cuantificación vectorial. Además, a las salidas de la codificación se puede realizar la implementación de un codificador entrópico que se caracteriza por no tener pérdidas y con lo cual se puede reducir el tamaño de la trama binaria. Posterior a esto, las tramas se empaquetan en conjunto con la información que contiene los datos de la

codificación necesarios para su decodificación y la detección de errores en caso de presentarse alguno [26].

Finalmente el decodificador desempaqueta la señal codificada obteniendo las tramas y la información auxiliar. Luego procede a decodificar y obtener cada una de las muestras y las convierte en el dominio temporal y así obtener el audio original con el formato PCM. Entre los estándares que utilizan este método de codificación tenemos el MPEG-1 ((MovingPicturesExpertsGroup).

3.5.2 Codificadores paramétricos.

Los codificadores paramétricos se caracterizan porque pueden representar la voz y el audio con tonos aislados, patrones armónicos y componentes ruidosas, los cuales requieren una mínima cantidad de bits para ser representados.

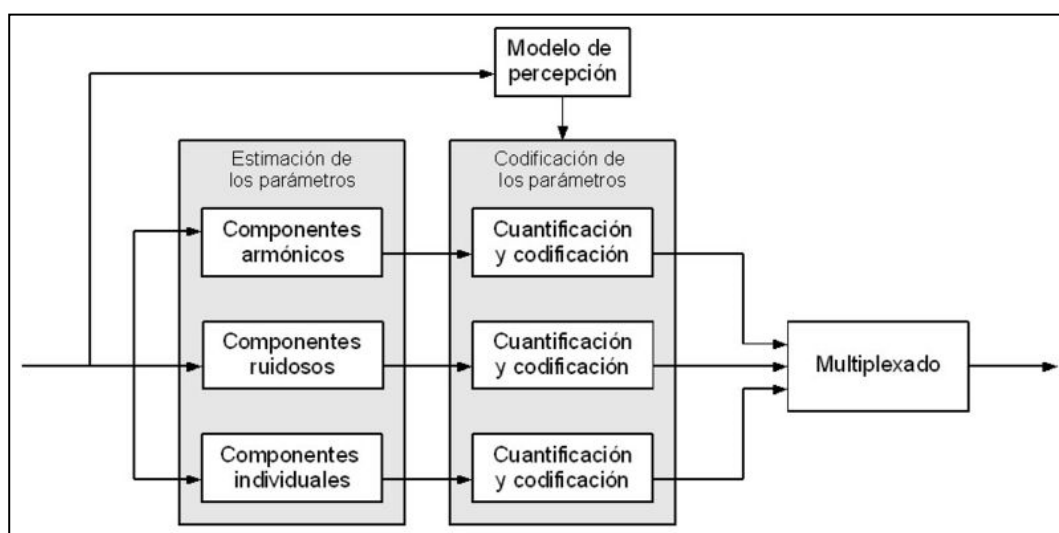


Fig.3.4 Codificación paramétrica[28]

Al iniciar la codificación primero se extrae la información de cada muestra de entrada implementando la transformada de Fourier para luego proceder a la estimación de los parámetros (27). Después se codifican estos parámetros basándose en el modelo psicoacústico y se multiplexan formando así la trama binaria para proceder a su transmisión como podemos observar en la Fig.3.4.

Para la decodificación se utiliza la información de un bloque de síntesis y utilizando los parámetros codificados se obtiene una señal con características similares a la original. Como un ejemplo de este método tenemos la codificación HVXC (Harmonic Vector Excitation Coding) para voz, que se implementa en el estándar MPEG-4

(nombre formal es ISO/IEC 44196) [28].

3.5.2.1 Vocoders.

Los Vocoders son codificadores paramétricos que específicamente se utilizan para la codificación de voz. Su funcionamiento se basa en analizar la señal de voz la cual corresponde a un segmento temporal y extraer los parámetros del modelo y la excitación. Esta información obtenida es la que se codificará. Para el proceso de decodificación se implementa un modelo de producción de voz que sirve para sintetizar los parámetros, todos estos procesos se resumen en la Fig.3.5 [27].

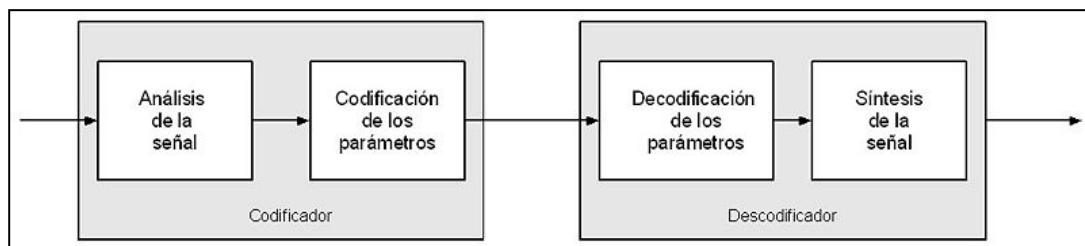


Fig.3.5 Esquema de un Vocoder[28]

Actualmente existen dos tipos de vocoders:

- Vocoder por predicción lineal.
- LPC-10.

3.5.2.2 Codificadores de forma de onda.

Los codificadores de forma de onda se caracterizan por realizar un estudio de la señal de audio. Su finalidad es reproducir la forma de la señal de entrada. Este tipo de codificadores son diseñados para ser independientes al tipo de la señal, por lo cual se utilizan para codificar una gran variedad de señales. Su funcionamiento se basa en utilizar la redundancia de la señal e implementar un sistema de predicción lineal para después proceder a codificar el canal auditivo. De esta forma aunque las señales fueran muy redundantes y aparentemente nulas se pueden conseguir tasas de compresión muy altas[28].

Esta forma de codificación se puede realizar tanto en el dominio frecuencial como en el dominio temporal.

a) 3.5.2.2.1 Codificadores en el dominio frecuencial:

- Codificación en sub-bandas.
- Codificación por transformada.

b) 3.5.2.2.2 Codificadores en el dominio temporal:

- Modulación por codificación de impulsos (PCM, Pulse Code Modulation).
- Modulación por codificación de impulsos diferenciales (DPCM, Differential PCM).
- Modulación por codificación de impulsos diferenciales adaptativos (ADPCM, Adaptive PCM) [28].

3.5.2.3 Codificadores híbridos.

Los codificadores híbridos (Fig.3.6) son codificadores de análisis por síntesis que combinan las técnicas de codificación de los vocoders y los codificadores de forma de onda. Su fin es obtener una voz de alta calidad a tasas de bit muy bajas, inferiores a los 8KHz. Su funcionamiento consiste en realizar el análisis de un conjunto de muestras como si se tratase de una sola, para de esta forma obtener los parámetros de la señal [29].

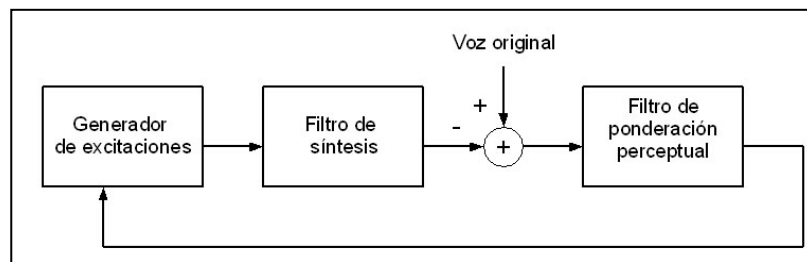


Fig.3.6 Codificador Híbrido[28]

A continuación tenemos algunos ejemplos de los codificadores híbridos más conocidos:

- Codificador RELP.
- Codificador multipulso MPC.
- Codificador CELP.
- Codificador VSELP.
- Codificador RPE-LTP

CAPÍTULO 4

Códecs de compresión de audio.

Este capítulo va a tratar de forma detallada las características principales de los codecs de compresión utilizados para el desarrollo de esta tesis como son: el OggVorbis (MDCT³) y G.722 (ADPCM⁴).

4.1 OggVorbis (MDCT)

En esta parte se va a tratar las principales especificaciones del códec OggVorbis, de igual manera su proceso de codificación y decodificación, terminando con algunas ventajas y desventajas al utilizar este tipo de código de compresión.

4.1.1 Introducción

Vorbis es un códec de audio perceptivo de fines generales previsto para permitir flexibilidad máxima del codificador, permitiéndole escalar competitivamente sobre una gama excepcionalmente amplia de tasas o velocidades de bits. Se crea en respuesta a la demanda de codecs de compresión de licencia gratuita, especialmente para integrarse a sistemas operativos como Linux. Soporta hasta 255 canales. Las frecuencias de muestreo en que trabaja son 44.1 KHz y 48KHz y más de 16 bits de resolución. Su código es abierto, razón por la cual no es posible delinear sus características con exactitud ya que puede adaptarse a las necesidades de uso. Es perfectamente comparable con otros formatos usados para almacenar y reproducir música digital, como por ejemplo MP3, VQF o AAC⁵. Su popularidad está en auge principalmente porque su versión definitiva es reciente, y ya es soportado por los principales reproductores del mercado. Además consta en los paquetes de códecs que se pueden instalar en Windows, Linux y en MAC OS24 [30].

Vorbis también está pensado para frecuencias de muestreo bajas desde telefonía de 8KHz y hasta alta definición de 192KHz, y una gama de representaciones de canales (monoaural, polifónico, estéreo, cuadrafónico, 5.1, ambisónico o hasta 255 canales discretos).

³MDCT (Transformada discreta del coseno modificado), una transformada lineal ortogonal.

⁴ADPCM (Adaptive Differential Pulse Code Modulation), codificador de forma de onda.

⁵MP3 (MPEG-1 Audio Layer III o MPEG-2 Audio Layer III) Compresor de audio con pérdidas, VQF (extensión de un fichero TwinVQ), TwinVQ (Transform-domain Weighted Interleaved Vector Quantization) y AAC (Advanced Audio Coding) otro formato de compresión de sonido basado en MPEG-2.

Ogg es el nombre del formato contenedor desarrollado por Xiph.org para audio, video y metadatos. Vorbis es el nombre de un esquema específico de compresión de audio que fue designado para ser contenido en Ogg. Cabe aclarar que Ogg es capaz de contener otros formatos, como FLAC y Speex.

OggVorbis es totalmente abierto, libre de patentes y de regalías; la biblioteca de referencia (libVorbis) se distribuye bajo una licencia tipo BSD⁶ por lo que cualquiera puede implementarlo ya sea tanto para aplicaciones propietarias como libres. Además utiliza un algoritmo de compresión con pérdidas lo que significa que descarta datos para poder comprimir mejor un archivo de lo que sería posible normalmente [30].

Existen varios proyectos multimedia que usan Ogg que están siendo desarrollados por la Fundación Xiph.org, pero entre los más conocidos están:

- Vorbis: Códec de audio general.
- Theora: Códec de video basado en VP3⁷.
- FLAC: Códec de audio sin pérdida.
- Speex: Códec de voz humana.
- Icecast: Servidor de streaming [31].

4.1.2 Características y Funcionamiento.

OggVorbis utiliza una codificación variable de la frecuencia de bits (abreviado *VBR*). VBR es un método de codificación que modifica el número de bits utilizados por segundo para codificar datos de audio en función de la complejidad de la transmisión de audio en un momento determinado. En otras palabras, esto significa que la cantidad de datos que se utiliza para codificar una porción determinada de audio, no será igual para el silencio que para un concierto polifónico [30].

4.1.2.1 Codificador.

OggVorbis sólo define su decodificador, esto quiere decir que, cualquier codificador que produzca una trama decodificable por Vorbis es considerado un codificador

⁶ BSD (Berkeley Software Distribution). Es una licencia de software libre permisiva

⁷VP3 era originalmente un códec de vídeo propietario, desarrollado por On2 Technologies. Es comparable en calidad y bitrate al códec de vídeo MPEG-4 Parte-2

Vorbis. Esto permite que se implementen mejoras o cambios en el codificador sin cambiar el decodificador. A continuación se define el proceso para codificar una trama de audio (Fig. 4.1).

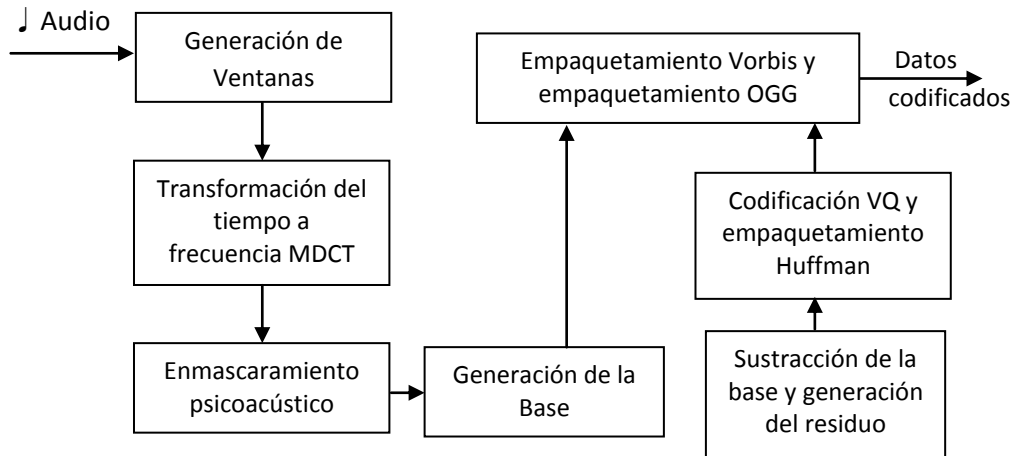


Fig.4.1 Codificación OggVorbis⁸

4.1.2.2 Generación de ventanas.

La trama de audio entrante PCM es dividida en bloques llamados ventanas, lo que se hace para efectos de reducción del pre-eco producido por la posterior transformación de dominio con MDCT. La extensión de las ventanas deben ser una potencia de dos entre 64 y 8192 muestras. Generalmente se utiliza dos tamaños comunes de ventanas: las grandes con 2048 y las pequeñas con 256. La razón de utilizar varios tamaños de ventanas se debe a que el uso de ventanas pequeñas reduce el efecto del pre-eco y las ventanas grandes permiten una mejor resolución del espectro. Estas a su vez son creadas con un solapamiento del 50% para cumplir con los requerimientos de la MDCT, como se muestra en la Fig. 4.2 y 4.3 [30].

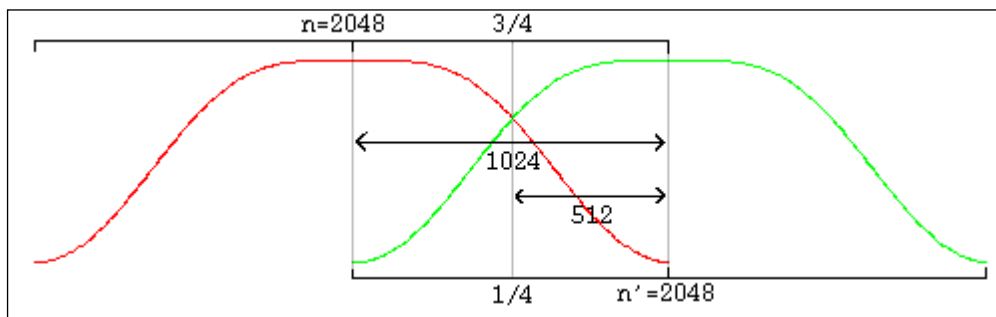


Fig.4.2 Solapamiento de dos ventanas con la misma extensión⁹

⁸ Esquema realizado en base a la descripción de un codificador vorbis general descrito en la especificación Vorbis 1 [30]

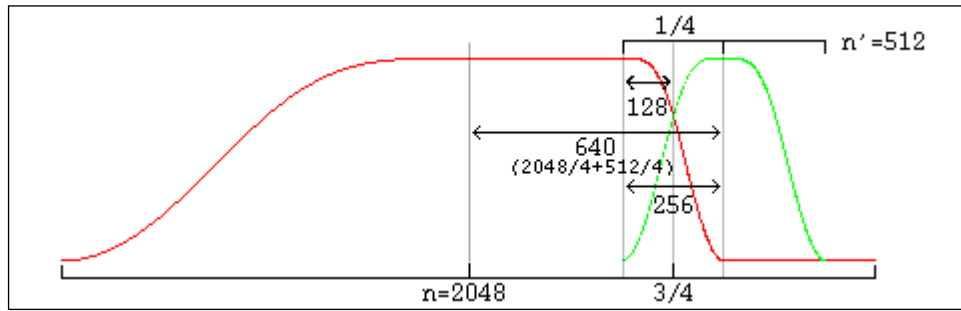


Fig.4.3 Solapamiento de una ventana grande y una pequeña⁹

Vorbis utiliza el tipo de ventana descrito en la ecuación 4.1 para el solapamiento [30].

$$y = \text{sen} \left(\frac{\pi}{2} \cdot \text{sen}^2 \left(\frac{x+0.5}{n} \pi \right) \right) \quad (4.1)$$

4.1.2.3 Transformación de dominio MDCT.

Debido a que el estándar del codificador en Vorbis es abierto, es posible utilizar cualquier tipo de transformada. En el caso de la primera versión de Vorbis, se utiliza la MDCT, la cual permite la transformación de una señal temporal al dominio de la frecuencia. Esta transformada ha sido adoptada por los códec de alta calidad debido a que elimina el fenómeno conocido como aliasing¹⁰ [30].

La MDCT es muestreada críticamente. El muestreo crítico mantiene fija la cantidad de datos para reconstruir la señal, es decir, para reconstruir una señal de tamaño N se necesita N coeficientes de la MDCT.

El proceso de la transformada describe dos fases: análisis y síntesis. En la fase de análisis se toman bloques de N muestras con un traslape de 50%. A estos bloques se los multiplica por una ventana de análisis válida que debe cumplir con el principio de Princen-Bradley que dice: "La suma al cuadrado de las partes de la ventana que se solapan debe ser igual a 1", luego se calcula la MDCT a través de la ecuación 4.2 para cada una de los bloques [30].

$$X_i[k] = \sum_{n=0}^{N-1} w_a^i[n] x_i[n] \cos \left[\frac{2\pi}{N} (n + n_0) \left(k + \frac{1}{2} \right) \right] \quad (4.2)$$

⁹Tomado de las especificaciones Vorbis I [30]

¹⁰Aliasing es el efecto que causa que señales continuas distintas se tornen indistinguibles cuando se les muestrea digitalmente.

Para $k = 0, \dots, \frac{N}{2} - 1$

Donde:

N = Número de muestras

$$n_0 = \frac{\left(\frac{N}{2} + 1\right)}{2} \quad (4.3)$$

w_a^i = ventana de análisis del i -ésimo bloque.

4.1.2.4 Enmascaramiento Psicoacústico.

Vorbis utiliza el modelo psicoacústico humano para descartar información no audible, característica que lo ubica en el grupo de códec con pérdidas. Pero a diferencia de los otros códecs que utilizan un modelo con un volumen fijo durante la duración de la trama de audio, Vorbis asume que el volumen se ajusta dinámicamente con un máximo situado en el umbral del dolor [30].

4.1.2.5 Cabecera de identificación.

La cabecera de identificación contiene campos que son usados para declarar la trama como Vorbis y proveer información acerca de la trama de audio como el número de canales, la tasa de bits (máxima, nominal y mínima), la longitud de cada campo se define a continuación:

- [vorbis_version] = 32 bits enteros sin signo, versión de Vorbis.
- [audio_channels] = 8 bits enteros sin signo, número de canales de audio.
- [audio_sample_rate] = 32 bits enteros sin signo, tasa de muestreo.
- [bitrate_maximum] = 32 bits enteros con signo, máxima tasa de bits.
- [bitrate_nominal] = 32 bits enteros con signo, tasa de bits normal.
- [bitrate_minimum] = 32 bits enteros con signo, mínima tasa de bits.
- [framing_flag] = 1 bit, bandera señalando el fin de la cabecera [30].

4.1.2.6 Cabecera de comentarios.

Esta cabecera consta de ocho campos de 32 bits cada uno. Es utilizada para proveer información sobre el archivo de audio como el autor, nombre del archivo, año de producción, etc. Estos campos están abiertos para que se ubique cualquier información que se crea conveniente. A continuación se describe el proceso de decodificación de esta cabecera:

- [vendor_length] = 32 bits enteros sin signo, longitud del campo del nombre del proveedor.
- [vendor_string] = leer vector UTF-8, nombre o comentario del proveedor o creador.
- [user_comment_list_length] = 32 bits enteros sin signo, espacio para comentarios.
- [length] = leer 32 bits enteros sin signo.
- [comment] = leer un vector UTF-8 }
- [framing_flag]= 1 bit, bandera de fin de trama [30].

4.1.2.7 Cabecera de configuración.

La cabecera de configuración contiene toda la información necesaria para el proceso de decodificación. Esta cabecera contiene lo siguiente: la lista de códigos, configuraciones de la transformada, configuraciones de la base, configuraciones del residuo, y, configuraciones de los canales y del modo.

En este caso los campos son de longitud variable, así, el campo de los códigos contiene ocho bits sin signo por paquete. Estos se almacenan en una matriz llamada [vorbis_codebook_configurations] [30].

El campo de la base o floor [vorbis_floor_types] contiene 16 bits enteros sin signo, indicando el tipo de base a decodificar (0 o 1). Esta configuración se almacena en un campo llamado [vorbis_floor_configurations]. Si el campo lee un número mayor a uno la trama no es decodificable. El campo del residuo [vorbis_residue_types] indica el tipo de residuo a decodificar tal como el caso anterior [vorbis_residue_configurations] almacena las configuraciones [30].

El campo modo contiene lo siguiente:

- [vorbis_mode_blockflag] = 1 bit, es la bandera de inicio.
- [vorbis_mode_windowtype] = 16 bits, indica el tipo de ventana.
- [vorbis_mode_transformtype] = 16 bits, indica el tipo de transformada [30].

4.1.2.8 Proceso de decodificación.

Las tramas se decodifican a través del proceso detallado en la Fig.4.4.

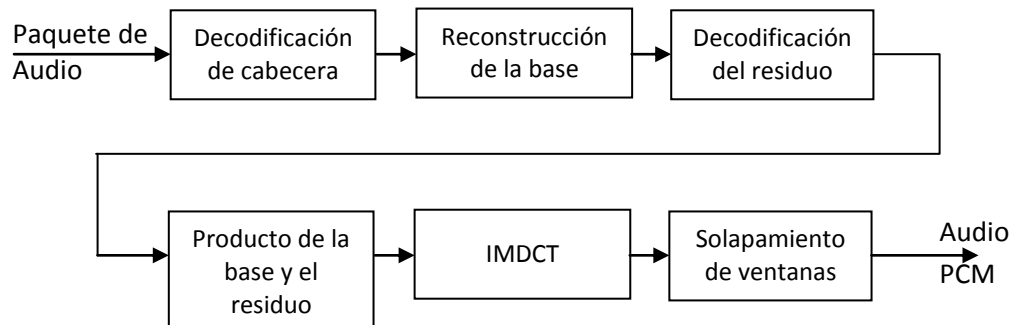


Fig.4.4 Proceso de decodificación¹¹

- Decodificación de cabecera: En este bloque se decodifica la información necesaria para el proceso de reconstrucción del audio, se realiza los siguientes pasos:
 - Decodificación de la bandera que señala el tipo de paquete.
 - Decodificación del número de modo.
 - Decodificación del tipo de ventana.
- Reconstrucción de la base: Se decodifica los vectores correspondientes a la base y se la reconstruye a través del algoritmo de Bresenham.
- Decodificación de los residuos en vectores de residuos.
- Cálculo del producto de la base y de los residuos generando un vector del espectro del audio.
- Transformada inversa monolítica del vector del espectro de audio, siempre de tipo MDCT en Vorbis I.
- Solapamiento de las ventanas: Finalmente se realizan los siguientes pasos:

¹¹ Realizado en base a la especificación Vorbis I [30]

- Solapar/adicionar la salida de parte izquierda de los bloques de la transformada con la salida de la parte derecha del anterior bloque.
- Guardar los datos de la parte derecha del bloque actual para el siguiente solapamiento.
- Si no es el primer bloque, devuelve los resultados del proceso de solapar/adicionar como audio resultante del bloque actual [30].

4.1.2.12 Códigos.

Vorbis no tiene un modelo probabilístico estático, es así que se debe empaquetar toda la configuración de decodificación para descifrar la información codificada con Huffman, en la cabecera de configuración [30].

4.1.2.13 Reconstrucción del audio.

El espectro de audio es transformado al dominio del tiempo en PCM (Modulación por Codificación de Pulsos) mediante la IMDCT (Transformada Inversa Discreta del Coseno Modificada). La señal PCM producida por la IMDCT no es todavía el audio final. La señal aún está dividida en bloques producto de la operación producida por el división en ventanas. Estos bloques deben ser solapados de nuevo entre ellos para lograr la reconstrucción correcta del audio. Cabe recalcar que este proceso demanda la utilización de un buffer para preservar los datos de la parte derecha de los bloques para que se pueda solapar con la parte izquierda de los nuevos bloques entrantes [30].

La transformada inversa IMDCT es descrita por la ecuación 4.4 [30]:

$$\tilde{x}_i[n] = w_s^i[n] \frac{4}{N} \sum_{k=0}^{\frac{N}{2}-1} X_i[k] \cos \left[\frac{2\pi}{N} (n + n_o) \left(k + \frac{1}{2} \right) \right] \quad (4.4)$$

Para $n = 0, \dots, N - 1$

w_s^i = ventana de síntesis para el i -ésimo bloque

4.1.3 Detalles técnicos

- Frecuencia de Muestreo: 8 – 192 kHz.
- Frecuencia máxima de la señal: 4000 – 95000 Hz.
- Tasa de bits: 45 - 500 kbits/s.
- Latencia (Retraso): >100 ms.
- Relación de compresión: 35%, 50%, 63%, 75%.
- Factor de Compresión: 2.84, 2, 1.6, 1.33 respectivamente.

4.1.4 Ventajas y desventajas.

En esta parte se van a detallar cada una de las ventajas y desventajas de éste códec de acuerdo a sus características y su funcionamiento.

4.1.4.1 Ventajas.

- Los archivos Vorbis pueden comprimir a un tamaño de archivo menor y seguir con un sonido de muy buena calidad. La mejor compresión de Vorbis hace que los costos de ancho de banda disminuyan y los resguarda de ser víctimas de su propia popularidad.
- Vorbis es abierto, y las personas son libres de usarlo en su plataforma preferida. Vorbis trata de mejorar la calidad del sonido en un archivo más comprimido que en los otros formatos de audio conocidos.
- Vorbis puede ser fácilmente editado y cortado, soporta un gran número de canales de sonido y permite ser encadenado lógicamente [32].

4.1.4.2 Desventajas.

- El formato Ogg-Vorbis no es compatible como el formato MP3. Esto significa que el usuario tiene que utilizar un reproductor de sonido que admita el formato o instalar un códec para reproducir los archivos Ogg-Vorbis.
- Todos los sistemas operativos aún no contienen en sus instaladores este códec que permite reproducir este tipo de archivos Ogg-Vorbis.
- A tratarse de un formato nuevo, no existe mucha información, que en algunas aplicaciones, como la elaboración de proyectos más avanzados, serian muy importantes.

4.2 G.722 (ADPCM).

En esta parte se va a tratar las principales especificaciones del códec G.722, detallando su proceso de codificación y decodificación, terminando con ventajas y desventajas al utilizar este tipo de código de compresión en archivos de audio.

4.2.1 Introducción

El códec G.722, lanzado en 1988, es utilizado para aplicaciones de voz de banda ancha con frecuencias desde 50 hasta 7000 Hz, y ofrece mayor calidad y claridad en audio a coste del ancho de banda. Este códec es útil en aplicaciones en una red de VoIP fija (voz transmitida sobre redes IP), donde el ancho de banda no es limitado. Este códec de compresión utiliza el método de Modulación por Pulsos Codificados Diferencial Adaptativo (ADPCM), el cual cambia el tamaño de la etapa de cuantificación, teniendo un factor adaptativo de escala [33].

Hoy en día cada vez más terminales que incorporan la característica de transmisión de voz en la red, utilizan como soporte el códec G.722, el cual es una evolución natural del conocido G.711(códec de compresión de audio para señales de frecuencias que caracterizan a la voz, modulación PCM, utilizado en videoconferencias), que se encuentra exclusivamente en la transmisión de voz sobre las redes de internet (VoIP) y que se desmarca (en cuanto a calidad) a la telefonía tradicional. Las patentes G.722 han expirado, así que está libremente disponible [33].

Entre los programas que lo utilizan tenemos sólo dedicados a realizar llamadas sobre la red como por ejemplo, Asterisk 1.6 (programa utilizado para VoIP) sí que soporta G.722 en su totalidad, siendo capaz de grabar y reproducir este tipo de archivos [34].

Como otros códecs profesionales, existen distintas versiones o distintas formas, existiendo en nuestro caso los siguientes:

- ITU G.722 – 48/56/64 Kbps ADPCM, ancho de banda de audio 7KHz.
- ITU G.722.1 – 24/32 Kbps, ancho de banda de audio 7KHz (basado en Polycom's SIREN codec)
- ITU G.722.1C – 32 Kbps, a Polycom extension, ancho de banda de audio 14KHz.

- ITU G.722.2 – 6.6Kbps to 23.85Kbps. También conocido como AMR-WB. CELP ancho de banda de audio 7KHz[35].

Existen otras implementaciones como el Siren7 de Polycom, una versión del G.722.1 cuyo soporte acaba de ser añadido y que ofrece una calidad similar al de una emisora de radio a la vez que reduce el consumo de procesamiento y de ancho de banda necesario [36].

4.2.2 Características y Funcionamiento

En la Fig.4.5 se puede observar de forma conjunta los procesos de codificación y decodificación del códec G.722.

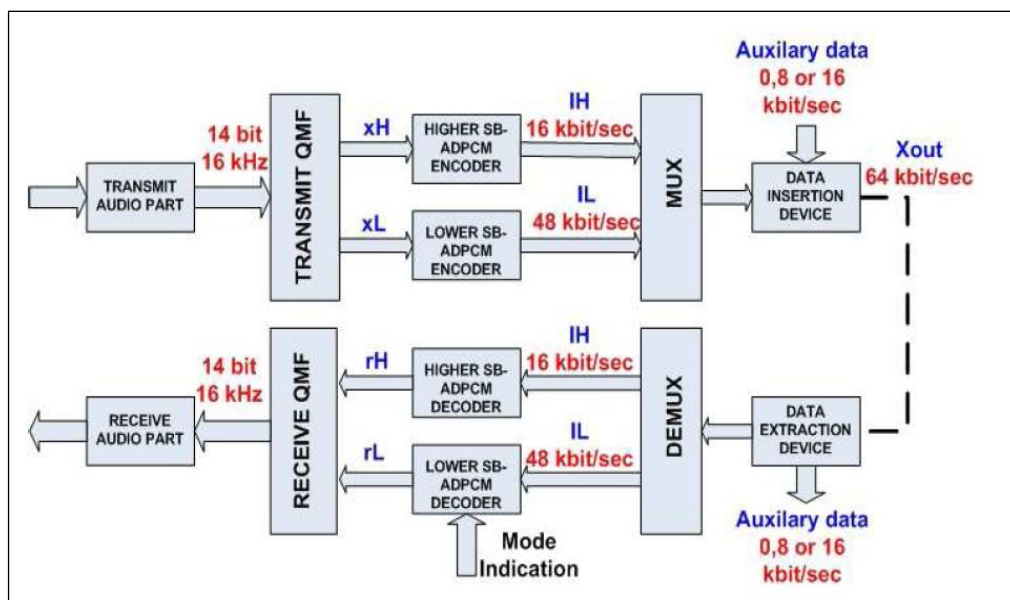


Fig.4.5 Diagrama de bloques Códec G.722¹²

4.2.2.1 Codificador y Decodificador

El códec G.722 utiliza la forma matemática del ADPCM, el mismo que es un codificador de forma de onda basado en DPCM que añade algunas funcionalidades. Antes de la digitalización se toma la señal analógica y se divide en bandas de frecuencia gracias a los filtros QMF (“QuadratureMirrorFilter”) en donde se obtienen sub-bandas de señal. Cada sub-banda es tratada de modo distinto utilizando las

¹² Tomada de “G.722 WIDEBAND SPEECH CODEC IMPLEMENTATION ON BF-533 DSP” tesis realizada por Avi Peretz y Cagdas Gumusoglu, Master’s Thesis Department of Telecommunications and Signal Processing Blekinge Institute of Technology, Sweden

propiedades de DPCM, es decir, se lleva a cabo el proceso de muestreo, cuantificación del error de predicción y finalmente se codifica. Un vez se obtiene la sucesión de bits (“bitstream”) de cada sub-banda, se multiplexan los resultados y ya se puede proceder a almacenar los datos o bien transmitirlos (Fig.4.6) [33].

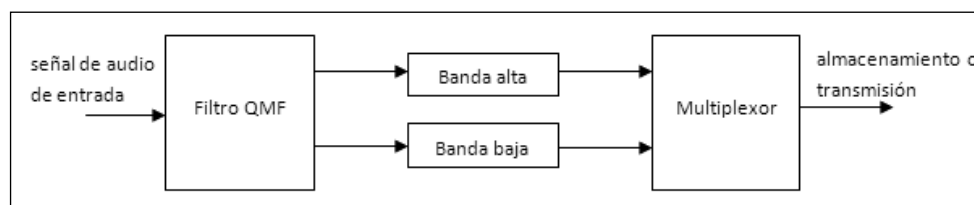


Fig.4.6 Diagrama de bloques Codificador G.722 [33]

Por la forma como realiza la codificación este códec, puede ser conveniente en ciertas aplicaciones, como por ejemplo en la codificación de voz, donde lo que se puede hacer es codificar la sub-banda que incluye la voz con más bits que en las otras sub-bandas que no son de tanto interés. Es una manera de reducir el tamaño del fichero resultante [35].

El decodificador ADPCM realiza el proceso inverso, es decir, demultiplexar y decodificar cada sub-banda del “bitstream” o sucesión de bits, como se muestra en la figura 4.7, para reconstruir de mejor manera la señal original.

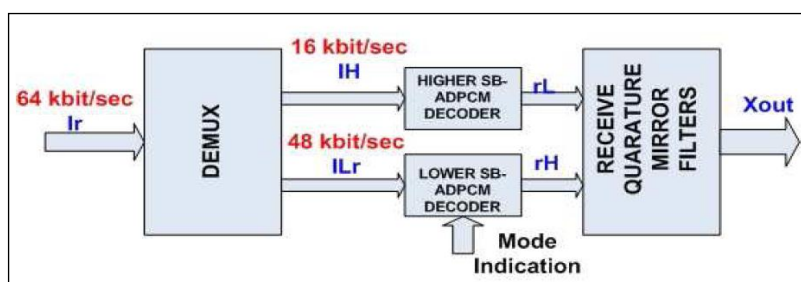


Fig.4.7 Diagrama de bloques Decodificador G.722¹³

4.2.3 Detalles técnicos

- Frecuencia de Muestreo: 8 KHz
- Frecuencia máxima de la señal: 7000 Hz
- Tasa de bits: 32 hasta 64 kbits/s

¹³ Tomada de “G.722 WIDEBAND SPEECH CODEC IMPLEMENTATION ON BF-533 DSP” tesis realizada por Avi Peretz y Cagdas Gumusoglu, Master’s Thesis Department of Telecommunications and Signal Processing Blekinge Institute of Technology, Sweden

- Latencia (Retraso): 4 ms
- Relación de compresión: 25%
- Factor de Compresión: 4

En la tabla 4.1 se observa las principales características del códec G.722.1, las mismas que ayudan a realizar comparaciones con el códec OggVorbis.

Tabla.4.1 Características generales G.722.1 [36]

Parámetros	Valor
Audio sample rate	16 kHz
Bit rate	16, 24, 32 kbps
Ancho de banda de audio	50 Hz - 7 kHz.
Tamaño del marco de audio	20 ms
Retraso del Algoritmo	40 ms
RAM	< 7.5 k bytes
ROM espacio de tabla	~ 20 k bytes

4.2.4 Ventajas y desventajas

En esta parte se van a detallar cada una de las ventajas y desventajas de éste códec de acuerdo a sus características y su funcionamiento.

4.2.4.1 Ventajas

- El G.722 es de muy buena calidad al compararse con la telefonía normal, al realizar llamadas de VoIP, que es su mayor aplicación.
- Las principales ventajas de la utilización de el uso de G.722 en VoIP son la integración de servicios en una misma estructura, y la reducción de costes.
- Los niveles de frecuencia con los que trabaja G.722 contienen a las frecuencias en las que se encuentran los sonidos implementados (100KHz a 1500KHz), además tiene un gran porcentaje de compresión, superior a 90%, ideal para la compresión de los sonidos utilizados en este proyecto.
- Las ventajas que ofrece G.722 en VoIP son suficientes para plantearse su uso dentro de una organización. Incluso los problemas de ruido que inicialmente

planteaba esta tecnología han sido mejorados, consiguiéndose una alta calidad en la comunicación. Sin embargo, no todo son ventajas.

4.2.4.2 Desventajas

- El códec G.722 sólo es utilizado para la transmisión de voz sobre IP, por ende no es comercializado como un tipo de compresor para cualquier archivo de audio como en el caso del Ogg-Vorbis también analizado.

- Un inconveniente primordial es la seguridad, ya que los sonidos circulan por la red de datos, lo que hace que tenga los mismos problemas de seguridad. Si alguien consigue acceso a los datos de una red, podría perfectamente captar una comunicación, y adueñarse de esta información.

- Otro inconveniente es que hasta el momento no se ha desarrollado un códec de almacenamiento con el estándar G.722, solamente ha sido implementado en programas de transmisión en tiempo real, por lo cual la información para el desarrollo de un códec de estas es casi nula.

CAPÍTULO 5

Descripción General

En este capítulo se va a detallar la parte teórica que ha sido aplicada en el software realizado para esta tesis, tanto para desarrollar el códec OggVorbis como para la implementación del códec G.722 a partir de un software establecido.

5.1 Desarrollo de codificador MDCT (Ogg-Vorbis)

El codificador de audio desarrollado para el tratamiento de señales obtenidas a través del estetoscopio fue diseñado en el software Matlab, en el cual sus principales características son el uso de curvas de enmascaramiento para determinar la asignación de bits de una forma adecuada para realizar la cuantificación de los coeficientes MDCT. De esta manera se puede utilizar el menor ancho de banda para la transmisión de estas señales en una red de Telemedicina, en comienzo en la red Tutupaly – Hospital UTPL.

5.1.1 Entramado y segmentación

El primer paso en el diseño de un codificador de audio consiste en segmentar el flujo de audio en tramas. A frame is a short section of audio, typically less than 50ms each. Una trama es una pequeña sección de audio que tiene aproximadamente una duración de menos de 50ms. At a sampling frequency of 44.1kHz, a frame of 2048 samples is about 46ms long. Cada trama tiene una frecuencia de muestreo de 44,1 KHz, y están constituidas por 2048 muestras. Las longitudes más largas de las tramas permiten una mayor resolución de frecuencia, gracias a que la longitud FFT es también mayor. Shorter frame sizes may be used for more transient sounds, to avoid an artifact known as pre-echo. En lo que respecta a tamaños más cortos de las tramas, estas pueden ser usadas para los sonidos más transitorios (sonidos percibidos usualmente), para evitar un artefacto conocido como pre-eco. El pre-eco es un fenómeno psicoacústico que se da cuando se produce una transición de audio abrupta, y causa un ataque borroso al sonido.

En porcentaje un estimado de 2 frames x Kb

427K	—————>	212 frames
829K	—————>	413 frames

5.1.2 SMR y aplicación de la FFT.

La relación señal a máscara (SMR) es la medida de la potencia de la señal respecto a la del nivel de la máscara utilizada en codificación sub-banda. Se calcula como el cociente entre la máxima energía de la señal y el umbral de enmascaramiento global. El umbral de enmascaramiento global se compara con el máximo nivel de señal para una sub-banda produciendo esta relación. Hay muchas maneras de calcular el umbral de enmascaramiento, el cualn general, the masking threshold varies depending on the frequency and intensity of the masker signal. varía en función de la frecuencia e intensidad de la señal enmascaradora. The width of the masking curve typically extends farther in the direction of higher frequencies than toward lower frequencies, and the amplitude depends on both the frequency and the tonality of the masker. Noise tends to mask much more than tones do. El ancho de la curva de enmascaramiento normalmente se extiende en la dirección de las frecuencias más altas que hacia frecuencias más bajas, y la amplitud depende de la frecuencia y la tonalidad de la señal enmascaradora.

In order to calculate the masking threshold, the first step is to calculate the FFT of the frame, and find the spectral peaks. Con el fin de calcular el umbral de enmascaramiento, el primer paso es calcular la FFT¹⁴ de la trama, y encontrar los picos espectrales. To find the peaks, simply search for every point where the slope changes from positive to negative. Para encontrar los picos, sólo se tiene que buscar todos los puntos donde la pendiente cambia de positivo a negativo. Each of these peaks corresponds to individual frequencies in the signal. Cada uno de estos picos corresponde a las frecuencias individuales de la señal. The next step is to calculate the approximate SPL of each peak. One way to do this is to normalize such that a full-amplitude tone at 1kHz is equal to 96dB. El siguiente paso es calcular el SPL (SoundPressureLevel) aproximado de cada pico. Una forma de hacerlo es normalizar de manera que un tono lleno de amplitud a 1 KHz es igual a 96 dB. The SPL of each peak can then be easily calculated according to the following formula: El SPL de cada pico se puede calcular fácilmente con la siguiente fórmula 5.2:

¹⁴FFT es la abreviatura de transformada rápida de Fourier, que es algoritmo que permite calcular la transformada de Fourier discreta (DFT) y su inversa. La FFT es de gran importancia en el tratamiento digital de señales y filtrado digital.

$$SPL = 96 + 20 \log \left(\frac{|peak|}{|FFT(1kHz)|} \right) \quad (5.2)$$

From these SPL values, a masking curve can be created for each peak. A partir de estos valores SPL, una curva de enmascaramiento se puede crear para cada pico. There are several masking functions that do this. This coder uses the function suggested by Schroeder: Hay varias funciones de enmascaramiento que hacen esto. Este codificador utiliza la función sugerida por Schroeder:

$$10 \log F(dz) = 15.81 + 7.5(dz + 0.474) - 17.5 \sqrt{1 + (dz + 0.474)^2} \quad (5.3)$$

$$dz = 13 \tan^{-1}(0.76f) + 3.5 \tan^{-1} \left(\left(\frac{f}{7.5} \right)^2 \right) \quad (5.4)$$

where dz is the bark frequency, defined by the following

The next step is to combine all the masking curves with the Threshold in Quiet

El siguiente paso es combinar todas las curvas de enmascaramiento con el umbral silencioso (tiq). The TIQ is the minimum SPL that a person can hear at a given frequency, and is typically defined by the following equation: El tiq es el SPL mínimo que una persona puede oír a una frecuencia dada, y se define por la siguiente ecuación:

$$\frac{TIQ(f)}{dB} = 3.64 f^{-0.8} - 6.5 e^{-0.6(f-3.3)^2} + 10^{-3} f^4 \quad (5.6)$$

where f is the frequency in kHz. donde f es la frecuencia en KHz.

There are a variety of ways to combine the masking curves and the TIQ which correspond to different values of alpha in the equation. Hay una variedad de maneras de combinar las curvas de enmascaramiento y tiq los que corresponden a diferentes valores de alfa en la ecuación

$$I_N = \left(\sum_{n=0}^{N-1} I_n^a \right)^{\frac{1}{a}} \quad 1 \leq a \leq \infty \quad (5.7)$$

In this coder, alpha is zero, which corresponds to using the highest masking curve (or the threshold in quiet). The signal-to-mask ratio (SMR) can easily be calculated by dividing the SPL of the signal by the masking threshold. En este codificador, alfa es cero, lo que corresponde al uso de la curva de enmascaramiento. La relación señal a

máscara, se puede calcular fácilmente dividiendo el SPL de la señal para el umbral de enmascaramiento.

5.1.3 Asignación de bits.

Con los datos obtenidos en la SMR puede determinarse que bandas de frecuencia deben recibir la mayoría de los bits. **As a general rule, each bit increases signal-to-noise ratio by about 6dB.** La asignación dinámica de bits sería garantizar que el ruido de cuantización está por debajo del umbral de enmascaramiento, y por lo tanto inaudible. However, there may not be enough bits available to do this, bits must be allocated to where they are needed most. Sin embargo, puede que no haya suficientes bits disponibles para hacer esto. Los bits deben ser asignados en donde más se necesitan. The "water-filling" bit allocation algorithm is used to allocate bits by looking for the maximum value of the SMR, allocating a bit to that subband, subtracting 6dB from the SMR at that frequency, and repeating as long as bits are available to allocate. El "water-filling" que es el algoritmo que se utiliza para asignar los bits buscando el valor máximo de la SMR.

5.1.4 Cuantización y Codificación.

After determining where bits should be allocated, the next step is to quantize the audio signal to the appropriate number of bits. Después de determinar dónde se deberían asignar los bits, el siguiente paso es cuantificar la señal de audio hasta el número adecuado de bits.

La MDCT es una transformada lineal ortogonal, basada en la idea de la cancelación del aliasing del dominio de tiempo, la MDCT se muestrea críticamente ya que, aunque el 50% de datos es solapado, los datos de la secuencia después de la MDCT tienen el mismo número de coeficientes que muestras antes de la transformada. Esto significa que un solo bloque de datos de la IMDCT no corresponden al bloque original en la cual se aplicó la MDCT; a continuación algunas fórmulas aplicadas para aplicar la MDCT en el software desarrollado:

$$\text{function } y = \text{mdct}(x) \quad (5.8)$$

$$y = \text{real}(X(1:N/2) .* \exp(-1i*2*\pi*n0*([0:N/2-1]'+0.5)/N)) \quad (5.9)$$

5.1.5 Almacenamiento.

Una vez que los coeficientes MDCT están cuantificados, se pueden guardar en un archivo el cual tendrá la extensión .sec (Sonidos Estetoscópicos Comprimidos), el archivo estará conformado por todas las características de este codificador establecidas anteriormente. In addition to the MDCT coefficients, the gain factor must also be specified as well as the number of bits allocated to each band. Además de los coeficientes MDCT, el factor de ganancia también se debe especificar, así como el número de bits asignados a cada banda. In this coder, a file header is also included which contains information such as the sampling frequency, frame length, bit rate, number of bits used for writing the gain factor, and the number of frames in the file. También se incluye un archivo de cabecera que contiene información tal como la frecuencia de muestreo, longitud de imagen, velocidad de bits, número de bits utilizados en la escritura del factor de ganancia, y número de tramas en el archivo.

En la Fig.5.1 se puede observar las diferentes líneas de códigos en el software aplicadas para el almacenamiento del archivo codificado. Because only a few bits are to be used to represent the gain factor, the logarithm of the gain is written to the file.

```

coded_filename = sprintf('normal.sip');
fid = fopen(coded_filename,'w');
fwrite(fid, Fs, 'ubit16'); % Frecuencia de Muestreo
fwrite(fid, N, 'ubit12'); % Longitud del Frame
fwrite(fid, bitrate, 'ubit8'); % Bit Rate
fwrite(fid, scalebits, 'ubit4'); % Número de bits de escala
fwrite(fid, length(FRAMES(:,1)), 'ubit26'); % Número de frames

```

Fig. 5.1 Parte de código utilizado **Fig1. Información almacenada en el archivo**

Información almacenada en el archivo

Fig1. Información almacenada en el archivo

5.1.6 Diagrama de Bloques.

En esta se observa de forma resumida en diagramas de bloques, el proceso de la codificación y decodificación aplicada para el software de compresión desarrollado para archivos de audio.

5.1.6.1 Codificación.

En la Fig.5.2 se observa de forma general cada paso del proceso de codificación tomando como paso principal la utilización de la MDCT, la misma que ayuda a comprimir los archivos de audio, ya que se descompone en diferentes coeficientes los mismos que son almacenados en la parte final.

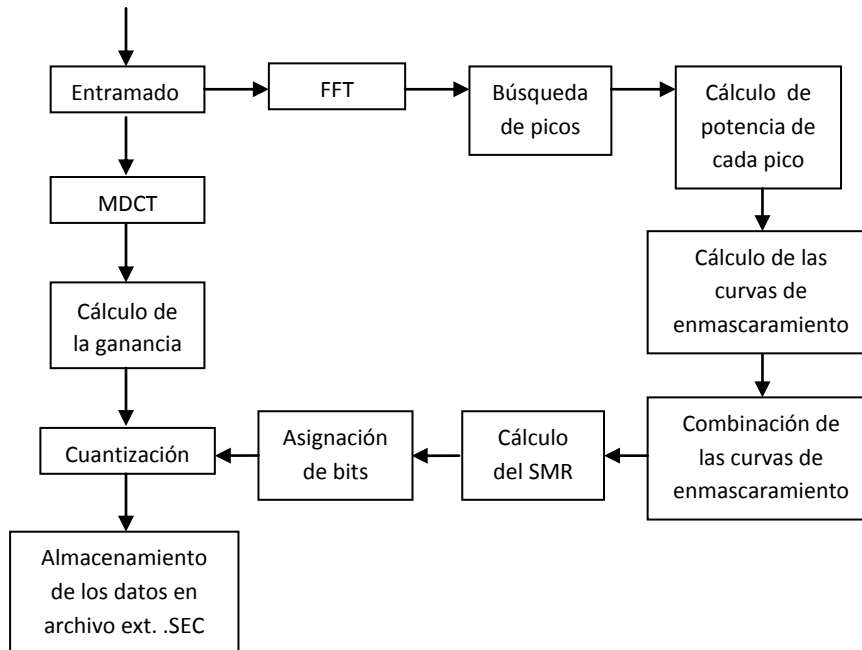


Fig.5.2 Proceso de codificación del archivo de audio

5.1.6.2 Decodificación

En la Fig.5.3 observamos el proceso inverso de la figura anterior es decir de forma general cada paso del proceso de decodificación y en este caso aplicando la IMDCT, en este caso transformamos los coeficientes del archivo comprimido al archivo de audio original.

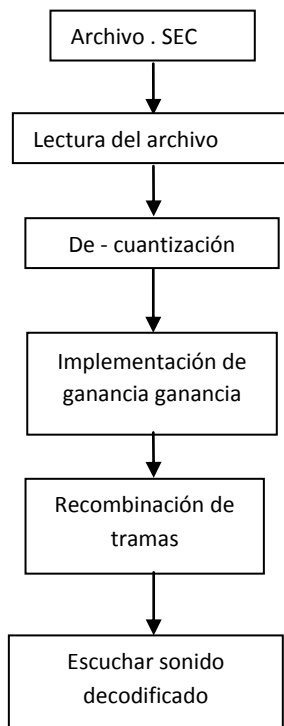


Fig.5.3 Proceso de decodificación del archivo de audio Fig3. Proceso de decodificación del archivo de audio

5.2 Implementación del codificador G.722.

La implementación del códec G.722 consistió en la búsqueda de un software en el cual se pueda configurar sus características principales. El estándar G.722 solo es utilizado para comunicaciones VoIP en tiempo real, por este motivo, no existe en la actualidad ningún programa de tratamiento de sonido que contenga este tipo de codificación. Además la información existente lo describe de forma muy generalizada, la misma que es insuficiente para desarrollar un tipo de codificador con este estándar en cualquier programador.

5.2.1 Codificador ADPCM.

Para conseguir un compresor que se asemeje a las características del G.722, principalmente se parte de su forma de codificación ADPCM, por esta razón, luego de analizar una variada y numerosa cantidad de programas, de los cuales se preseleccionaron cinco finales que se los puede observar en el anexo A6, finalmente se escogió el programa Audio MusicBatchConverter [37] (Fig.5.4), el mismo que permite implementar las principales características técnicas del G.722, las cuales son:

Frecuencia de Muestreo: 8000Hz
 Tipo de sonido: Mono
 Codificación: ADPCM
 Bitrate: 32-64 Kbps

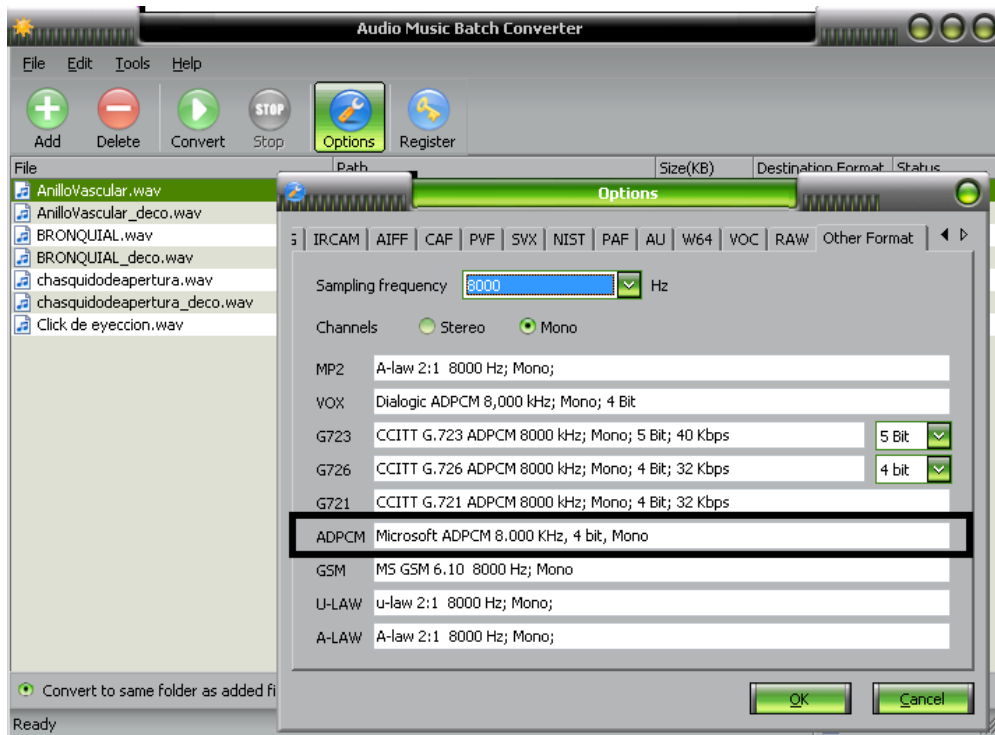


Fig.5.4. Configuración del programa

Una de las principales características del programa es que al ejecutarlo, todas las características que han sido modificadas en el mismo se mantienen, por lo cual, sólo se necesitará escoger el sonido a codificar y seleccionar la función "Convert". Automáticamente se creará el sonido comprimido en un destino previamente establecido.

5.2.2 Inclusión al Programa Principal.

Luego de haber realizado la configuración necesaria en el codificador ADPCM, procedemos a realizar su implementación en el software Matlab, logrando así anexarlo al programa principal donde se encuentra desarrollado también el

codificador MDCT. Esto lo se realiza mediante la ayuda del Guide-Matlab, en el cual, mediante la interfaz gráfica (Fig.5.5) llamamos de forma directa al programa.

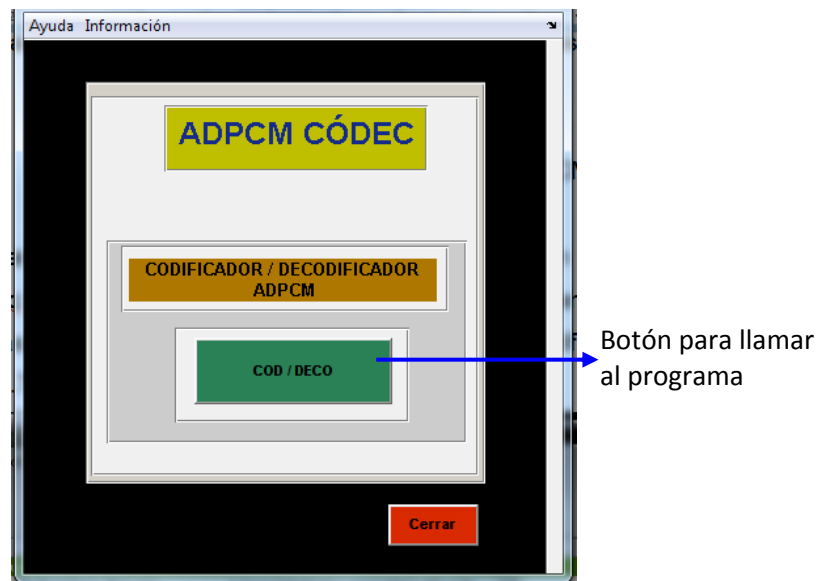


Fig.5.5. Llamado del programa desde Matlab

Después de realizar la codificación de los archivos deseados, como resultado obtendremos archivos comprimidos de extensión .WAV, que contienen todas las características pre-configuradas en el programa con una codificación ADPCM.

CAPÍTULO 6

Resultados.

En el presente capítulo describimos de manera detallada el funcionamiento del software realizado mediante el Guide de Matlab y también las diferentes comparaciones de los sonidos estetoscópicos comprimidos por los dos codificadores antes descritos.

6.1 Procedimiento general del software utilizado.

El codificador se realizó utilizando la herramienta Guide de Matlab para crear una interfaz gráfica que sea amigable y de fácil uso para el usuario. Con este programa se incorporó todas las herramientas necesarias para realizar los dos tipos de codificación: MDCT y ADPCM.

A continuación se describe su funcionamiento e interacción con el usuario:

Pantalla Principal: Esta es la interfaz gráfica inicial del programa. En esta parte se selecciona el tipo de codificador que se desea utilizar, tanto ADPCM como MDCT, como se muestra en la Fig.6.1.

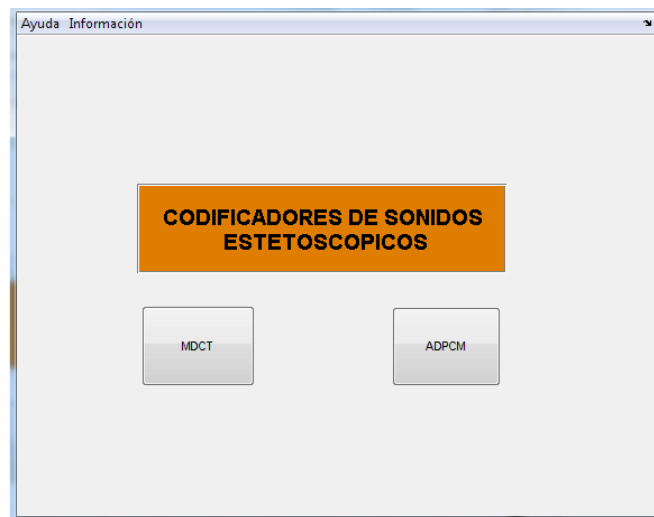


Fig. 6.1 Pantalla principal del software

Codificación / Decodificación ADPCM: Al seleccionar esta opción se desplegará una nueva ventana en la cual se selecciona el archivo que se desee procesar, como se observa en la Fig.6.2.

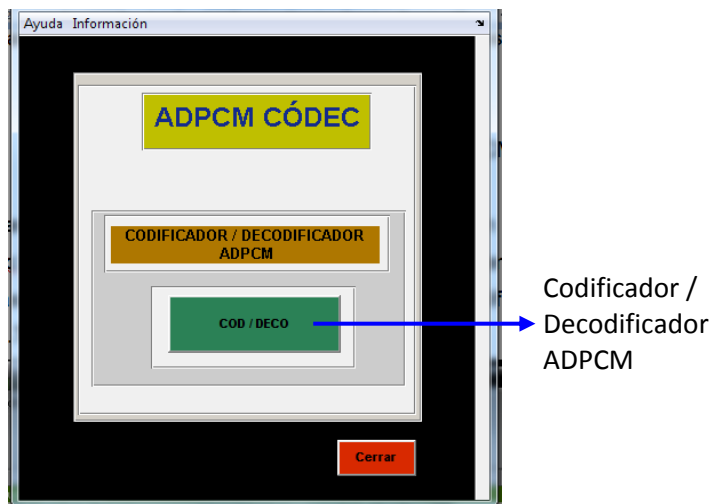


Fig. 6.2 Pantallacódec ADPCM

Para realizar los procesos de codificación y decodificación se escogió el programa Audio MusicBatchConverter (Fig.6.3) que a diferencia de otros programas nos da la posibilidad de variar las características de la codificación ADPCM modificándolas de acuerdo al estándar G.722.

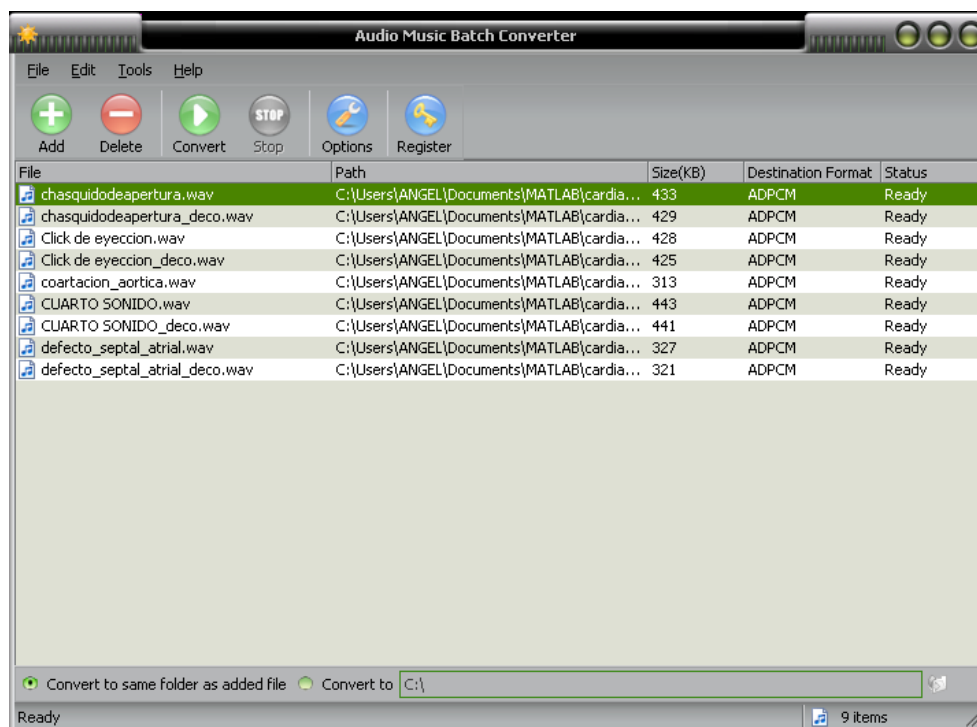


Fig.6.3 Pantalla del programa Audio MusicBatchConverter

Codificación MDCT: Al seleccionar el codificador MDCT de la pantalla principal, se despliega la siguiente ventana (Fig.6.4).



Fig. 6.4 Pantalla códec MDCT

Al seleccionar “codificador” se despliega la ventana de “Codificar”, Cuando el proceso de codificación este terminado aparece el mensaje de “Codificación_terminada” en recuadro rojo (figura 6.5).



Fig.6.5 Parte de codificación MDCT

Decodificación MDCT: Al seleccionar “decodificador” se despliega la ventana de “decodificar”, Al final del proceso aparece el mensaje de “Decodificación_terminada”, y además podemos observar los tamaños original, codificado y decodificado de los archivos (Fig.6.6).

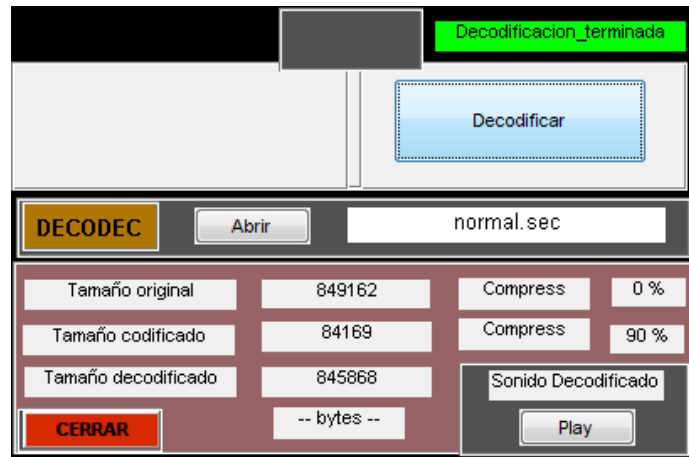


Fig.6.6 Parte de decodificación MDCT

6.2 Comparación de sonidos estetoscópicos utilizando los dos codificadores.

En esta sección se realizará una comparación de los archivos utilizados (originales y decodificados), tanto en el dominio del tiempo y frecuencia.

6.2.1 Comparación en Frecuencia.

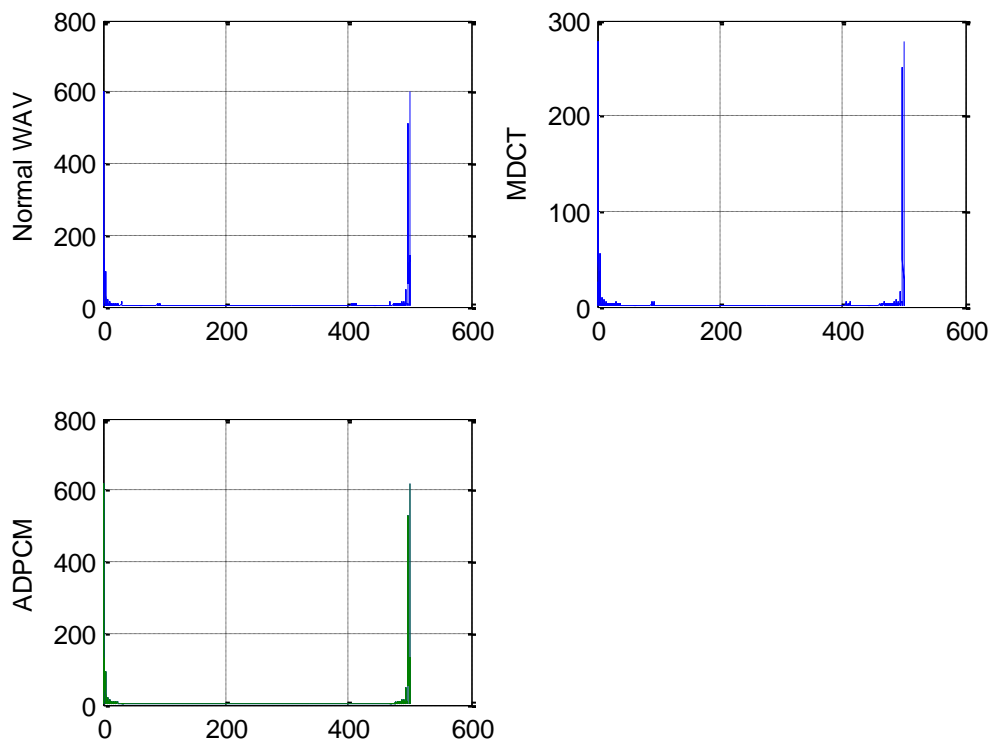


Fig.6.7 Espectros de la señal cardiaca anormal estenosismitral

Como se observa en la Fig. 6.7 las gráficas del espectro tanto de la señal original, la señal MDCT y la ADPCM son muy similares y no sufren ningún cambio significativo que dificultara el reconocimiento de la señal de audio. Por esta razón, podemos indicar que los dos codificadores están funcionando de una manera correcta.

6.2.2 Comparación Temporal.

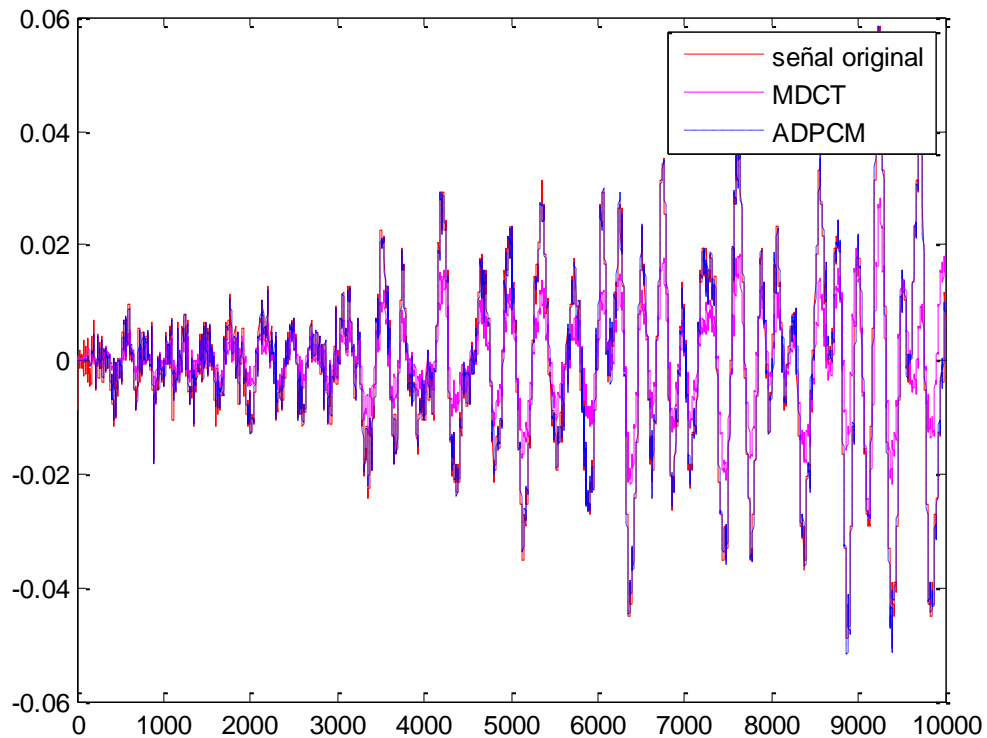


Fig.6.8 Comparación de las señales

En esta comparación temporal de las señales de la Fig. 6.8 no se presentan retrasos en la señal y la reconstrucción es muy precisa, podemos notar que en algunos segmentos no se distinguen las señales esto es debido al solapamiento de las mismas.

6.2.3 Comparación en Fase.

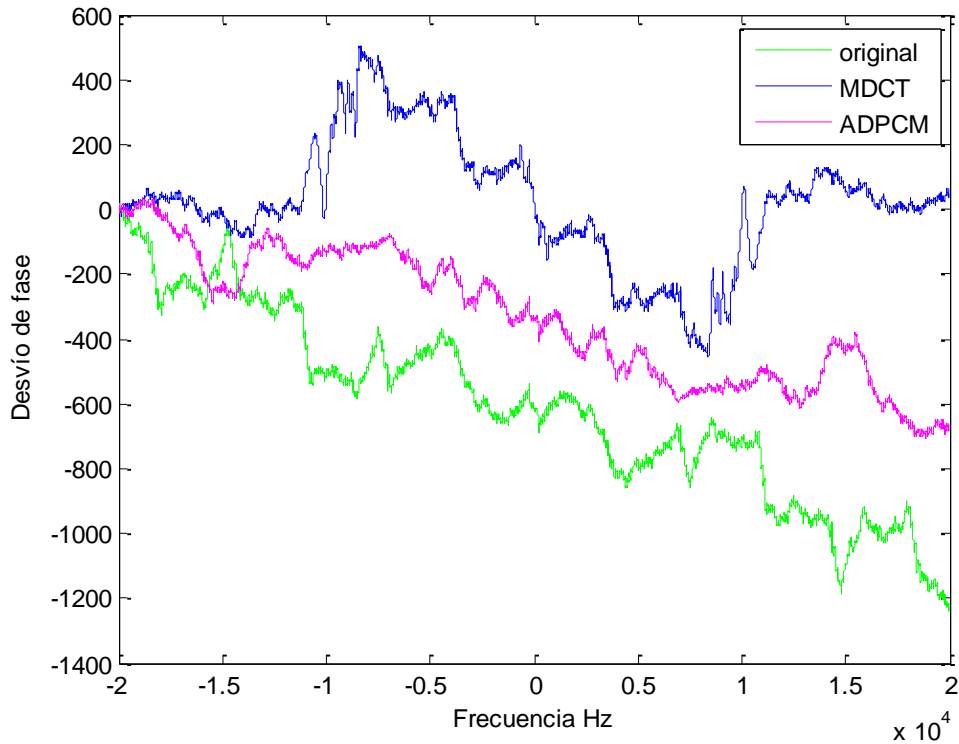


Fig.6.9 Comparación en fase de las señales

Una de las principales características de las señales es su respuesta en fase (Fig. 6.9) para cada uno de los formatos. Como se observa tanto la señal original como la señal ADPCM tiene una respuesta muy similar sin embargo la de la señal MDCT se encuentra dentro de un rango aceptable. Gracias a esto las señales pueden ser utilizadas en sistemas de transmisión en tiempo real, los cuales utilizan tasas de bits menores a 46 Kbps. Considerando la señal ADPCM como la más óptima para sistemas de transmisión en tiempo real.

Conclusiones.

1. Luego de analizar los códecs Ogg-Vorbis y G.722, se definió que la compresión y reconstrucción de la señal se la realiza de manera óptima mediante la MDCT, ya que se obtienen sonidos muy semejantes a los originales, con una ínfima pérdida de información.
2. En base a las pruebas graficas que realizaron señales, se concluye que la señales comprimidas con la MDCT, no han tenido alteraciones tanto en el dominio de la frecuencia , como en el dominio del tiempo, además no presentan variación en su fase.
3. La relación de compresión de acuerdo a los detalles técnicos del códec Ogg-Vorbis es de 75% máximo, sin embargo, al aplicarlo en este proyecto, y más precisamente en el software desarrollado en Matlab, se alcanzó una relación de compresión promedio de 82%.
4. Luego de realizar la compresión con los dos codificadores, se aprecia que el códec ADPCM tiene un mejor porcentaje de compresión 95%, que el códec MDCT, pero concluimos que el códec MDCT es el mejor para el tratamiento de estos sonidos debido a que al realizar la decodificación es el que mejor mantiene la calidad de los sonidos.
5. La diferencia que existe entre los sonidos estetoscópicos originales y codificados es mínima en base a las comparaciones gráficas realizadas en el dominio del tiempo y frecuencia, lo que indica que los codificadores están funcionando óptimamente.

Recomendaciones.

1. Debido a que se trabaja con sonidos estetoscópicos que deben mantener una calidad óptima, para poder identificar si se trata de un sonido normal o anormal, se debe tener cuidado en el momento de realizar el tratamiento de estas señales para que no se alteren las mismas, ya que esto impediría la correcta identificación de los sonidos.
2. Es necesario conocer los estándares internacionales referentes a las telecomunicaciones, en este caso normas y requisitos de compresión de sonido, esto para facilitar la compatibilidad con otros programas, donde puede ser implementados como una herramienta.
3. Tener en cuenta si la información y las licencias de aplicación de un códec son de libre uso, debido a que esto sería un gran obstáculo al momento de querer implementarlo y no tener la respectiva certificación.
4. Al tener archivos comprimidos en MP3, y querer transformarlos al formato Ogg-Vorbis o viceversa; se recomienda, primeramente transformarlos al formato WAV que es uno formato sin pérdidas ya que estos dos códecs son compresores con pérdidas, y de ahí pasarlo al formato deseado; si no se considera esta recomendación las pérdidas de información se multiplicaran, reduciendo la calidad del sonido.

Bibliografía.

- [1] Estetoscopios, Idóneos.com, [base de datos en línea], Medicina, <http://oficinas.idoneos.com/index.php/Medicina/Estetoscopios>, [2010, 16 de septiembre].
- [2] Telemedicina: Tecnología y Modelos de Organización, [base de datos en línea], Introducción a la Telemedicina, <http://www.itelemedicina.com/index.asp?p=intro/intro.asp>, [2010, 16 de septiembre].
- [3] Qué es la Telemedicina, [base de datos en línea], www.imaginar.org/index_archivos/telemed/telemedicina.pdf, [2010, 17 de septiembre].
- [4] Tele-Educación: Campaña de Salud Preventiva en Yacuambi, 2007, <http://blogs.utpl.edu.ec/blogtutupaly/>, [2010, 18 de septiembre].
- [5] Telemedicina en el Ecuador, 2009, <http://www.galenosdelastillero.com/2009/07/telemedicina-en-el-ecuador.html>, [2010, 15 de septiembre].
- [6] Telemedicina: Un intento de aproximación desde la Gestión Sanitaria, Revista Española de Administración Sanitaria, Vol. V, Nro. 19, 2001, <http://www.itelemedicina.com/EPalauRAS.pdf>, [2010, 19 de septiembre].
- [7] La Telemedicina, su estructura, objetivo y ventajas, [base de datos en línea], <http://www.ilustrados.com/publicaciones/EEEIkypulpeUbxCoz.php>, Ilustrados.com [2010, 13 de septiembre].
- [8] Infrasonido, 2010, <http://es.wikipedia.org/wiki/Infrasonido>, [2010, 19 de septiembre].
- [9] Ultrasonido, 2010, <http://es.wikipedia.org/wiki/Ultrasonido>, [2010, 23 de septiembre].
- [10] Mono, monofónico o monaural, 2009, <http://www.duiops.net/hifi/enciclopedia/mono.htm>, [2010, 17 de septiembre].
- [11] Sistema para la Afinación Automática de Sonidos Polifónicos (II), 2010, <http://babyroller.blogspot.com/2010/03/sistema-para-la-afinacion-automatica-de.html>, [2010, 17 de septiembre].
- [12] ¿Cómo funciona el sonido estereofónico?, Gonzalez F., http://tecnologia.idoneos.com/index.php/Sonido_estereofonico, [2010, 18 de septiembre].
- [13] Sonido estereofónico, 2010, http://es.wikipedia.org/wiki/Sonido_estereo_f%C3%B3nico, [2010, 18 de septiembre].
- [14] SUROS, J. y SUROS, A. Semiología médica y técnicas de exploración. Octava Edición, 2001, 1156 p.
- [15] ARGENTE, H. y ALVAREZ, M Semiología Medica Fisiopatología, semiotecnia, y propedéutica, Buenos Aires, 2007, 1603 p.

- [16] Fonocardiografía (FCG), 2005, <http://www.dalcame.com/fono.htm>, [2010, 22 de septiembre].
- [17] Auscultación, 2004, http://www.saludalia.com/docs/Salud/web_saludalia/pruebas_diagnosticas/doc/doc_auscultacion_respiratorio.htm, [2010, 20 de septiembre].
- [18] Compresión de Audio Digital, 2009, <http://www.ocw.espol.edu.ec/facultad-de-ingenieria-en-electricidad-y-computacion/procesamiento-de-audio-y-video-1/4.2-compresion-de-audio-digital>, [2010, 21 de septiembre].
- [19] Tratamiento digital del sonido, 2010, http://www.zator.com/Hardware/H10_2.htm, [2010, 21 de septiembre].
- [20] Compresor (sonido), 2009, http://es.wikipedia.org/wiki/Compresor_%28sonido%29, [2010, 24 de septiembre].
- [21] ¿Qué es Streaming?, http://translate.flossmanuals.net/MuSE_es/StreamingIntro, [2010, 28 de septiembre].
- [22] MP3, 2010, http://es.wikipedia.org/wiki/MP3#El_modelo_psicoac.C3.BAstico, [2010, 18 de septiembre].
- [23] Psicoacústica, <http://members.fortunecity.com/alex1944/psicoacustica/psycoasp.html>, [2010, 14 de septiembre].
última visita 14 de Septiembre 2010.
- [24] Bandas críticas y tasa de bandas críticas, http://www.labc.usb.ve/EC4514/AUDIO/PSICOACUSTICA/BANDAS_CRITICAS.html, [2010, 17 de septiembre].
- [25] Normas de compresión para audio digital, http://www.paradorfederal.com.ar/.../Normas_de_compresión_para_audio_digital.pdf, [2010, 19 de septiembre].
- [26] Codificación Perceptual, http://members.fortunecity.es/obiwankenobi2000/docs/ee424/cod_porcen.htm, [2010, 21 de septiembre].
- [27] Bitstream, <http://www.dspace.epn.edu.ec/bitstream/123456789/744/15/T10505CAP2.pdf>, [2010, 26 de septiembre].
- [28] Códec de Audio, 2010, http://es.wikipedia.org/wiki/C%C3%B3dec_de_audio, [2010, 28 de septiembre].
- [29] Codificación, http://www.physionet.cps.unizar.es/~eduardo/.../Tema_6-codificacion.pdf, [2010, 29 de septiembre].
- [30] Xiph.org Foundation, Xiph open source community, 2009, <http://xiph.org/vorbis/doc/>, [2010, 30 de septiembre].
- [31] Contenedores Multimedia, 2008, <http://www.herramientaslibres.es/artlos-mainmenu-27/22-herramientas-libres/52-formatos-multimedia-libres.html>, [2010, 18 de septiembre].

- [32] OggVorbis, <http://hernaaan.laconsolablog.com/wiki/vorbis.html>, [2010, 24 de septiembre].
- [33] ADPCM, 2010, http://es.wikipedia.org/wiki/ADPCM_7_kHz_Audio-Coding_Within_64_kbit/s. ITU-T Recommendation G.722, 1988, [2010, 30 de septiembre].
- [34] El códec G.722 y cómo ofrecer audio de calidad, Rojano E., 2009, <http://www.sinologic.net/blog/2009-11/codec-g722-ofrecer-audio-calidad> [2010, 01 de octubre]
- [35] G.722, <http://www.worldlingo.com/ma/enwiki/es/G.722>, [2010, 01 de octubre]
- [36] Polycom® Siren™/G 722.1 FAQs, http://www.polycom.com/company/about_us/technology/siren_g7221/faq.html, [2010, 3 de octubre].
- [37] Descarga del programa Audio MusicBatchConverter, http://download.cnet.com/Audio-Music-Batch-Converter/3000-2141_4-10785589.html, [2010, 30 de septiembre].
- [38] La Salud Virtual. Vistazo [en línea], 2010 <http://www.vistazo.com/ea/pais/imprimir.php?Vistazo.com&id=2940>, [2010, 27 de noviembre].
- [39] Física, Acústica, Velocidad del sonido, [base de datos en línea], Wikilibros, http://es.wikibooks.org/wiki/F%C3%ADsica/Ac%C3%BAstica/Velocidad_del_sonido, [2010, 19 de septiembre].
- [40] Grabación digital, [base de datos en línea], Grabación de audio, http://www.lpi.tel.uva.es/~nacho/docencia/ing_ond_1/trabajos_04_05/io8/public_html/digital.htm, [2010, 12 de octubre].

Glosario de términos.

Termino o Abreviaturas	Definición
Carótida	Cada una de las dos arterias, que por uno y otro lado del cuello llevan la sangre a la cabeza.
Códec	Codificador/decodificador
ECG	Electrocardiograma
ELF	Frecuencias extremadamente bajas
Estenosis	Estrechamiento de los conductos de las arterias
Eyectivo	Impulsar con fuerza
Miocardio	Parte muscular del corazón de los vertebrados, situada entre el pericardio y el endocardio.
Mitral	válvula
Pericardio	Envoltura del corazón, que está formada por dos membranas, una externa y fibrosa, y otra interna y serosa.
Prolapso	Caída o descenso de una víscera, o del todo o parte de un órgano.
Protodiástole	Primera fase de la diástole cardiaca
Sistólico	Movimiento de contracción del corazón y de las arterias para empujar la sangre que contienen
Streaming	Tecnología que permite reproducir contenidos en tiempo real en internet.
Tricúspide	válvula
Ultrasonido (medicina)	Técnica diagnóstica en la que un sonido de frecuencia muy alta es dirigido hacia el organismo; se conoce como ecografía.
Ventricular	Cavidad del corazón que recibe sangre de una aurícula y la impulsa por el sistema arterial

ANEXOS.**A1. CARACTERÍSTICAS DE COMPRESIÓN MDCT**

A.1.1. Características de compresión de los sonidos cardiacos utilizando la MDCT.

Nombre del archivo .wav	Tamaño original (kBytes)	Tamaño archivo codificado (kB)	Tamaño archivo decodificado (kB)	Porcentaje de compresión
Chasquido de apertura	431	77	428	82%
Normal	829	82	826	90%
Click de eyeccion	427	77	424	82%
Defecto septal atrial	325	57	320	82%
Defecto septal ventricular	135	22	130	83%
Ducto arterioso persistente	329	59	324	82%
Estenosis mitral	332	59	328	82%
Hearth1	386	70	382	82%
Hearth2	395	69	392	82%
Hearth3	429	75	426	83%
Insuficiencia aortica	171	30	166	82%
Mediawebserver	442	80	440	82%
Murmullo pansistolico	455	82	452	82%
Murmullo sistólico precoz	438	80	436	82%
Murmullo sistólico tardio	445	81	442	82%
Normal2	332	60	328	82%
Primer Sonido	386	70	382	82%
Regurgitacion mitral	155	28	152	82%
Regurgitacion tricuspidea	366	66	364	82%
Segundo Sonido	395	69	392	82%
Soplo diastólico	396	71	394	82%
Tercer Sonido	429	75	426	83%
Cuarto Sonido	442	80	440	82%

A.1.2. Características de compresión de los sonidos pulmonares utilizando la MDCT.

Nombre del archivo .wav	Tamaño original (kBytes)	Tamaño archivo codificado (kB)	Tamaño archivo decodificado (kB)	Porcentaje de compresión
Bronquial	872	158	866	82%
Crackle1	872	157	866	82%
Estenosis pulmonar	298	54	294	82%
Normalr	872	157	866	82%
Traqueal	863	155	858	82%
Weezbaby	854	154	849	82%
Wheeze1	872	158	866	82%
CrepitoNino	974	162	972	83%
EstridorNino	874	153	872	82%
AnilloVascular	949	172	946	82%

A2. CARACTERÍSTICAS DE COMPRESIÓN ADPCM

A.2.1. Características de compresión de los sonidos cardiacos utilizando la ADPCM.

Nombre del archivo .wav	Tamaño original (kBytes)	Tamaño archivo codificado (kB)	Tamaño archivo decodificado (kB)	Porcentaje de compresión
Chasquido de apertura	431	21	431	95%
Normal	829	20	830	97%
Click de eyeccion	427	21	426	95%
Defecto septal atrial	325	16	324	95%
Defecto septal ventricular	135	7	130	94%
Ducto arterioso persistente	329	16	324	95%
Estenosis mitral	332	16	329	95%
Hearth1	386	19	383	95%
Hearth2	395	19	394	95%
Hearth3	429	21	426	95%
Insuficiencia aortica	171	8	167	95%
Mediawebserver	442	21	442	95%
Murmullo pansistolico	455	22	453	95%
Murmullo sistólico precoz	438	21	437	95%
Murmullo sistólico tardio	445	21	442	95%
Normal2	332	16	329	95%
Primer Sonido	386	19	383	95%
Regurgitacion mitral	155	8	151	95%
Regurgitaciontricuspidea	366	18	367	95%
Segundo Sonido	395	19	394	95%
Soplo diastólico	396	19	394	95%
Tercer Sonido	429	21	426	95%
Cuarto Sonido	442	21	442	95%

- Fórmula utilizada para calcular el porcentaje de compresión:

$$\% \text{ de compresión} = \frac{100\% * \text{nuevo tamaño (kbytes)}}{\text{tamaño original (kbytes)}}$$

A.2.2 Características de compresión de los sonidos pulmonares utilizando la ADPCM.

Nombre del archivo .wav	Tamaño original (kBytes)	Tamaño archivo codificado (kB)	Tamaño archivo decodificado (kB)	Porcentaje de compresión
Bronquial	872	41	867	95%
Crackle1	872	41	867	95%
Estenosis pulmonar	298	14	297	95%
Normalr	872	41	867	95%
Traqueal	863	41	856	95%
Weezbaby	854	40	851	95%
Wheeze1	872	41	867	95%
CrepitoNino	974	46	970	95%
EstridorNino	874	41	873	95%
AnilloVascular	949	45	948	95%

A3. CÓDIGO FUENTE

Software Matlab 2008b

Código Principal

```
% CODIFICACIÓN
scalebits = 4; % escala de bits
bitrate = 128000;
N = 2048; % framelength
J = get(handles.text2,'string');
[y,fs]= wavread(J); %#ok<NASGU>
original_filename = sprintf(J);
coded_filename = sprintf('normal.sip');
[Y,Fs,NBITS] = wavread(original_filename); %#ok<NASGU>
tone = Y;
num_subbands = floor(fftbank(N/2,N/2,Fs))+1;
bits_per_frame = floor(((bitrate/Fs)*(N/2)) - (scalebits*num_subbands));
sig=sin(2*pi*1000*[1/Fs:1/Fs:(N/2)/Fs]);
win=(0.5 - 0.5*cos((2*pi*([1:(N/2)]-0.5))/(N/2)));
fftmax = max(abs(fft(sig.*win))); % definido para 96dB

% Entramado de Audio
FRAMES = enframe(tone,N,N/2);

% Datos principales del archivo
fid = fopen(coded_filename,'w');
fwrite(fid, Fs, 'ubit16'); % Frecuencia de Muestreo
fwrite(fid, N, 'ubit12'); % Longitud del Frame
fwrite(fid, bitrate, 'ubit18'); % Bit Rate
fwrite(fid, scalebits, 'ubit4'); % Número de bits de escala
fwrite(fid, length(FRAMES(:,1)), 'ubit26'); % Número de frames

% Codificación
forframe_count=1:length(FRAMES(:,1))
if mod(frame_count,10) == 0
outstring = sprintf('Codificando Frame %i de %i', frame_count, length(FRAMES(:,1)));
```

```

disp(outstring);
end
fft_frame = fft(FRAMES(frame_count,:));
ifft_frame == zeros(1,N)
Gain = zeros(1,floor(fftbank(N/2,N/2,Fs))+1);
bit_alloc = zeros(1,floor(fftbank(N/2,N/2,Fs))+1);
else
len = length(fft_frame);
peak_width = zeros(1,len);
peak_points = cell(len,len); %#ok<NASGU>
peak_min_value = zeros(1,len); %#ok<NASGU>

% Se encuentra los picos
centers = find(diff(sign(diff( abs(fft_frame).^2 )) == -2) + 1);
spectral_density = zeros(1,length(centers));
for k=1:length(centers)
peak_max(k) = centers(k) +2; %#ok<AGROW>
peak_min(k) = centers(k) - 2; %#ok<AGROW>
peak_width(k) = peak_max(k) - peak_min(k);
for j=peak_min(k):peak_max(k)
if (j > 0) && (j < N)
spectral_density(k) = spectral_density(k) + abs(fft_frame(j))^2;
end
end
end

% Cálculo de SPL
modified_SD = spectral_density / ((N^2)/8);
SPL = 96 + 10*log10(modified_SD); %#ok<NASGU>

% Se aplica la fft a los valores del SPL
fft_spl = 96 + 20*log10(abs(fft_frame)/fftmax);

% Se combinan todas las curvas de enmascaramiento
f_kHz = [1:Fs/N:Fs/2];
f_kHz = f_kHz/1000;
A = 3.64*(f_kHz).^(-0.8) - 6.5*exp(-0.6*(f_kHz - 3.3).^2) + (10^(-3))*(f_kHz).^4;

```

```

% Enmascaramiento de Espectro
big_mask = max(A,Schroeder(centers(1)*(Fs/2)/N,fft_spl(centers(1)),...
14.5+bark(centers(1)*(Fs/2)/N)));
for peak_count=2:length(centers)
try
big_mask
=max(big_mask,Schroeder(centers(peak_count)*(Fs/2)/N,fft_spl((peak_count)),...
14.5+bark(centers(peak_count)*(Fs/2)/N)));
catch %#ok<CTCH>
peak_count=peak_count; %#ok<FXSET,ASGSL>
end
end
New_FFT = fft_spl(1:N/2)-big_mask;
New_FFT_indices = find(New_FFT > 0);
New_FFT2 = zeros(1,N/2);
for i=1:length(New_FFT_indices)
New_FFT2(New_FFT_indices(i)) = New_FFT(New_FFT_indices(i));
end
bit_alloc = allocate(New_FFT2,bits_per_frame,N,Fs);
[Gain,Data] = p_encode(mdct(FRAMES(frame_count,:)),Fs,N,bit_alloc,scalebits);
end

% Se guarda los datos de audio en un archivo
qbits = sprintf('ubit%i', scalebits);
fwrite(fid, Gain, qbits);
fwrite(fid, bit_alloc, 'ubit4');
for i=1:25
indices = find((floor(fft_bark([1:N/2],N/2,Fs))+1)==i);
qbits = sprintf('ubit%i', bit_alloc(i));
if ((bit_alloc(i) ~= 0) && (bit_alloc(i) ~= 1))
fwrite(fid, Data(indices(1):indices(end)) ,qbits);
end
end
end
fclose(fid);

```

```

% DECODIFICACIÓN
% Datos extraídos del archivo
coded_filename = sprintf('normal.sip');
decoded_filename = sprintf('normal_deco.wav');
fid = fopen(coded_filename,'r');
Fs      = fread(fid,1,'ubit16'); % Frecuencia de Muestreo
framelength = fread(fid,1,'ubit12'); % Longitud de la trama
bitrate    = fread(fid,1,'ubit18'); %#ok<NASGU> % Bit Rate
scalebits  = fread(fid,1,'ubit4' ); % Número de bits de escala
num_frames = fread(fid,1,'ubit26'); % Número de tramas
forframe_count=1:num_frames
qbits = sprintf('ubit%i', scalebits);
gain = fread(fid,25,qbits);
bit_alloc = fread(fid,25,'ubit4');
for i=1:floor(fftbank(framelength/2,framelength/2,Fs))+1
indices = find((floor(fftbank([1:framelength/2],framelength/2,Fs))+1)==i);
if ((bit_alloc(i) ~= 0) & (bit_alloc(i) ~= 1)) %#ok<AND2>
qbits = sprintf('ubit%i', bit_alloc(i));
InputValues(indices(1):indices(end)) = fread(fid, length(indices) ,qbits);
else
InputValues(indices(1):indices(end)) = 0;
end
end
end

%Dequantización
for i=1:length(InputValues)
if InputValues(i) ~= 0
if max(bit_alloc(floor(fftbank(i,framelength/2,Fs))+1),0) ~= 0
InputValues(i) = midtread_dequantizer(InputValues(i),...
max(bit_alloc(floor(fftbank(i,framelength/2,Fs))+1),0)); %#ok<AGROW>
end
end
end
for i=1:25
gain2(i) = 2^gain(i); %#ok<AGROW>
end

```

```

%Se aplicanancia
for i=1:floor(fftbank(framelength/2,framelength/2,Fs))+1
indices = find((floor(fftbank([1:framelength/2],framelength/2,Fs))+1)==i);
InputValues(indices(1):indices(end)) = InputValues(indices(1):indices(end)) * gain2(i);
end

%MDCT inversa (IMDCT)
x2((frame_count-1)*framelength+1:frame_count*framelength) =
imdct(InputValues(1:framelength/2));
end
status = fclose(fid); %#ok<NASGU>
%Se recombina las frames
x3 = zeros(1,(length(x2)-1)/2+1);
for i=0:0.5:floor(length(x2)/(2*framelength))-1
x3(i*framelength+1 : (i+1)*framelength) = x3(i*framelength+1 : (i+1)*framelength) +
x2((2*i)*framelength+1 : (2*i+1)*framelength);
end
set(handles.text13,'visible','on');
Cod='Decodificacion_terminada';
set(handles.text13,'string',Cod);

%Se crea el archivo .wav
wavwrite(x3/2,Fs,decoded_filename);

```

Código función entramado (enframe)

```

function f=enframe(x,win,inc)
% El número de tramas esta dado por fix((length(X)-LEN+INC)/INC)
nx=length(x);
nwin=length(win);
if (nwin == 1)
len = win;
else
len = nwin;
end
if (nargin< 3)
inc = len;
end
end

```

```

nf = fix((nx-len+inc)/inc);
f=zeros(nf,len);
indf= inc*(0:(nf-1)).';
inds = (1:len);
f(:) = x(indf(:,ones(1,len))+inds(ones(nf,1),:));
if (nwin> 1)
w = win(:)';
f = f .* w(ones(nf,1),:);
end

```

Código función MDCT

```

function y = mdct(x)
x=x(:);
N=length(x);
n0 = (N/2+1)/2;
wa = sin([(0:N-1]'+0.5)/N*pi);
y = zeros(N/2,1); %#ok<NASGU>
x = x .* exp(-1i*2*pi*[0:N-1]'/2/N) .* wa;
X = fft(x);
y = real(X(1:N/2) .* exp(-1i*2*pi*n0*([0:N/2-1]'+0.5)/N));
y=y(:);

```

Código función IMDCT

```

function y = imdct(X)
X=X(:);
N = 2*length(X);
ws = sin([(0:N-1]'+0.5)/N*pi);
n0 = (N/2+1)/2;
Y = zeros(N,1);
Y(1:N/2) = X;
Y(N/2+1:N) = -1*flipud(X);
Y = Y .* exp(j*2*pi*(0:N-1)*n0/N);
y = ifft(Y);
y = 2*ws .* real(y .* exp(j*2*pi*([0:N-1]'+n0)/2/N));

```

A4. Relación entre la escala Bark y Hertzios

	Frecuencia (Hz)	Frecuencia central (Hz)
1ª Banda	0 - 100	50
2ª Banda	100 - 200	150
3ª Banda	200 - 300	250
4ª Banda	300 - 400	350
5ª Banda	400 - 510	450
6ª Banda	510 - 630	570
7ª Banda	630 - 770	700
8ª Banda	770 - 920	840
9ª Banda	920 - 1080	1000
10ª Banda	1080 - 1270	1170
11ª Banda	1270 - 1480	1370
12ª Banda	1480 - 1720	1600
13ª Banda	1720 - 2000	1850
14ª Banda	2000 - 2320	2150
15ª Banda	2320 - 2700	2500
16ª Banda	270 - 3150	2900
17ª Banda	3150 - 3700	3400
18ª Banda	3700 - 4400	4000
19ª Banda	4400 - 5300	4800
20ª Banda	5300 - 6400	5800
21ª Banda	6400 - 7700	7000
22ª Banda	7700 - 9500	8500
23ª Banda	9500 - 12000	10500
24ª Banda	12000 - 15500	13500
25ª Banda	15500 - 22050	19500

A6. Pre selección de códecs ADPCM

No.	Nombre Codificador
1	ACM Converter 1.0
2	Audio Converter 8.20
3	Audio MusicBatchConverter
4	Sound Converter ActiveX 1.2.6
5	Axara Audio Converter 3.1.4

A6. COMPROBACIÓN DE CALIDAD DE CODECS UTILIZADOS EN SONIDOS ESTETOSCOPICOS PARA TELEMEDICINA



COMPROBACIÓN DE CALIDAD DE CODECS UTILIZADOS EN SONIDOS ESTETOSCOPICOS PARA TELEMEDICINA

La presente evaluación consiste en una valoración de los sonidos decodificados, por un especialista en cada campo.

Médico: Dr. Fco. Gerardo Cueva Especialidad: Neumólogo
 Lugar de Trabajo: IESS - Hospital Los Rios

Sonidos Respiratorios

Decodificación realizada mediante el códec MDCT.

Sonidos	Muy Buena	Buena	Regular	Mala
Bronquial	X			
Traqueal		X		
Crepito Nino			X	
Anillo Vascolar				X
Normal	X			

Decodificación realizada mediante el códec ADPCM.

Sonidos	Muy Buena	Buena	Regular	Mala
Bronquial	X			
Traqueal	X			
Crepito Nino			X	
Anillo Vascolar				X
Normal	X			

Observaciones: El Anillo Vascolar que se escuchó no ha sido copiado de forma correcta por lo que es difícil para de obtener.

Instituto Ecuatoriano de Seguridad Social
 HOSP. "MANUEL Y. MONTEROS Y."

Dr. Gerardo Cueva
 Médico Especialista
 IESS NEUMÓLOGO
 C.D. 301127-001
 CEDULA: 1100611048