

CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE
EDUCACIÓN SUPERIOR DE ENSENADA



PROGRAMA DE POSGRADO EN CIENCIAS
EN ELECTRÓNICA Y TELECOMUNICACIONES

Análisis y adaptación de codecs para la transmisión de audio cardíaco
sobre redes inalámbricas con tasas de datos bajas/medianas

TESIS

que para cubrir parcialmente los requisitos necesarios para obtener el grado de

MAESTRO EN CIENCIAS

Presenta:

LILIANA CASTORENA LARA

Ensenada, Baja California, México, 2012.

TESIS DEFENDIDA POR

Liliana Castorena Lara

Y APROBADA POR EL SIGUIENTE COMITÉ



Dr. Salvador Villarreal Reyes

Director del Comité



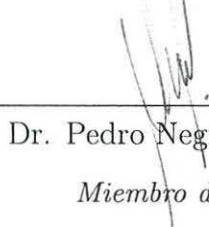
Dr. Roberto Conte Galván

Miembro del Comité



Dr. Raul Rivera Rodriguez

Miembro del Comité



Dr. Pedro Negrete Regagnon

Miembro del Comité



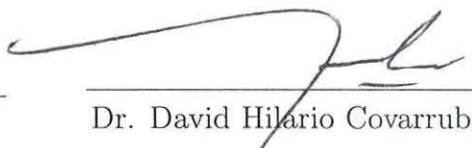
M. C. Humberto Cervantes De Ávila

Miembro del Comité



Dr. Roberto Conte Galván

*Coordinador del programa de
posgrado en Electrónica y Telecomunicaciones*



Dr. David Hilario Covarrubias Rosales

Director de Estudios de Posgrado

10 de febrero de 2012

RESUMEN de la tesis de **LILIANA CASTORENA LARA**, presentada como requisito parcial para la obtención del grado de **MAESTRO EN CIENCIAS en ELECTRÓNICA Y TELECOMUNICACIONES** con orientación en **TELECOMUNICACIONES**. Ensenada, Baja California, febrero de 2012.

**ANÁLISIS Y ADAPTACIÓN DE CODECS PARA LA TRANSMISIÓN
DE AUDIO CARDIACO SOBRE REDES INALÁMBRICAS CON
TASAS DE DATOS BAJAS/MEDIANAS**

Resumen aprobado por:



Dr. Salvador Villarreal Reyes

Director de Tesis

Una de las prioridades de cualquier sociedad debe ser la de proveer servicios de salud adecuados para la población. Dentro del perímetro de un área urbana, no existe gran problema para recibir atención médica básica basada en auscultación con estetoscopio, el cual es la herramienta de diagnóstico más utilizada por los médicos generales. Sin embargo, en muchas localidades de nuestro país no se cuenta con acceso a dicha atención médica básica. Esta problemática puede ser resuelta estableciendo sistemas de telemedicina sobre la redes de datos disponibles en dichas localidades. Sin embargo, una limitante para el establecimiento efectivo de un sistema de telemedicina son las tasas de datos ofrecidas por dichas redes. Por ejemplo, aun y cuando en la actualidad existen estetoscopios digitales que permiten la transmisión de señales de audio cardiaco sobre redes de datos, el ancho de banda requerido hace que su utilización sea prohibitiva cuando se consideran las capacidades ofrecidas por las redes de datos disponibles en localidades remotas de nuestro país.

El presente trabajo de investigación tiene como objetivo encontrar un mecanismo que permita la transmisión de señales de audio cardiaco en redes inalámbricas con tasas de datos bajas/medianas. Se decidió enfocarse en este tipo de señales debido a que las enfermedades cardiovasculares son la segunda causa de mortalidad en México. Para ello se analizaron códecs de audio existentes para aplicarlos a señales de audio cardiorrespiratorio y realizar su compresión y transmisión. El uso de los códecs para comprimir/transmitir audio cardiorrespiratorio no es trivial; esto debido a que la mayoría de los códecs fueron desarrollados para otro tipo de aplicaciones y podrían introducir distorsiones en la señal que conlleven a un mal diagnóstico. En este sentido se evaluó el desempeño de los códecs al codificar/decodificar señales de audio cardiorrespiratorio para después comparar la señal resultante con la original mediante pruebas objetivas y preceptuales estandarizadas. Como resultado de estas pruebas se descartaron varios códecs, seleccionando únicamente aquellos que arrojaron resultados satisfactorios en las pruebas. Posteriormente se analizó la factibilidad de utilizar estos códecs para la transmisión de audio cardiorrespiratorio sobre redes inalámbricas con tasas de datos bajas/medianas. Para esto se analizaron las tramas generadas por dichos códecs para

posteriormente determinar un mecanismo de empaquetamiento y transmisión sobre una red punto a punto con transceptores IEEE 802.15.4. Se evaluó la utilización de los códecs en este tipo de redes por medio de simulación y utilizando una cama de pruebas física con transceptores XBee Pro. Finalmente se realizaron pruebas sobre las señales recibidas para dar una recomendación final respecto a los códecs a utilizar para este tipo de aplicaciones. Las contribuciones de esta tesis pueden utilizarse como referencia para el establecimiento de un sistema de teleauscultación en redes de datos con anchos de banda limitados.

Palabras Clave: códec, compresión de audio, audio cardíaco, audio cardiorrespiratorio, redes inalámbricas de área personal, redes de sensores, telemedicina

ABSTRACT of the thesis presented by **LILIANA CASTORENA LARA**, in partial fulfillment of the requirements of the **MASTER IN SCIENCES** degree in **ELECTRONICS AND TELECOMMUNICATIONS** with orientation in **TELECOMMUNICATIONS**. Ensenada, Baja California, February 2012.

ANALYSIS AND ADAPTATION OF CODECS FOR HEARTH SOUND TRANSMISSION OVER LOW/MEDIUM DATA RATES WIRELESS NETWORKS

Within the perimeter of an urban area, there is no mayor problem to receive basic medical services based on auscultation with a stethoscope (which is an almost ubiquitous diagnostic tool used by general physicians). However, many locations in our country do not have access to such basic medical services. One way to solve this problem is by establishing telemedicine systems over the existing data networks available in these locations. However, the data rates offered by these data networks present a challenge for the effective deployment of any telemedicine solution. For example, even though digital stethoscopes that enable transmission of cardiorespiratory audio over data networks can be found in the market, the bandwidth required to implement such systems overcomes the capabilities offered by the data networks found in remote regions of our country.

The aim of this research work is to find a mechanism that enables the transmission of heart sound signals over wireless networks with low/medium data rates. The fact that cardiovascular diseases are the second cause of death in Mexico is the reason why this work focused in this kind of signals. In order to address this issue, different audio codecs are analyzed to determine its suitability for the compression/transmission of cardiorespiratory signals. The use of codecs for transmitting such signals is not a trivial task. This is because most of the codecs were developed for other applications and may introduce significant distortions to the signal that could lead to wrong diagnosis. Therefore, each codec is evaluated by encoding/decoding cardiorespiratory audio signals and then comparing the decoded and the original audio signals. The comparison is performed by using standardized perceptual (subjective) and objective audio tests. As a result of these tests several popular audio codecs are discarded, choosing only those that provided satisfactory results in the tests. The selected codes are further analyzed to determine its suitability for cardiorespiratory audio transmission over wireless networks with low/medium data rates. The analysis focuses in the data framing provided by the codec, such that a mechanism for the signal transmission over an IEEE 802.15.4 peer to peer network can be set. The use of the codecs over these kind of networks is evaluated by means of simulation and a testbed using XBee Pro transceivers. In order to provide a final recommendation, audio quality tests are further performed over the compressed signals after they are received in the XBee Pro transceiver. The contributions of this thesis can be used as a guideline for the design of a teleauscultation system over data networks with low data rates.

Keywords: codec, audio compression, heart sound, WPAN, sensor networks, IEEE 802.15.4, telemedicine

A mis padres, David y Cecilia

A mi hermana, Cecilia Marisol

Al amor de mi vida, Edwin

Agradecimientos

A Dios, por guiarme en esta etapa de mi vida y ayudarme a cumplir con este objetivo.

A mis padres David y Cecilia. Gracias por forjar la persona que soy y permanecer a mi lado, aun en la distancia.

A mi hermana Marisol que siempre será un “pedacito” de mi vida.

A Edwin, gracias por ser mi amigo y compañero de vida. Te amo.

A los miembros de mi comité por su disposición y sus consejos, tanto en el aspecto académico como en el personal. En especial al Dr. Salvador Villarreal por brindarme sin reservas su tiempo, conocimiento y experiencia.

Al Dr. Miguel Ángel Alonso por sus comentarios y observaciones durante el desarrollo de este trabajo de investigación.

Al Grupo de Comunicaciones Inalámbricas (GCI) del CICESE por su apoyo y compañía. En especial al Dr. David Covarrubias por brindarme sus consejos y apoyo.

Al enfermero de CICESE, Angel Rico, por su disposición y apoyo en la obtención de muestras de audio para este trabajo.

Al personal de PLAMEX Tijuana, en especial al Ing. Luis Luna y al técnico Isaías Ramírez por su apoyo en la obtención de algunos resultados plasmados en esta tesis.

A mis amigos estudiantes de maestría y doctorado, gracias por convertirse en mi otra familia.

A mis maestros y a todo el personal del CICESE por facilitar mi estancia en este plantel.

Al CONACYT por la beca otorgada para la realización de este proyecto.

Contenido

	Página
Resumen en Español	i
Resumen en Inglés	iii
Dedicatoria	iv
Agradecimientos	v
Contenido	vi
Lista de Figuras	ix
Lista de Tablas	xi
I. Introducción	1
I.1 Antecedentes	1
I.1.1 Codificador - Decodificador (Códec)	5
I.1.2 Redes Inalámbricas de Área Personal (WPANs)	5
I.2 Planteamiento del Problema	6
I.3 Objetivo General	8
I.4 Objetivos Particulares	9
I.5 Justificación	9
I.6 Metodología de la investigación	10
I.7 Organización de la tesis	11
II. Sonidos cardiacos y respiratorios	12
II.1 Introducción	12
II.2 Ciclo cardiaco	12
II.2.1 Fonocardiograma	13
II.2.2 Soplos cardiacos	15
II.2.3 Análisis espectral del fonocardiograma	16
II.3 Sonidos respiratorios	19
II.3.1 Origen de los ruidos respiratorios	19
II.3.2 Ruidos respiratorios normales	20
II.3.3 Ruidos adventicios	21
II.3.4 Análisis espectral de los ruidos respiratorios	22
III. Introducción a esquemas de codificación de audio	24
III.1 Introducción	24

Contenido (continuación)

	Página
III.2 Psicoacústica	25
III.2.1 Umbral absoluto de audición	27
III.2.2 Umbral de detección del ruido	27
III.2.3 Enmascaramiento sonoro	28
III.2.4 Bandas críticas	30
III.3 Digitalización de datos	31
III.3.1 PCM	32
III.4 Sistemas de codificación de audio	39
III.4.1 Códec de audio	41
III.4.2 Codificadores de forma de onda	44
III.4.3 Codificadores perceptuales	45
III.4.4 Codificadores paramétricos	48
III.4.5 Codificadores híbridos	51
IV. Tecnologías de redes de área personal	53
IV.1 Introducción	53
IV.2 Bluetooth	55
IV.2.1 Transmisión de audio sobre redes Bluetooth	58
IV.2.2 Perfil A2DP	59
IV.2.3 Análisis de la tasa efectiva de transmisión para enlaces Bluetooth	61
IV.2.4 Estetoscopio digital con enlace Bluetooth	63
IV.3 IEEE 802.15.4	65
IV.3.1 Componentes de red y topologías	65
IV.3.2 Capa física	68
IV.3.3 Sub-capa MAC	70
IV.3.4 Análisis de la tasa efectiva de transmisión para IEEE 802.15.4	72
IV.4 Escenario de aplicación y esquema propuesto	76
V. Análisis de códecs y evaluación de calidad	81
V.1 Introducción	81
V.2 Códecs considerados	82
V.2.1 Códec G.711	82
V.2.2 Códec G.726	83
V.2.3 Códec G.729	84
V.2.4 Códec GSM	86
V.2.5 Códec Speex	88
V.2.6 Códec MP2	88
V.2.7 Códec MP3	90

Contenido (continuación)

	Página
V.2.8 Códec AAC	94
V.2.9 Códec FLAC	97
V.2.10 Códec Ogg Vorbis	99
V.3 Evaluación de calidad de los códec de audio	102
V.3.1 Evaluación objetiva de códec	102
V.3.2 Evaluación subjetiva de códec	105
V.3.3 Conclusión	114
VI. Aplicación de los códec seleccionados en el escenario propuesto	116
VI.1 Introducción	116
VI.2 Consideraciones	116
VI.3 Análisis de tramas FLAC	118
VI.4 Análisis de tramas AAC	121
VI.5 Análisis de tramas MP3	122
VI.6 Etapa de simulación	122
VI.7 Etapa de experimentación	127
VII. Conclusiones y trabajo futuro	129
VII.1 Introducción	129
VII.2 Conclusiones	129
VII.3 Trabajo futuro	131
Referencias bibliográficas	133

Lista de Figuras

Figura	Página
1	Estetoscopio digital comercial con enlace inalámbrico Bluetooth. 4
2	Características de las redes de área personal. 6
3	Flujo sanguíneo a través de las válvulas cardiacas (Chiras (2011)). 13
4	Ocurrencia de los sonidos cardiacos dentro de un ciclo cardiaco (Echeverry <i>et al.</i> (2007)). 14
5	Espectrograma de un PCG, analizado con una ventana Hamming de 256 muestras (Djebbari y Bereksi Reguig (2000)). 16
6	Diagrama de contorno del espectrograma de un PCG, analizado con una ventana Hamming de 256 muestras (Djebbari y Bereksi Reguig (2000)). 17
7	Densidad espectral de energía de un PCG de frote pericárdico. 18
8	Densidad espectral de energía de un PCG normal. 19
9	Densidad espectral de energía de un PCG de sonidos respiratorios. 23
10	Intervalo de frecuencia de la capacidad auditiva del ser humano. 25
11	Umbral absoluto de audición del ser humano (Spanias <i>et al.</i> (2007)). 27
12	Enmascaramiento temporal (Figueiras Vidal y Rodríguez, 2002). 29
13	Muestreo de una señal analógica. 33
14	Diagrama a bloques de un convertidor analógico-digital. 35
15	Diagrama general de un codec. 44
16	Diagrama a bloques de un codificador perceptual. 45
17	Diagrama a bloques de un codificador paramétrico. 48
18	Diagrama a bloques de un codificador de voz. 51
19	Diagrama a bloques de un codificador híbrido de voz. 52
20	Relación entre tasa de datos y alcance de algunas tecnologías inalámbricas. 54
21	Tipo de redes Bluetooth que pueden ser establecidas. a) Picored de un solo esclavo b) Picored con varios esclavos c) Scatternet. 57

Lista de Figuras (continuación)

Figura		Página
22	Estructura de un paquete Bluetooth.	62
23	Estetoscopio electrónico con enlace Bluetooth de la marca 3M Littmann (Littmann, 2010).	64
24	Pila de protocolos IEEE802.15.4 y ZigBee.	66
25	Topologías de red en IEEE 802.15.4 a) Estrella b) Igual-a-igual c)Árbol (Mahalik (2007)).	67
26	Bandas de operación en el estándar IEEE 802.15.4.	68
27	Estructura de los paquetes ACK en IEEE 802.15.4.	74
28	Escenarios de aplicación para la transmisión de audio cardiaco codificado.	78
29	Esquema propuesto para la codificación y transmisión de audio cardio- rrrespiratorio.	79
30	Diagrama a bloques del codificador G.711.	83
31	Diagrama a bloques del codificador G.726.	83
32	Diagrama a bloques del codificador G.729.	85
33	Diagrama a bloques del codificador GSM Full-Rate.	87
34	Diagrama a bloques del codificador MP3.	91
35	Diagrama a bloques del codificador AAC.	95
36	Diagrama a bloques del codificador Vorbis I.	100
37	Procedimiento empleado para hacer la evaluación objetiva.	103
38	Procedimiento empleado para hacer la evaluación subjetiva con PEMO-Q.	108
39	Diagrama del proceso de digitalización, codificación y transmisión del audio cardiorrespiratorio sobre una red IEEE 802.15.4.	117
40	Formato de tramas codificadas con FLAC.	119
41	Topología de red del escenario de transmisión bajo interferencia.	124
42	Diagrama de transmisión de audio codificado con módulos XBee-PRO.	127

Lista de Tablas

Tabla	Página
I Caracterización de los ruidos cardiacos.	15
II Principales recomendaciones de la ITU para la codificación de voz. . . .	50
III Frecuencias de operación de Bluetooth en algunas regiones.	56
IV Codecs que se definen en el perfil A2DP de Bluetooth.	60
V Paquetes ACL (Bruno <i>et al.</i> (2002)).	63
VI Paquetes SCO (Bruno <i>et al.</i> (2002)).	63
VII Características de capa física para cada banda de frecuencia en IEEE 802.15.4.	69
VIII Evaluación de códecs por medio del coeficiente de correlación de Pearson.	104
IX Valores ODG muestreo 8KHz, cuantificación 16 bits.	108
X Valores ODG muestreo 8KHz, cuantificación 16 bits (continuación). . .	109
XI Valores ODG muestreo de 8KHz, cuantificación 8 bits.	109
XII Valores ODG muestreo de 4KHz, cuantificación 16 bits.	110
XIII Parámetros de evaluación con software ACQUA, muestras a 8KHz con cuantificación de 16 bits.	113
XIV Parámetros de evaluación con software ACQUA, muestras a 8KHz con cuantificación de 8 bits.	114
XV Tamaño de las tramas FLAC con 192 muestras de audio PCM por bloque codificado para el archivo de audio número 73.	120
XVI Valores ODG de señales obtenidas de la simulación de transmisión del audio cardiaco sobre una red IEEE 802.15.4 sin interferencia.	125
XVII Valores ODG de señales obtenidas de la simulación de transmisión del audio cardiaco sobre una red IEEE 802.15.4 con 3 interferentes. . . .	126
XVIII Parámetros utilizados para la codificación de los archivos transmitidos en la etapa experimental.	128
XIX Parámetros de evaluación final con software ACQUA.	128

Capítulo I

Introducción

I.1 Antecedentes

Una de las prioridades de cualquier sociedad debe ser la de proveer servicios de salud adecuados para la población. Sin embargo, muchas veces no es posible proporcionar dichos servicios debido a las limitantes impuestas por la geografía, infraestructura, cantidad de personal, entre otras. Esto es particularmente problemático en poblaciones alejadas en las cuales no se dispone de atención médica local.

Problemas tales como la ubicación y accesibilidad han provocado el surgimiento de las áreas de telemedicina y telesalud, las cuales buscan apoyarse en los desarrollos logrados en el ámbito de las tecnologías de la información y comunicaciones (TIC) para proveer servicios de salud básicos a la mayoría de la población.

En la actualidad existe un gran interés en el desarrollo de tecnología enfocada a las áreas de telemedicina y telesalud. La telemedicina puede ofrecer un medio alternativo de atención, que potencialmente permitirá brindar una atención adecuada sin necesidad de costosos traslados que, adicionalmente, puedan empeorar el estado de salud de los pacientes. Los escenarios de aplicación de este tipo de tecnología son múltiples, variando desde el cuidado preventivo en hogares de ancianos hasta pre-diagnósticos remotos en casos de emergencia.

Todavía existen diversos retos a resolver en el área de telemedicina, los cuales van desde el desarrollo de dispositivos completamente digitales hasta el diseño adecuado

de las redes de comunicaciones a ser utilizadas. Esto ha propiciado el surgimiento de trabajos de investigación y desarrollo que tratan con temas como: el desarrollo de herramientas de software para la realización de auscultaciones remotas a través de redes cableadas (Johanson *et al.*, 2002); la caracterización de dispositivos médicos utilizados el diseño de servicios de telesalud (Gortzis *et al.*, 2008); la precisión en la detección de ritmos cardiacos irregulares vía telemedicina (Zenk *et al.*, 2004); la utilización de redes de sensores basadas en el estándar Bluetooth para monitoreo remoto de señales fisiológicas de un paciente (Zhang y Xiao, 2009); entre otros.

Por otro lado, las enfermedades cardiovasculares son la segunda causa de mortalidad en México, de acuerdo con las estadísticas del Sistema Nacional de Información en Salud (SINAIS). Debido a este tipo de enfermedades mueren más de 60 mil personas al año ¹. Desafortunadamente, la tendencia a la mortandad por dichas enfermedades es a la alza, debido a la epidemia de obesidad que atraviesa actualmente nuestro país ². La situación en el resto del mundo no es muy diferente, pues según el Índice de Mortalidad de la Organización Mundial de la Salud y el Comité de Estadísticas de la American Heart Association, las enfermedades cardiovasculares son una de las principales causas de mortalidad en el mundo tanto en niños como en adultos (Rosamond *et al.*, 2007).

En la actualidad existen técnicas sofisticadas para diagnosticar el estado cardiovascular de un paciente, e.g. el ultrasonido cardiaco (ecocardiografía), la cateterización cardiaca, y la imagen por resonancia magnética (IMR). Aunque todas estas técnicas proporcionan una gran cantidad de información, desafortunadamente tienen un alto costo, lo que las vuelve inaccesibles para el grueso de la población de países en de-

¹SINAIS (2010). Sistema Nacional de Información en Salud. Principales causas de mortalidad general. <http://sinais.salud.gob.mx/mortalidad/index.html>. Consultado en agosto 2011.

²OECD (2010). Organisation for Economic Co-operation and Development. Health: Obesity rising in developing countries, warns OECD. <http://bit.ly/r9hVWQ>. Consultado en agosto de 2011.

sarrollo como México. Sin embargo, existe una técnica sencilla y de bajo costo que ha auxiliado a los médicos en el diagnóstico de enfermedades cardiovasculares durante décadas: la **auscultación caridaca**.

La auscultación es el examen médico que se realiza con el estetoscopio (también llamado foneidoscopio). Este procedimiento, consiste en escuchar los ruidos internos del cuerpo humano, generalmente ruidos cardíacos o ruidos respiratorios, aunque algunas veces también se usa para objetivar ruidos intestinales o soplos por flujos sanguíneos anómalos en arterias y venas (Surós Batlló y Surós Batlló, 2001).

Dentro del perímetro de un área urbana no existe gran problema para recibir atención médica básica basada en auscultación con estetoscopio. Adicionalmente, es relativamente sencillo pedir la opinión de un médico especialista en caso de que se considere necesario. Sin embargo, en muchas localidades de nuestro país no se cuenta con acceso a dicha atención médica básica, e incluso hay lugares donde a pesar de que se cuenta con acceso a un médico general no se cuenta con servicios de especialidad médica. En este contexto, un uso adecuado de las redes de datos permitiría proveer de servicios médicos básicos a este tipo de localidades. Sin embargo, uno de los retos a enfrentar es que en su mayoría las prestaciones de las redes de datos disponibles dichas localidades son bajas, motivo por el cual es necesario encontrar mecanismos que permitan usar eficientemente el poco ancho de banda disponible.

En sus orígenes el estetoscopio fue desarrollado con base en principios mecánico-acústicos. Sin embargo, actualmente existen estetoscopios digitales los cuales poseen: mejor respuesta a la frecuencia que sus precursores, mayor sensibilidad al sonido y control de volumen. Adicionalmente, estos instrumentos tienen la capacidad de grabar los sonidos escuchados en la auscultación, ya sea para guardarlos en el expediente clínico del paciente o para que el médico pueda pedir la valoración de otro colega. La mayoría

de los estetoscopios digitales requieren estar conectados por un medio cableado a una PC para poder almacenar y grabar dichos sonidos cardiacos. Sin embargo, en el último año se ha observado que algunos modelos incluyen conectividad inalámbrica basada en el estándar Bluetooth. Un ejemplo es el modelo 3200 de la marca 3M Littmann, el cual se muestra en la Figura 1.



Figura 1. *Estetoscopio digital comercial con enlace inalámbrico Bluetooth.*

El contar con conectividad inalámbrica ofrece ventajas evidentes en aplicaciones de telemedicina, ya que el médico puede indicar remotamente en que punto colocar el diafragma del estetoscopio y escuchar los sonidos cardiacos/respiratorios de forma remota. Aun y cuando los estetoscopios comerciales basados en Bluetooth ofrecen varias ventajas en términos de conectividad, cabe mencionar que el ancho de banda requerido para su utilización en redes de telemedicina no es despreciable (e.g. el modelo 3200 de la marca 3M Littmann requiere 128 kbps dedicados). Por este motivo, es necesario encontrar mecanismos que permitan reducir dicho ancho de banda, por ejemplo, con la

utilización de codificadores de audio que compriman la señal. Adicionalmente, debido a la sobresaturación del uso de la banda de frecuencia en la que trabaja Bluetooth (2.4 GHz), sería conveniente que dicho mecanismo pueda ser adaptado para su utilización con otras soluciones tecnológicas para redes inalámbricas con tasas de datos bajas-medianas.

I.1.1 Codificador - Decodificador (Códec)

Básicamente, un códec (codificador-decodificador) es un mecanismo por medio del cual se comprimen señales de audio y video para reducir el ancho de banda antes de que dichas señales sean transmitidas a través de una red de telecomunicaciones. Los códecs son comúnmente utilizados cuando se requiere almacenar o transmitir gran cantidad de información. Dichos dispositivos realizan la compresión y descompresión de señales digitales para reducir el tamaño de la señal y hacer que esta pueda ser almacenada y/o transmitida con mayor facilidad a través de la red de comunicaciones.

Existen códecs para distintas aplicaciones, tales como texto, imágenes, audio y video. Sin embargo, el principio general de los códecs es buscar la reducción del tamaño de un archivo eliminando la redundancia que existe en las señales.

Dentro de los códecs de audio existen algunos que han sido adaptados a las características de la voz y otros que son de propósito general. Estos últimos son mayormente utilizados para comprimir señales de música.

I.1.2 Redes Inalámbricas de Área Personal (WPANs)

Actualmente existe una tendencia a la eliminación de todo tipo de cableado entre dispositivos, con lo cual han cobrado una mayor relevancia las redes inalámbricas de área

personal (WPAN, por sus siglas en inglés). Este tipo de redes ofrece bajas y medianas tasas de transmisión, mientras que los dispositivos asociados a ellas presentan baja complejidad y bajo consumo de potencia.

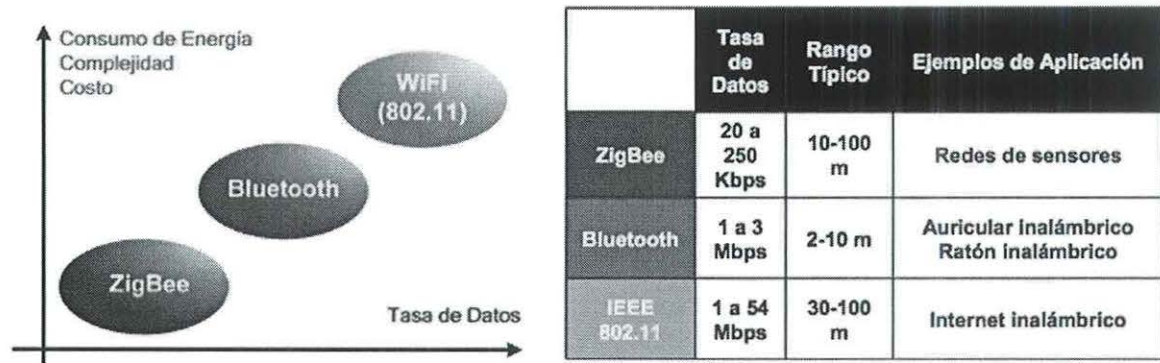


Figura 2. Características de las redes de área personal.

Las aplicaciones de las WPAN, van desde la comunicación entre dispositivos de oficina tales como impresoras, teclados, ratones y computadoras; hasta domótica, redes de sensores orientadas al control industrial y sensado de variables biomédicas.

Para estas aplicaciones existen redes basadas en diversos estándares para WPAN, entre los cuales se encuentran IEEE 802.15.4 y Bluetooth. La Figura 2 muestra una comparación de IEEE 802.15.4 y Bluetooth con respecto al estándar IEEE 802.11 ampliamente utilizado para conectividad inalámbrica.

I.2 Planteamiento del Problema

Como se explicó anteriormente, una herramienta de diagnóstico que es prácticamente ubicua es el estetoscopio. El intervalo de frecuencia de auscultación, es decir, con-

siderando audio cardio-respiratorio es de 20 Hz a 3000 Hz (Vannuccinni *et al.*, 2000), sin embargo, el presente trabajo de investigación pretende centrarse en señales de audio cardiaco, las cuales se encuentran por debajo de los 1000 Hz de acuerdo con los trabajos de (Romero *et al.*, 2009), (Mahabuba *et al.*, 2009), (Debbal y Bereksi-Reguig, 2008), (Noponen *et al.*, 2007), (Gauthier *et al.*, 2007). Este límite superior en la frecuencia, es una característica que puede ser aprovechada para la adaptación de los códecs ya existentes, que abarcan un intervalo mayor de frecuencia que el ocupado por el audio cardiaco.

Actualmente existen varias tecnologías y estándares que permiten el establecimiento de redes inalámbricas de sensores, dentro de los cuales podemos identificar a los estándares IEEE 802.15.4 y Bluetooth como los más prominentes (Bing, 2008). Aun y cuando Bluetooth puede proveer el ancho de banda suficiente para transmitir audio cardio-respiratorio sin comprimir y con una calidad aceptable, utilizar el enlace de esta manera técnicamente consumiría todo el ancho de banda disponible, por lo cual no quedarían recursos para otras aplicaciones. En lo que respecta a el estándar IEEE 802.15.4, este provee un menor ancho de banda que Bluetooth, por lo que su utilización hasta el momento se ha centrado en el monitoreo de variables tales como el ritmo cardiaco, la temperatura, etc. En este contexto, una área de investigación que surge de manera natural consiste en encontrar mecanismos que permitan transmitir audio cardio-respiratorio sobre enlaces inalámbricos con tasas de datos bajas/medianas. Este problema no es trivial, ya que aun y cuando Bluetooth e IEEE 802.15.4 ofrecen tasas pico de transmisión de hasta 3 Mbps y 250 Kbps respectivamente, la tasa de transmisión efectiva será mucho menor, tal cual se analizará en el Capítulo IV de esta tesis.

Es peculiar que aun y cuando se han hecho diversos estudios acerca del proceso

de digitalización de audio cardio-respiratorio, con el mejor de nuestro conocimiento, hasta la fecha no existe un estándar internacional que determine como transmitir dicha información a través de enlaces de comunicaciones inalámbricas. Como se mencionó en la introducción, el estudiar mecanismos que permitan la transmisión de dichas señales en este tipo de redes es muy importante para la implementación efectiva de sistemas de telemedicina. Una posible solución consiste en utilizar un códec de audio estándar que permita comprimir la señal de audio. Sin embargo, la selección del códec a utilizar no es trivial, dado que se requiere que la señal de audio conserve la información necesaria para que un médico pueda realizar un diagnóstico. Similarmente, se debe de cuidar que el enlace inalámbrico no degrade de manera significativa la señal.

Basados en la discusión anterior, se puede concluir que la utilización de códecs para la transmisión de audio cardiaco sobre redes inalámbricas con tasas bajas-medianas es un problema abierto, por lo cual este proyecto de investigación se enfoca en estudiar cuáles códecs se pueden utilizar para realizar dicha transmisión tomando como criterio de evaluación la calidad de la señal de audio cardiaco.

I.3 Objetivo General

El objetivo de esta tesis es evaluar diferentes códecs de audio para proponer al menos uno que permita la transmisión de audio cardiaco digitalizado sobre redes de inalámbricas con tasas bajas-medianas. El códec seleccionado debe de mantener la calidad de la señal de tal manera que la realización de un diagnóstico sea factible. Adicionalmente, el códec debe de permitir la paquetización de los datos comprimidos para permitir su transmisión sobre un enlace inalámbrico IEEE 802.15.4.

I.4 Objetivos Particulares

- Evaluar la factibilidad de uso de diversos códecs de audio conocidos para la compresión de audio cardiaco.
- Analizar y adaptar códecs de audio conocidos para que sea posible utilizarlos en la transmisión de audio cardiaco sobre enlaces inalámbricos de redes de área personal.
- Evaluar el desempeño del esquema propuesto en una red típica con varios sensores mediante simulación.
- Recomendar un códec que se adapte a las necesidades de ancho de banda de la red inalámbrica, así como a la calidad de audio requerida para poder realizar un diagnóstico adecuadamente.

I.5 Justificación

La justificación de este trabajo de investigación surge del hecho que, con el mejor de nuestro conocimiento, no existe un estándar que indique que códec o códecs son los adecuados a utilizar cuando se transmite audio cardio-pulmonar en redes de datos inalámbricas con tasas de datos bajas-medianas. La realización de este trabajo de investigación permitirá hacer una recomendación respecto a que códec utilizar para la transmisión de audio cardiaco en enlaces IEEE 802.15.4, considerando que la realización de un diagnóstico sea factible.

I.6 Metodología de la investigación

- Se estudiaron los principales códecs de audio utilizados en aplicaciones de telecomunicaciones para identificar aquellos que fueran adecuados para los requerimientos del proyecto.
- Se analizaron las capacidades ofrecidas por los estándares IEEE 802.15.4 y Bluetooth.
- Se hizo el análisis de algunas muestras de audio cardiaco y respiratorio para obtener su firma espectral y comprobar que los resultados fueran consistentes con la documentación. Este análisis fue desarrollado en Matlab.
- Tomando en consideración el escenario de transmisión de audio cardiaco, se propusieron adecuaciones a los códecs, enfocándose en mantener una calidad de audio aceptable para poder realizar un diagnóstico y así como en la reducción del ancho de banda necesario para transmitir la señal.
- Se realizó una evaluación objetiva y subjetiva de los códecs mediante software, para determinar cuales serían utilizados en la etapa de simulación.
- Se evaluó el desempeño sobre un enlace inalámbrico mediante una simulación realizada sobre la plataforma NS-2.
- Se evaluó la factibilidad de uso de los códecs al transmitir varias señales comprimidas sobre un enlace IEEE 802.15.4 físico.

I.7 Organización de la tesis

De acuerdo con la metodología expuesta anteriormente, en el Capítulo II se describen el origen y caracterización del audio cardíaco y el audio pulmonar, así como la forma en que dicho audio puede ser digitalizado sin perder información relevante.

Posteriormente, el Capítulo III se presenta una introducción a los esquemas de codificación de los códecs de audio, su clasificación, así como las características principales de cada tipo de códec.

En el Capítulo IV se aborda el tema de las redes de área personal (WPAN), y se describen los dos estándares principales de este tipo de redes: Bluetooth e IEEE 802.15.4. Un análisis del caudal eficaz máximo para cada estándar es presentado en este capítulo. En este capítulo también se detalla el escenario de aplicación para la transmisión de señales cardíacas en redes de área personal y se describe una aplicación comercial ya existente.

El Capítulo V muestra el análisis de varios códecs existentes; la descripción de su funcionamiento y sus diagramas a bloques. También se presentan los resultados que se obtuvieron de la compresión de las señales de audio cardíaco y pulmonar. Se expone el resultado de una evaluación objetiva y subjetiva de estas muestras.

En el Capítulo VI se exponen los resultados del análisis del desempeño de los códecs seleccionados mediante la simulación de redes de área personal en el programa NS-2. También se presentan resultados experimentales de la transmisión de paquetes de audio codificado sobre una red IEEE 802.15.4 con y sin interferencia.

Finalmente, en el Capítulo VII, se exponen las conclusiones del presente trabajo de investigación.

Capítulo II

Sonidos cardiacos y respiratorios

II.1 Introducción

En este capítulo se presenta una breve descripción del origen de los sonidos cardiacos y respiratorios. También se hace una caracterización de dichos sonidos, centrando el interés en su espectro frecuencia. Posteriormente, se presenta el análisis de muestras de audio cardiaco y respiratorio de individuos en condiciones normales, y muestras de individuos con alguna cardiopatía y sonidos adventicios.

II.2 Ciclo cardiaco

Los eventos que ocurren desde el inicio de un latido hasta el inicio del siguiente se conocen como ciclo cardiaco. Cada ciclo cardiaco consta de un periodo de relajación, llamado diástole, durante el cual el corazón se llena de sangre, seguido de un periodo de contracción llamado sístole (Ganong, 1996), durante la sístole la sangre es bombeada nuevamente. El corazón cuenta con cuatro válvulas, las cuales se clasifican en dos tipos: auriculoventriculares (A-V) y sigmoideas. Las válvulas A-V (mitral y tricúspide) impiden el flujo retrógrado de la sangre de los ventrículos a las aurículas durante la sístole, y las válvulas sigmoideas (aórtica y pulmonar) impiden que la sangre de las arterias aorta y pulmonar regrese a los ventrículos durante la diástole (Ganong, 1996).

En la Figura 3 se observa la dirección que sigue el flujo sanguíneo a través de las válvulas cardiacas.

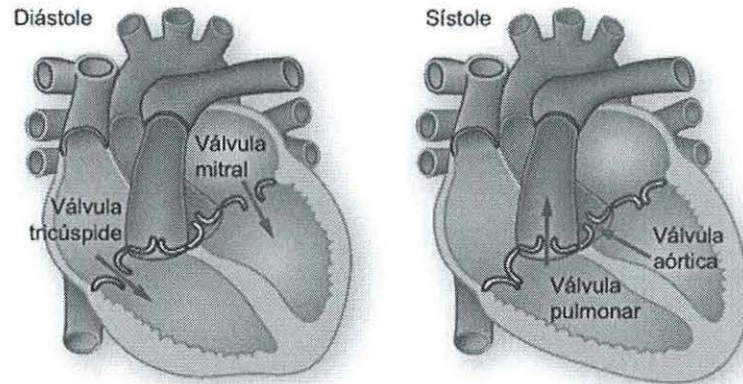


Figura 3. *Flujo sanguíneo a través de las válvulas cardiacas (Chiras (2011)).*

II.2.1 Fonocardiograma

El corazón humano puede generar cuatro sonidos durante la actividad de un ciclo cardíaco. Estos sonidos son conocidos como S1, S2, S3 y S4, y tienen la peculiaridad de que no son todos audibles (Djebbari, 1999),(Feigen, 1971),(Rangayan y Lehner, 1988). El conjunto de sonidos mencionado constituye una señal llamada fonocardiograma (PCG, por sus siglas en inglés).

Generalmente, en una persona con una condición de salud normal, sólo son audibles los sonidos S1 y S2, aunque en ocasiones también se pueden presentar los sonidos S3 y S4. Sin embargo, en adultos mayores S3 está asociado con un cambio patológico en el funcionamiento ventricular (Bates *et al.*, 2008). En la Figura 4 se ilustra la ocurrencia de los cuatro sonidos dentro de un ciclo cardíaco.

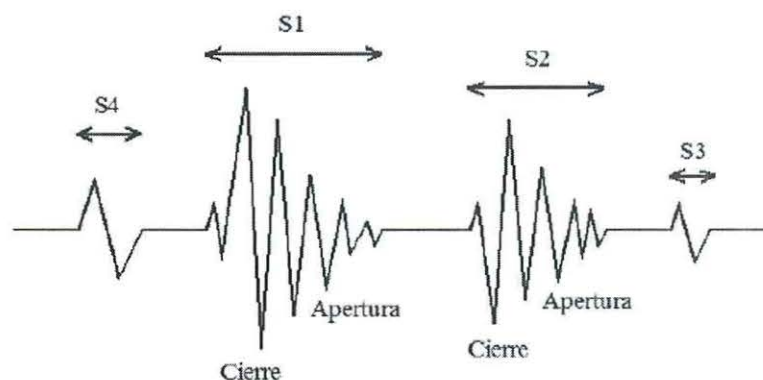


Figura 4. Ocurrencia de los sonidos cardíacos dentro de un ciclo cardíaco (Echeverry et al. (2007)).

Los cuatro sonidos anteriormente mencionados ocurren como consecuencia del cierre o apertura las válvulas cardíacas. Durante la sístole ventricular, los ventrículos se contraen para expulsar la sangre hacia las arterias. En ese momento las aurículas están en la fase de diástole auricular (relajadas), y las válvulas mitral y tricúspide se cierran para evitar que la sangre retroceda hacia ellas, mientras que las aórtica y pulmonar se abren. Esto da lugar al primer sonido cardíaco, conocido como S1.

Durante la sístole auricular, las dos aurículas se contraen para impulsar la sangre a los ventrículos, los cuales están en ese momento en la fase de diástole ventricular. En esos instantes las válvulas aórtica y pulmonar se cierran, mientras que las válvulas mitral y tricúspide se abren. Esto origina el segundo sonido, llamado S2. Finalmente, se produce una fase de relajación de las aurículas y ventrículos (diástole general), a partir de la cual se iniciará un nuevo ciclo (Chiras, 2011).

En algunos casos se presenta un tercer ruido cardíaco S3, que corresponde a la fase de llenado ventricular. Es un ruido de baja frecuencia, y se puede presentar entre 0.12 a 0.16s después de S2 (Cheitlin *et al.*, 1993). El cuarto ruido cardíaco (S4) es producido

en la final de la diástole, y se genera por la contracción de las aurículas desplazando el flujo dentro de los ventrículos.

Las características de frecuencia de los sonidos cardiacos se describen en la Tabla I (Ewing, 1998).

Tabla I. *Caracterización de los ruidos cardiacos.*

Ruido	Duración (s)	Rango frecuencial (Hz)
S1	0.1 - 0.12	20 - 150
S2	0.08 - 0.14	50 - 60
S3	0.04 - 0.05	20 - 50
S4	0.04 - 0.05	< 25

II.2.2 Soplos cardiacos

Los soplos cardiacos son vibraciones causadas por turbulencias que se forman en la sangre cuando esta fluye a través de un orificio estrecho o algún tubo. Un soplo es uno de los fenómenos anormales más comunes en el PCG y puede ser detectado con el estetoscopio. Los soplos pueden ser distinguidos de los sonidos básicos del corazón (S1, S2, S3, S4) porque son ruidosos y tienen una duración mayor. Mientras que los sonidos cardiacos básicos presentan frecuencias bajas las cuales están caracterizadas mayormente por debajo de los 200 Hz (Mahabuba *et al.*, 2009), (Debbal y Bereksi-Reguig, 2008), (Ewing, 1998); los soplos cardiacos presentan componentes de frecuencia más altos, extendiéndose estos hasta 1000 Hz (Mahabuba *et al.*, 2009).

II.2.3 Análisis espectral del fonocardiograma

De acuerdo con Djebbari y Bereksi Reguig (2000), la mayor parte de la energía de S1 se caracteriza entre los 100 y los 150 Hz, mientras que para S2 una parte importante de la energía se localiza entre los 50 y los 200 Hz. El segundo sonido cardiaco está distribuido de una manera más uniforme que el sonido S1, como se puede apreciar en las Figuras 5 y 6.

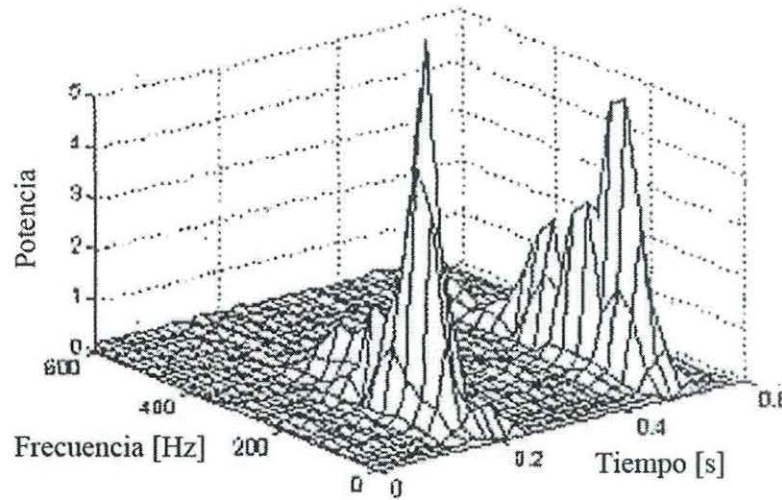


Figura 5. Espectrograma de un PCG, analizado con una ventana Hamming de 256 muestras (Djebbari y Bereksi Reguig (2000)).

Varias cardiopatías, entre ellas las que están asociadas con soplos cardiacos, ocasionan que en el PCG se presenten componentes de frecuencia mayores que las que son generadas por los cuatro sonidos básicos del corazón. Sin embargo, de acuerdo con Mahabuba *et al.* (2009), Debbal y Bereksi-Reguig (2008), Noponen *et al.* (2007) y Gauthier *et al.* (2007) el rango de frecuencia encontrado en dichos casos no excede los 1000 Hz.

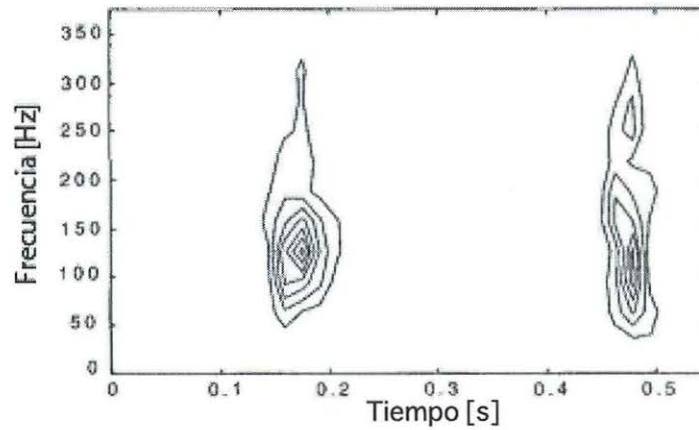


Figura 6. Diagrama de contorno del espectrograma de un PCG, analizado con una ventana Hamming de 256 muestras (Djebbari y Bereksi Reguig (2000)).

Análisis espectral de muestras de PCG

Fueron analizadas 127 muestras de audio cardiaco con el fin de obtener su espectro en frecuencia. Dichas muestras proceden de varias fuentes: los portales de internet de las compañías Thinklabs y Littmann; los portales de internet de la Universidad de Dundee y el Texas Heart Institute y un CD desarrollado por la compañía Littmann para fines educativos. Las muestras mencionadas son archivos de audio en formato digital de auscultaciones realizadas por médicos.

El análisis de dichas muestras fue realizado con el programa Matlab de la forma que a continuación se describe. Se calculó la transformada rápida de Fourier (FFT) a las muestras de la señal de audio para obtener su espectro de frecuencia. La transformada está dada por:

$$X(f) = \sum_{t=0}^{N-1} x(t)e^{-j2\pi ft} \quad (1)$$

Después se calculó la densidad espectral de energía:

$$\Psi_g(f) = |X(f)|^2 \quad (2)$$

Posteriormente se calculó el valor de la $\Psi_g(f)$ en decibeles normalizada a uno

$$\text{EDS [dB]} = 10\log\Psi_{g\text{norm}}(f) \quad (3)$$

Para todas las muestras se calculó la frecuencia bajo la cual se localiza el 99% de la densidad espectral de energía $\Psi_g(f)$. Se encontró que el resultado mayor fue 668.09 Hz y se obtuvo para la muestra 73, cuya fuente es la Universidad de Dundee y corresponde al PCG de un *frote pericárdico*³.

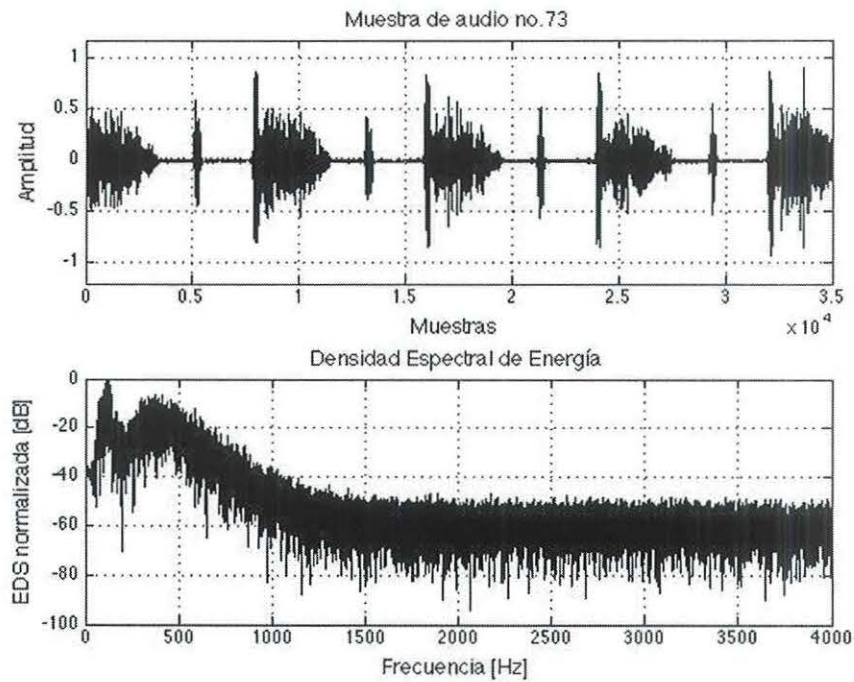


Figura 7. Densidad espectral de energía de un PCG de frote pericárdico.

Con base en este estudio se puede asegurar que la energía de las muestras cardíacas se encuentra por debajo de los 1000 Hz. En la Figura 7 se muestran las gráficas de la muestra 73. En la parte superior se presenta la señal en el dominio del tiempo y en la parte inferior se muestra la gráfica de la densidad espectral de energía en dB.

³Los frotos pericárdicos son una de las principales manifestaciones de pericarditis aguda.

En la Figura 8 se muestran las gráficas de una señal de audio de un PCG normal, de grupo de muestras de la misma fuente que la muestra 73.

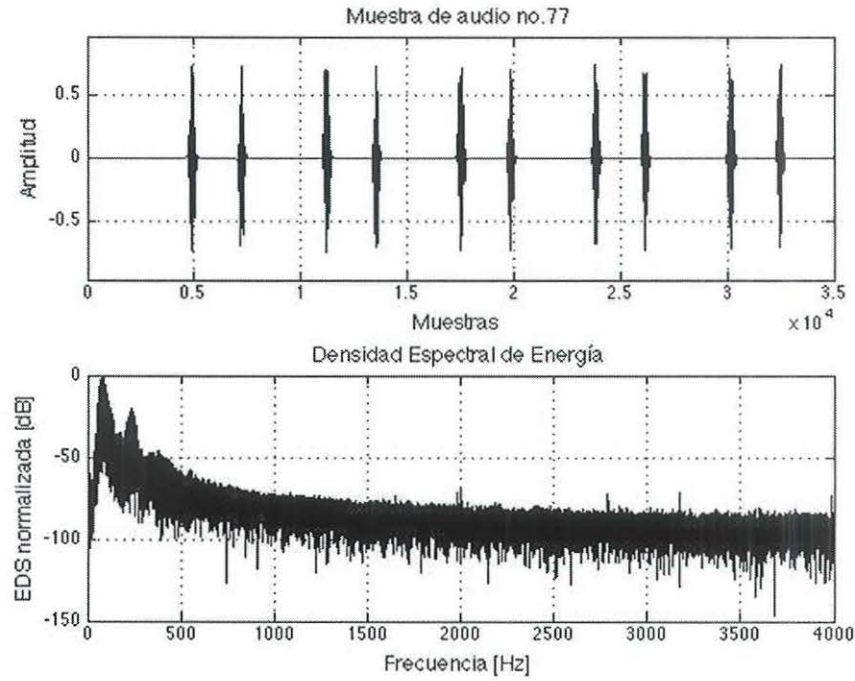


Figura 8. *Densidad espectral de energía de un PCG normal.*

II.3 Sonidos respiratorios

II.3.1 Origen de los ruidos respiratorios

El flujo de aire a través de las vías aéreas causa turbulencias, lo cual origina las vibraciones que percibimos como ruidos respiratorios. Las turbulencias se producen en las zonas donde la velocidad del aire es mayor y en aquellas con condiciones geométricas

que dificultan un flujo laminar ⁴. Esto ocurre principalmente en la laringe y en las bifurcaciones de los bronquios mayores, lobulares y segmentarios (Moreno Bolton, 2000).

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 (Kelley, 1993).

Los ruidos respiratorios se originan en los bronquios mayores, principalmente lobulares y segmentarios. Este ruido es transmitido a través de las vías aéreas tanto hacia la boca como hacia la periferia. El análisis de los ruidos registrados en la boca muestra una amplia variedad de frecuencias, mientras que los obtenidos en la pared torácica son principalmente de baja frecuencia (Quintana, 2004). Lo anterior se debe a que las vías aéreas normales, rodeadas de tejido pulmonar normal, actúan como un filtro que sólo permite transmitir las frecuencias bajas.

II.3.2 Ruidos respiratorios normales

Según Kelley (1993) a nivel de la tráquea, el sonido normal, denominado ruido traqueal, es de alta frecuencia y alta intensidad. El ruido corrientemente se ausculta durante toda la inspiración y toda la espiración, con una separación neta entre ambas fases, que se produce cuando el flujo disminuye y cae a cero.

A nivel periférico, por ejemplo en la cara lateral de la base del tórax, el sonido percibido, denominado murmullo pulmonar, es de baja frecuencia e intensidad, debido a la filtración del ruido. Normalmente se ausculta durante toda la inspiración y pasa casi sin pausa a la espiración, que se percibe usualmente sólo en su primera

⁴Se llama flujo laminar o corriente laminar, al movimiento de un fluido cuando éste es ordenado, estratificado, suave.

mitad, ya que al caer el flujo también cae la intensidad del ruido.

En las zonas paravertebrales y paraesternales el ruido tiene características intermedias entre el ruido traqueal y el murmullo pulmonar: de mayor intensidad y frecuencia que este último y tiene una duración más similar al ruido traqueal. Este ruido es denominado ruido traqueobronquial (Kelley, 1993).

II.3.3 Ruidos adventicios

El tejido pulmonar o la pleura anormales pueden generar ruidos anormales, mas no siempre es el caso, pues las crepitaciones, por ejemplo, también pueden auscultarse en individuos normales, como lo señala Moreno Bolton (2000) y se describe a continuación.

Crepitaciones. Son ruidos discontinuos, cortos, numerosos, de escasa intensidad, que ocurren generalmente durante la inspiración. Estos ruidos se producen cuando existen condiciones que favorecen el colapso de las pequeñas vías aéreas. Al producirse la inspiración en zonas con obstrucción completa de las vías aéreas, aumenta exageradamente la presión negativa del territorio alveolar distal a la obstrucción. Al continuar la inspiración, finalmente se produce una apertura de las vías aéreas, lo que determina una brusca entrada de aire, que hace vibrar el tejido pulmonar.

Frotes pleurales. Son ruidos discontinuos que se producen por el frote de las superficies pleurales inflamadas, cubiertas de exudado. En las zonas con derrame pleural este ruido no se percibe, ya que las pleuras están separadas por el líquido.

Sibilancias y roncus. Son ruidos continuos, generalmente múltiples, de alta frecuencia las primeras y de baja frecuencia los segundos, como silbidos o ronquidos, respectivamente. Se producen por obstrucción de las vías aéreas, debido a que la pared de éstas vibra al abrirse y cerrarse rápidamente debido al flujo de aire.

Respiración ruidosa. Normalmente, es casi imposible escuchar los ruidos respiratorios de una persona durante la respiración tranquila, a menos que el médico se acerque mucho a ésta. En los pacientes con obstrucción bronquial aumenta la intensidad de estos ruidos, por lo que con frecuencia es posible escucharlos, signo de gran sensibilidad y especificidad.

Cornaje. Es un fenómeno similar al anterior, pero que se produce por obstrucción de las vías aéreas superiores, laringe o tráquea, por lo que el ruido producido es generalmente de alta frecuencia.

II.3.4 Análisis espectral de los ruidos respiratorios

Se han realizado ya algunos estudios sobre el análisis espectral de los ruidos respiratorios, para hacer posible su procesamiento digital. De acuerdo con Vannuccinni *et al.* (2000), el ancho de banda utilizado más frecuentemente para grabar sonidos respiratorios va de los 60-100 Hz a los 4 KHz, cuando el sonido es proveniente de la tráquea. Para sonidos adventicios en el pecho, la banda utilizada para grabar es de 60-100 Hz a 6 KHz (Vannuccinni *et al.*, 2000). Es decir, no se espera encontrar ruidos traqueales con frecuencia superior a 2 KHz ni ruidos adventicios con frecuencias superiores a 3 KHz.

Análisis espectral de muestras de ruidos respiratorios

Fueron analizadas 9 muestras de audio respiratorio, correspondientes todas a pacientes con patologías. Estas muestras se obtuvieron de un CD con fines educativos desarrollado por la compañía Littman. Se hizo el mismo análisis realizado para las muestras de audio cardíaco. Se encontró, que para las muestras de audio respiratorio, la frecuencia mayor

en la que se encontró el 99% del EDS fue 902 Hz. En la Figura 9 se muestran las gráficas correspondientes a dicha muestra. Entonces, se puede asegurar que la información de las muestras de audio respiratorio estudiadas se encuentra por debajo de los 1000 Hz, según la literatura analizada se sabe que los ruidos respiratorios están caracterizados por debajo de los 3000 Hz.

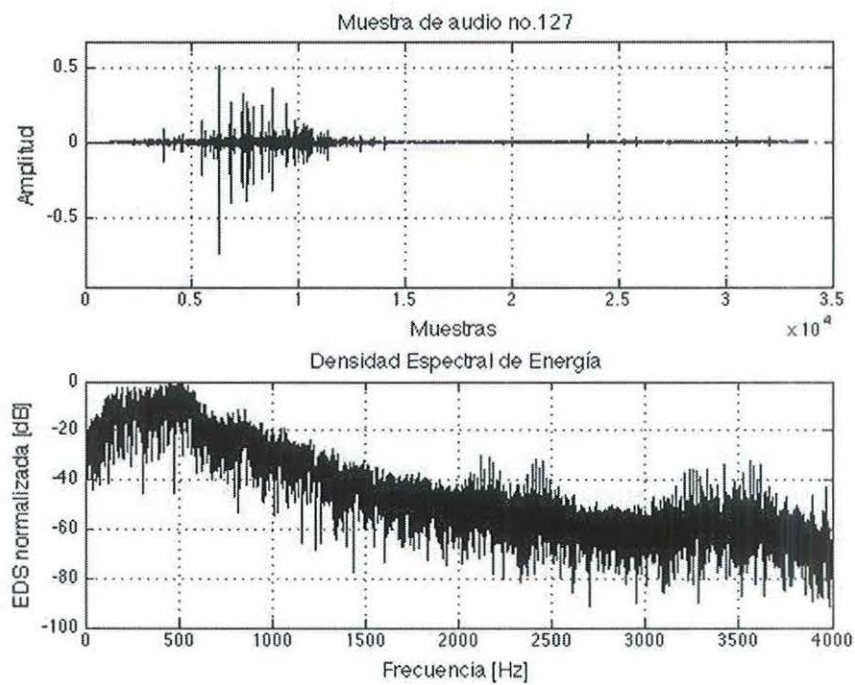


Figura 9. Densidad espectral de energía de un PCG de sonidos respiratorios.

Capítulo III

Introducción a esquemas de codificación de audio

III.1 Introducción

El sonido puede entenderse como una perturbación física en el aire u otro medio. Dicha perturbación da lugar a una serie de ondas producidas por variaciones en la presión del aire ocasionadas por el movimiento de determinados átomos o moléculas. Estas ondas se convierten en ondas mecánicas dentro del oído para ser posteriormente interpretadas por el cerebro.

No todas las personas perciben los sonidos de la misma forma, debido a que, para ello se involucran aspectos físicos y psicológicos (Roederer, 2009). Aunque la percepción auditiva sea distinta para cada persona, es un hecho que el oído humano percibe solamente sonidos que se encuentran dentro de un intervalo limitado de frecuencias. El intervalo de capacidad auditiva va inicialmente de los 20 Hz a los 20 KHz, pero tiende a ser más estrecho conforme una persona va envejeciendo, ya que se presenta una menor sensibilidad para las altas frecuencias.

En la Figura 10 se observan los rangos de frecuencia de distintas señales comparados con la capacidad auditiva del ser humano. Se aprecia que el rango de los sonidos cardiacos va desde las frecuencias bajas hasta aproximadamente 1000 Hz, de acuerdo con los estudios de Mahabuba *et al.* (2009), Debbal y Bereksi-Reguig (2008), Noponen *et al.* (2007) y Gauthier *et al.* (2007). Los sonidos pulmonares y traqueales se tienen

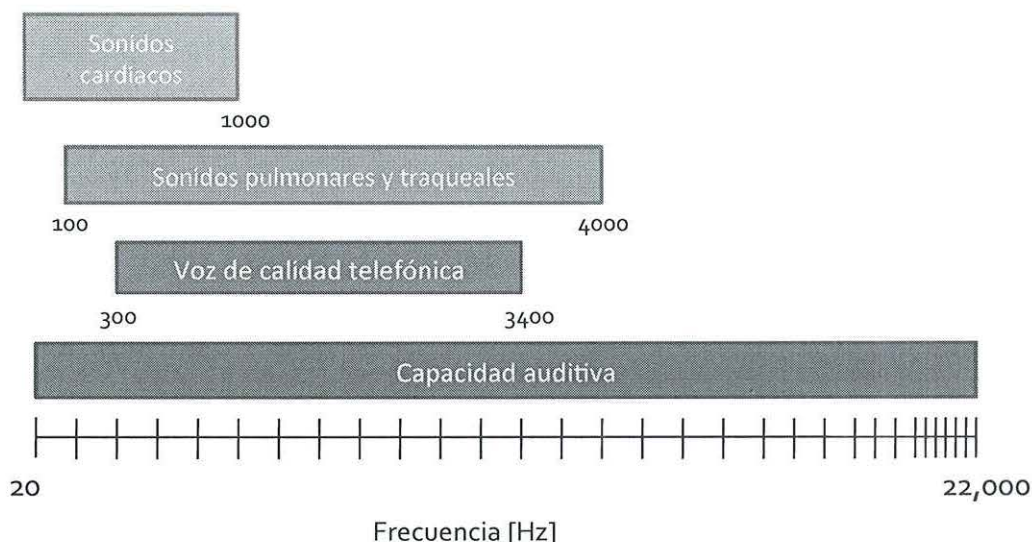


Figura 10. Intervalo de frecuencia de la capacidad auditiva del ser humano.

caracterizados desde las bajas frecuencias (aproximadamente 100 Hz) hasta los 4 KHz de acuerdo con los estudios de Vannuccinni *et al.* (2000). También se muestra el rango de frecuencia que ocupa la voz de calidad telefónica que va de los 300 a los 3400 Hz.

Existen varias características relacionadas con la precepción auditiva que son aprovechadas por los codecs. En este capítulo se pretende ofrecer una visión general acerca de la compresión de audio. En primer lugar se describen los fundamentos básicos de la misma, para posteriormente analizar los distintos algoritmos utilizados para realizar dicha compresión y así como las características principales de estos algoritmos.

III.2 Psicoacústica

La audición no es un fenómeno puramente mecánico de la propagación de una onda, sino que también es un evento sensorial y perceptual. Cuando escuchamos un sonido

podemos percibir distintas sensaciones. Por ejemplo, estas sensaciones nos permiten diferenciar entre sonidos graves y agudos. Percibimos también sensaciones de magnitud relacionadas con fuerza, volumen o intensidad sonora, donde nuestro oído puede diferenciar entre un sonido fuerte y uno débil. Es importante señalar que el volumen no solo depende de la presión del sonido, sino también de la frecuencia.

La psicoacústica es una rama de la psicofísica, y liga el estímulo acústico con las sensaciones auditivas (Roederer, 2009). Se puede ver también a la psicoacústica como la rama de la ciencia que estudia las respuestas psicológica y fisiológica asociadas con el sonido. La psicoacústica es aplicada en diversas áreas, tales como: acústica musical, acústica arquitectónica, acústica ambiental, música, electroacústica, medicina, etc.

Los parámetros psicoacústicos más relevantes son:

- **Sonoridad.-** Es la percepción subjetiva de la intensidad (amplitud) sonora.
- **Altura.-** Está ligada a la percepción del tono, en concreto, con la frecuencia fundamental de la señal sonora. Es lo que nos hace percibir un sonido como grave o agudo.
- **Timbre.-** Es la capacidad que nos permite diferenciar los sonidos. El timbre está caracterizado por la forma de la onda, es decir, por su componente armónico.

Se conoce como sensibilidad auditiva a la eficiencia de la respuesta en frecuencia que presenta el oído humano. Debido a esta sensibilidad, los parámetros anteriormente mencionados, en el contexto de la psicoacústica no son totalmente independientes, ya que los tres se influyen mutuamente. De hecho al modificar un parámetro, cambian los otros y con ello, la percepción del sonido también cambia.

III.2.1 Umbral absoluto de audición

El umbral absoluto de audición es la intensidad mínima que debe tener un tono puro para ser detectado por un oyente en un ambiente sin ruido (España Boquera, 2003). Genralmente se expresa en dB de NPS (decibeles de nivel de presión del sonido).

Este umbral depende de la frecuencia del tono aplicado. El valor de este umbral ha sido obtenido mediante estudios experimentales realizados a una amplia gama de oyentes. En la Figura 11 se muestra el resultado de dichos estudios.

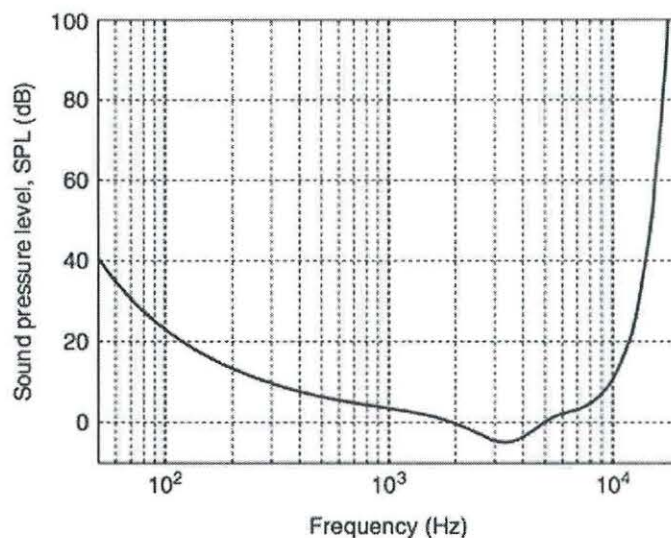


Figura 11. *Umbral absoluto de audición del ser humano (Spanias et al. (2007)).*

Se aprecia que la máxima sensibilidad auditiva se produce alrededor de 3-4 KHz, mientras que el umbral de audición crece de forma rápida a partir de 10-20 KHz.

III.2.2 Umbral de detección del ruido

Aplicado a la compresión de señales, el umbral de detección del ruido se interpreta como la máxima energía permitida a las distorsiones introducidas por la codificación (ruido

de cuantificación) (España Boquera, 2003).

Aunque el umbral absoluto del audición proporciona una primera orientación sobre la capacidad del oído para detectar señales, por sí solo, es de limitado valor en el contexto del ruido generado por la cuantificación. Debido a esto, se utiliza el umbral de detección del ruido. Las razones principales para ello son las siguientes:

1. Los umbrales absolutos están asociados a tonos puros, mientras que el ruido de cuantificación es de naturaleza espectral más compleja que la tonal.
2. El valor del umbral de detección para el ruido de cuantificación depende del estímulo de señal presente. Puesto que el estímulo es variable en el tiempo, también lo será el umbral de detección del ruido.

III.2.3 Enmascaramiento sonoro

Cuando el oído está expuesto a dos o más sonidos simultáneos, existe la posibilidad de que uno de ellos enmascare a los demás. Esto es debido a un efecto producido en la percepción sonora cuando se escuchan dos sonidos de diferente intensidad al mismo tiempo. Al suceder esto, el sonido más débil resultará inaudible, ya que el cerebro sólo procesará el sonido enmascarador. Por ejemplo, el sonido de nivel alto posee un efecto de enmascaramiento mayor si el suave tiene una frecuencia cercana.

Podemos dividir el enmascaramiento sonoro en dos tipos: temporal y frecuencial.

Enmascaramiento temporal.- Sucede cuando dos estímulos sonoros llegan a nuestro oído de forma cercana en el tiempo. El estímulo enmascarante hará que el otro resulte inaudible. En esta situación, el tono más intenso tiende a enmascarar al tono más débil. Según el instante de tiempo en que se produce el estímulo enmascarante respecto al instante en el que se produce el enmascarado, podremos distinguir entre

postenmascaramiento y preenmascaramiento (Figueiras Vidal y Rodríguez, 2002).

El efecto de enmascaramiento debido a determinado sonido enmascarador, precede a este y perdura tras su extinción. Es decir, el umbral absoluto de audición se eleva antes (preenmascaramiento), durante (enmascaramiento simultáneo) y después (postenmascaramiento) de que aparezca el sonido enmascarador. Mientras que el pre-enmascaramiento tiende a durar alrededor de 5 ms, el postenmascaramiento se extiende entre 50 y 300 ms, dependiendo de la fuerza y la duración del sonido enmascarador (Figueiras Vidal y Rodríguez, 2002). Este concepto se ilustra en la Figura 12.

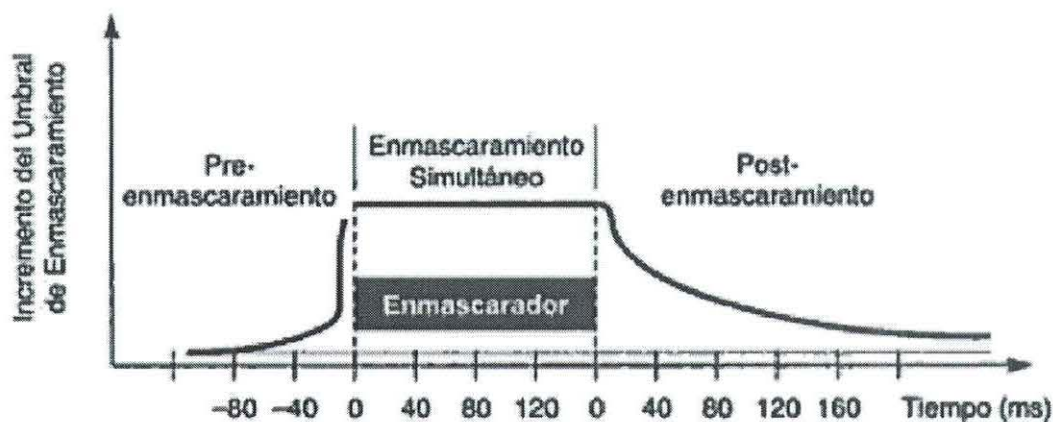


Figura 12. *Enmascaramiento temporal (Figueiras Vidal y Rodríguez, 2002).*

La explicación de esta “anticipación” se basa en que la información que llega a la corteza auditiva del cerebro humano se procesa por ráfagas. Asimismo, es sabido que los sonidos intensos son procesados de forma más rápida por el cerebro que los sonidos suaves, facilitando de esta forma el fenómeno psicoacústico del preenmascaramiento (Hartmann, 1997).

Enmascaramiento frecuencial.- Es la disminución de la sonoridad de un tono a una cierta frecuencia, en presencia de otro tono simultáneo a una frecuencia diferente. Es decir, cuando el oído es expuesto a dos o más sonidos de diversas frecuencias, existe la posibilidad que uno de ellos camufle a los demás y por tanto, que éstos no se oigan.

Este fenómeno perceptivo puede explicarse de manera simplificada considerando como varía la excitación de la membrana basilar del oído según la frecuencia. Esta membrana vibra, en función de la tonalidad, más cerca o más lejos de la ventana oval. Por dicha razón este fenómeno no es simétrico. Un tono grave enmascara a uno agudo con mayor facilidad (Hartmann, 1997).

III.2.4 Bandas críticas

Diversos estudios han demostrado que en las bajas frecuencias, los tonos con unos cuantos hertz de separación pueden ser distinguidos. Sin embargo, en las altas frecuencias para poder discriminar los tonos se necesita que estén separados por cientos de hertz (Hz). En cualquier caso, 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.

Una banda crítica es la banda máxima de frecuencias que al ser de ruido blanco, es capaz de enmascarar a un sonido puro con la misma intensidad, centrado a la banda del ruido (Salesa *et al.*, 2005). La banda crítica de frecuencias más bajas activa la partición del extremo apical o final de la membrana basilar y la banda de frecuencias más altas, la partición del extremo basal o inicial.

Cualquier banda crítica hasta la frecuencia de 500 Hz es aproximadamente de 100 Hz (Salesa *et al.*, 2005). Las bandas críticas desde 500 Hz hasta 16 KHz son aproxi-

madamente de un tercio de octava (aproximadamente 23% de la frecuencia central de la banda), por lo tanto el ancho de la banda crítica es progresivamente creciente al aumentar la frecuencia central.

Más importante aún es el hecho de que las bandas críticas no son fijas; son continuamente variables en frecuencia y cualquier tono audible creará una banda crítica centrada en él. Visto desde otro punto de vista, el concepto de la banda crítica es un fenómeno empírico: una banda crítica es el ancho de banda al cual las respuestas subjetivas cambian abruptamente.

III.3 Digitalización de datos

Una señal analógica es aquella cuya amplitud (que típicamente es la tensión de una señal que proviene de un transductor y amplificador) puede tomar en principio cualquier valor, es decir, su nivel en cualquier muestra no está limitado a un conjunto finito de niveles predefinidos. En cambio, una señal digital es aquella cuyas dimensiones (tiempo y amplitud) no son continuas sino discretas, lo que significa que la señal necesariamente ha de tomar unos determinados valores fijos predeterminados en momentos también discretos.

La digitalización o conversión analógica-digital del sonido (CAD) consiste en la transcripción de señales analógicas en señales digitales, con el propósito de facilitar su procesamiento (codificación, compresión, etc.) y lograr con esto que la señal resultante sea más inmune al ruido.

III.3.1 PCM

Para explicar el proceso de la digitalización del sonido, se describirá el funcionamiento de la Modulación por Impulsos Codificados (PCM, Pulse Code Modulation). PCM es la modulación utilizada típicamente para voz en sistemas cableados de telefonía.

Para realizar la conversión analógica-digital existen cuatro procesos que intervienen en PCM:

- **Muestreo:** el muestreo consiste en tomar muestras periódicas del valor instantáneo de una señal analógica. La velocidad con que se toma esta muestra, es decir, el número de muestras por segundo, es lo que se conoce como frecuencia de muestreo f_m . La condición que debe de cumplir la frecuencia de muestreo está descrita por el teorema de muestreo de Nyquist.

De acuerdo con este teorema, se tiene por ejemplo que las señales analógicas de frecuencia vocal⁵, se han de muestrear a una frecuencia igual o superior a 6800 Hz (2×3400). Sin embargo, en la práctica se suele tomar una frecuencia de muestreo $f_m = 8000$ Hz. Al haber sobremuestreo se pueden evitar algunas de las pérdidas por digitalización, ya que se tienen más muestras para la misma señal de audio. En la Figura 13 se presenta la forma que tiene una señal muestreada.

- **Retención:** las muestras tomadas han de ser retenidas por un circuito de retención, el tiempo suficiente para permitir evaluar su nivel. Desde el punto de vista matemático este proceso no se contempla, ya que se trata de un recurso técnico debido a limitaciones prácticas, y carece, por tanto, de modelo matemático.

⁵El rango de frecuencia de la voz llega hasta los 7 KHz, sin embargo el audio de calidad telefónica es tomado de 300 a 3400 Hz, debido a que en este rango de frecuencia es en donde hay mayores cambios en la intensidad, es decir, donde se encuentra la mayor parte de la información.

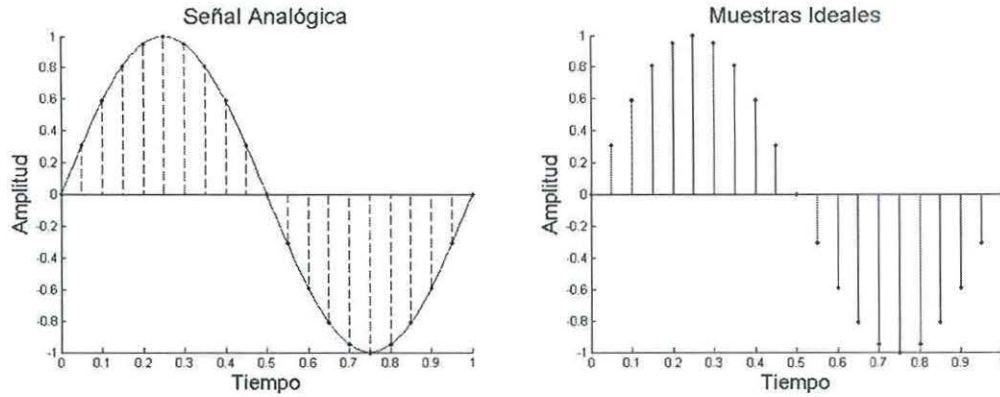


Figura 13. *Muestreo de una señal analógica.*

- **Cuantificación:** en el proceso de cuantificación se mide el nivel de voltaje de cada una de las muestras. Consiste en asignar un margen de valor de una señal analizada a un único nivel de salida. En el procesamiento digital de señales, la cuantificación es la discretización de un rango continuo de amplitudes por aproximación o truncamiento de valores. El resultado será un grupo más reducido de amplitudes discretas. En este proceso, las muestras toman el valor del nivel más próximo dentro del rango de niveles establecidos. Así, se define el paso de cuantificación como la diferencia que se tendrá entre dos niveles consecutivos como lo define la ecuación 4.

$$\Delta = \frac{MD}{2^{NB}} \quad (4)$$

Donde MD corresponde al margen dinámico de la señal (diferencia entre la amplitud máxima y la mínima) y NB representa el número de bits con el que se cuantifica. Existen diferentes métodos para establecer los niveles a los cuales se tendrán que ajustar las muestras. En general, los métodos más utilizados para hacer la cuantificación de las muestras son los siguientes:

1. *Cuantificación uniforme:* Se caracteriza por tener un paso de cuantificación uniformemente distribuido. El número de niveles en la salida se determinará por el número de bits de que se disponga.
2. *Cuantificación no uniforme:* Se caracteriza por disponer de mayor número de niveles para aquellas amplitudes de señal que son más probables, y viceversa. Se utiliza para las señales de voz, en donde los valores de las amplitudes pequeñas son los más probables. El método más común de PCM no uniforme es el denominado PCM logarítmico, definido por los algoritmos de ley μ en América y Japón, y ley A en Europa y el resto del mundo.
3. *Cuantificación vectorial:* La particularidad de este tipo de cuantificación radica, en que, en lugar de cuantificar las muestras retenidas individualmente, se cuantifican por bloques de muestras. Con ello, se logra una cuantificación más eficaz. Cada bloque de muestras será tratado como si se tratara de un vector; de ahí el nombre de este tipo de cuantificación. La cuantificación vectorial es la más eficiente de todas las modalidades de cuantificación en lo referente al error de cuantificación. No obstante, está más predispuesta a verse afectada por errores de transmisión. Otro inconveniente, es que los procesos informáticos para lograr esta codificación resultan muy complejos.

El proceso de cuantificación, incluso en su versión ideal, añade, como resultado, una señal indeseada a la señal de entrada: el ruido de cuantificación, el cual se detalla más adelante.

- **Codificación:** Es el proceso mediante el cual se representa una muestra cuantificada mediante código, generalmente el binario. Normalmente en telefonía se utilizan números binarios con 256 intervalos de cuantificación para representar todas

las posibles muestras, por tanto se necesitarán números binarios de 8 bits (PCM) para representar 256 valores ($2^8 = 256$). Otros codecs que usan ADPCM (Adaptive Differential Pulse Code Modulation) o cuantificación delta utilizan menos intervalos y por tanto menos bits.

El dispositivo que realiza la cuantificación y la codificación se llama codificador. La decodificación es el proceso mediante el cual se reconstruyen las muestras a partir de la señal numérica procedente de línea. Este proceso se realiza en un dispositivo denominado decodificador.

Durante el muestreo y la retención, la señal aún es analógica, puesto que aún puede tomar cualquier valor. No obstante, a partir de la cuantificación, cuando la señal ya toma valores finitos, se puede considerar que la señal ya es digital. En la Figura 14 se puede observar un diagrama a bloques del proceso de conversión A/D.

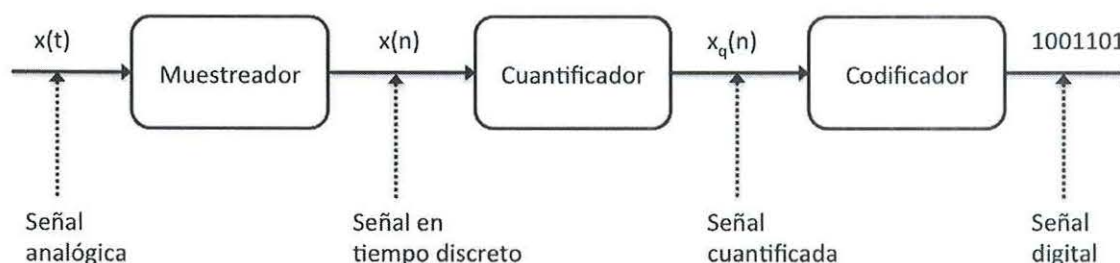


Figura 14. Diagrama a bloques de un convertidor analógico-digital.

Ruido de cuantificación

Se define como error de cuantificación a la señal en tiempo discreto y amplitud continua introducida por el proceso de cuantificación, que resulta de igualar los niveles de las muestras de amplitud continua a los niveles de cuantificación más próximos.

La diferencia que resulta de restar la señal de entrada del cuantificador a la de salida es el error de cuantificación, esto es, la medida en la que ha sido necesario cambiar el valor de una muestra para igualarlo a su nivel de cuantificación más próximo. Esta diferencia, entendida como una secuencia de muestras de tiempo discreto pero de amplitud continua, puede ser interpretado en la práctica como una señal indeseada añadida a la señal original, de modo que se cumple:

$$e_q(n) = x_q(n) - x(n) \quad (5)$$

donde:

$x(n)$ es la secuencia de muestras de amplitud continua a la entrada del cuantificador.

$x_q(n)$ es la secuencia de muestras de amplitud discreta a la salida del cuantificador.

$e_q(n)$ es la secuencia de muestras de amplitud continua del error de cuantificación.

La reconstrucción de las muestras originales de amplitud continua no es posible sólo a partir de las muestras cuantificadas, pues falta la información necesaria para distinguir el error de la señal analógica.

El error de cuantificación $e_q(n)$ queda dentro del rango de $-\Delta/2$ a $\Delta/2$ mientras que la señal analógica de entrada se encuentre dentro del rango del cuantificador:

$$\frac{-\Delta}{2} < e_q(n) < \frac{\Delta}{2} \quad (6)$$

donde Δ es el tamaño del escalón de cuantificación que viene dado por:

$$\Delta = \frac{R}{L} \quad (7)$$

donde R es el rango del cuantificador y L el número de niveles de cuantificación.

Teóricamente, la cuantificación de las señales analógicas resulta siempre en una pérdida de información, incluso en su caso ideal. De hecho, la cuantificación es un

proceso no reversible, dado que a todas las muestras a un intervalo inferior a $\Delta/2$ de un determinado nivel se les asignan el mismo valor. Sin embargo, discretizar una señal en su otra dimensión (el tiempo) mediante el proceso de muestreo, no es irreversible tal y como demuestra el teorema de muestreo y si se cumplen los criterios que impone el propio teorema debido a la naturaleza periódica y por tanto determinista de las señales que se someten a este proceso y a la limitación del ancho de banda. Dicho de otro modo, una onda periódica muestreada cumpliendo los criterios de Nyquist sólo puede comportarse de un único modo entre dos muestras contiguas y este comportamiento es totalmente deducible a partir de la serie completa de muestras de amplitud continua de la señal. La cuantificación es, por tanto, el único proceso que introduce un error teórico sobre la señal original en todo el procedimiento completo de digitalización de una señal.

El error de cuantificación es expresado también como un valor promedio de la potencia del ruido en relación con la potencia promedio de la señal. Por lo tanto la tasa de señal a ruido de cuantificación SQR (Signal-to-quantizing noise ratio) puede ser determinada como:

$$SQR = \frac{E\{x^2(t)\}}{E\{[y(t) - x(t)]^2\}} \quad (8)$$

Donde:

$E\{.\}$ es el promedio o valor esperado

$x(t)$ es la señal análoga de entrada

$y(t)$ es la señal decodificada de salida

Si suponemos un nivel de resistencia de 1Ω , la potencia promedio del ruido de cuantificación queda determinada por:

$$P_{\text{ruido de cuantificación}} = \frac{q^2}{12} \quad (9)$$

donde q es la altura del intervalo de cuantificación.

Si se considera una cuantificación uniforme (los niveles tienen longitudes iguales), el ruido de cuantificación es independiente del valor de las muestras y la tasa de señal a ruido de cuantificación se determina como:

$$\begin{aligned} SQR(dB) &= 10\log_{10} \left(\frac{v^2}{q^2/12} \right) \\ &= 10.8 + 20\log_{10} \left(\frac{v}{q} \right) \end{aligned} \quad (10)$$

donde v es el valor rms de la amplitud de la señal de entrada. Para el caso de una señal senoidal como entrada, la tasa de señal a ruido de cuantificación producida por una cuantificación uniforme es:

$$\begin{aligned} SQR(dB) &= 10\log_{10} \left(\frac{A^2/2}{q^2/12} \right) \\ &= 7.78 + 20\log_{10} \left(\frac{A}{q} \right) \end{aligned} \quad (11)$$

Para un sistema que utiliza PCM con cuantificación uniforme de n bits se tiene que:

$$q = \frac{2A_{max}}{2^n} \quad (12)$$

donde A_{max} es la amplitud máxima.

Sustituyendo la ecuación 12 en la ecuación 11 tenemos que la ecuación que describe el desempeño de una codificación PCM uniforme es:

$$SQR = 1.76 + 6.02n + 20\log_{10} \left(\frac{A}{A_{max}} \right) \quad (13)$$

Los primeros dos términos de la ecuación 13 corresponden a la SQR que se obtiene al codificar una señal senoidal de rango completo. El último término de la ecuación indica una pérdida en la SQR cuando se codifica una señal de nivel bajo.

Con la ecuación 13 podemos calcular la tasa de señal a ruido de cuantificación para PCM con cuantificación uniforme de 8 bits:

$$SQR = 1.76 + 6.02n \quad (14)$$

$$= 1.76 + 6.02 \times (8) = 49.92dB \quad (15)$$

Y para PCM con cuantificación uniforme de 16 bits de precisión es:

$$SQR = 1.76 + 6.02n \quad (16)$$

$$= 1.76 + 6.02 \times (16) = 98.08dB \quad (17)$$

Del cálculo anterior se concluye que la diferencia entre realizar una cuantificación con 8 y 16 bits es muy grande. Es preferible por lo tanto utilizar la mayor cantidad posible de bits de cuantificación que un sistema nos permita.

III.4 Sistemas de codificación de audio

La compresión consiste en la reducción de la cantidad de datos a transmitir o grabar, pues hay que tener en cuenta que la capacidad de almacenamiento de los soportes es finita, de igual modo que los sistemas de transmisión tienen un límite en la tasa de datos. Existen dos tipos de compresión:

- **Compresión sin pérdidas:** en esencia se transmite toda la información, eliminando la información repetida y agrupando la información que no se repite para que ocupe un menor espacio.
- **Compresión con pérdidas:** se desprecia cierta información considerada irrelevante. Este tipo de compresión puede producir pérdida de calidad en el resultado final.

Las técnicas de compresión sin pérdidas se basan en algoritmos matemáticos que permiten la reducción de los bits que es necesario almacenar o transmitir. Como por ejemplo la llamada codificación de longitud de secuencias, muy utilizada en las técnicas de transmisión digital, mediante la cual se sustituye las secuencias de bits repetidos por la codificación de la longitud de la secuencia.

También es utilizada la codificación relativa o incremental que codifica las diferencias entre dos valores consecutivos, en vez de los valores absolutos. Es decir, para representar el valor absoluto de una muestra de una señal con un gran valor dinámico se necesita un elevado número de bits, pero si la señal no tienen transitorios muy bruscos, se puede codificar el rango de la diferencia entre dos muestras consecutivas con un número menor de bits. Un ejemplo de este tipo de codificación es DPCM (Differential Pulse Code Modulation), en donde se comparan dos muestras consecutivas y solo se codifica la diferencia entre ellas. Las técnicas de codificación mencionadas son de gran utilización en los sistemas de transmisión digital. Sin embargo, en lo que se refiere al tratamiento digital de imagen y sonido, dada la aleatoriedad de este tipo de señales, son poco efectivos en cuanto a la reducción del tamaño de los archivos resultantes.

Por eso, los algoritmos para compresión de sonido e imagen que muestran un mejor compromiso entre la reducción del tamaño de archivo y la calidad de la señal, se basan principalmente en el conocimiento del funcionamiento de nuestros sentidos. Son técnicas que asumen pérdidas de información, de ahí su nombre de compresión con pérdidas, pero están diseñados de modo que las pérdidas no sean perceptibles por la vista y el oído humano.

III.4.1 Códec de audio

Un códec de audio es un sistema que incluye un conjunto de algoritmos que permiten codificar y decodificar los datos auditivos, lo cual significa reducir la cantidad de bits que ocupa el archivo de audio. Sirve para comprimir señales o archivos de audio con un flujo de datos (stream) que ocupe el menor espacio posible, consiguiendo una calidad final aceptable, para luego descomprimirlos y hacer posible su reproducción o manipulación. Se implementa en software, hardware o una combinación de ambos.

Los códecs de audio se caracterizan por los siguientes parámetros:

- **Número de canales.** Se refiere al número de señales de audio simultáneos que contiene el flujo de datos. Dependiendo de la señal de audio, puede ser mono (un canal), estéreo (dos canales) o multicanal 5.1 (6 canales) ó 7.1 (8 canales).
- **Frecuencia de muestreo.** De los trabajos de Nyquist, Kotelnikov y Shannon, se sabe que la selección de la frecuencia de muestreo está directamente relacionada con el ancho de banda de la señal: si la máxima frecuencia presente en $x(t)$ es f , basta con emplear una frecuencia de muestreo mayor que $2f$. En la práctica también entran en juego algunas consideraciones preceptuales, que son las que determinan el mínimo número de muestras que han de tomarse para preservar determinado nivel de calidad. Por ejemplo, para codificar sonido con calidad CD nunca se usan frecuencias de muestreo superiores a 44.1 kHz, ya que el oído humano no es capaz de escuchar frecuencias superiores a 22 kHz.
- **Pérdida de información.** Algunos codecs pueden eliminar frecuencias de la señal original que teóricamente, son inaudibles para el ser humano. De esta ma-

nera se puede reducir la frecuencia de muestreo.

- **Número de bits por muestra.** Determina la precisión con la que se reproduce la señal original y el rango dinámico de la misma. Se suelen utilizar 8 (para un rango dinámico de hasta 45 dB), 16 (para un rango dinámico de hasta 90 dB como el formato CD) o 24 bits por muestra (para 109 a 120 dB de rango dinámico). El más común es 16 bits.
- **Tipo de compresión.** Distingue entre compresión con pérdidas (lossy) y sin pérdidas (lossless).
- **Tasa de bits.** Determina el número de bits de información necesarios por unidad de tiempo. La tasa de bits no se puede deducir de los parámetros anteriores puesto que la compresión puede ser con pérdidas o sin pérdidas. Además, puede ser constante (CBR, Constant Bit Rate), variable (VBR, Variable Bit Rate) o promedio (ABR, Average Bit Rate). En audio, es frecuente el uso de una tasa de bits variable, puesto que es más eficiente que CBR cuando hay silencios o segmentos donde la complejidad es baja.
 - **Tasa de bits constante (CBR).** La codificación CBR implica que la tasa de salida del codificador es constante. CBR es muy útil para flujo de datos multimedia cuando los canales son de capacidad limitada. Sin embargo, CBR no es la mejor opción para el almacen ya que no asigna suficientes bits para las secciones complejas (resultantes en la degradación de la calidad) y

por el contrario usará bits innecesarios en secciones de menor complejidad.

- **Tasa de bits variable (VBR).** Este método de compresión consigue una mayor calidad de sonido para un tamaño de archivo determinado, en contraste con CBR. Con el método de VBR se otorga la tasa de bits necesaria a cada parte del archivo que se pretende codificar, consiguiendo una calidad mayor en archivos de un tamaño reducido. Si los archivos son codificados en VBR, el codificador asigna tasas de bits que varían según la complejidad de la onda de audio a lo largo del archivo.

Actualmente existe una gran variedad de codecs de audio. Aunque cada vez son más complejos y se añaden características adicionales, se pueden clasificar en cuatro grandes grupos dependiendo del método de codificación o reducción de la tasa de bits:

- Codificadores perceptuales.
- Codificadores paramétricos.
- Codificadores de forma de onda.
- Codificadores híbridos.

Para la mayoría de los códecs se utiliza como entrada el audio ya digitalizado en formato PCM. Se tiene entonces que, de manera general, la codificación de audio puede ser representada como se muestra en el diagrama de la Figura 15.

Para voz en formato PCM, por ejemplo se utilizan 8KHz de muestreo y 8 bits por muestra, lo que da como resultado una tasa de bits de 64 kbps (8KHz x 8 bits por muestra).

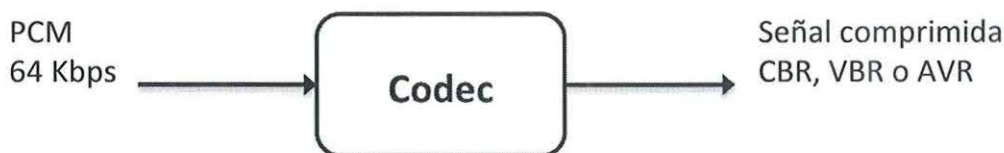


Figura 15. *Diagrama general de un codec.*

III.4.2 Codificadores de forma de onda

Los codificadores de forma de onda se basan en el estudio de la señal, de forma que intentan reproducir la forma de la señal de entrada. Generalmente se diseñan para ser independientes de la señal, de manera que se utilizan para codificar una gran variedad de señales. Estos codificadores aprovechan la redundancia de la señal, y a partir de una predicción lineal realizan la codificación; de esta forma se consiguen tasas de compresión elevadas cuando las señales son muy redundantes y prácticamente nulas cuando no es así.

Algunos codificadores de forma de onda son:

- Modulación por impulsos codificados (PCM, Pulse Code Modulation). Este tipo de codificación, como se mencionó anteriormente, es la base para la mayoría de los codificadores y está basado a su vez en el procedimiento para la digitalización de las señales.
- Modulación diferencial de impulsos codificados (DPCM, Differential Pulse Code Modulation). Este tipo de codificación está basado en el principio de PCM, pero en vez de codificar el valor completo de las muestras se codifica solamente la diferencia entre la muestra presente y la muestra anterior. Esto es basado en que para las señales de voz hablada existe una variación mínima de una muestra a

otra. DPCM permite enviar una mayor cantidad de información con un número más reducido de bits si se compara con PCM. Sin embargo, la desventaja de este tipo de codificación es que si existe un error en una trama codificada con DPCM, este error afectará no solo a la muestra errónea sino también a las subsecuentes, pues ellas dependen de dicha muestra para ser decodificadas.

- Modulación diferencial adaptativa de impulsos codificados (ADPCM, Adaptive Pulse Code Modulation). Es una variante de DPCM, en la cual el paso de cuantificación es variable. La variación en el paso de cuantificación permite la reducción del ancho de banda para una tasa de señal a ruido dada.

III.4.3 Codificadores perceptuales

Los codificadores perceptuales aprovechan las limitaciones en la percepción del sistema auditivo humano para codificar el flujo de datos. En la Figura 16 se puede observar el diagrama a bloques de un codificador perceptual.

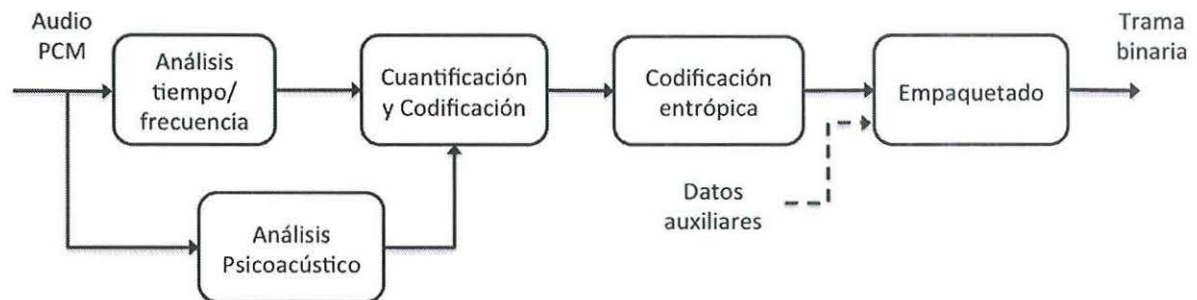


Figura 16. Diagrama a bloques de un codificador perceptual.

Para codificar las muestras, que inicialmente se encuentran en formato PCM, se realiza una transformada al dominio de la frecuencia. En esta etapa, la señal puede ser

descompuesta en sub-bandas. En MP3 se utiliza un banco de filtros para descomponer la señal en sub-bandas. El banco de filtros utilizado en esta capa es el llamado banco de filtros híbrido polifase/MDCT. Este banco de filtros se encarga de realizar la conversión del dominio del tiempo al de la frecuencia tanto para el codificador como para los filtros de reconstrucción del decodificador. Las muestras de salida del banco están cuantificadas y proporcionan una resolución en frecuencia variable, 6×32 ó 18×32 subbandas, ajustándose mucho mejor a las bandas críticas de las diferentes frecuencias. Usando 18 puntos, el número máximo de componentes frecuenciales es: $32 \times 18 = 576$. Dando lugar a una resolución frecuencial de: $24000/576 = 41,67$ Hz (si $f_s = 48$ kHz.). Si se usan 6 líneas de frecuencia la resolución frecuencial es menor, pero la temporal es mayor, y se aplica en aquellas zonas en las que se espera efectos de pre-eco (transiciones bruscas de silencio a altos niveles energéticos).

Después de la etapa de análisis, se cuantifican y codifican las muestras a partir de un conjunto de datos obtenidos de un modelo psicoacústico. La compresión se basa en la reducción del margen dinámico irrelevante, es decir, en la incapacidad del sistema auditivo para detectar los errores de cuantificación en condiciones de enmascaramiento. La implementación de este modelo psicoacústico es lo que determina la calidad final y es dónde se aprovecha el enmascaramiento temporal o frecuencial, siendo el umbral de enmascaramiento el que controla la cuantificación.

En la cuantificación de audio se utilizan distintos tipos de cuantificadores: uniforme, no-uniforme, logarítmico (ley μ o ley A), diferencial o vectorial. Generalmente se utiliza cuantificación no-uniforme para música, cuantificación diferencial o logarítmica para señales de voz y cuantificación vectorial para procesado avanzado del señal de voz.

A la salida del bloque de cuantificación y codificación se puede incorporar un codificador entrópico que comprime sin pérdidas para reducir el tamaño de la trama binaria.

Un algoritmo utilizado comúnmente para este fin es el de codificación Huffman.

Después de hacer la codificación entrópica, la trama de datos se empaqueta junto con información adicional que contiene datos imprescindibles para la decodificación como el tamaño de la trama, el número de bits de codificación, el margen dinámico, entre otros. También se puede añadir corrección de errores.

El decodificador desempaqueta la señal codificada para obtener las muestras y la información auxiliar. Posteriormente, decodifica las muestras y las convierte en el dominio temporal para obtener audio en formato PCM, aunque éste no será igual al inicial, puesto que es un proceso de codificación con pérdidas.

Codificadores en el dominio frecuencial:

- Codificación en sub-bandas.- La banda de la señal se divide en sub-bandas de frecuencia, usando filtros pasa banda. La salida de cada filtro se muestrea (o submuestrea) y se codifica independientemente. Adicionalmente, este codificador puede hacer uso de las propiedades del oído y asignar un mayor número de bits a las bandas de frecuencia donde este sea más sensible (bandas de baja frecuencia) (España Boquera, 2003).
- Codificación por transformada.- En este tipo de codificador se codifican las componentes de una transformada unitaria de la señal y el decodificador realiza la transformada inversa. El potencial de reducción del régimen binario estriba en el hecho de que las transformadas unitarias tienden a generar componentes casi incorreladas entre sí, que pueden codificarse de forma independiente, mientras que la variación de una misma componente en el tiempo presenta ciertas redundancias (España Boquera, 2003).

Las distintas versiones de MPEG-1, que es un estándar de codificación de audio

y video desarrollado por el Grupo de Expertos en Imágenes Móviles (Moving Picture Experts Group), utilizan este método de codificación.

III.4.4 Codificadores paramétricos

Los codificadores paramétricos se basan en que el audio y la voz se pueden representar y sintetizar con tonos aislados, patrones armónicos (representados con señales senoidales) y componentes ruidosas. Estos, se representan con parámetros como la amplitud, la frecuencia fundamental o los componentes espectrales y requieren pocos bits para representarlos.

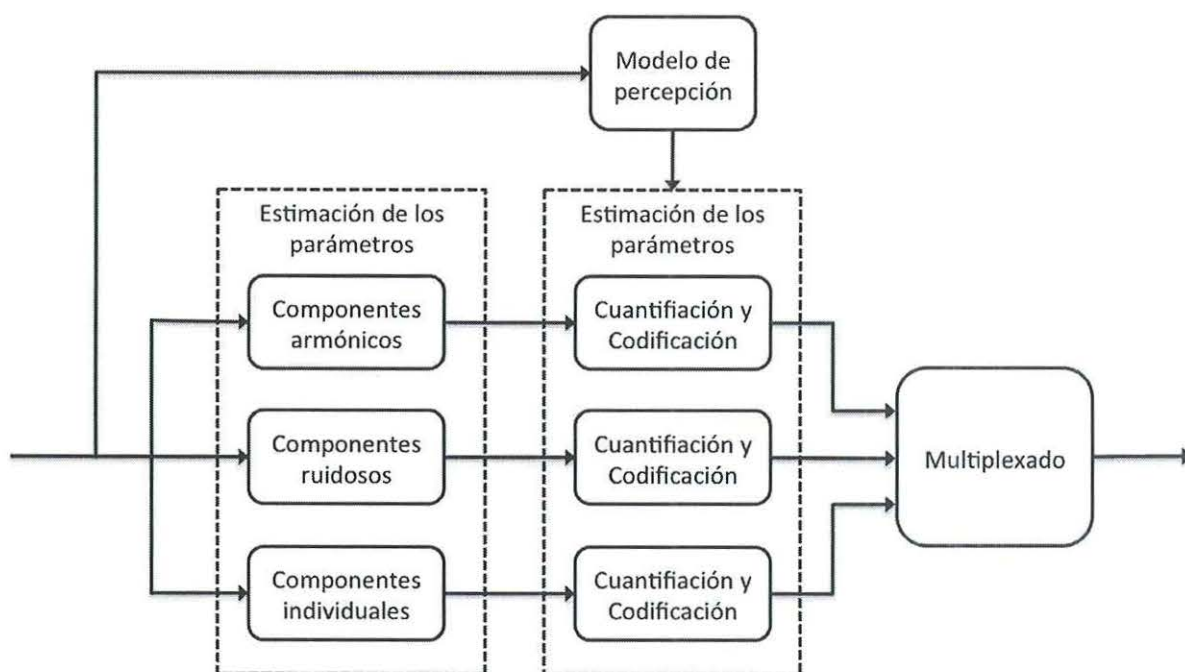


Figura 17. Diagrama a bloques de un codificador paramétrico.

El esquema más simple basado en un modelo paramétrico es el codificador por predicción lineal (LPC). En la etapa de análisis, se obtienen los parámetros representa-

tivos del modelo de producción del habla, que consiste en dos tipos de informaciones: referente a los coeficientes del filtro predictor y la correspondiente señal de excitación.

Coeficientes de predicción

Se calcula un predictor lineal de, por ejemplo, orden $p = 10$ a intervalos regulares de unos 25 ms (entre 15 y 40 ms)(Faúndez Zanuy, 2000). Este predictor proporciona, para cada segmento analizado, el conjunto de coeficientes del filtro así como la potencia del error de predicción $\sigma_{e_p}^2$.

Señal de excitación

En el esquema LPC no se realiza una codificación de forma de onda muestra a muestra, sino que se realiza una simplificación. Primeramente, se clasifican las tramas en dos grandes tipos, según presenten sonoridad o no. En los segmentos sonoros se obtiene el valor de la frecuencia fundamental de vibración de las cuerdas vocales, o valor del pitch. Dado que la fisiología del aparato fonador humano establece unos límites superior e inferior del valor del pitch (entre 50 y 800 Hz equivalente a un margen de 10 a 160 muestras a 8KHz), se asigna un valor del pitch nulo a las tramas etiquetadas como sordas. De este modo, se establece un conjunto de 12 parámetros reales cada 25 ms que deberán ser transmitidos al sistema receptor. Para que su representación sea eficiente, deberán ser cuantificados y codificados de forma conveniente (Faúndez Zanuy, 2000).

Otros ejemplos de codificación paramétrica son HVXC (Harmonic Vector Excitation Coding) para voz y HILN (Harmonic and Individual Lines plus Noise) para música, ambos dentro del estándar MPEG-4.

Codificadores de voz

Los codificadores específicos para voz, también conocidos como vocoders, se basan en modelos de generación, y originan, por tanto, representaciones estructuradas. Mediante estos modelos se trata de reproducir la voz de manera inteligible para el oído, sin que necesariamente la forma de onda se ajuste a la de la señal original. Estos modelos tienen la ventaja de que pueden funcionar a tasas de datos muy bajas, pero tienden a producir voz de calidad sintética y su desempeño para señales de entrada distintas de la voz es muy pobre.

Tabla II. *Principales recomendaciones de la ITU para la codificación de voz.*

	Tipo de código	Tasa de bits	Tam. trama/ Tiempo adi.	Complejidad
G.711	PCM (ley A y ley μ)	64 kbps	0 / 0 ms	0 MIPS
G.722	Banda ancha (7KHz)	64, 56 ó 48 kbps	0.125 / 0 ms	5 MIPS
G.726	ADPCM	40, 32, 24 ó 16 kbps	0.125 / 0 ms	2 MIPS
G.727	ADPCM	40, 32, 24 ó 16 kbps	0.125 / 0 ms	2 MIPS
G.728	LD-CELP ⁶	16 kbps	0.625 / 0 ms	20 MIPS
G.729	CS-ACELP ⁷	8 kbps	10 / 5 ms	20 MIPS
G.729 A	CS-ACELP	8 kbps	10 / 5 ms	11 MIPS
G.723.1	MP-MLP ⁸	6, 4, 5, 3, kbps	30 / 7.5 ms	16 MIPS

⁶Low Delay Code Excited Linear Prediction Coder.

⁷Conjugate Structure Algebraic Code Excited Linear Prediction Coder.

⁸Multipulse Excitation with a Maximum Likelihood Quantizer.

En la Tabla II (España Boquera, 2003) se pueden observar las características de las principales recomendaciones de la ITU para la codificación de voz. En la Figura 18 se presenta un diagrama a bloques para un codificador de voz.

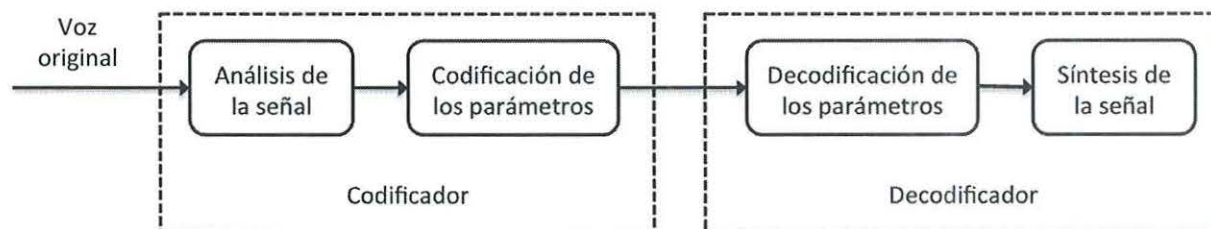


Figura 18. *Diagrama a bloques de un codificador de voz.*

III.4.5 Codificadores híbridos

Los codificadores híbridos, también conocidos como codificadores de análisis-por-síntesis combinan las técnicas de los codificadores de forma de onda con los vocoders. El objetivo de estos es obtener voz de alta calidad con tasas de bit bajas (inferiores a 8kHz).

El funcionamiento de los códecs híbridos está basado en el análisis de un conjunto de muestras como si se tratase de una sola para obtener los parámetros de la señal. Al decodificar la trama, se sintetizan los parámetros para conseguir que se parezca al original. Algunos codificadores híbridos son:

- Codificador RELP (Residual-Excited Linear Prediction).
- Codificador CELP (Code Excited Linear Prediction).
- Codificador VSELP (Vector Sum Excited Linear Prediction).
- Codificador RPE-LTP (Regular Pulse Excitation - Long Term Prediction).

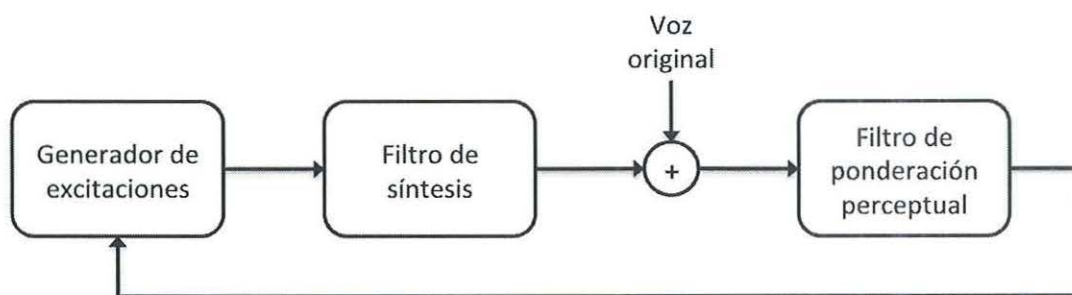


Figura 19. *Diagrama a bloques de un codificador híbrido de voz.*

En este capítulo se hizo una descripción del principio de la codificación de audio, así como de la clasificación de los codificadores de audio. Adicionalmente, en el Capítulo V se hará una descripción más detallada de los códecs evaluados en el presente proyecto de tesis.

Capítulo IV

Tecnologías de redes de área personal

IV.1 Introducción

La tendencia a la movilidad y la ubicuidad han propiciado que cada vez sean más utilizados los sistemas inalámbricos. Uno de los objetivos en el diseño actual de sistemas de comunicación y de electrónica de consumo⁹ es eliminar la conexión mediante cables.

La tendencia general de las tecnologías inalámbricas anteriormente había sido tener mayores tasas de transmisión y calidad de servicio (QoS, por sus siglas en inglés), como principales métricas de desempeño. El costo y el consumo de potencia eran considerados como características secundarias. Con la nueva tendencia a eliminar todo tipo de cableado, han cobrado una mayor relevancia las redes inalámbricas de área personal (WPAN, Wireless Personal Area Networks). Las WPAN proporcionan tasas de datos altas, bajas y medianas.

Entre las ventajas de las redes que ofrecen tasas bajas y medianas, se pueden destacar la menor complejidad de los dispositivos y por consecuencia, menor costo. Otra característica importante de este tipo de redes es la baja potencia que se requiere para la comunicación entre los dispositivos de la red.

Como es de suponerse, existe un compromiso entre las prestaciones que pueden brindar redes inalámbricas y la complejidad de las tecnologías que las hacen posibles. En la Figura 20 se puede observar dicho compromiso.

⁹Se dice de todos los equipos eléctricos utilizados cotidianamente, generalmente enfocados al entretenimiento, la comunicación y la oficina.

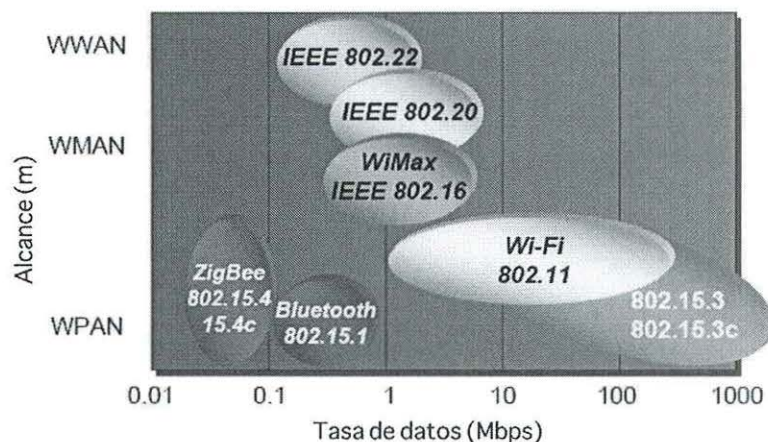


Figura 20. Relación entre tasa de datos y alcance de algunas tecnologías inalámbricas.

La estandarización de las redes inalámbricas ha estado a cargo de la IEEE (Institute of Electrical and Electronics Engineers) que es una asociación técnico-profesional mundial sin fines de lucro dedicada a la estandarización. Dentro de esta asociación existen grupos de trabajo enfocados a tareas específicas y denominados por un número. En lo correspondiente a las redes inalámbricas, el grupo 802.11 se enfoca al desarrollo de estándares para las redes de área local (WLAN) y el grupo 802.15 está orientado a la estandarización de las WPAN.

El grupo de trabajo IEEE 802.15 ha definido 3 clases de WPAN, diferenciadas por tasa de datos, consumo de batería, y QoS.

- High-data rate WPAN (IEEE 802.15.3) para aplicaciones multimedia que requieren una alta QoS.
- Medium-rate WPAN (IEEE 802.15.1), como reemplazo de cables para dispositivos electrónicos de consumo centrados en teléfonos móviles y PDAs con QoS adecuada

para aplicaciones de voz.

- Low-rate WPAN(IEEE 802.15.4) fueron ideadas para servir a aplicaciones que requieran bajo consumo y costos que no son ofrecidos por otras WPANs.

Se prevee que en los próximos años, las redes inalámbricas de área personal con bajas tasas de datos (LR-WPAN, Low Rate Wireless Personal Area Networks) serán utilizadas en una amplia variedad aplicaciones embebidas, tales como automatización del hogar, control y sensado industrial, monitoreo ambiental y monitoreo de signos vitales de deportistas y personas de la tercera edad.

En este capítulo se abordan los estándares más prominentes para WPAN: IEEE 802.15.1 (Bluetooth) e IEEE 802.15.4. Además se describe el escenario de aplicación hacia el que está enfocado el presente proyecto de investigación.

IV.2 Bluetooth

Bluetooth es un estándar para redes inalámbricas de área personal. Este estándar fue hecho para conectar dispositivos entre sí, sin cables de por medio, dentro de una red de corto alcance.

La tecnología Bluetooth originalmente fue desarrollada por la compañía Ericsson en 1994. Sin embargo, en febrero de 1998, se formó un grupo llamado Bluetooth Special Interest Group (SIG) con más de 200 compañías del área de las telecomunicaciones, tales como Ericsson, IBM, Intel, Microsoft y Motorola. En el año 2010 el SIG estaba formado por 13,000 compañías de las áreas de telecomunicaciones, informática, automovilismo, música, textil, automatización industrial y tecnologías de red¹⁰.

¹⁰<http://www.bluetooth.com/Pages/History-of-Bluetooth.aspx>.

El objetivo de Bluetooth es transmitir voz o datos entre equipos mediante circuitos de radio de bajo costo, a través de un rango aproximado de entre diez y cien metros, utilizando poca energía. Los dispositivos Bluetooth operan en la banda ISM¹¹ de 2.4 GHz. Dentro de esta banda de transmisión el radio Bluetooth realiza un esparcimiento por salto de frecuencia en 79 saltos desplazados por 1 MHz, en el intervalo de frecuencia que va de 2.402 GHz a 2.480 GHz.

Para cumplir con la reglamentación de cada país, en ambos sistemas se utiliza una guarda en los límites superior e inferior de la banda de transmisión. En la Tabla III se presentan algunas de las frecuencias de operación del estándar Bluetooth en diversas regiones del mundo.

Tabla III. *Frecuencias de operación de Bluetooth en algunas regiones.*

País	Rango de frecuencia (MHz)	Canales RF (MHz)	
Europa* y EEUU	2400 - 2483.5	$f = 2402 + k$	$k = 0, \dots, 78$
Francia	2446.5 - 2483.5	$f = 2454 + k$	$k = 0, \dots, 22$
España	2445 - 2475	$f = 2449 + k$	$k = 0, \dots, 22$

Existen varias versiones de la especificación Bluetooth. La versión 1.0 ofrece una tasa pico menor a 1 Mbps; mientras que la versión 2.0 ofrece tasas pico de hasta 3 Mbps. El rango de alcance depende de la clase del equipo utilizado:

- Clase 1: diseñada para dispositivos de largo alcance (rango de hasta 100 m) tales como puntos de acceso. Su potencia de transmisión puede ser de hasta 20 dBm (100 mW).

¹¹Las bandas ISM son bandas reservadas internacionalmente para el uso no comercial de radiofrecuencia electromagnética destinado a las áreas industrial, científica y médica.

- Clase 2: para dispositivos normales (rango de hasta 10 m) conectados a dispositivos tales como laptops y PCs. Su potencia de transmisión máxima es de 4 dBm (2.5 mW).
- Clase 3: para dispositivos de baja potencia con rango menor a 1 m. Su potencia de transmisión máxima es de 0 dBm (1 mW).

El estándar Bluetooth (IEEE802.15.1, 2002) se basa en el modo de operación maestro/esclavo. Se utiliza el término piconet (piconet) para referirse a la red formada por un dispositivo y todos los dispositivos que se encuentran dentro de su rango de alcance. Teóricamente pueden coexistir hasta 10 piconets dentro de una sola área de cobertura.

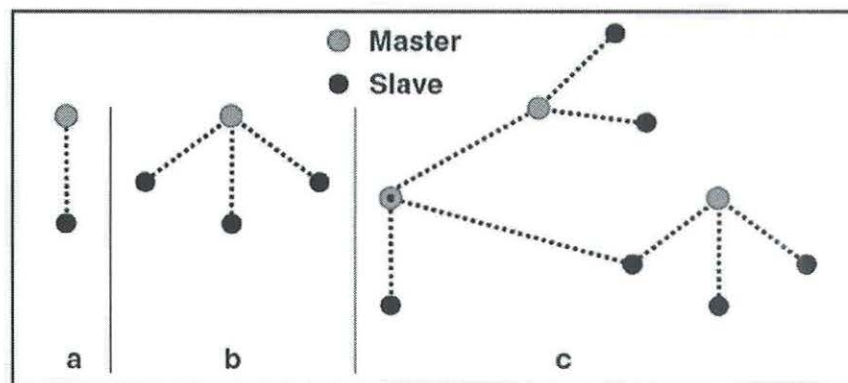


Figura 21. Tipo de redes Bluetooth que pueden ser establecidas. a) Piconet de un solo esclavo b) Piconet con varios esclavos c) Scatternet.

Un dispositivo maestro se puede conectar simultáneamente con hasta 7 dispositivos esclavos activos (200 en modo inactivo). Los dispositivos en una piconet poseen una dirección lógica de 3 bits, para un máximo de 8 dispositivos. Los dispositivos que se encuentran en el modo en espera se sincronizan, pero no tienen su propia dirección física en la piconet. La capacidad adaptativa de saltos en frecuencia (AFH) fue diseñada para

reducir la interferencia entre las tecnologías inalámbricas que comparten el espectro de frecuencia de los 2.4 GHz con los dispositivos Bluetooth. AFH opera dentro del espectro para tomar ventaja de las frecuencias disponibles. Esto se hace mediante la tecnología de detección de energía en el espectro, con la cual, un dispositivo Bluetooth evita utilizar las frecuencias que ya están siendo utilizadas.

El transceptor del radio cambia la frecuencia de salto en un esquema pseudoaleatorio que es determinado por el maestro. Esta técnica es conocida como espectro esparcido por saltos de frecuencia (FHSS, Frequency Hopping Spread Spectrum). El salto es adaptable entre 79 frecuencias en intervalos de 1 MHz, lo que permite una transmisión más eficiente dentro del espectro. La tasa de saltos es de 1600 saltos por segundo. Cada canal queda definido por la secuencia de saltos específica que le fue asignada.

Otros mecanismos utilizados para lograr una mayor confiabilidad en los enlaces Bluetooth son: ARQ (Automatic Repeat Request), CRC (Cyclic Redundancy Check) y FEC (Forward Error Correction).

IV.2.1 Transmisión de audio sobre redes Bluetooth

Para la transmisión de datos en tiempo real, como lo es la voz, se utilizan enlaces SCO (Synchronous Connection Oriented). Para los demás tipos de datos se utilizan enlaces ACL (Asynchronous Connectionless Link).

Cuando se establece un enlace SCO, si hay pérdidas de datos, no se intenta introducir mecanismos de control de error para retransmitir. El objetivo es mantener una latencia constante en el flujo de datos. Se establece un canal bidireccional entre el maestro y el esclavo punto a punto y el maestro reserva unas ranuras de tiempo constantes para poder transmitir un flujo de datos de forma permanente.

El ancho de banda que se reserva por conexión SCO es de 64 kbps (8 bits por muestra a 8 KHz de muestreo). El audio que se transmite sobre los enlaces SCO está en formato PCM, con cuantificación hecha con ley A o ley μ , ya que los enlaces SCO se utilizan principalmente para la transmisión de voz. Otro esquema de codificación que se puede utilizar es CVSD (Continuously Variable Slope Delta).

El maestro reserva ranuras de tiempo para la transmisión de paquetes SCO.

- Los paquetes HV1 contienen 1.25 ms de audio a 64 kb/s y deben ser transmitidos cada 2 ranuras de tiempo.
- Los paquetes HV2 contienen 2.5 ms de audio a 64 kb/s y deben ser transmitidos cada 4 ranuras de tiempo.
- Los paquetes HV3 contienen 3.75 ms de audio a 64 kb/s y deben ser transmitidos cada 6 ranuras de tiempo.

En una piconet se pueden tener hasta tres conexiones de un maestro con sus esclavo, en las cuales se transmitirían paquetes HV1.

IV.2.2 Perfil A2DP

Adicionalmente a la transmisión de audio con enlaces SCO, se ha definido el perfil¹² A2DP (Advanced Audio Distribution Profile). A partir de versión 2 del estándar Bluetooth puede realizarse la transmisión de audio de alta calidad con el perfil A2DP. Este perfil define los protocolos y procedimientos mediante los cuales se realiza la distribución de audio de alta calidad en canales mono o estéreo sobre enlaces ACL.

¹²Dentro del estándar Bluetooth se encuentran definidos varios perfiles. Los perfiles (definidos por Bluetooth SIG) fueron definidos para asegurar la interoperabilidad entre las aplicaciones de Bluetooth y los dispositivos de diversos fabricantes.

Un caso típico del uso del perfil A2DP es el streaming¹³ de contenido de música desde un reproductor estereo hacia audífonos o bocinas. El audio es comprimido para hacer un uso eficiente del ancho de banda. En la Tabla IV se señalan los codecs contemplados por el perfil A2DP.

Tabla IV. *Codecs que se definen en el perfil A2DP de Bluetooth.*

Codec	Obligatorio
SBC	Sí
MPEG-1,2 audio	No
MPEG-2,4 AAC	No
ATRAC	No

Cabe mencionar que los enlaces SCO utilizan un método de modulación más robusto respecto al ruido comparado con A2DP. De hecho el perfil A2DP considera la adaptabilidad de la calidad de audio con respecto a la calidad del enlace que se tiene. El motivo por el cual se utilizan tasas de transmisión mayores para A2DP (2 a 3 Mbps, mientras que para enlaces SCO la máxima tasa es de 1 Mbps) es el requerimiento de ancho de banda para el audio de alta calidad. Dado que en el presente trabajo interesa el análisis de códecs para la transmisión de audio cardiaco cuyo contenido de frecuencia llega hasta los 1000 Hz e incluso el audio respiratorio que ha quedado demostrado no supera los 3000 Hz, es conveniente hacer un análisis de la tasa efectiva máxima para enlaces Bluetooth que utilicen la versión 1 del estándar.

¹³Transmisión de audio o video para ser reproducido en tiempo real, esto gracias al almacenamiento temporal en *buffers*.

IV.2.3 Análisis de la tasa efectiva de transmisión para enlaces Bluetooth

Los estándares de comunicaciones, generalmente definen tasas pico de transmisión, las cuales son sólo una referencia. Sin embargo, hay que tener cuidado cuando se diseña un sistema de comunicaciones, pues si se toman esas tasas como base, seguramente se tendrán problemas en la operación de estos sistemas. En la práctica los sistemas no alcanzan las tasas pico mencionadas en el estándar, pues existen limitantes, por ejemplo, los bits utilizados por los encabezados, el uso de retransmisiones y la interferencia de otros sistemas de comunicación que utilizan el mismo espectro de frecuencia.

A continuación se detalla el análisis de la tasa máxima efectiva para el estándar Bluetooth, la cual contempla sólo la información que es enviada por el usuario (carga útil).

Se sabe que el canal Bluetooth está dividido en ranuras (slots) de tiempo. Además se sabe que de acuerdo con la estructura del paquete Bluetooth, el tamaño de la carga máxima es de 2745 bits (Figura 22) y que los enlaces que permiten una mayor carga en el paquete son los ACL, sin embargo, estos no tienen slots de tiempo garantizados. De los bits de la carga se utilizan 16 bits para encabezado y 16 para el CRC.

El tipo de paquete para enlace ACL que permite tener una carga mayor (puesto que ocupa más ranuras de tiempo) es el DH5. Los paquetes DH5 utilizan la carga máxima disponible en un paquete. Entonces se tiene que la carga máxima efectiva para DH5 se reduce a:

$$\text{Carga máxima efectiva} = 2745 - 32 = 2713 \text{ bits} = 339 \text{ bytes}$$

Un paquete DH5 es transmitido durante 5 ranuras de tiempo, por lo tanto, en

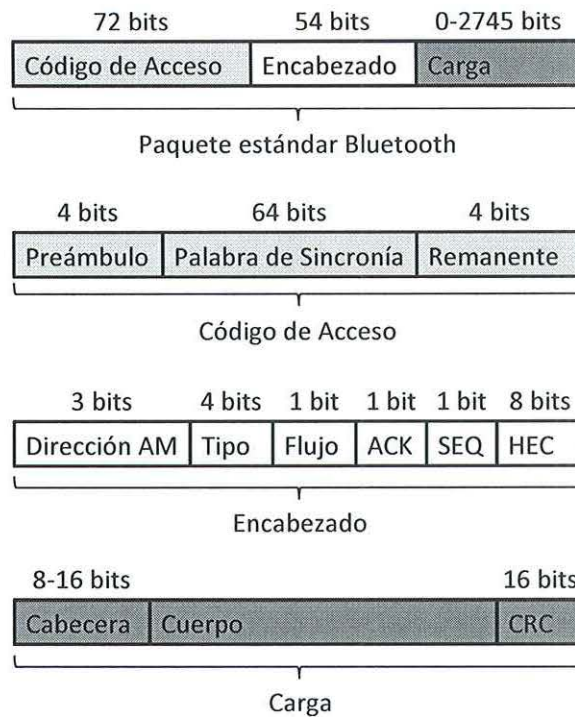


Figura 22. Estructura de un paquete Bluetooth.

el enlace asimétrico 5 ranuras son utilizadas por el enlace de ida y 1 ranura, por el enlace de retorno. Considerando condiciones ideales para maximizar el caudal eficaz (*throughput*), es decir, que sólo existen dos dispositivos en la piconet, con un enlace asimétrico se tiene que:

$$\text{Throughput máximo BT} = \frac{339 \text{ bytes}}{6(625\mu s)} = 723.2 \text{ kbps}$$

Del mismo modo que se hizo el cálculo anterior, se puede llegar a los resultados que se exponen en Tabla V.

Para los paquetes SCO la tasa de transmisión es fija, lo que varía de uno a otro tipo es la carga del usuario y el tipo de FEC utilizado. Las características de los paquetes para enlaces SCO se muestran en la Tabla VI.

Tabla V. *Paquetes ACL (Bruno et al. (2002)).*

Tipo (bytes)	Carga usuario	FEC	CRC	Max. tasa simétrica (kbps)	Max. tasa asimétrica (kbps)	
					Forward	Reverse
DM1	0-17	2/3	sí	108.8	108.8	108.8
DM3	0-121	2/3	sí	258.1	387.2	54.4
DM5	0-124	2/3	sí	286.7	447.8	36.3
DH1	0-27	no	sí	172.8	172.8	172.8
DH3	0-183	no	sí	390.4	585.6	86.4
DH5	0-339	no	sí	433.9	723.2	185.6

Tabla VI. *Paquetes SCO (Bruno et al. (2002)).*

Tipo	Carga (bytes)	FEC	CRC	Max. tasa simétrica (kbps)
HV1	10	1/3	no	64
HV2	20	2/3	no	64
HV3	30	no	no	64

IV.2.4 Estetoscopio digital con enlace Bluetooth

Durante el desarrollo de esta tesis inició la comercialización de un estetoscopio electrónico con enlace Bluetooth, desarrollado por la marca 3M Littmann. Este instrumento es el designado con el modelo 3200. Algunas de sus características son:

- Display LCD en el que se indica el nivel de sonido, el modo en que está operando el dispositivo (grabación o escucha directa).

- Capacidad de grabar los sonidos para ser reproducidos después.
- Transferencia de sonidos en tiempo real hacia una computadora o tableta, para su posterior análisis o almacenamiento.
- Tecnología Bluetooth para el enlace inalámbrico.
- Capacidad de la transmisión en tiempo real a través de Internet, para lo cual se pide un **enlace dedicado de 128 kbps**.

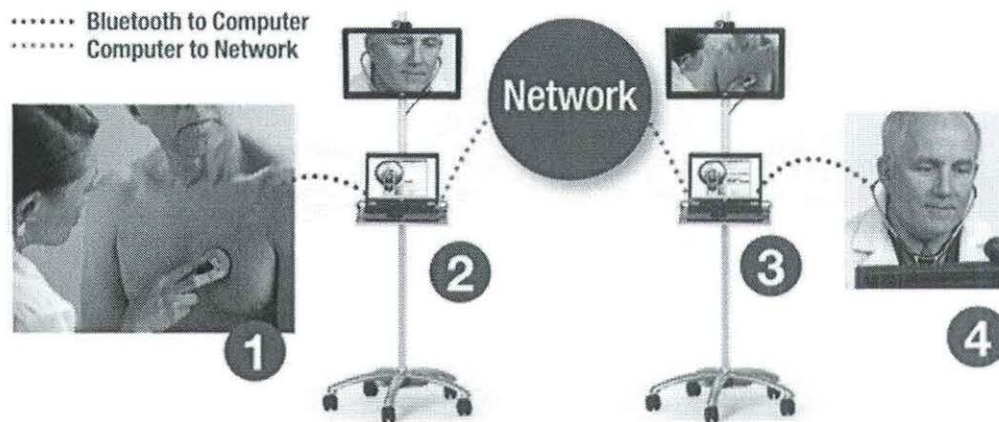


Figura 23. *Estetoscopio electrónico con enlace Bluetooth de la marca 3M Littmann (Littmann, 2010).*

Desafortunadamente la arquitectura de este estetoscopio es cerrada, y no hay un enlace directo al flujo de datos que proporciona el dispositivo. Sin embargo, se supone que el tipo de enlace utilizado es SCO, debido al ancho de banda dedicado que se pide, pues como se explicó en la sección anterior, los enlaces SCO son de 64 kbps en una dirección.

Los archivos de audio que entrega el estetoscopio Littmann están muestreados a 4 KHz con y codificados con 16 bits por muestra. Se supone que el hecho de tomar esa

tasa de muestreo, se debe a que el mayor contenido energético de estas señales está por debajo del rango de los 2 KHz y que se decidió asignar 16 bits a cada muestra para disminuir el ruido de cuantificación (ver cálculo de SQR en el Capítulo III Sección 3).

Cabe mencionar que la necesidad de 128 kbps dedicados puede ser considerada un tanto elevada si hablamos de llevar aplicaciones de telemedicina a localidades apartadas que muy posiblemente no cuenten con conectividad de banda ancha.

IV.3 IEEE 802.15.4

El estándar IEEE 802.15.4 (IEEE802.15.4, 2006) especifica la sub-capa MAC (Medium Acces Control) y la capa física (PHY) para LR-WPAN (de aquí en adelante llamada PAN). El protocolo IEEE 802.15.4 está muy asociado con el protocolo ZigBee, el cual especifica los protocolos para las capas por encima de IEEE 802.15.4, para proporcionar una pila de protocolos completa para redes inalámbricas de comunicaciones con baja potencia, bajo costo y bajas tasas de datos. Esta arquitectura de capas se muestra en la Figura 24.

IV.3.1 Componentes de red y topologías

En el protocolo IEEE 802.15.4 se definen tres tipos de nodo:

- **Coordinador PAN.** Es el controlador principal de la red (maestro), el cual identifica a su PAN (Personal Area Network), y con el que se asocia el resto de los nodos de la red. Este nodo también proporciona los servicios de sincronización global a los demás nodos de la red, por medio de la transmisión de paquetes faro (*beacons*). Los *beacons* contienen la identificación de la PAN y otra información relevante.

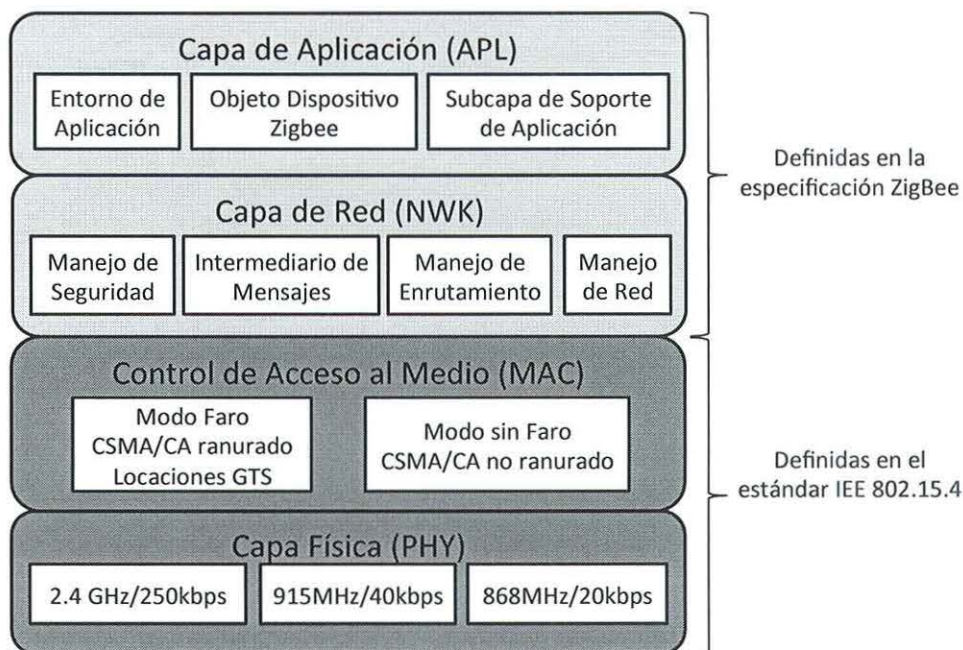


Figura 24. Pila de protocolos IEEE802.15.4 y ZigBee.

- **Coordinador.** Este tipo de nodo tiene las mismas funcionalidades que el coordinador PAN, con la excepción de que no crea su propia PAN. Un coordinador está asociado con un coordinador PAN y provee de servicios de sincronización local a los nodos que se encuentran dentro de su rango.
- **Dispositivo (nodo simple).** Este tipo de nodo no tiene funcionalidades de coordinador. Se asocia como esclavo con el coordinador PAN (o coordinador correspondiente) para sincronizarse con los demás nodos de la red.

En el estándar IEEE 802.15.4, los primeros dos tipos de nodo se refieren a dispositivos de función completa (FFD, Full Function Devices), lo cual significa que estos pueden implementar todas las funcionalidades del protocolo IEEE 802.15.4 para asegurar la sincronización y la administración de la red. El tercer tipo de nodos se refiere a los

dispositivos de funciones reducidas (RFD, Reduced Function Device), lo que significa que este nodo opera con una implementación mínima del protocolo IEEE 802.15.4. Un FFD también puede ser nodo simple.

Una PAN debe de incluir al menos un FFD como coordinador PAN, para que se encargue de los servicios de sincronización global y administre a los demás coordinadores potenciales y nodos esclavos que están dentro de su rango de alcance.

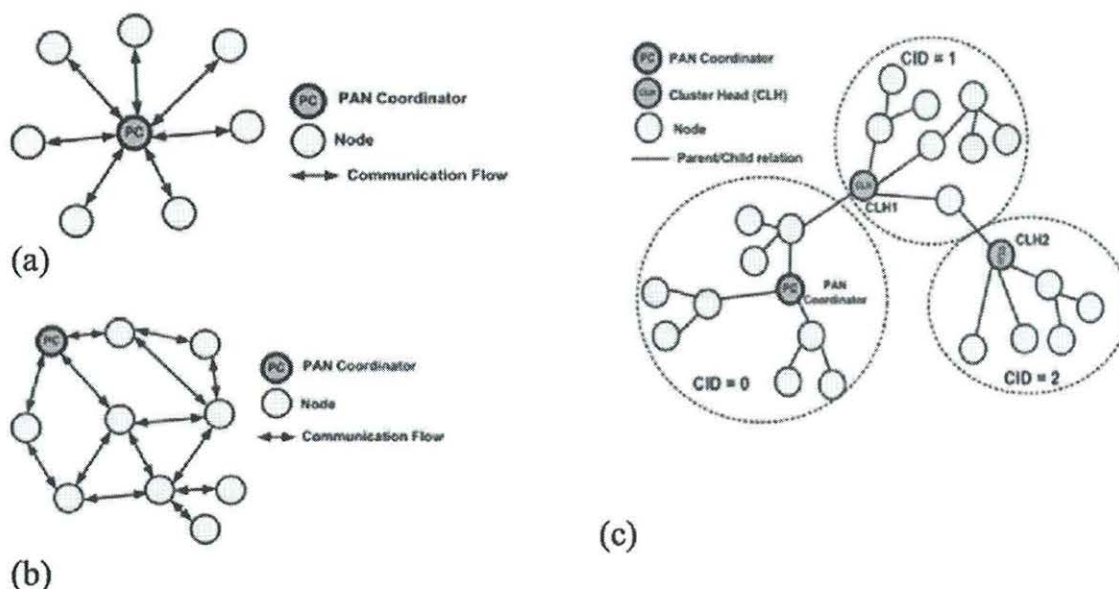


Figura 25. Topologías de red en IEEE 802.15.4 a) Estrella b) Igual-a-igual c)Árbol (Mahalik (2007)).

Una vez que la PAN es iniciada, debe de ser mantenida por su coordinador PAN, por medio de la generación y envío de *beacons*, coordinando las asociaciones y disociaciones de otros nodos a la PAN, proporcionando servicios de sincronización, coordinando la asignación y gestión de GTS (Guaranteed Time Slot), etc.

Se definen dos topologías básicas de red en el estándar IEEE 802.15.4: la topología estrella y a topología igual-a-igual (peer-to-peer). La topología de árbol (cluster-tree)

puede ser considerada un caso particular de la topología igual-a-igual. En la Figura 25 se observan los tipos de topología descritos en el estándar (IEEE802.15.4, 2006).

IV.3.2 Capa física

El estándar IEEE 802.15.4 ofrece tres bandas de frecuencia de operación: 2.4 GHz, 915 MHz y 868 MHz. Existe un canal entre los 868 y los 868.6 MHz, 10 canales entre 902 y 928 MHz, y 16 canales entre 2.4 y 2.4835 GHz. La distribución de los canales se puede observar en la Figura 26.

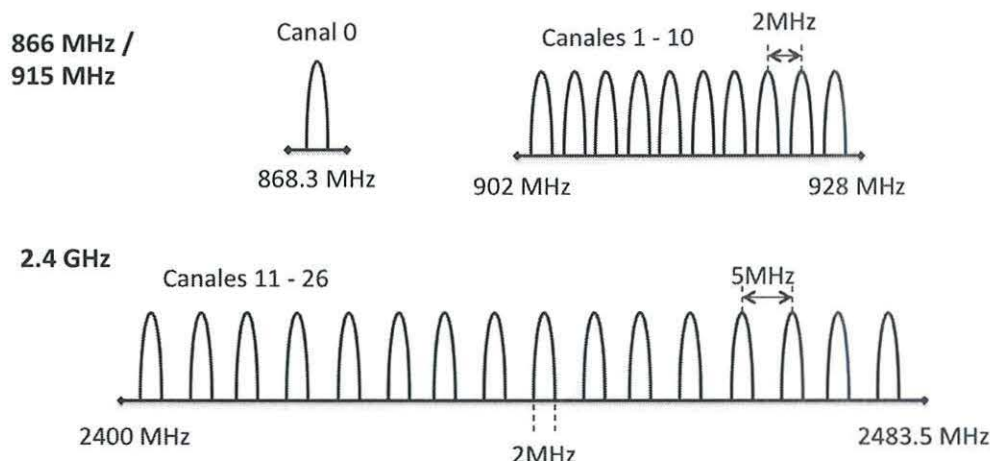


Figura 26. Bandas de operación en el estándar IEEE 802.15.4.

Las tasas de operación son 250 kbps en 2.4 GHz, 40 kbps en 915 MHz y 20 kbps en 868 Mhz. Las frecuencias más bajas son más convenientes para rangos de transmisión largos, debido a que las pérdidas por propagación son menores. La ventaja de las tasas de datos mayores es que proveen de un caudal eficaz mayor, una menor latencia y ciclos de trabajo menores.

Tabla VII. *Características de capa física para cada banda de frecuencia en IEEE 802.15.4.*

Banda de frecuencia	Parámetros de esparcimiento		Parámetros de datos		
	Tasa de chips (kchip/s)	Modulación	Tasa de bits (kbps)	Tasa de símbolos (ksímbolos/s)	Símbolos
868	300	BPSK	20	20	Binaria
915	600	BPSK	40	40	Binaria
2400	2000	O-QPSK	250	62.5	16-aria

La transmisión en todas estas bandas de frecuencia está basada en la técnica de esparcimiento DSSS (Direct Sequence Spread Spectrum). Las características de cada banda de frecuencia se resumen en la Tabla VII. Un símbolo es equivalente a cuatro bits.

Las tasas de bits señaladas en la Tabla VII son obligatorias. Sin embargo, en la versión 2006 del estándar se introdujeron tasas opcionales de hasta 250 Kbps para las bandas de 868 y 915 MHz.

Además de definir estos parámetros, la capa física también se encarga de las siguientes tareas:

- Activación y desactivación del transceptor.
- Realizar la detección de energía en el receptor (ED, Energy Detection) o la detección de portadora (CS, Carrier Sense).
- Indicación de calidad del enlace (LQI, Link Quality Indication).

- Evaluación de canal libre (CCA, Clear Channel Assessment). La operación CCA es responsable de reportar el estado de actividad del medio, ya sea ocupado o libre. Para esto, CCA tiene tres modos operacionales:
 - Modo de detección de energía (ED).
 - Modo de sensado de portadora (CS).
 - Modo de sensado de portadora con detección de energía.
- Selección de frecuencia del canal.

IV.3.3 Sub-capa MAC

El protocolo MAC soporta dos modos operacionales que pueden ser seleccionados por el coordinador PAN.

- El modo sin paquetes faro (*non beacon-enabled mode*) habilitados, en el cual el control de acceso al medio se rige por un CSMA/CA (Carrier Sense Multiple Access / Contention Avoidance) no ranurado.
- El modo de paquetes faro (*beacon-enabled mode*) habilitados, en el cual los *beacons* son enviados periódicamente por el coordinador PAN para sincronizar a los nodos que están asociados a él y para identificar a la PAN. Un paquete faro delimita el inicio de una supertrama, definiendo un intervalo de tiempo durante el cual las tramas son intercambiadas entre diferentes nodos de la PAN. El acceso al medio está regido por CSMA/CA ranurado. Además, el modo de paquetes faro habilitados permite la asignación de algunas ranuras de tiempo garantizadas en la supertrama llamadas GTs (Guaranteed Time Slots) para nodos que requieren servicios garantizados.

Al utilizar paquetes fano se incrementa el overhead, por esta razón al momento de hacer el análisis de la tasa efectiva se considerará una red que no utiliza paquetes fano.

Mecanismos de CSMA/CA

Como se describió anteriormente, existen dos versiones del mecanismo de acceso al medio CSMA/CA: el algoritmo CSMA/CA ranurado y el algoritmo CSMA/CA no ranurado. En ambos casos CSMA/CA utiliza una unidad de básica de tiempo de desistimiento (Backoff Period, BP) llamada *UnitBackoffPeriod* = 20 símbolos (0.32 ms).

En CSMA/CA ranurado cada operación (acceso al canal, conteo de desistimiento, CCA) puede ocurrir en el límite de un BP. Adicionalmente, los límites de BP deben ser alineados con los límites de la ranura de la supertrama.

En CSMA/CA no ranurado los periodos de desistimiento de un nodo son completamente independientes de los periodos de desistimiento de cualquier nodo de la PAN. Los algoritmos de desistimiento tanto del CSMA/CA ranurado, como del CSMA/CA no ranurado, dependen principalmente de tres variables.

- **El exponente de desistimiento (BE, Backoff Exponent).** Permite el cálculo del retardo de desistimiento, el cual es el tiempo antes de que se lleven a cabo los CCAs. El retardo de desistimiento es una variable aleatoria entre 0 y $(2^{BE} - 1)$.
- **La ventana de contención (CW, Contention Window).** Representa el número de periodos de desistimiento durante los cuales el canal debe de ser sentido libre (*idle*) antes de acceder al canal. La ventana de contención es utilizada en la versión ranurada de CSMA/CA. El estándar IEEE 802.15.4 marca por omisión un valor de inicialización de $CW = 2$, que corresponde a 2 CCAs. En cada periodo de desistimiento el canal es sentido durante los 8 primeros símbolos del BP.

- El número de desistimientos (NB, Number of Backoff). Representa el número de veces que debe de ser llevado a cabo el algoritmo de CSMA/CA para desistir de la transmisión mientras se intenta acceder al canal. Se inicializa a cero antes de cada nuevo intento de transmisión.

IV.3.4 Análisis de la tasa efectiva de transmisión para IEEE 802.15.4

Como se mencionó al analizar la tasa efectiva brindada por Bluetooth, las tasas pico de los estándares de comunicación son tan sólo una referencia, pues en la práctica no se obtienen en el sistema de comunicaciones. En el estándar IEEE 802.15.4, aunque la tasa de transmisión pico que señala ya es baja (250 kbps), esta disminuye aun más al contemplar parámetros como los encabezados en el paquete 802.15.4.

A continuación se presenta el análisis de la tasa efectiva de transmisión para el estándar IEEE 802.15.4 en el modo sin paquetes fijo habilitados, que como fue descrito en secciones anteriores, utiliza el mecanismo de acceso al medio CSMA/CA no ranurado.

En el algoritmo de CSMA/CA no ranurado, cada vez que un dispositivo requiere transmitir, espera un número aleatorio de unidades de periodo de desistimiento que queda dentro del rango de $\{0, 2^{BE} - 1\}$ antes de llevar a cabo el CCA.

- Si el canal está inactivo (*idle*), el dispositivo transmite.
- Si el canal está ocupado, el dispositivo espera otro periodo aleatorio antes de intentar acceder nuevamente al canal.

Inicialmente, el exponente de desistimiento (BE) se iguala a *macMinBE*. Utilizando por omisión el valor 3 para *macMinBE* y suponiendo que el canal se encuentra libre, el

tiempo de acceso al canal en el peor de los casos puede ser calculado como:

$$\begin{aligned}
 InitialbackoffPeriod + CCA &= (2^3 - 1) \times aUnitBackoffPeriod + CCA \\
 &= 7 \times 320\mu s + 128\mu s \\
 &= 2.368 \text{ ms}
 \end{aligned} \tag{18}$$

Donde:

El tiempo de detección de CCA está definido como el periodo de 8 símbolos.

$aUnitBackoffPeriod$ está definida como el periodo de 20 símbolos.

El periodo de un símbolo es igual a $16 \mu s$.

Máxima carga útil

El estándar IEEE 802.15.4 sepecifica que el máximo número de bytes que pueden ser transmitidos en la cargaútil de datos MAC es 102 bytes:

$$aMaxMACFrameSize = aMaxPHYPacketSize - aMaxFrameOverhead \tag{19}$$

Donde:

$$aMaxFrameOverhead = 25$$

$$aMaxPHYPacketSize = 127$$

El estándar IEEE 802.15.4 2003 define $aMaxMACFrameSize$ utilizando el máximo encabezado de trama sin importar el tamaño de encabezado de la trama actual. La versión 2006 del estándar permite que la máxima carga útil de datos sea mayor cuando se utilizan menos campos de la trama para el direccionamiento. Con esto, el encabezado de la trama se reduce a 13 bytes, dejando la carga útil para datos en 114 bytes.

Tiempo de transferencia de una trama

El paquete de la unidad de datos del protocolo MAC (MPDU) considera un encabezado de 6 bytes, los cuales corresponden al preámbulo, inicio de delimitación de trama (SHR) y la longitud de la trama (PHR). Si se añaden los 6 bytes y dada una tasa de datos fundamental de 250 kbps, el tiempo de transferencia de trama es calculado como:

$$\frac{(aMaxPHYPacketSize + SHR + PHR) \times 8}{250 \times 10^3} = \frac{(127 + 5 + 1) \times 8}{250 \times 10^3} = 4.256 \text{ ms} \quad (20)$$

Para una máxima carga útil de datos igual a 114 bytes.

Tiempo de transmisión de reconocimientos (ACK)

La trama correspondiente a un reconocimiento (paquete ACK) es de 11 bytes, y su estructura se muestra en la Figura 27. Dada una tasa de datos fundamental de 250 kbps, la transmisión de esta trama toma 0.325 ms. La transmisión de los paquetes ACK no utiliza el mecanismo CSMA/CA.

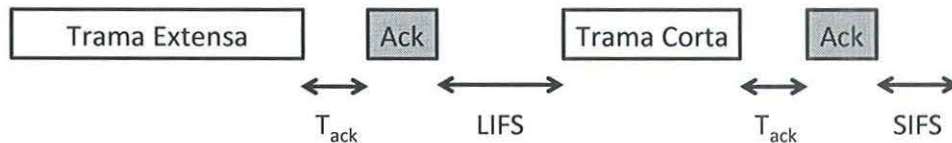


Figura 27. Estructura de los paquetes ACK en IEEE 802.15.4.

La transmisión de una trama de reconocimiento (en una red sin paquetes faro habilitados) comienza a los *TurnaroundTime* símbolos después de la recepción de la trama de datos, donde un *TurnaroundTime* es igual a 192 μ s. Este es el tiempo que se le da de margen al dispositivo para cambiar su estado de transmisión a recepción o viceversa.

Un reconocimiento siempre va seguido de un periodo mínimo de separación entre tramas (IFS, Inter Frame Spacing). La longitud del periodo de IFS depende del tamaño de la trama.

Las tramas (MPDUs) que tienen una longitud de hasta 18 bytes, deben de ser seguidas por un periodo SIFS (Short Inter Frame Spacing) de al menos 12 símbolos. Las tramas con una longitud mayor de 18 bytes deben ser seguidas por un periodo LIFS (Long Inter Frame Spacing) de al menos 40 símbolos. Este retardo en el tiempo es usualmente absorbido por el temporizado de acceso al canal de CSMA-CA.

Tasa de datos efectiva

Con la información detallada anteriormente, se puede calcular la tasa efectiva de datos, tomando en cuenta las siguientes consideraciones:

- Una red con paquetes fijo no habilitados.
- El algoritmo CSMA/CA nunca encuentra el canal ocupado.
- Se considera la recepción de paquetes de reconocimiento ACKs.
- No se solicitan retransmisiones.
- La carga útil máxima es 114 bytes.

Haciendo la suma de los tiempos calculados anteriormente se tiene que:

Tiempo para CSMA/CA = 2.368 ms

Transmisión de la trama de datos = 4.256 ms

Tiempo de cambio del transceptor = 0.192 ms

Transmisión de ACKs = 0.352 ms

$$T_{total802.15.4} = T_{CSMA} + T_{Tx} + T_{Tx/Rx} + T_{ACK} = 7.168ms \quad (21)$$

$$\text{Tasa efectiva de datos} = (114 \times 8)/(7.168 \times 10^{-3}) = 127 \text{ kbps} \quad (22)$$

Con esta tasa efectiva cabe la posibilidad de que se pueda transmitir audio cardiaco de buena calidad sobre una red IEEE 802.15.4. Cabe mencionar que el cálculo tanto de la tasa efectiva en Bluetooth como en IEEE 802.15.4 fue hecho bajo consideraciones tales como un ambiente sin interferencia ni ruido y una transmisión sin pérdida de paquetes. Tomando en cuenta lo anterior sigue siendo conveniente reducir el ancho de banda requerido para la transmisión del audio cardiaco.

IV.4 Escenario de aplicación y esquema propuesto

Como se mencionó en el Capítulo I, la problemática que implican las enfermedades cardiovasculares en nuestro país y la necesidad de brindar un servicio de salud a la sociedad son las principales motivaciones de este trabajo de tesis.

El estetoscopio es una herramienta básica para el diagnóstico médico y la auscultación, un procedimiento accesible para la evaluación de la condición cardiovascular de una persona. Y aunque en la actualidad ya existen estetoscopios digitales, no es fácil saber las características de la tecnología que utilizan, pues generalmente son de arquitectura cerrada.

Con respecto al estetoscopio Littmann con enlace Bluetooth descrito anteriormente, se sabe que posibilita la transmisión del audio cardiorespiratorio en tiempo real. Sin embargo, dicho dispositivo demanda un enlace dedicado de 128 kbps, requisito que es inviable en muchas localidades de nuestro país.

Otro factor importante es el uso del canal y el número de dispositivos que se pueden asociar en una misma WPAN. Si como se supone, en dicho dispositivo se utilizan enlaces SCO, esto implica que se está utilizando una parte importante de la capacidad del canal en brindar servicio a éste dispositivo. Además, bajo este esquema se incrementa la probabilidad de colisiones de paquetes debida a dispositivos operantes en la misma banda de frecuencia.

Una ventaja de utilizar Bluetooth es el nivel de penetración con el que ya cuenta. Sin embargo, se puede considerar que su consumo de energía es elevado, con lo cual se puede sugerir el uso de otro tipo de redes. En el caso del presente trabajo se considera la utilización de radios IEEE 802.15.4 como una alternativa para el enlace inalámbrico de estetoscopios digitales.

Además del ahorro de energía al utilizar redes IEEE 802.15.4 en lugar de redes Bluetooth existe otra ventaja con respecto a la susceptibilidad a interferencia. Como se describió anteriormente, los dispositivos IEEE 802.15.4 pueden operar además de la banda de 2.4 GHz en otras dos bandas de frecuencia, con lo que pueden evitar la saturación de la banda de 2.4 GHz, en la que también operan las redes WiFi, los hornos de microondas y el mismo Bluetooth.

Con respecto al audio cardiorrespiratorio, se busca que este tenga un ancho de banda mucho menor tal que incluso se pueda transmitir dicho audio con una calidad aceptable utilizando trancéptores IEEE 802.15.4. Aun y cuando estos trancéptores ofrecen una menor tasa efectiva pico que Bluetooth, se tiene la ventaja de que tienen un menor consumo de energía y un rango de transmisión más amplio.

Adicionalmente, es factible que si se logra encontrar un códec adecuado para realizar la transmisión sobre redes IEEE 802.15.4, este códec pueda ser utilizado también con la versión 4 del estándar Bluetooth recientemente liberada.

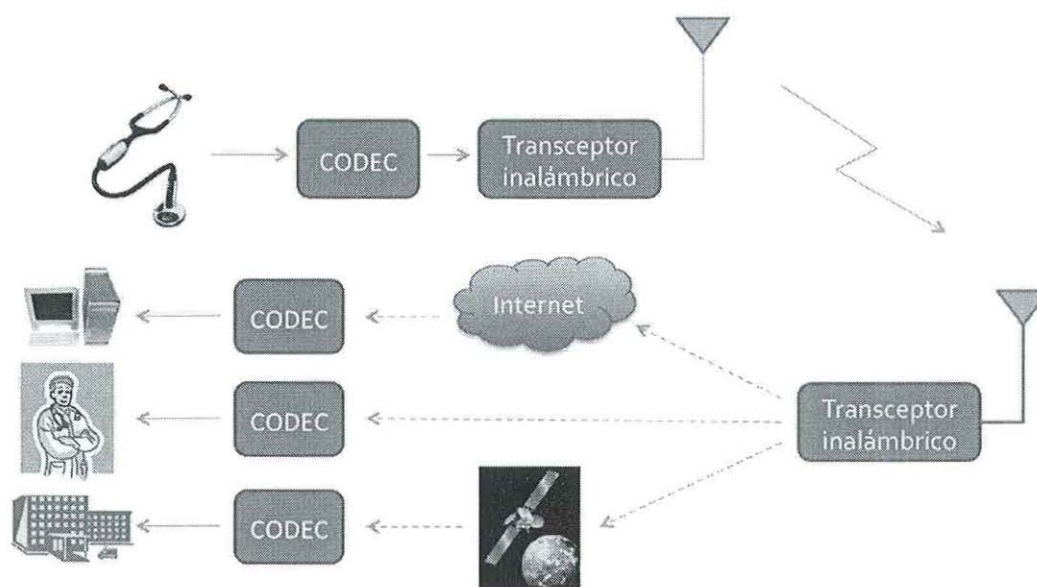


Figura 28. Escenarios de aplicación para la transmisión de audio cardíaco codificado.

En la Figura 28 se observa el escenario de aplicación del presente proyecto de tesis. En un consultorio médico se ausculta a un paciente con un estetoscopio digital. El audio en el estetoscopio es comprimido utilizando un códec para ser transmitido mediante un transceptor inalámbrico. La señal es recibida por otro transceptor inalámbrico. Tres escenarios posibles contemplados para el uso de la señal de audio recibida son:

- La señal es escuchada en por el profesional de la salud que realiza la auscultación en ese mismo momento.
- La señal se transmite a través de la red de internet a otro médico para su evaluación en tiempo real o su posterior análisis.
- La señal es transmitida vía satélite hacia un hospital de otra localidad para su análisis.

El enfoque principal de este proyecto de tesis es facilitar el envío de señales adquiridas con un estetoscopio digital en redes inalámbricas que tienen anchos de banda de transmisión limitados, i.e. WPANS. Debido a que este proyecto se centra en un enlace inalámbrico de tasas bajas o medianas no se hará un análisis detallado de los otros tipos de redes. Por lo tanto, topológicamente se analizará la utilización de códecs para la transmisión de audio cardíaco en una red que consta de dos nodos inalámbricos.

Se propone el esquema mostrado en la Figura 29 para llevar a cabo la transmisión de señales de audio cardiorrespiratorio sobre redes inalámbricas tipo IEEE 802.15.4.

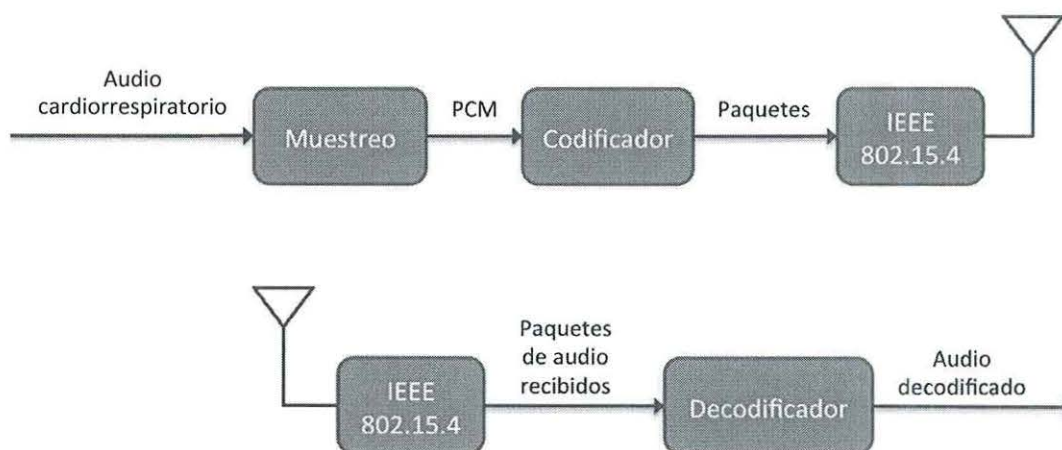


Figura 29. Esquema propuesto para la codificación y transmisión de audio cardiorrespiratorio.

Inicialmente se tiene audio cardíaco o respiratorio de forma analógica que requiere ser digitalizado para su procesamiento. Debido a los requerimientos de entrada que tienen los codificadores de audio se tienen dos opciones para la digitalización del audio cardiorrespiratorio:

- Muestreo a 8 KHz con cuantificación de 16 bits de precisión.
- Muestreo a 8 KHz con cuantificación de 8 bits de precisión.

Enseguida el audio que ya fue digitalizado es ahora comprimido mediante un códec.

A la salida del bloque del codificador, se obtienen paquetes de audio codificado, los cuales son transmitidos a través de una red IEEE 802.15.4 que consta de sólo dos dispositivos. Un dispositivo transmite el audio codificado Los paquetes de audio son recibidos en el segundo dispositivo de la red inalámbrica, para posteriormente ser decodificados.

Con las pruebas realizadas durante el desarrollo de este proyecto, se analizan dos cuestiones fundamentales:

- ¿Qué códecs pueden funcionar en términos de calidad para comprimir señales de audio cardiorrespiratorio?
- ¿Qué códecs presentan un desempeño aceptable al hacer una transmisión inalámbrica de audio cardiorrespiratorio?

Capítulo V

Análisis de códecs y evaluación de calidad

V.1 Introducción

Existen numerosos formatos para la compresión de audio y estos pueden ser divididos (además de las clasificaciones descritas en el Capítulo III) en dos grandes grupos: los que pertenecen a algún estándar internacional (que por lo tanto son abiertos¹⁴) y los que son diseños propietarios.

Para el desarrollo de este proyecto de tesis se buscó trabajar con códecs que fueran abiertos principalmente por la facilidad de obtener la información correspondiente a la formación de las tramas del audio salida. Lo anterior es algo sumamente importante para la siguiente etapa puesto que el audio debe de ser paquetizado para poder ser transmitido sobre una red de comunicaciones.

En el presente capítulo se describen los códecs evaluados durante este proyecto. Además se presentan los tipos de evaluación que se puede hacer a los códecs de audio para medir la calidad de la señal que estos entregan.

Adicionalmente aquí se presentan los resultados obtenidos al evaluar la calidad de muestras codificadas al ser comparadas con las originales. Esto permitió descartar varios de ellos antes de analizar con más detalle la factibilidad de su implementación en el escenario propuesto.

¹⁴Libres de patentes y de regalías.

V.2 Códecs considerados

V.2.1 Códec G.711

El códec G.711 está definido en la recomendación ITU-T-G.711 (1998). Es el formato nativo de las redes telefónicas digitales. Utiliza un esquema de codificación PCM sin compresión, proporcionando así una tasa de bits de 64kbps. Especifica dos algoritmos para la cuantificación de la señal analógica. Dichos algoritmos son conocidos como ley A, para el estándar E1 usado en Europa; ley μ , que corresponde al estándar T1 usado en Estados Unidos y Japón. Tanto en la ley A como en la ley μ la cuantificación es logarítmica.

La función que define a la ley μ es la siguiente:

$$F(x) = \text{sgn}(x) \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)} \in -1 \leq x \leq 1 \quad (23)$$

mientras que la ley A está definida por:

$$F(x) = \text{sgn}(x) \begin{cases} \frac{A|x|}{1+\ln(A)} & |x| < \frac{1}{A} \\ \frac{1+\ln(A|x|)}{1+\ln(A)} & \frac{1}{A} \leq |x| \leq 1 \end{cases} \quad (24)$$

donde A es el parámetro de compresión. En Europa $A = 87.7$, aunque también se utiliza el valor 87.6.

G.711 proporciona la mejor calidad ya que no utiliza un algoritmo de compresión y por lo tanto no introduce pérdidas. Este códec también tiene menor latencia debido al menor procesamiento que requiere por no hacer compresión. Se utiliza en las redes PSTN y líneas ISDN. Si se utiliza este códec, se necesita un mínimo de 128 kbps para una comunicación bidireccional.

En la Figura 30 se muestra el diagrama a bloques del codificador G.711.

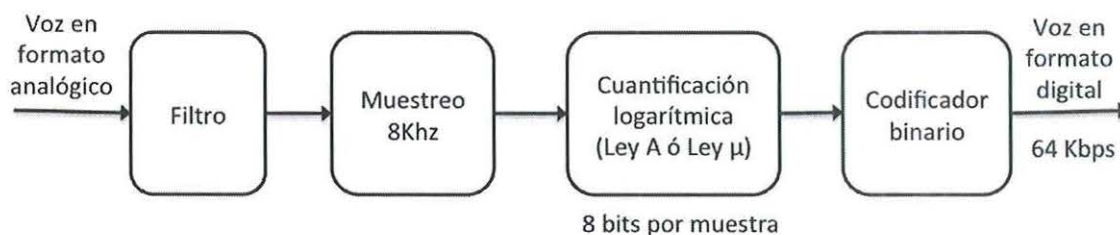


Figura 30. Diagrama a bloques del codificador G.711.

V.2.2 Códec G.726

ADPCM (Adaptive Differential Pulse Code Modulation) es un esquema de codificación conocido también como códec G.726, ya que está definido en la recomendación ITU-T-G.726 (1990). Esta recomendación define tasas de 16, 24, 32 y 40 kbps (que corresponden a tamaños de muestra de 2, 3, 4 y 5 bits respectivamente).

ADPCM es un codificador de forma de onda basado en DPCM que añade algunas funcionalidades.

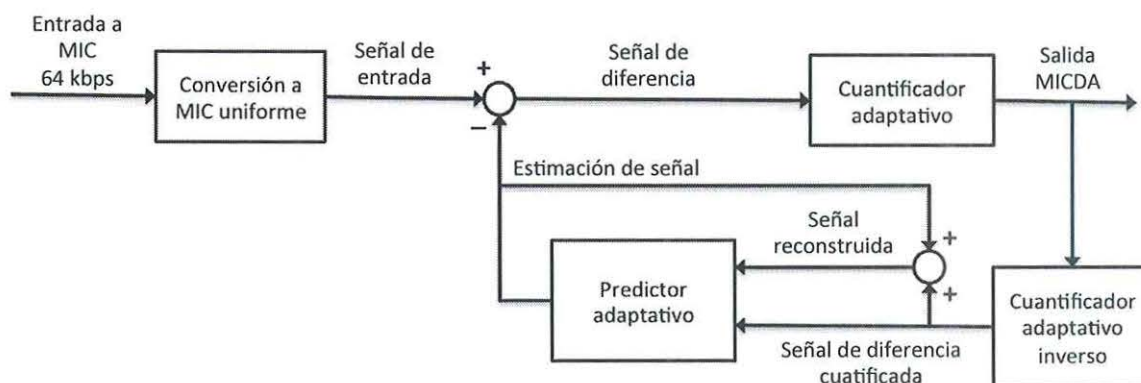


Figura 31. Diagrama a bloques del codificador G.726.

En la Figura 31 se muestra un diagrama a bloques del codificador G.726. El uso de este codificador puede ser conveniente en ciertas aplicaciones, 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.

V.2.3 Códec G.729

El códec G.729 fue definido por la ITU en la recomendación ITU-T-G.729 (2007). Es un algoritmo de compresión que comprime el audio de la voz en trozos de 10 ms. Señales distintas de la voz como lo son música y tonos DTMF (Dual Tone Multi-Frequency) no pueden ser transmitidos con integridad utilizando este códec.

G.729 se utiliza mayormente en aplicaciones de Voz sobre IP (VozIP) por sus requerimientos de bajo ancho de banda y está basado en el algoritmo CS-ACELP (Conjugate-Structure Algebraic-Code-Excited Linear Prediction), aprobado en el estándar [ITU-T-G.729 (2007)].

El estándar G.729 opera a 8 kbps, pero existen extensiones. Las mismas proveen además soporte a 6.4 kbps y 11.8 kbps dando una mejor o peor calidad respectivamente. G.729A es también muy común y compatible con G.729, pero requiere menos procesos. Esta baja complejidad tiene una consecuencia ya que la calidad de la señal codificada es mucho menor. En la Figura 32 se muestra un diagrama a bloques del codificador G.729.

El anexo B de G.729 es un esquema de compresión de silencios, el cual tiene un módulo VAD (Voice Activity Detection) que se utiliza para detectar actividad en la voz hablada o no hablada.

Además, incluye un módulo DTX (Discontinuous Transmission) que se encarga de

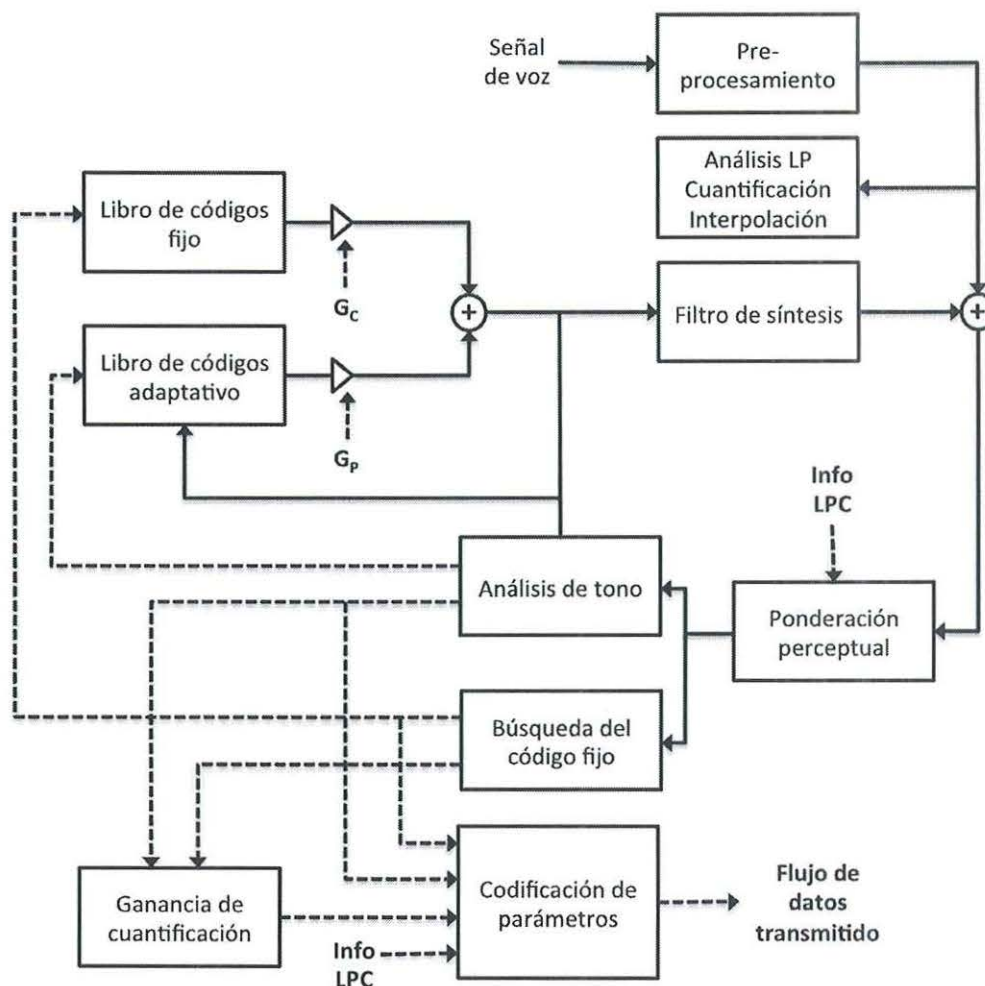


Figura 32. Diagrama a bloques del codificador G.729.

mejorar los parámetros del sonido ambiental.

El anexo B también incluye un generador de sonido de confort (CNG, Comfort Noise Generation) pues si en un canal de comunicación se detiene la transmisión por no haber voces, el receptor puede suponer que la transmisión ha sido cortada.

Existen otros anexos al estándar G.729 los cuales son: C, D, E, F, G, H, I y J. Es importante señalar que no todos son compatibles entre sí.

V.2.4 Códec GSM

GSM (Global System for Mobile communications) es un sistema de telefonía celular popular fuera de los Estados Unidos. GSM incluye un códec que usualmente es llamado códec GSM. Este códec utiliza la técnica RPE-LTP (Regular Pulse Excitation Long-Term Prediction)[Hanzo *et al.* (2008)].

El códec GSM puede trabajar con los siguientes modos de operación: tasa media (Half Rate, HR), tasa completa (Full Rate, FR), tasa completa mejorada (Enhanced Full Rate, EFR), tasa de datos adaptable (Adaptative Multi-Rate, AMR), tasa de datos adaptable para banda ancha (AMR-WB) y tasa de datos extendida adaptable para banda ancha (AMR-WB+).

En el codificador de tasa completa se procesan bloques de voz de 20 ms. Cada bloque contiene 260 bits como se muestra en la Figura 33 ($188+36+36 = 260$). Por lo cual este códec entrega una tasa de datos de 13 kbps ($260 \text{ bits} / 20 \text{ ms} = 13 \text{ kbps}$). El codificador tiene tres partes principales:

1. Análisis de predicción lineal (predicción de corto plazo).
2. Predicción de largo plazo.
3. Análisis de excitación.

La predicción lineal utiliza una función de transferencia de orden 8:

$$A(z) = 1 + \sum_{k=1}^8 a_k z^{-k} \quad (25)$$

En total, la parte de predicción lineal utiliza 36 bits. El predictor de largo plazo estima el *emphpitch* y la ganancia cuatro veces en intervalos de 5 ms. Cada estimación proporciona un coeficiente de retraso (*lag*) y un coeficiente de ganancia de 7 bits y 2 bits

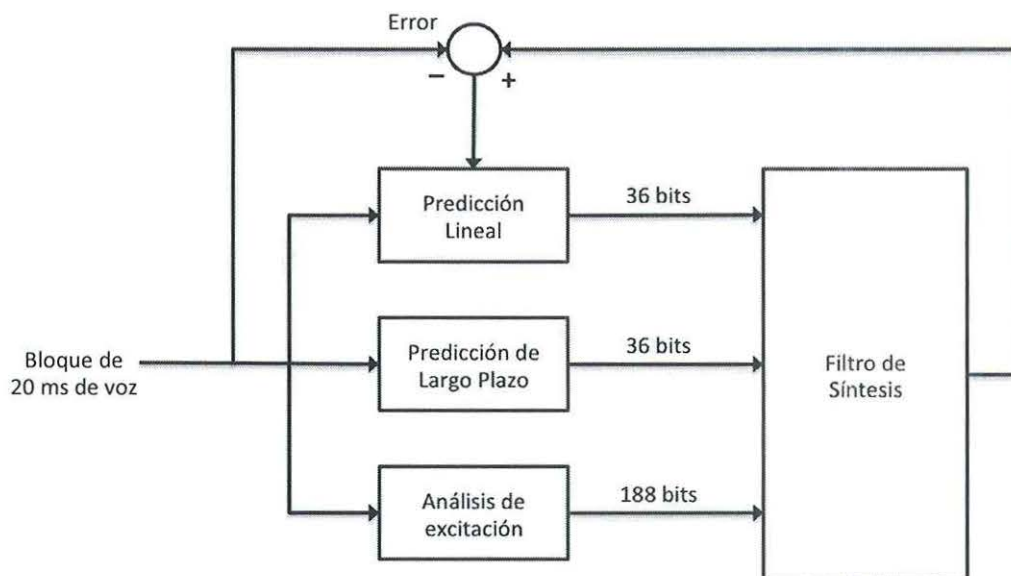


Figura 33. *Diagrama a bloques del codificador GSM Full-Rate.*

respectivamente. Las cuatro estimaciones en total requieren $4 \cdot (7+2) \text{ bits} = 36 \text{ bits}$. El factor de ganancia en la muestra de predicción de voz asegura que la voz sintetizada tendrá el mismo nivel de energía que la señal original de voz.

Los 188 bits restantes son derivados del análisis de excitación de pulso regular. Después del filtrado de las predicciones de corto y largo plazo, la señal residual es cuantificada para cada sub-trama de 5 ms.

En resumen, este códec usa la información de muestras anteriores (esta información no cambia rápidamente) para poder predecir la muestra actual. La señal de voz es dividida en bloques de 20 ms. Estos bloques son enviados al codificador de voz, el mismo que tiene una velocidad de 13kbps, para poder obtener bloques de 260 bits.

V.2.5 Códec Speex

Es un códec de voz de código abierto. En esencia Speex es para voz lo que Vorbis es para música. Fue desarrollado especialmente para redes de paquetes y aplicaciones de VoIP. Speex está diseñado para ser muy flexible y soportar una amplia variedad de calidades de voz y tasas de datos. Esto significa que Speex puede codificar voz de banda ancha es decir *wideband* (tasa de muestreo de 16 KHz) además de soportar banda estrecha (*narrowband*, 8KHz de muestreo). También soporta la tasa de muestreo de 32 KHz (*ultra-wideband*). El hecho de que este códec haya sido diseñado para VoIP y no para teléfonos móviles hace que Speex sea robusto contra la pérdida de paquetes, pero no contra los paquetes corruptos. Esto se basa en la suposición de que en VoIP los paquetes llegan inalterados o no llegan. Este códec utiliza un algoritmo CELP [Ganguly y Bhatnagar (2008)].

V.2.6 Códec MP2

El grupo de expertos dedicados a la parte de audio en el Moving Picture Experts Group (MPEG) han definido tres niveles distintos de desempeño y complejidad para el estándar de audio denominado MPEG. Estos niveles son llamados capas (*layers*).

Las técnicas utilizadas en las capas I y II (MP2) del estándar MPEG para codificación de audio están basadas en el principio de la codificación de sub-bandas. Este principio imita el mecanismo de análisis en frecuencia del oído separando el espectro de audio en un gran número de bandas de frecuencia distintas. Las señales que se encuentran dentro de estas bandas son cuantificadas de forma independiente.

El número de sub-bandas que es utilizado para un codificador depende de las otras herramientas que son combinadas con la codificación de sub-bandas para realizar la

compresión. Si se pretende optimizar la compresión basándose en el enmascaramiento auditivo, las sub-bandas deben ser preferiblemente más estrechas que las bandas críticas del oído (Watkinson, 2004), por lo que se requerirá un gran número de bandas. Frecuentemente no se cumple este requisito, por ejemplo, para las capas I y II de MPEG se utilizan sólo 32 bandas.

En MP2 después de hacer la división en sub-bandas, se calcula la FFT de estas bandas, para proporcionarle al modelo de enmascaramiento una mejor resolución espectral. La FFT utilizada tiene 1024 puntos. La longitud de cada bloque de entrada al codificador es de 1152 muestras, lo cual corresponde a 24 ms a 48 KHz. Las sub-bandas son clasificadas posteriormente en tres rangos de frecuencia: bajo, medio y alto, para que cada rango sea recuantificado de la siguiente manera:

- Las muestras del rango de baja frecuencia pueden ser cuantificadas con 15 longitudes de palabra distintas (4bits).
- Las muestras de frecuencias medias pueden ser cuantificadas con 7 longitudes de palabra distintas (3 bits).
- Las muestras de frecuencias altas pueden ser cuantificadas con sólo 3 longitudes de palabra distintas (2 bits).

Para cada caso existe una palabra que indica que no se enviaron bits de esa sub-banda.

Los bloques de 1152 muestras son divididos en tres bloques de 384 muestras para utilizar el mismo esquema de *companding* que se utiliza en la capa I. Con este esquema, cada una de las bandas es procesada de acuerdo a su propio nivel.

Las bandas en las que hay poca energía son codificadas con palabras más cortas, puesto que la variación en los niveles de amplitud es menor. Así, cada banda tiene

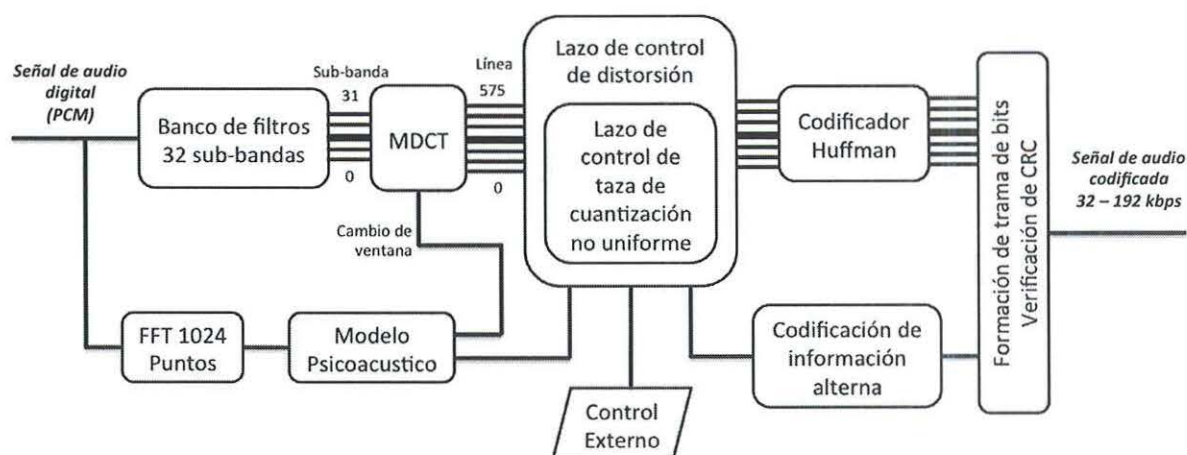


Figura 34. Diagrama a bloques del codificador MP3.

Banco de filtros

Un proceso utilizado muy frecuentemente en compresión de audio es el que se conoce como sub-división en bandas (*band splitting*), el cual divide el espectro del audio en una serie de rangos de frecuencia. Este es un proceso complejo y requiere de bastantes cálculos. Un método de sub-división en bandas que es muy útil es el denominado filtro de espejo en cuadratura (QMF, Quadrature Mirror Filter).

Un QMF es un filtro que divide la señal de entrada en dos bandas que posteriormente suelen ser sub-muestreadas por un factor de 2. Además, ambas bandas (superior e inferior) de frecuencias se intercambian entre sí. Es decir, las frecuencias bajas se codifican como frecuencias altas y viceversa.

Un QMF es un filtro que divide la señal de entrada en dos bandas. Las señales resultantes de frecuencias altas y bajas son nuevamente reducidas por un factor de 2, dando como resultado una representación de dos canales críticamente muestreados de la señal original.

Los filtros de análisis son representados mediante la siguiente fórmula:

$$|H_0(e^{j\Omega})|^2 + |H_1(e^{j\Omega})|^2 = 1 \quad (26)$$

donde Ω es la frecuencia y la tasa de muestreo es normalizada a 2π .

En MP3 la señal de entrada es dividida en 32 sub-bandas de frecuencia. La subdivisión en estas bandas simula el mecanismo de análisis en frecuencia del oído. Una vez dividida en bandas una señal, esta puede ser cuantificada independientemente.

Modelo psicoacústico

La compresión de audio en MP3 se basa en la reducción del margen dinámico irrelevante, es decir, se basa en la incapacidad del sistema auditivo para detectar los errores de cuantificación en condiciones de enmascaramiento. Para ello la señal es dividida en bandas de frecuencia que se aproximan a las bandas críticas, después cada sub-banda es cuantificada en función del umbral de detección del ruido dentro de esa banda.

El modelo psicoacústico es una modificación del empleado en la capa II del estándar, y utiliza un método denominado predicción polinómica. Analiza la señal de audio y calcula la cantidad de ruido que se puede introducir en función de la frecuencia, es decir, calcula el umbral de enmascaramiento en función de la frecuencia. El codificador usa esta información para decidir la mejor manera de utilizar los bits disponibles.

Cuantificación y codificación

Los procesos de cuantificación y codificación se llevan a cabo en dos ciclos de iteración, uno interno y uno externo. Se examinan tanto las muestras de salida del banco de filtros como el SMR (Signal-to-Mask Ratio) proporcionado por el modelo psicoacústico, y se ajusta la asignación de bits o ruido según el esquema utilizado para satisfacer

simultáneamente los requisitos de tasa de bits y de enmascaramiento. Dichos ciclos consisten en:

- **Ciclo interno**

El ciclo interno realiza la cuantificación no-uniforme de acuerdo con el sistema de punto flotante (cada valor espectral MDCT se eleva a la potencia $3/4$). El ciclo escoge un determinado intervalo de cuantificación y a los datos cuantificados se les aplica codificación de Huffman en el siguiente bloque. El ciclo termina cuando los valores cuantificados que han sido codificados con Huffman usan menor o igual número de bits que la máxima cantidad permitida.

- **Ciclo externo**

El ciclo externo se encarga de verificar si el factor de escala para cada sub-banda tiene más distorsión de la permitida (ruido en la señal codificada), comparando cada factor de escala de la sub-banda con los datos previamente calculados en el análisis psicoacústico.

El ciclo externo se termina cuando una de las siguientes tres condiciones se cumple:

1) Ninguno de los factores de escala de la banda tiene mucho ruido; 2) La siguiente iteración amplifica a una de las bandas más de lo permitido; 3) Todas las bandas han sido amplificadas al menos una vez.

En resumen, MP3 es un codificador de propósito general que basa su algoritmo de compresión en un modelo psicoacústico el cual toma ventaja de la respuesta del oído al enmascaramiento frecuencial y temporal.

V.2.8 Códec AAC

El códec de audio avanzado (AAC, Advanced Audio Coding) corresponde al estándar internacional (ISO/IEC-13818-7, 2004) como una extensión de MPEG-2, por lo que también es llamado MPEG-2 AAC. AAC es un códec perceptual y se encuentra en el núcleo del MPEG-4, 3GPP y 3GPP2 (Watkinson, 2004). Este códec es utilizado para conexiones inalámbricas y radio difusión digital. AAC también ha sido elegido por la compañía Apple como formato principal para los iPods y para el software iTunes.

AAC es un codificador con pérdidas y soporta hasta 48 canales de audio. Por omisión este códec soporta audio monofónico, estéreo y de 5.1 canales. El concepto de AAC está basado en herramientas de codificación llamados módulos, los cuales pueden ser combinados en distintas maneras para producir tramas de datos en tres perfiles diferentes.

El perfil principal (*main profile*) requiere el codificador más complejo, y hace uso de todos los módulos. El perfil de baja complejidad (LC, low complexity) omite ciertas herramientas y restringe la potencia de otras para reducir los requerimientos de procesamiento y memoria. Las herramientas que se utilizan en el perfil LC son idénticas a sus homólogas en el perfil principal, así que un decodificador del perfil principal puede también decodificar tramas de datos del perfil de baja complejidad.

El perfil de tasa de muestreo escalable (SSR, Scaleable Sampling Rate) divide la entrada de audio en cuatro bandas iguales de frecuencia cada una de las cuales resulta en una trama de datos auto-contenida.

La complejidad mayor de AAC (comparada con los estándares MPEG anteriores) permite la introducción de herramientas de codificación sin pérdidas. Esto permite que se tengan tasas de bits menores para una calidad igual o mejor, en donde se reduce la

dependencia de la pérdida de codificación.

En la Figura 35 se muestra un diagrama a bloques del codificador AAC de perfil principal.

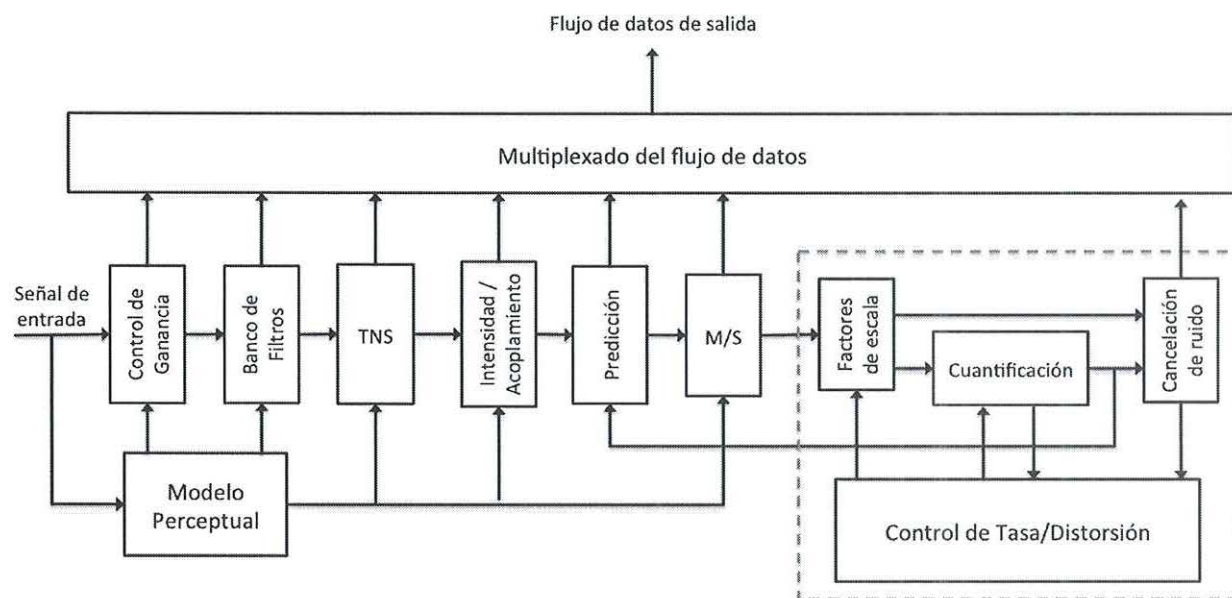


Figura 35. Diagrama a bloques del codificador AAC.

AAC presenta algunos cambios comparados con la capa III que ayudan a tener la misma calidad en tasas de datos bajas. Estos cambios son:

- Resolución de frecuencia mayor. El número de líneas de frecuencia en AAC es de hasta 1024, comparado con los 576 que se utilizan en la capa III.
- Predicción. Una predicción hacia atrás (backward) opcional, calculada línea por línea, logra una mejor eficiencia de codificación, especialmente para señales tonales (por ejemplo, un diapasón). Esta característica sólo está disponible en el perfil principal y rara vez se utiliza.

- Codificación de *joint stereo* mejorada. Comparado con MP3, tanto la codificación de medio lado (mid/side) como la codificación de intensidad son más flexibles, permitiendo que estas sean aplicadas para reducir la tasa de datos más frecuentemente.
- Codificación Huffman mejorada. En AAC, se aplica más seguido la codificación de líneas cuádruples de frecuencia, además la asignación de las tablas de código Huffman a las particiones del codificador puede ser mucho más flexible.

Existen otros cambios en AAC con respecto a MP3, los cuales ayudan a mantener la alta calidad para tipos de señales más complicados.

- Conmutación de bloques mejorada. Además del banco de filtros híbrido (en casacada) que se tiene en la capa III, AAC utiliza un banco de flitros MDCT estándar conmutado con una respuesta al impulso (para bloques cortos) de 5.3 ms a una frecuencia de muestreo de 48 KHz. Esto se compara favorablemente con la capa III a 18.6 ms y reduce la cantidad de artefactos¹⁵ de pre-eco.
- Temporal Noise Shaping (TNS). Esta técnica hace la formación de ruido en el dominio del tiempo mediante un lazo abierto de predicción en el dominio de la frecuencia. TNS es una técnica que ha demostrado ser especialmente eficaz en la mejora de la calidad de señales de voz a bajas tasas de bits.

En resumen AAC sigue el mismo paradigma básico de codificación que MP3, i.e. banco de filtros de alta resolución, cuantificación no uniforme, codificación Huffman, estructura de ciclos de iteración utilizando análisis por síntesis. Sin embargo, AAC utiliza nuevas herramientas de codificación para mejorar la calidad en tasas de datos bajas.

¹⁵En este contexto, artefacto es el ruido perceptible introducido por un proceso de compresión.

V.2.9 Códec FLAC

El códec libre de audio sin pérdidas (FLAC, Free Lossless Audio Codec) es parte del proyecto Ogg de código abierto de la fundación Xiph.org¹⁶. Los algoritmos con pérdida pueden comprimir a más de 1/10 del tamaño inicial, a costa de descartar información; FLAC, en su lugar, usa la predicción lineal para convertir las muestras en series de pequeños números no correlativos (conocido como residuos), que se almacenan eficientemente usando la codificación Golomb-Rice. Además de esto, para aprovechar los silencios usa codificación por longitud de pista (RLE, Run-Length Encoding) para muestras idénticas.

FLAC compila en muchas plataformas computacionales, incluyedo Linux y Windows, y define un formato de flujo, referencia codificadores y decodificadores así como un programa de línea de comandos para codificar y decodificar archivos. FLAC, como su nombre lo dice, es un códec sin pérdidas, y además es asimétrico, por lo que el proceso de decodificación es más rápido que el de codificación.

Las tramas FLAC están protegidas con un CRC de detección de errores de 2 bytes y por una capacidad de firma digital que aseguran la integridad de los datos.

Las tramas pueden ser transmitidas independientemente. Además se pueden insertar bloques de *metadatos*¹⁷ para etiquetar o identificar los parámetros de los flujos, como lo son la tasa de muestreo y el número de canales. La pequeña compresión que realiza FLAC está basada en la explotación, mediante predicción, de la correlación que existe entre tramas subsecuentes de audio y la explotación de la correlación entre canales, e.g. entre los canales izquierdo y derecho en el modo estéreo (Weinstein, 2005). Los

¹⁶La fundación Xiph.org es una corporación no lucrativa dedicada a la producción de herramientas de dominio público para el manejo de archivos multimedia. <http://xiph.org/>

¹⁷Los metadatos son datos que describen otros datos. El concepto de metadatos es análogo al uso de índices para localizar objetos en vez de datos.

métodos principales de predicción utilizados para modelar la señal de audio entregada son predicciones lineales fijas y predicciones lineales de respuesta finita al impulso (FIR, Finite Impulse Response).

El primer método utiliza predictores lineales fijos como *Shorten*, el cual es eficiente computacionalmente y puede ser extendido en FLAC. El algoritmo *Shorten* (Robinson, 1994) (que también puede hacer compresión con pérdidas) utiliza un modelo predictivo simple de forma de onda seguido por codificación Huffman de signo residual, mientras que FLAC utiliza un código alternativo llamado código Golomb-Rice.

La predicción FIR utiliza predictores hasta de orden 32, con coeficientes derivados de los coeficientes de autocorrelación de la señal y cuantificados antes de calcular la señal residual (Weinstein, 2005). La precisión de cuantificación puede variar de sub-trama (*subframe*) a sub-trama, dependiendo del tamaño del bloque (el número de muestras que son codificadas por bloque varía en el rango de 16 a 65,535) y el rango dinámico de la señal entregada.

Los pasos en la codificación FLAC son:

- Separar el archivo en bloques (*blocking*). Se forman bloques de muestra de un tamaño fijo para cada canal.
- Decorrelación intercanal. Cuando el archivo está en modo estéreo, se calcula un promedio de los canales izquierdo y derecho y la diferencia entre las señales de ambos canales.
- Predicción. Se predicen las muestras futuras y se genera una señal residual, que es la diferencia entre la señal actual y la predicción.
- Codificación residual.

V.2.10 Códec Ogg Vorbis

Vorbis es un códec perceptual de audio de propósito general, es un códec abierto y realiza una compresión con pérdida. Forma parte del proyecto Ogg de la fundación Xiph.org y entonces es llamado Ogg Vorbis o sólo Ogg por ser el códec más comúnmente encontrado en el contenedor Ogg.

Ogg

Ogg es un formato contenedor para multimedia de código abierto. Su funcionalidad es encapsular datos en crudo (*raw*) comprimidos y permitir el entrelazado (*interleaving*) de datos de video y audio en un sólo formato. Además de estas funciones, Ogg provee de empaquetamiento de tramas, detección de errores y marcas de tiempo (*timestamps*) para la búsqueda. En 2003 fueron publicados dos RFCs para el formato Ogg. El RFC 3533 define la estructura de la trama y el RFC 3534 define el tipo de contenido de internet.

Vorbis

Ogg Vorbis utiliza el formato del contenedor Ogg y puede contener una variedad distinta de esquemas de compresión de audio y video. Vorbis es un esquema específico de compresión que fue diseñado especialmente para música para ser contenido en Ogg.

La especificación de Vorbis es pública, sus bibliotecas están bajo una licencia BSD (Berkeley Software Distribution) modificada, y sus herramientas están bajo una licencia pública general de GNU.

Vorbis puede ser utilizado para codificar una variedad de tasas de datos desde 8 KHz hasta 192 KHz y una variedad de representaciones de canal de hasta 255 canales. Es un

códec de transformada monolítica adaptativa (forward-adaptive monolithic transform) basado en la MDCT, al igual que MP3 y AAC.

Los datos de dominio de frecuencia son divididos en ruido de fondo y componentes residuales, y después son cuantificados y la codificados con codificación entrópica utilizando un libro de códigos (*codebook*) basado en el algoritmo de cuantificación vectorial.

Ogg Vorbis codifica con una tasa VBR, por esta razón al seleccionar los parámetros de codificación no se da una lista de tasas de datos sino una escala de calidad, para seleccionar un valor de codificación. Dicha escala va en enteros del 1 al 10, siendo 10 el número que da una mejor calidad y menor compresión.

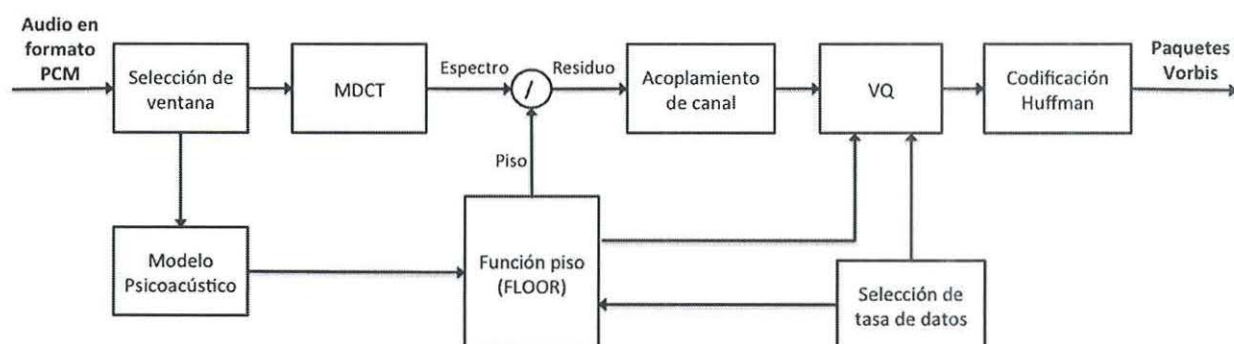


Figura 36. Diagrama a bloques del codificador Vorbis I.

En la Figura 36 se puede observar una de las posibles implementaciones del codificador Vorbis I. Esta implementación se basa en la cuantificación vectorial (VQ) y la transformación con superposición de ventanas, llamada MDCT. Las ventanas utilizadas pueden tener una de las longitudes especificadas (2048 o 512 muestras). Se utiliza la ventana más corta cuando se está codificando una señal de música que es crítica con cambios rápidos en el tiempo, e.g. percusión. Después de la transformación al dominio de la frecuencia, la señal es analizada por un modelo psicoacústico y la parte del espectro

que se considera inaudible por el modelo es removida. Entonces se genera el vector de piso (*floor vector*) para cada uno de los canales. Este vector es una representación de baja resolución del espectro de audio para el canal dado en la trama actual. El piso (*floor*) representa la curva como una representación de interpolación lineal por partes en una escala de amplitud en dB y una escala de frecuencia lineal.

El siguiente paso es el proceso de codificación es restar la curva de piso del espectro de audio de salida. Después de esta operación solo la estructura delicada del espectro permanece, la cual es conocido como *residuo*. Los vectores de residuo de ambos canales son transformados de la forma cartesiana a una representación polar. Este proceso es llamado acoplamiento de canal (*channel coupling*).

Después se utiliza cuantificación vectorial en cascada. Los resultados (incluyendo los libros de código VQ) son codificados con un algoritmo Huffman para eliminar aun más la redundancia.

El resultado final de todo el proceso es un paquete Vorbis de datos en crudo. Por último, estos paquetes son encapsulados en el contenedor Ogg y están listos para ser transmitidos. El decodificado de Vorbis no necesita tantos recursos computacionales como los formatos de MPEG, sin embargo, tiene más requerimientos de memoria. Esto es debido a que el decodificador no incluye ningún libro de código fijo. Los libros de código son específicos para cada flujo de datos y están incluidos en el encabezado de inicio. Los libros de código deben de ser almacenados en la memoria durante todo el tiempo de decodificación.

V.3 Evaluación de calidad de los códecs de audio

Una forma de onda de audio digital es sencillamente una variable que cambia con el tiempo, dentro de un espectro en frecuencia limitado, el cual depende de la tasa de muestreo. Si se toma en cuenta lo anterior, cualquier error introducido por un dispositivo determinado podría ser en principio extraído y analizado. Sin embargo, como ya se revisó en el Capítulo III, en el desarrollo de los códecs se incluyen características psicoacústicas, lo cual hace que la parte subjetiva juegue un papel importante.

Para evaluar la calidad del audio en los sistemas de telecomunicaciones existen dos tipos de pruebas: las evaluaciones objetivas y subjetivas; las cuales se describen a continuación. Además de la descripción de ambos tipos de evaluación se presentan los resultados obtenidos al evaluar el desempeño de los códecs estudiados para las señales de audio cardiaco y respiratorio.

V.3.1 Evaluación objetiva de códecs

La evaluación objetiva de señales de audio consiste en hacer mediciones que indiquen la exactitud con la cual se están reproduciendo dichas señales. Cuando las medidas son estandarizadas y repetibles estas forman una base para hacer la comparación.

Para evaluar la calidad de las pérdidas mediante una prueba objetiva, se registra la diferencia entre las formas de onda de la señal a medir y la original. La diferencia medida entre una señal de referencia y la señal que se pretende evaluar se considera una señal de error. Cualquier error en la forma de onda de una señal de audio puede ser considerada como una señal indeseada la cual ha sido linealmente agregada a la señal deseada (Spanias *et al.*, 2007).

Durante la etapa de análisis de los códecs se realizó la evaluación de estos com-

parando el coeficiente de correlación de Pearson para cada par de señales: las señales originales y las que pasaron por el proceso de codificación-decodificación. La expresión matemática del coeficiente de correlación de Pearson está dada por:

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (27)$$

Donde:

σ_{XY} es la covarianza de (X,Y)

σ_X es la desviación típica de la variable X

σ_Y es la desviación típica de la variable Y

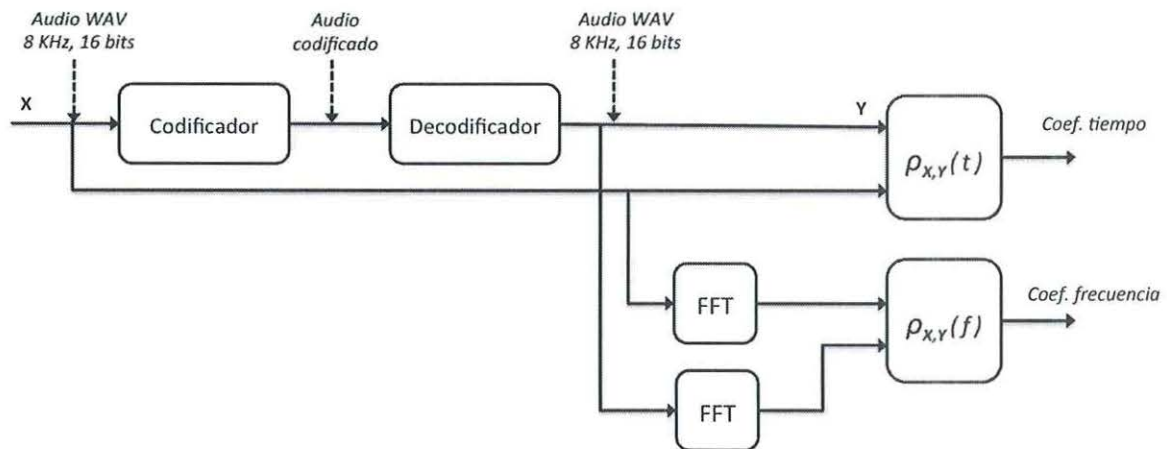


Figura 37. Procedimiento empleado para hacer la evaluación objetiva.

En la Figura 37 se muestra el esquema que se siguió para comparar las muestras. Para obtener los coeficientes de correlación de las señales en el dominio del tiempo $\rho_{X,Y}(t)$ se comparó la señal original con la señal que pasó por el proceso de codificación y decodificación. Para obtener los coeficientes de correlación de las señales en el dominio de la frecuencia se comparó el espectro de ambas señales.

Los cálculos fueron hechos con Matlab y los resultados se presentan en la Tabla VIII. En las columnas 4 y 5 de la Tabla VIII se muestra el promedio de los coeficientes de correlación de Pearson de 22 muestras de audio cardiaco, de las cuales sólo una es de un paciente en estado de salud cardiaca normal y el resto corresponde a ciertas patologías cardiacas.

Tabla VIII. *Evaluación de códecs por medio del coeficiente de correlación de Pearson.*

Códec	Bitrate entrada	Bitrate salida	Correlación tiempo	Correlación frecuencia	Tasa de compresión
FLAC	128 kbps	32 kbps	1.0000	1.0000	3.8
OGG-10	128 kbps	29 kbps	0.9999	0.9999	3.6
G.711 (Ley μ)	64 kbps	64 kbps	0.9998	0.9999	1.0
AAC	128 kbps	28 kbps	0.9996	0.9998	4.6
MP2	128 kbps	48 kbps	0.9994	1.0000	2.7
G.726	128 kbps	32 kbps	0.9980	0.9993	4.0
OGG-3	128 kbps	20 kbps	0.9958	0.9974	6.5
MP2	128 kbps	16 kbps	0.9956	0.9977	8.0
MP3	128 kbps	32 kbps	0.9995	0.9980	4.0
GSM	128 kbps	13 kbps	0.9838	0.9893	9.7
G.729A	128 kbps	8 kbps	0.0250	0.7171	16.0

Como era de esperarse, para FLAC ambos coeficientes de correlación fueron iguales a 1, puesto que éste es un códec sin pérdidas. Los códecs que obtuvieron mejores coeficientes de correlación en ambos dominios fueron Ogg Vorbis, AAC y G.711, siendo

códecs perceptuales con pérdida los dos primeros y un codificador de forma de onda el último. Con este primer acercamiento se buscó descartar algunos códecs, sin embargo, la variación entre los coeficientes de correlación de los distintos códecs es mínima (exceptuando el caso de G.729), lo cual no permite hacer una selección clara de códecs a descartar.

La correlación entre dos señales define el grado de similitud que existe entre ellas, y para algunos procesos estocásticos esta puede tener gran validez. Sin embargo, puesto que en las señales de audio el factor psicoacústico juega un papel importante, fue necesario recurrir a pruebas de evaluación subjetiva para hacer una mejor selección de los códecs a usar en la siguiente etapa de pruebas.

V.3.2 Evaluación subjetiva de códecs

La evaluación subjetiva de un sistema consiste en juzgar la calidad del sistema o de alguno de sus componentes desde el punto de vista de sus usuarios. Las pruebas subjetivas de calidad de audio se basan en la evaluación de la calidad del audio hecha por individuos a quienes se les presentan las muestras. El proceso de evaluación subjetiva de audio consiste en que el sujeto escuche un par de muestras de audio, una que es considerada como referencia y la muestra a evaluar. Después de escuchar dichas muestras, el sujeto pondera la calidad de acuerdo con una escala que le es dada previamente. Para las pruebas de evaluación existen algunos estándares internacionales de la ITU-T (Telecommunication Standardization Sector of International Telecommunication Union).

Para la evaluación subjetiva de los códecs se utilizaron dos programas:

- PEMO-Q. Basado en el Modelo de Percepción Oldenburg.
- ACQUA(Advanced Communication Quality Analysis system).- Está basado en

los estándares P.800 (1996), P.800.1 (2006) y P.834.1 (2010) de la ITU-T.

Ambos programas requieren del pago de una licencia para su uso.

PEMO-Q

El software PEMO-Q sirve para predecir la diferencia entre la calidad percibida de un par de señales de audio. Está basado en el trabajo de Huber y Kollmeier (2006), el cual a su vez sigue el planteamiento de Hansen (2000), utilizando el Modelo de Percepción Oldenburg (“PEMO”, por sus siglas en inglés) de Dau *et al.* (1996) para calcular las “representaciones internas” de los pares de señales. Estos pares son comparados cuantitativamente basado en el cálculo del coeficiente de correlación cruzada. El valor de correlación resultante sirve como una *medida objetiva* de la similaridad perceptual entre dos señales de audio.

El método de Hansen y Kollmeier fue originalmente desarrollado con la finalidad de predecir la calidad del audio de voz limitado dentro de la banda telefónica con las distorsiones causadas por los codecs de voz de bajas tasas de datos. Para esto, el software cuenta con un módulo llamado *SpeechQual*.

El módulo *AudioQual* se utiliza para predecir las degradaciones de señales generales de audio de banda ancha. La sensibilidad para distorsiones de menor calidad puede ser incrementada seleccionando un banco de filtros en lugar de un filtro pasa-bajas para el procesamiento de modulación en el modelo de percepción. Sin embargo, esto también incrementa el tiempo de cómputo considerablemente. *AudioQual* además de calcular la correlación global entre las representaciones internas (valor PSM de salida), también calcula un estimado de la calidad de audio instantánea (vector PSM_inst de salida).

En el presente trabajo de investigación se utilizó el módulo *AudioQual* del software PEMO-Q para evaluar las muestras de audio codificado. PEMO-Q corre sobre línea

de comandos de MS-DOS. *AudioQual* es un esquema que predice el resultado de las pruebas de evaluación subjetiva en forma de un Grado de Diferencia Objetiva (ODG, por sus siglas en inglés).

El parámetro ODG expresa el deterioro en la calidad de la señal de acuerdo con la siguiente escala de ponderación:

- 0: imperceptible
- -1: perceptible pero no molesto
- -2: un poco molesto
- -3: molesto
- -4: muy molesto

Antes de realizar las pruebas con el software PEMO-Q fue necesario sobremuestrear las señales a comparar (*upsampling*) a una frecuencia de muestreo de 32 KHz, pues el programa requiere que el audio esté en dicho formato. El proceso de *upsampling* fue realizado con el programa SoX v14.3.0 que corre sobre la plataforma LINUX. En la Figura 38 se puede ver un diagrama a bloques del procedimiento empleado para realizar las pruebas con el software PEMO-Q. En este proceso se evaluaron en total 135 muestras, las cuales corresponden tanto a PCGs como a audio respiratorio.

En la Tabla IX se observan los códecs que tuvieron mejores resultados para un muestreo de 8KHz y una cuantificación de 16 bits. Se aprecia que los valores de ODG son de un deterioro en la calidad imperceptible hasta perceptible pero no molesto, de acuerdo con la escala presentada anteriormente. En esta prueba, al igual que en la de evaluación objetiva, el códec FLAC mostró un mejor desempeño, solo que en esta

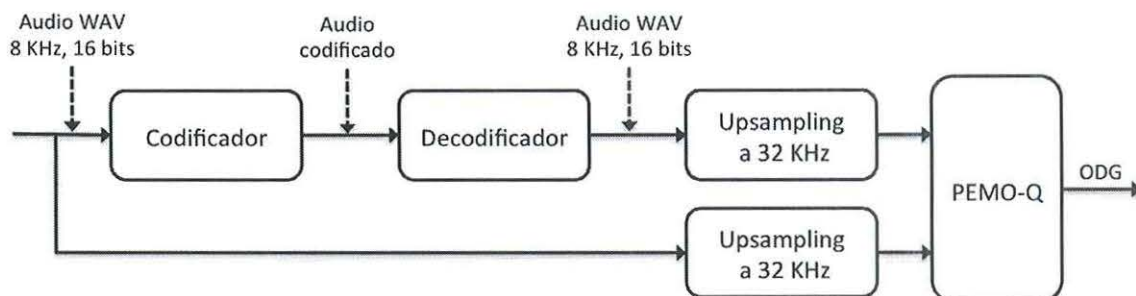


Figura 38. Procedimiento empleado para hacer la evaluación subjetiva con PEMO-Q.

ocasión el siguiente códec fue OGG (con calidad de compresión de 10) y posteriormente MP3 y AAC.

Tabla IX. Valores ODG muestreo 8KHz, cuantificación 16 bits.

	FLAC	FLAC	OGG	MP3	AAC	OGG
	1152	192	10			3
ODG prom	-0.0160	-0.0160	-0.2160	-0.4576	-0.8054	-1.4500
ODG max	-0.0003	-0.0004	-0.0297	-0.0022	-0.1325	-0.4806
ODG min	-0.1084	-0.1064	-1.6390	-2.5010	-2.5354	-3.1849
σ^2	0.0005	0.0005	0.0479	0.1735	0.2360	0.3366
Bitrate[kbps]	33-38	36-70	31-40	32	28	9-21
	VBR	VBR	VBR	CBR	CBR	VBR

En la Tabla X se presentan los códecs que tuvieron un desempeño más pobre para un muestreo de 8KHz y 16 bits de precisión. De acuerdo con la escala de ponderación, los valores obtenidos por estos códecs señalan un deterioro en la calidad que va de un poco molesto a molesto.

Tabla X. Valores ODG muestreo 8KHz, cuantificación 16 bits (continuación).

	G726	GSM	SPEEX	G729	MP2
ODG prom	-2.7535	-3.0999	-3.1327	-3.3061	-3.3523
ODG max	-0.3401	-2.5775	-2.1395	-3.0622	-3.2000
ODG min	-3.3846	-3.4025	-3.3958	-3.4095	-3.4155
σ^2	0.5028	0.0401	0.0670	0.0053	0.0016
Bitrate[kbps]	32	13	16	8	16
	CBR	CBR	CBR	CBR	CBR

Tabla XI. Valores ODG muestreo de 8KHz, cuantificación 8 bits.

	FLAC 192	Ley μ	MP3	AAC	Microsoft ADPCM	PCM	FLAC 1152
ODG prom	-3.2277	-3.2449	-3.2872	-3.2993	-3.3268	-3.3663	-3.3780
ODG max	-3.0229	-3.0113	-3.2182	-3.1822	-3.2449	-3.2874	-3.2379
ODG min	-3.3563	-3.3526	-3.3431	-3.3434	-3.3720	-3.4064	-3.4109
σ^2	0.0048	0.0043	0.0012	0.0011	0.0011	0.0006	0.0009
Bitrate[kbps]	19-24	64	17-20	27-37	33	64	33-38
	VBR	CBR	VBR	VBR	CBR	CBR	VBR

Es importante señalar que los códecs presentes en la Tabla X han sido diseñados para señales de voz, excepto el códec MP2, que es de propósito general.

Los valores de ODG para señales con 8KHz de muestreo y 8 bits de precisión en la cuantificación son presentados en la Tabla XI. Los resultados no fueron los esperados,

puesto que para PCM se esperaría un deterioro nulo en la calidad, sin embargo el valor de ODG obtenido fue de un promedio de -3.3663, lo cual significa que el deterioro en la calidad es molesto. Si embargo, hay que señalar que el software está diseñado solo para evaluar muestras con cuantificación de 16 bits de precisión, así que estos resultados no fueron considerados para la selección de códecs.

Se procedió a evaluar señales de audio con un muestreo de 4KHz y una cuantificación de 16 bits de precisión. En esta ocasión solo se evaluaron los tres códecs que habían mostrado el mejor desempeño en las pruebas anteriores: MP3, AAC y FLAC. Aunque OGG obtuvo también buenos resultados, no fue considerado para esta prueba puesto que el codificador no permite utilizar una tasa de muestreo por debajo de los 8 KHz. Los resultados de esta evaluación son los que se presentan en la Tabla XII.

Tabla XII. Valores ODG muestreo de 4KHz, cuantificación 16 bits.

	FLAC 192	AAC (iTunes)	MP3	AAC (nero)	AAC (faac)
ODG prom	-0.0044	-0.2623	-0.7725	-3.2584	-3.4100
ODG max	-0.0003	-0.0758	-0.0149	-2.3677	-3.3799
ODG min	-0.0270	-1.9626	-2.8131	-3.3944	-3.4213
σ^2	0.0000	0.0713	1.1373	0.0269	6
Bitrate[kbps]	20	11	8	8	12
	VBR	VBR	CBR	VBR	VBR

En el caso del códec AAC se probó con tres implementaciones de software que codifica archivos en éste formato, las cuales son: iTunes, Nero y Faac, siendo la primera un software propietario de Apple y las dos últimas programas de código abierto que

corren en línea de comandos. Los resultados muestran que aunque codifiquen en el mismo formato no necesariamente se obtiene el mismo resultado en calidad del audio cuando se utilizan distintas implementaciones de un códec. Por otro lado, FLAC y MP3 siguen mostrando un buen desempeño aunque se disminuya la tasa de datos, así como la implementación de AAC en iTunes.

MOS (Mean Opinion Score)

La nota media de opinión (MOS, Mean Opinion Score) se define en la recomendación de la ITU-T (P.10/G.100, 2010) del siguiente modo: valor de una escala predefinida que un sujeto asigna a su propia opinión sobre la calidad de funcionamiento del sistema de transmisión telefónica utilizado para una conversación o únicamente para una escucha de material hablado (P.800.1, 2006).

TOSQA (Telecommunications Objective Speech Quality Assessment)

Al igual que otros métodos utilizados para predecir la calidad de la voz, *TOSQA* y *TOSQA2001* son utilizados para modelar los resultados de los pruebas de audición, es decir, proporcionan los resultados que se podrían esperar de una prueba de audición si una señal de voz hablada fuera evaluada subjetivamente por personas. El método de TOSQA puede ser aplicado a un escenario en donde se transmite una señal de voz, por medio de un objeto de medición con grabación en una interfaz eléctrica. La evaluación de una grabación acústica por medio de una terminal no es posible, ya que en estos métodos una terminal está completamente integrada, y por lo tanto es parte del propio método. Esto significa que la terminal no puede ser incluida en la evaluación de calidad. Por el contrario, TOSQA2001 permite ésta evaluación de una terminal en la trayectoria de la transmisión y la grabación acústica de las señales transmitidas

como una opción adicional. Mientras que TOSQA se encuentra limitada a grabaciones eléctricas, TOSQA2001 permite tanto la evaluación de señales grabadas eléctricamente como acústicamente (por medio de terminales) (TOSQA, 2003).

Tanto en TOSQA como en TOSQA2001 se evalúan los siguientes parámetros:

- Valor TOSQA
- Valor TMOS
- Factor de deterioro por equipo (Equipment Impairment Factor IE)
- Retardo

El resultado más importante para uso práctico es el valor TOSQA, el cual describe la calidad de la señal de voz transmitida. Del valor máximo de TOSQA que es 100, se resta el grado de deterioro de la señal transmitida. Un valor de 90 indica voz de alta calidad, mientras que un valor por debajo de 50 corresponde a calidad de voz inaceptable. TMOS (TOSQA Mean Opinion Score) es derivado del valor TOSQA. Se basa en una prueba de audición en donde los sujetos de prueba dan una calificación después de haber escuchado una muestra una señal de voz. La escala es de 5 puntos, de acuerdo con (P.834.1, 2010), con la siguiente asignación de la puntuación:

- 5: excelente
- 4: bueno
- 3: regular
- 2: pobre
- 1: malo

El promedio de las puntuaciones dadas por todos los sujetos de prueba para las muestras de voz. Un valor de MOS aproximado de 4.2 típicamente corresponde a la calidad máxima posible vía sistemas de transmisión telefónicos de banda limitada. Interesantemente, el valor de MOS igual a 5.0 nunca es alcanzado en una prueba de audición.

Se realizaron pruebas con el software ACQUA que fue facilitado por la empresa PLAMEX Tijuana. Mediante este software es posible calcular el TOSQA, el factor MOS y el factor de deterioro que presenta una señal al compararla con otra señal de referencia. Los resultados de dichas pruebas son presentados en las Tablas XIII y XIV.

Tabla XIII. *Parámetros de evaluación con software ACQUA, muestras a 8KHz con cuantificación de 16 bits.*

Códec	TOSQA	Factor TMOS	Factor de deterioro
FLAC	100	4.22	-2.1
AAC	100	4.22	-2.1
MP3	100	4.22	-2.1
G.729	95.3	3.75	7.6
GSM	94.7	3.6	8.3
MP2	90.5	3.35	16.5
G.726	89.3	3.26	18.5
OGG	76.9	2.48	37.3
Ley μ	49.1	1	66.7

Como se aprecia en la Tabla XIII, nuevamente los mejores resultados son obtenidos al utilizar los códecs FLAC, AAC y MP3 para señales de audio con una frecuencia de muestreo de 8 KHz y una cuantificación de 16 bits.

Tabla XIV. *Parámetros de evaluación con software ACQUA, muestras a 8KHz con cuantificación de 8 bits.*

Códec	TOSQA	Factor TMOS	Factor de deterioro
FLAC 1152	100	4.22	-2.1
FLAC 192	100	4.22	-2.1
MP3	100	4.22	-2.1
AAC	97.9	4	2.4

Cabe mencionar que aunque el códec OGG había obtenido resultados satisfactorios en las pruebas de evaluación objetiva, en las pruebas subjetivas realizadas con el software ACQUA no presentó un desempeño aceptable. En la Tabla XIII se puede observar que OGG presenta calificaciones incluso por debajo de las obtenidas por los vocoders.

Al mantener la tasa de muestreo pero disminuir los bits de cuantificación, el desempeño de FLAC y MP3 sigue siendo el mejor, sin embargo, para AAC se ven puntuaciones menores aunque siguen siendo bastante aceptables.

V.3.3 Conclusión

Analizando los resultados obtenidos de las pruebas de evaluación de calidad a las que fueron sometidos los códecs estudiados se obtuvieron las siguientes conclusiones:

Los códecs que fueron originalmente diseñados para voz (G.726, G.729, Speex, GSM, Microsoft ADPCM) presentaron una evaluación de calidad pobre para señales de audio cardiorrespiratorio, por lo cual quedan descartados para la etapa de evaluación sobre redes inalámbricas. Se piensa que el bajo desempeño de este tipo de códecs para señales de audio cardiorrespiratorio es debido a que utilizan un modelo predictivo basado en la

voz humana.

Por otro lado los códecs que son de propósito general mostraron un mejor desempeño con para el audio cardiorrespiratorio. Sin embargo, en las pruebas subjetivas MP2 no obtuvo muy buenos resultados, por lo que también se descartó para la siguiente etapa.

Los códecs seleccionados para su análisis de desempeño en el escenario propuesto fueron: FLAC, AAC y MP3.

Capítulo VI

Aplicación de los códecs seleccionados en el escenario propuesto

VI.1 Introducción

En este capítulo se analiza el modo en que quedan conformadas las tramas para los tres códecs seleccionados con las pruebas anteriormente descritas. Estos códecs son: AAC, MP3 y FLAC.

Adicionalmente se muestran los resultados de la transmisión del audio cardiorrespiratorio codificado y enviado sobre una red inalámbrica de área personal bajo el estándar IEEE 802.15.4, tanto en simulación como de manera experimental.

VI.2 Consideraciones

De los códecs seleccionados para esta nueva etapa de análisis dos son codificadores perceptuales (MP3 y AAC) y el último es un codificador sin pérdidas. En el Capítulo V ya se explicó el funcionamiento de estos códecs, sin embargo, para transmitir el audio codificado aun es necesario agregar una etapa: la paquetización, es decir, adaptar las tramas de audio codificador al tamaño especificado para cada unidad de información que puede ser transmitida (paquete).

En la Figura 39 se muestra un diagrama a bloques que ilustra mejor esta idea.

El audio cardiorrespiratorio que originalmente se encuentra en modo analógico tiene

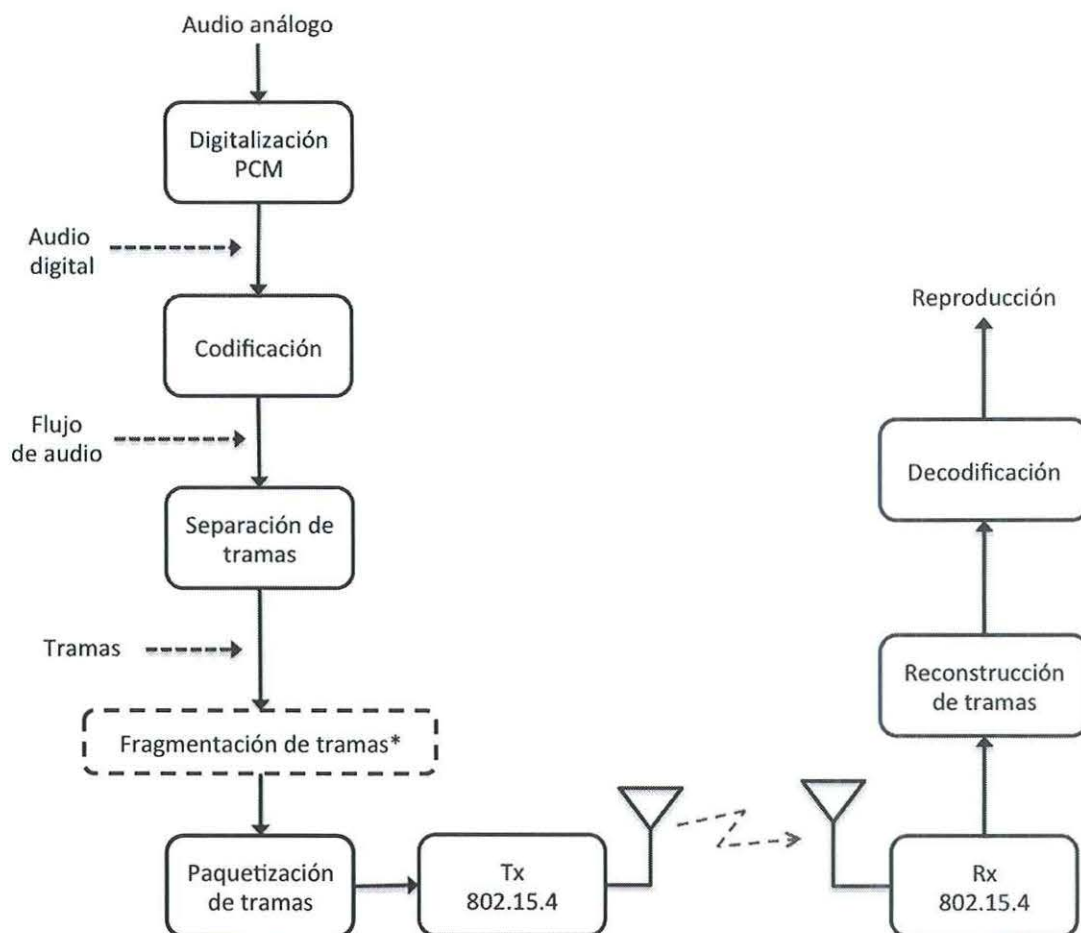


Figura 39. Diagrama del proceso de digitalización, codificación y transmisión del audio cardiorrespiratorio sobre una red IEEE 802.15.4.

que ser digitalizado para su procesamiento. Al digitalizarlo el audio queda en formato PCM, muestreado a 8KHz y con 16 u 8 bits de precisión en la cuantificación, según sea el requerimiento del codificador. Posteriormente el audio pasa por el proceso de codificación por medio del cual el audio queda comprimido.

A la salida del codificador se tiene un flujo de audio el cual va seccionado en partes llamadas tramas. Una trama de audio es la unidad mínima que puede ser decodificada y el tamaño de esta trama varía de un códec a otro. Por ejemplo, para FLAC se pueden

seleccionar 1152 muestras (entre otros valores) a la entrada del codificador para obtener a la salida una trama de audio.

Antes de pasar por el proceso de paquetización, el flujo de audio tiene que ser entonces separado en tramas y estas tramas ahora sí pueden ser paquetizadas. Este proceso es hecho de una forma similar a lo que ocurre con RTP. Si el tamaño de la trama es menor al tamaño máximo de la carga útil de un paquete entonces la trama se empaqueta de manera íntegra en el paquete. Si por el contrario, la trama es mayor al tamaño máximo de la carga útil la trama debe de ser fragmentada.

Una vez listos los paquetes se envían con un radio IEEE 802.15.4 y son recibidos por otro radio que opera bajo el mismo estándar. Posteriormente las tramas deben de ser reconstruidas para poder ser decodificadas y dejar el audio listo para ser reproducido.

A continuación se hace un análisis de la estructura de trama para cada codificador

VI.3 Análisis de tramas FLAC

Cada trama FLAC tiene el formato mostrado en la Figura 40. Como se observa en dicha figura, al inicio cada trama o *frame* hay un código de sincronía. Se dividió cada archivo FLAC en tramas para analizar el tamaño de cada una de ellas.

Para este fin se buscaron los códigos de sincronía dentro de cada archivo. Una vez encontrados estos códigos se procedió a hacer la fragmentación del archivo y el análisis de los tamaños de trama mediante un programa en C++.

Para un archivo codificado con FLAC en que cada trama codificada tiene a la entrada 192 muestras de audio en formato PCM muestreado a 8 KHz, se tienen 8000 muestras por segundo, por lo tanto:

$$T_{tramaFLAC} = \frac{192 \text{ muestras}}{8000 \text{ muestras/s}} = 24ms \quad (28)$$

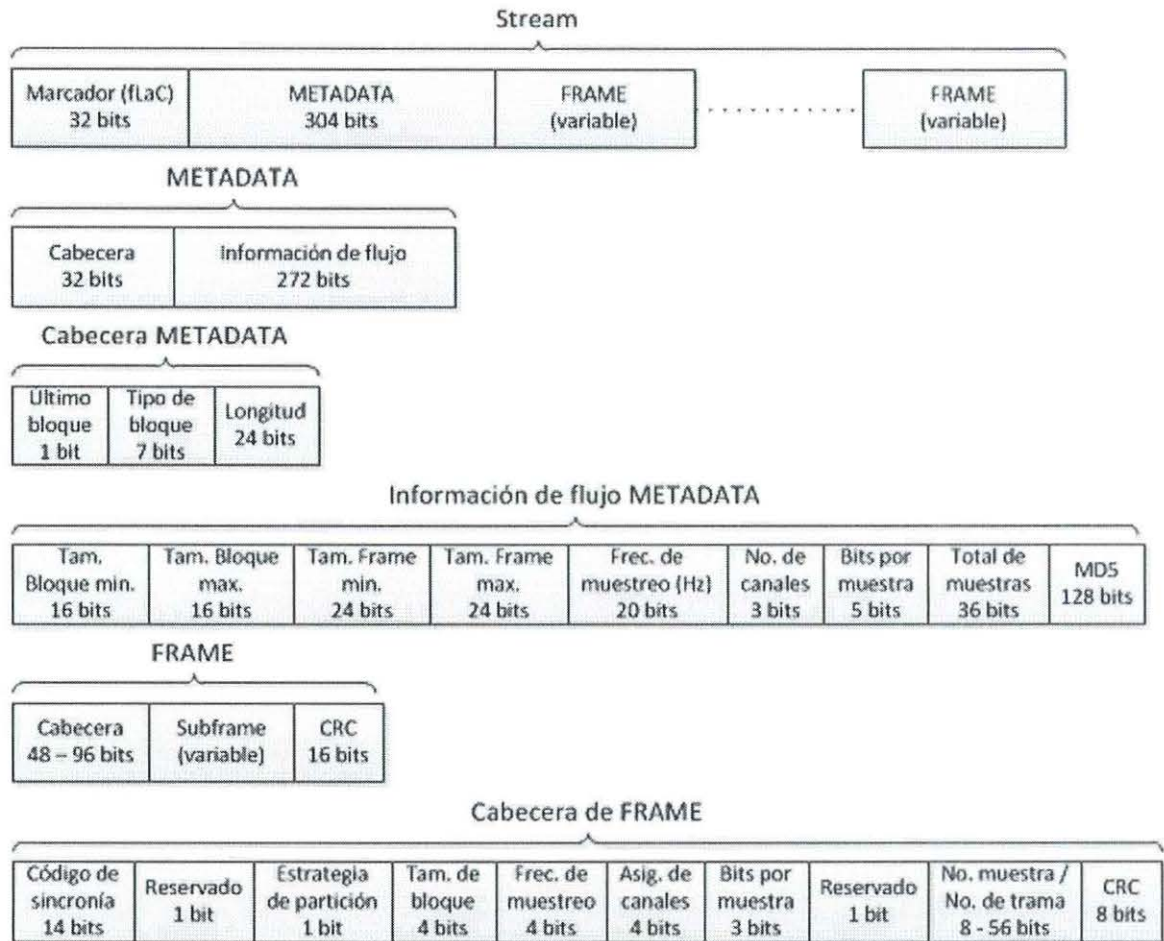


Figura 40. Formato de tramas codificadas con FLAC.

De la ecuación 21 (Capítulo IV Sección 3.4) se tiene que $T_{total802.15.4} = 7.168ms$ por lo tanto, de acuerdo con la ecuación 28 (Capítulo VI Sección 3), una trama FLAC de 192 muestras debe ser transmitida en un máximo de 3 paquetes IEEE 802.15.4 (24 ms / 7.168 ms) para que se considere una transmisión en tiempo real .

Además, el acomodo de los segmentos de trama FLAC en paquetes IEEE 802.15.4 fue hecho de la siguiente manera: un paquete IEEE 802.15.4 puede contener una fracción de trama FLAC o un número entero de tramas FLAC. Esta restricción está basada en la

forma en que hace la paquetización el códec Vorbis (RFC-5215, 2008). Esta restricción es establecida para proteger la información de la pérdida de paquetes, puesto que si se envían 2 o más fracciones de tramas distintas en un sólo paquete y este se pierde se tendrían que descartar los demás paquetes asociados a dichas fracciones de trama, pues no podría hacerse la decodificación de una trama incompleta en el receptor. Con lo anterior se perdería una cantidad de información aun mayor. En cambio, si sólo se permite que se transmita una fracción de trama por paquete sólo se tendrían que descartar los paquetes asociados a esa trama.

Considerando el archivo de audio número 73, el cual presentó la trama con la menor compresión (318 bytes en la trama 260), se presenta el siguiente análisis.

Tabla XV. *Tamaño de las tramas FLAC con 192 muestras de audio PCM por bloque codificado para el archivo de audio número 73.*

No. Trama	258	259	260	261	262	263
Tamaño (bytes)	297	312	318	302	297	298

Como se señaló en el Capítulo IV, se sabe que la carga útil máxima en los paquetes IEEE 802.15.4 es de 102 bytes, sin embargo, se limitará el tamaño de los paquetes a 90 bytes. Lo anterior es debido a que al realizar pruebas experimentales con los transeptores X-Bee este fue el tamaño de carga máxima permitida.

En la Tabla XV se presenta el tamaño de las tramas de audio comprimidas con FLAC tomando 192 muestras por bloque. Para paquetizar la trama 258 esta es dividida en 4 paquetes, 3 de 90 bytes y uno de 27 bytes, los cuales se exceden el límite de los 3 paquetes IEEE 802.15.4 por trama FLAC. Posteriormente, la muestra 259 tiene que ser dividida en 4 paquetes, 3 de 90 y uno de 42 bytes, lo cual tampoco cumple con la

restricción. Lo mismo ocurre con las muestras siguientes.

Se podría pensar en añadir un buffer en el receptor, el cual no debe superar los 100 ms (España Boquera, 2003) para que la transmisión pueda ser considerada en tiempo real. Considerando el archivo 73 y la fragmentación para los paquetes IEEE 802.15.4 mencionada anteriormente se tiene que este archivo debe ser segmentado en 906 paquetes. Si se sabe que cada paquete se transmite en aproximadamente 8 ms, el tiempo total para la transmisión de dicho archivo sería de 7.248s. Considerando que la duración del archivo de audio es de 6.576s se estima que el buffer mínimo necesario sería de 672 ms, lo cual supera el límite de 100 ms.

Por este motivo se concluye que no es posible utilizar el códec FLAC con 8 KHz de muestreo y 16 bits de cuantificación para la transmisión de señales cardiorespiratorias.

VI.4 Análisis de tramas AAC

Para analizar las tramas AAC se siguió el mismo procedimiento que se llevó a cabo con FLAC. Se buscaron varias implementaciones del códec AAC. Se codificaron los archivos con tres softwares, iTunes, nero y faac, siendo los dos últimos programas que corren en línea de comandos, tanto en la plataforma Windows como en Linux.

Ninguna de estas tres implementaciones codificó los archivos como CBR, pues aunque se especificó esto en los parámetros de entrada, lo que dan estos codificadores a la salida es un ABR, es decir, una tasa de datos promedio. Debido a esto fue necesario buscar la palabra de sincronía en los archivos.

Se hizo la búsqueda de la palabra de sincronía, sin embargo, no fue posible encontrarla en los archivos de ninguno de los 3 codificadores. Por este motivo se tuvo que descartar el códec AAC para las pruebas experimentales.

VI.5 Análisis de tramas MP3

Como se comentó en el Capítulo IV cada trama de MP3 contiene la información de 1152 muestras de audio, por lo tanto, el audio contenido en una trama MP3 equivale en tiempo a:

$$T_{tramaMP3} = \frac{1152 \text{ muestras}}{8000 \text{ muestras/s}} = 144ms \quad (29)$$

De la ecuación 21 se tiene que $T_{total802.15.4} = 7.168ms$, por lo tanto, de acuerdo con la ecuación 29 una trama MP3 puede ser transmitida con un máximo de 20 paquetes IEEE 802.15.4. (144 ms / 7.168 ms).

MP3 fue codificado con una tasa de bits constante. El tamaño cada trama es de 288 bytes. Entonces, si se tiene una capacidad de 90 bytes de carga máxima por paquete IEEE 802.15.4, esto implica que cada trama requiere de tan solo 4 paquetes, número que se encuentra muy por debajo del límite de 20 paquetes por trama. Debido a esto MP3 fue seleccionado para pasar a la etapa de experimentación.

VI.6 Etapa de simulación

La plataforma elegida para hacer la simulación de las redes inalámbricas de área personal fue NS-2, software que opera bajo la plataforma Linux. NS-2 cuenta con módulos para simular el comportamiento de las capas de distintos estándares de telecomunicaciones.

La simulación estará basada en la implementación del estándar IEEE 802.15.4 hecha por Zheng y Lee [2003]. Dicha implementación simula la capa de control de acceso al medio (MAC) de IEEE 802.15.4.

Se realizó una simulación en el programa NS-2 bajo dos escenarios. En ambos escenarios fueron consideradas las siguientes condiciones:

- Los paquetes faro (beacons) no están habilitados.
- Se habilita el envío de reconocimientos (paquetes ACK).
- Se permite un número máximo de 3 retransmisiones.
- La separación entre el coordinador PAN y el otro dispositivo es de 5 metros.

En el primer escenario se tienen 2 transceptores IEEE 802.15.4. Se considera bajo este escenario que uno de los nodos es el coordinador PAN, el cual recibe el audio codificado que es transmitido por el otro nodo. Se analizó el archivo de trazas resultante de la simulación y se encontró que la tasa de pérdida de paquetes (PLR por sus siglas en inglés) fue igual a cero, es decir, todos los paquetes fueron recibidos exitosamente.

En el segundo escenario se tienen 4 pares de transceptores operando todos en el estándar IEEE 802.15.4 sobre el mismo canal, pero cada par en una red independiente. Cada uno de los pares está formado por un coordinador PAN y un dispositivo final que requiere transmitir. Para el primer par de nodos, la fuente de datos es CBR con tasa de 30 kbps y se considera que en los paquetes es transportado el audio codificado de un sólo archivo. Los 3 pares de nodos restantes se comportan como interferentes. A cada dispositivo final de dichos pares de nodos le fue asignada una fuente de tráfico de tipo Poisson con tasa de 5 kbps. El modelo de canal utilizado (*propagation model*) para la simulación es el tipo *shadowing*. En la Figura 41 se muestra la topología de red del escenario de transmisión bajo interferencia. Los nodos 0 y 1 corresponden a la red en que se transmite el audio cardiorrespiratorio.

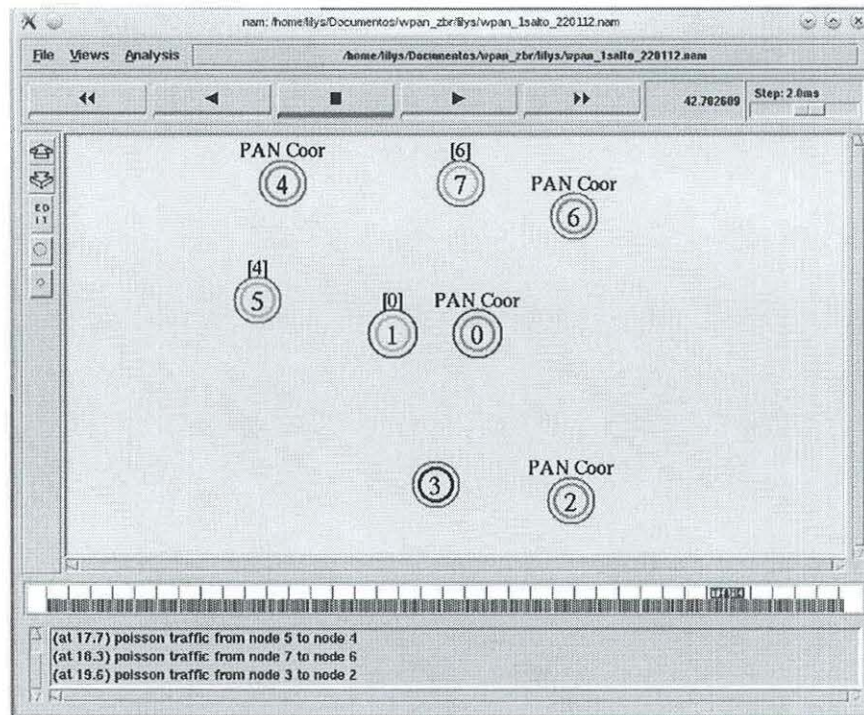


Figura 41. Topología de red del escenario de transmisión bajo interferencia.

Se transmitieron 500 paquetes y el resultado fue $PLR = 0.016$, es decir se perdieron el 1.6 % de los paquetes.

Es necesario puntualizar que NS-2 es un simulador basado en eventos, es decir como resultado de la simulación se obtiene un archivo en el cual se agendan los paquetes recibidos, enviados y descartados por cada nodo de la red. Por este motivo, no se pudieron obtener resultados de la capa física, es decir, sólo se sabe que un paquete ha llegado a su destino, pero no se sabe si ha llegado de un modo incorrupto o ha llegado con errores.

Partiendo de esto, se considera que todos los paquetes que se recibieron llegaron sin corrupción en la información y se procedió a eliminar del archivo codificado las tramas

correspondientes a los paquetes no recibidos. Es necesario puntualizar que como las tramas fueron fragmentadas, en caso de no recibir tan sólo un paquete correspondiente a un fragmento de la trama se tuvieron que descartar también los otros paquetes recibidos correspondientes a la misma trama, puesto que si se trata de decodificar una trama con información incompleta el resultado es un audio ruidoso.

El espacio de las tramas que fueron eliminadas del archivo codificado fue rellenado con ceros. Después de haber decimado¹⁸ el archivo se realizaron pruebas de evaluación subjetiva para comparar el archivo recibido con el original (antes de ser codificado).

Los resultados de esta evaluación se muestran en la Tablas XVI y XVII.

Tabla XVI. Valores ODG de señales obtenidas de la simulación de transmisión del audio cardíaco sobre una red IEEE 802.15.4 sin interferencia.

	FLAC	FLAC	MP3
ODG prom	-0.1138	-0.0163	-0.2160
ODG max	-0.0023	-0.0007	-0.0297
ODG min	-0.1237	-0.1062	-1.6390
σ^2	0.0012	0.0005	0.0479
f_m/q	8KHz 8b	4 KHz 16b	8 KHz 16b

Se puede observar en la Tabla XVI que los resultados son similares a los presentados en la Tabla IX del Capítulo V, lo que era de esperarse puesto que bajo este escenario se tuvo una pérdida de paquetes nula $PLR = 0$.

Por otro lado, la Tabla XVII muestra que la interferencia afectó en gran medida a las

¹⁸En este contexto, cuando hablamos de decimar una señal se hace referencia a la eliminación de muestras de la misma.

Tabla XVII. Valores ODG de señales obtenidas de la simulación de transmisión del audio cardíaco sobre una red IEEE 802.15.4 con 3 interferentes.

	FLAC	FLAC	MP3
ODG prom	-0.530	-0.2740	-0.4582
ODG max	-0.2480	-0.0971	-0.0587
ODG min	-1.1064	-0.8922	-0.690
σ^2	0.0105	0.0052	0.0034
f_m/q	8KHz 8b	4 KHz 16b	8 KHz 16b

señales, puesto que los niveles de ODG decayeron considerablemente. Sin embargo, es importante señalar que se colocaron 3 interferentes, lo cual es un ambiente muy hostil.

Otro aspecto importante es que para el codificador FLAC se obtuvieron mejores resultados con la combinación de 4KHz de tasa de muestreo y 16 bits de cuantificación. La disminución en la tasa de muestreo no afectó el desempeño de FLAC porque las señales que se tienen en nuestro banco de muestras cardiorrespiratorias están muy por debajo de 2 KHz.

Sin embargo al disminuir los bits de cuantificación y mantener la tasa de muestreo hubo resultados menos satisfactorios. Suponemos que esto es debido al ruido de cuantificación resultante de utilizar 8 bits para muestras que fueron cuantificadas originalmente en 16 bits.

VI.7 Etapa de experimentación

Se implementó una red IEEE 802.15.4 de 2 nodos. Uno de los nodos fue conectado a una computadora en donde se almacenaron previamente los archivos de audio codificados. Esta conexión fue hecha por el puerto RS-232. El nodo coordinador PAN a su vez fue conectado a otra computadora mediante el puerto USB como se muestra en la Figura 42.

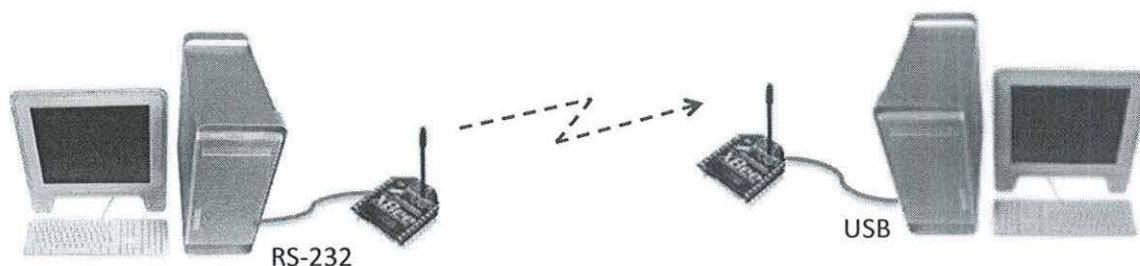


Figura 42. *Diagrama de transmisión de audio codificado con módulos XBee-PRO.*

Se transmitieron los 135 archivos codificados con los parámetros que se muestran en la Tabla XVIII.

Las condiciones del experimento fueron las siguientes:

- Distancia entre los dispositivos XBee-PRO igual a 1 metro.
- Interferencia de un ambiente de oficina, es decir tipo WiFi.
- Esquema de red sin paquetes fijo habilitados.
- Sin envío de ACKs, por lo tanto sin solicitud de retransmisiones.

Una vez recibidos todos los archivos y almacenados en la computadora, se procedió a decodificarlos y someter algunos de ellos a pruebas de evaluación subjetiva con el

Tabla XVIII. *Parámetros utilizados para la codificación de los archivos transmitidos en la etapa experimental.*

Códec	Tasa de muestreo	Bits de cuantificación
MP3	8 KHz	16 b
FLAC	8 KHz	8 b
FLAC	4 KHz	16 b

software ACQUA para evaluar la calidad del audio de los archivos recibidos. Esto debido a que, como se comentó anteriormente, el hecho de recibir un paquete no garantiza la incorruptibilidad del mismo.

Se evaluaron 5 archivos para los códecs FLAC y MP3. Los archivos evaluados fueron los más representativos del conjunto general de muestras de audio cardiaco y respiratorio. Los resultados señalaron que no hubo errores en los paquetes, puesto que se obtuvieron las puntuaciones máximas en todos los parámetros para todas las muestras de audio sometidas a la evaluación, como se muestra en la Tabla XIX.

Tabla XIX. *Parámetros de evaluación final con software ACQUA.*

Códec	f_m/q	TOSQA	Factor TMOS	Factor de deterioro
FLAC 192	8KHz 8b	100	4.22	-2.1
FLAC 192	4KHz 16b	100	4.22	-2.1
MP3	8KHz 16 b	100	4.22	-2.1

Capítulo VII

Conclusiones y trabajo futuro

VII.1 Introducción

Este capítulo tiene como objetivo mostrar las conclusiones y aportaciones generadas en el desarrollo de esta tesis, las cuales permiten determinar que el objetivo de la tesis ha sido cumplido en su totalidad. Adicionalmente se describen algunas propuestas para la realización de trabajos futuros sobre la misma línea de investigación.

VII.2 Conclusiones

Se analizaron códecs de audio existentes aplicados a la compresión de audio cardiorrespiratorio. De la evaluación de calidad y posteriormente, la evaluación del desempeño sobre una red IEEE 802.15.4 simulada y experimental, se obtuvieron las siguientes conclusiones:

- Mediante la evaluación subjetiva de las señales del audio cardiorrespiratorio se descartaron varios códecs, principalmente los que están diseñados para señales de voz (vocoders). Era de esperarse este resultado puesto que este tipo de códecs se basa en modelos que tratan de reproducir la voz de manera inteligible para el oído, los cuales no tienen un buen desempeño al ser aplicados a señales distintas a la voz.
- Los códecs de propósito general brindaron una mejor calidad para la compresión

de señales del audio cardiorrespiratorio. En particular, al hacer la evaluación subjetiva de audio se decidió seleccionar a los que presentaron las mejores valoraciones, los cuales fueron: MP3, AAC y FLAC.

- El códec AAC fue descartado para las pruebas experimentales. Esto fue debido a que se tuvieron dificultades para encontrar la palabra de sincronía en los archivos codificados. Esto aun y cuando se probó con varios programas para hacer la codificación en el formato AAC.
- Los códecos FLAC y MP3 ofrecen un buen desempeño para señales de audio cardíaco y respiratorio. Sin embargo, fue necesario tomar en cuenta la tasa de bits generada por trama para verificar su utilización en la red IEEE 802.15.4. Aunque el códec FLAC ofrece muy buena calidad de audio utilizando un muestreo de 8KHz y una precisión en la cuantificación de 16 bits, se encontró que no es factible utilizarlo sobre la red propuesta. Esto se debe a que el códec genera tramas de longitud variable. Para el caso de las señales analizadas, la longitud de estas tramas no permitiría su transmisión en tiempo real, lo que ocasionaría retrasos inaceptables en el proceso de decodificación. Es importante hacer notar que aun y cuando en apariencia el códec comprime la señal lo suficiente como para transmitirla sobre el enlace, mucha de la compresión se logra en los segmentos en los cuales no hay sonido, mientras que los segmentos con sonido generan tramas FLAC relativamente grandes.
- El códec FLAC puede ser utilizado con 4 KHz de muestreo y 16 bits de precisión y también con 8 KHz de muestreo y 8 bits de precisión, prefiriendo el primer par de valores para tener menor ruido de cuantificación. Para estos casos se demostró que se puede transmitir el audio cardíaco.

- El códec MP3 ofrece la ventaja de generar tramas de longitud fija, lo cual facilitó el proceso de empaquetamiento y transmisión. En este sentido podemos concluir que la utilización de este códec es adecuada para la transmisión de señales cardiorrespiratorias sobre redes inalámbricas con tasas de datos bajas/medianas.
- Para un ambiente de operación real se sugiere la utilización de códec que generen una tasa de bits constante (CBR) para la transmisión de audio sobre redes de datos, pues este tipo de codificación facilita la paquetización del audio.
- Aunque existan varias implementaciones de algún codificador, esto no significa que el resultado a la salida de cada una de ellas sea igual al seleccionar los mismos parámetros (tasa de muestreo, precisión en la cuantificación, tasa de datos). Esto quedó demostrado con las pruebas de evaluación subjetiva que se hicieron del códec AAC.

VII.3 Trabajo futuro

Como resultado del trabajo de investigación realizado en esta tesis y con la experiencia adquirida en el tema, se hacen las siguientes propuestas para la realización de trabajos futuros sobre la misma línea de investigación:

- Buscar otras alternativas para codificar el audio cardiorrespiratorio en formato AAC para que sea posible hacer la fragmentación en paquetes.
- Evaluar otros códec sin pérdida, para ver si dan un mejor desempeño que FLAC, que fue el códec sin pérdidas evaluado en el presente trabajo de investigación.

- Desarrollar una base de datos estandarizada de audio cardiaco y respiratorio con el estetoscopio Littman BT.
- Una vez formada la base de datos, hacer nuevas pruebas con los códecs elegidos en el presente trabajo de tesis y repetir el experimento planteado en este trabajo de investigación. Finalmente someter las muestras de audio resultantes a una evaluación hecha por profesionales de la salud.
- Publicar los resultados obtenidos de experimento descrito en el punto anterior.
- Implementar un sistema de estetoscopio digital que incorpore un codificador FLAC con 4 KHz 16 bits o un codificador MP3 con 8KHz 16 bits.
- Desarrollar un códec diseñado especialmente para audio de señales biomédicas.

Referencias bibliográficas

- Bates, B., Bickley, L., y Szilagyi, P. (2008). *Bates' guide to physical examination and history taking*. Guide to Physical Exam & History Taking. Wolters Kluwer Health/Lippincott Williams & Wilkins. ISBN 9780781780582. 964 p.
- Bing, B. (2008). *Emerging technologies in wireless LANs: theory, design, and deployment*. Cambridge University Press, USA, primera edición. 476 p.
- Bruno, R., Conti, M., y Gregori, E. (2002). Bluetooth: Architecture, protocols and scheduling algorithms. *Cluster Computing*, 5: 117–131.
- Cheitlin, M., Sokolow, M., y McIlroy, M. (1993). *Clinical cardiology*. Lange medical book. Prentice-Hall International. ISBN 9780838500125. 659 p.
- Chiras, D. (2011). *Human Biology*. Jones & Bartlett Learning. ISBN 9780763783457. 494 p.
- Dau, T., Püschel, D., y Kohlrausch, A. (1996). A quantitative model of the “effective” signal processing in the auditory system: I. model structure. *J. Acoust. Soc. Am.*, 6(99): 3615–3622.
- Debbal, S. y Bereksi-Reguig, F. (2008). Frequency analysis of the heartbeat sounds. *Biomedical Soft Computing and Human Sciences*, 13(1): 85–90.
- Djebbari, A. (1999). *Synthese des Methodes d'Analyse Temporelle, Spectrale et Spectro-Temporelle du signal Phonocardiogramme PCG*. Tesis de maestría, Abou Bekr Belkard University, Faculty of Science Engineering, Chetouane, Tlemcen, Algeria.
- Djebbari, A. y Bereksi Reguig, F. (2000). Short-time fourier transform analysis of the phonocardiogram signal. *Proceedings of the 7th IEEE International Conference on Electronics, Circuits and Systems (ICECS'00)*, páginas 844–847.
- Echeverry, J., López, A., y López, J. (2007). Reconocimiento de valvulopatías cardíacas en señales de fonocardiografía empleando la transformada gabor. *Scientia et Technica*, 13(3): 139 –143.
- España Boquera, M. (2003). *Servicios Avanzados de Telecomunicación*. Díaz de Santos, S.A., España, primero edición. 816 p.
- Ewing, G. J. (1998). *A New Approach to the Analysis of the Third Heart Sound*. Tesis de maestría, The University of Adelaide, Australia, Australia. 100 p.
- Faúndez Zanuy, M. (2000). *Tratamiento digital de voz e imagen y aplicación a la multimedia*. Marcombo. ISBN 9788426712448. 271 p.

- Feigen, L. P. (1971). Physical characteristics of sound and hearing. *American Journal of Cardiology*, **28**(1): 130–133.
- Figueiras Vidal, A. y Rodríguez, A. (2002). *Una panorámica de las telecomunicaciones*. Prentice Hall. ISBN 9788420531007. 394 p.
- Ganguly, S. y Bhatnagar, S. (2008). *VoIP: wireless, P2P and New Enterprise Voice Over IP*. Wiley. 258 p.
- Ganong, F. (1996). *Fisiología médica: El manual moderno*. Prentice Hall, México, 15mo edición. 962 p.
- Gauthier, D., Akay, Y. M., Paden, R. G., Pavlicek, W., Fortuin, F. D., Sweeney, J. K., Lee, R. W., y Akay, M. (2007). Spectral analysis of heart sounds associated with coronary occlusions. *6th International Special Topic Conference on ITAB, Tokyo*, páginas 49–52.
- Gortzis, L. G., Bakettas, I., Makropoulos, C., Grasczew, G., y Nikiforidis, G. (2008). Portable biomedical devices: a critical issue during telecare services design. *Informatics for Health and Social Care*, **33**(2): 91–98.
- Hansen, Martin; Kollmeier, B. (2000). Objective modeling of speech quality with a psychoacoustically validated auditory model. *J. Audio Eng. Soc*, **48**(5): 395–409.
- Hanzo, L., Somerville, F., Somerville, C., y Woodard, J. (2008). *Voice and Audio Compression for Wireless Communications*. IEEE Series on Digital & Mobile Communication. John Wiley & Sons. 880 p.
- Hartmann, W. M. (1997). *Signals, sound, and sensation*. AIP series in modern acoustics and signal processing. American Institute of Physics, New York, USA. 647 p.
- Huber, R. y Kollmeier, B. (2006). Pemo-q: A new method for objective audio quality assessment using a model of auditory perception. *Audio, Speech, and Language Processing, IEEE Transactions on*, **14**(6): 1902–1911.
- IEEE802.15.1 (2002). *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Wireless Personal Area Networks (WPANs)*.
- IEEE802.15.4 (2006). *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs)*.
- ISO/IEC-11172-3 (1993). Coding of moving pictures and associated audio for digital storage media at up to about 1.5 mbit/s, part 3: Audio. *Standard*.
- ISO/IEC-13818-3 (1995). Generic coding of moving pictures and associated audio information, part 3: Audio. *Standard*.

- ISO/IEC-13818-7 (2004). Information technology - generic coding of moving pictures and associated audio information - part 7: Advanced audio coding (aac). *Standard*.
- ISO/IEC-DIS13818-3 (1997). Generic coding of moving pictures and associated audio information, part 3: Audio.(revision of iso/iec 13818-3:1995). *Standard*.
- ITU-T-G.711 (1998). Modulación por impulsos codificados de frecuencias vocales. Reporte técnico, International Telecommunications Union (ITU), Ginebra. 28 p.
- ITU-T-G.726 (1990). Modulación por impulsos codificados diferencial adaptativa (micda) a 40, 32, 24, 16 kbit/s. Reporte técnico, International Telecommunications Union (ITU), Ginebra. 57 p.
- ITU-T-G.729 (2007). Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (cs-acelp). Reporte técnico, International Telecommunications Union (ITU), Geneva.
- Johanson, M., Gustafsson, M., y Johansson, L.-A. (2002). A remote auscultation tool for advanced home healthcare. *Telemedicine and Telecare*, **2**(8): 45-47.
- Kelley, W. (1993). *Medicina interna*. Medicina interna. Médica Panamericana. ISBN 9789500612401. 2003-2005.
- Mahabuba, A., Vijay Ramnath, J., y Anil, G. (2009). Analysis of heart sounds and cardiac murmurs for detecting cardiac disorders using phonocardiography. *Jl. of Instrum. Soc. of India*, **39**(1): 38-41.
- Mahalik, N. P. (2007). *Sensor Networks and Configuration. Fundamentals, Standards, Platforms, and Applications*. Springer.
- Moreno Bolton, R. (2000). Auscultación de pulmón. *Departamento de Enfermedades Respiratorias*, **1**(5).
- Noponen, A.-L., Lukkarinen, S., Angerla, A., y Sepponen, R. (2007). Phono-spectrographic analysis of heart murmur in children. *BMC Pediatrics*, **23**(7).
- P.10/G.100 (2010). Amendment 2: New definitions for inclusion in recommendation itu-t p.10/g.100. Reporte técnico, Telecommunication Standardization Sector of ITU (ITU-T), Geneva.
- P.800 (1996). P.800: Telephone transmission quality methods for subjective determination of transmission quality. Reporte técnico, Telecommunication Standardization Sector of ITU (ITU-T), Geneva.
- P.800.1 (2006). Telephone transmission quality telephone installations local line networks methods for objective and subjective assessment of quality. Reporte técnico, Telecommunication Standardization Sector of ITU (ITU-T), Geneva.

- P.834.1 (2010). *Methods for objective and subjective assessment of speech quality*. Geneva.
- Quintana, B. (2004). *Monitorización En Anestesia, Medicina de Urgencias y Cuidados Intensivos*. Elsevier - Health Sciences Division. ISBN 9788481747089.
- Rangayan, R. M. y Lehner, R. J. (1988). Phonocardiogram signal analysis: A review. *Critical Reviews in Biomedical Engineering*, **15**(3): 55–62.
- RFC-5215 (2008). *RTP Payload Format for Vorbis Encoded Audio*.
- Robinson, T. (1994). Shorten: Simple lossless and near-lossless waveform compression. Reporte técnico, Cambridge University Engineering Department, United Kingdom. 104 p.
- Roederer, J. G. (2009). *The Physics and Psychophysics of Music*. Springer, USA, cuarta edición. 9.
- Romero, A., Velázquez, J., y Concha, J. (2009). Estudio acústico de sonidos ligados a patologías cardíacas. *Sonido y Acústica*, **3**(3): 27–33.
- Rosamond, W., Flegal, K., Friday, G., Furie, K., Go, A., Greenlund, K., Haase, N., Ho, M., Howard, V., Kissela, B., Kittner, S., Lloyd-Jones, D., McDermott, M., Meigs, J., Moy, C., Nichol, G., O'Donnell, C. J., Roger, V., Rumsfeld, J., Sorlie, P., Steinberger, J., Thom, T., Wasserthiel-Smoller, S., y Hong, Y. (2007). Heart disease and stroke statistics 2007 update: a report from the american heart association statistics committee and stroke statistics subcommittee. *Circulation*, páginas e69– e171.
- Salesa, E., Perelló, E., y Bonavida, A. (2005). *Tratado de Audiología*. Elsevier, Barcelona, España, primera edición. 88, 89.
- Spanias, A., Painter, T., y Atti, V. (2007). *Audio signal processing and coding*. Wiley-Interscience. ISBN 9780471791478.
- Surós Batlló, A. y Surós Batlló, J. (2001). *Semiología médica y técnica exploratoria*. Elsevier España, España, octava edición. 56.
- TOSQA, O. (2003). Telecommunications objective speech quality assessment. Reporte técnico, HEAD acoustics GmbH.
- Vannuccinni, L., Earis, J., Helistö, P., Cheetham, B., Rossi, M., Sovijärvi, A., y Vanderschoot, J. (2000). Capturing and preprocessing of respiratory sounds. *European Respiratory Review*, **10**(77): 616–620.
- Watkinson, J. (2004). *The MPEG handbook: MPEG-1, MPEG-2, MPEG-4*. Broadcasting and communications. Elsevier/Focal Press. ISBN 9780240805788. 451 p.

- Weinstein, S. (2005). *The multimedia Internet*. Information technology—transmission, processing, and storage. Springer. ISBN 9780387236810. 255 p.
- Zenk, B., Bratton, R., y Flipse, T. (2004). Accuracy of detecting irregular cardiac rhythms via telemedicine. *Telemed Telecare*, páginas 55–58.
- Zhang, Y. y Xiao, H. (2009). Bluetooth-based sensor networks for remotely monitoring the physiological signals of a patient. *IEEE Transactions on Information Technology in Biomedicine*, **13**(6): 1040–1048.