

**CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE EDUCACIÓN  
SUPERIOR DE ENSENADA, BAJA CALIFORNIA**



---

**PROGRAMA DE POSGRADO EN CIENCIAS  
EN ELECTRÓNICA Y TELECOMUNICACIONES**

---

**Desarrollo de un códec para la transmisión de audio cardiaco  
sobre redes de bajas tasas de datos.**

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de  
Maestro en Ciencias

Presenta:

**Roilhi Frajo Ibarra Hernández**

Ensenada, Baja California, México

2014

Tesis defendida por

**Roilhi Frajo Ibarra Hernández**

y aprobada por el siguiente comité

---

Dr. Miguel Ángel Alonso Arévalo  
*Director del Comité*

---

Dr. Salvador Villarreal Reyes  
*Miembro del Comité*

---

Dr. Roberto Conte Galván  
*Miembro del Comité*

---

Dr. Jonás de Dios De Basabe Delgado  
*Miembro del Comité*

---

Dr. César Cruz Hernández  
*Coordinador del Programa de  
Posgrado en Electrónica y Telecomunicaciones*

---

Dr. Jesús Favela Vara  
*Director de Estudios de Posgrado*

Octubre, 2014

Resumen de la tesis que presenta Roilhi Frajo Ibarra Hernández como requisito parcial para la obtención del grado de Maestro en Ciencias en Electrónica y Telecomunicaciones.

## **Desarrollo de un códec para la transmisión de audio cardiaco sobre redes de bajas tasas de datos.**

Resumen elaborado por:

---

Roilhi Frajo Ibarra Hernández

Durante los últimos años la principal causa de mortalidad en México y a nivel mundial está dada por las enfermedades cardiovasculares (CVD's) debido a los altos índices de obesidad y problemas de hipertensión entre la población. La auscultación es un método económico y sencillo de diagnóstico para la atención de anomalías cardíacas, sobre todo aquellas relacionadas al mal funcionamiento en alguna de las válvulas asociadas al corazón.

Los límites del oído humano a las bajas frecuencias que presenta una señal de fonocardiograma (PCG) obtenida mediante la auscultación han originado el surgimiento de estetoscopios electrónicos, con los cuales es posible visualizar la forma de onda eléctrica del audio cardiaco. Con este avance es posible el análisis tiempo-frecuencia del PCG complementando y reforzando el diagnóstico de patologías cardíacas.

No obstante en la actualidad los codificadores tradicionales de audio tienen soporte y buena resolución para las altas frecuencias, proporcionando un exceso de recursos en la codificación de un fonocardiograma y complicando su transmisión sobre redes de bajas tasas de datos. Por tal motivo, se propone en este trabajo de tesis implementar un codificador-decodificador adaptado al audio cardíaco, a su rango definido de frecuencias y a baja tasa de datos, proporcionando compresión y la información necesaria para que sea reconstruido con calidad.

En este trabajo de tesis se han explotado las técnicas de Matching Pursuit con diccionarios de Gabor y de codificación predictiva lineal (LPC) para el modelado, representación y codificación paramétrica de una señal de fonocardiograma. El codificador es evaluado mediante pruebas subjetivas y objetivas verificando la calidad de la señal reconstruida.

Palabras Clave: **audio cardiaco, fonocardiograma, codificación, codificación predictiva lineal, Matching Pursuit.**

Abstract of the thesis presented by Roilhi Frajo Ibarra Hernández as a partial requirement to obtain the Master of Science degree in Master in Sciences in Electronics and Telecommunications.

## **Design of a Codec for Cardiac Sound Transmission Over Low Rate Data Networks.**

Abstract by:

---

Roilhi Frajo Ibarra Hernández

During the last years the main cause of mortality in Mexico and around the world is due to cardiovascular diseases (CVD's), because of the high levels of obesity and hypertension problems among the population. The auscultation is an easy and economic method to diagnose cardiac anomalies, mostly those diseases associated with the malfunction of some of the heart valves.

The phonocardiogram signal (PCG), recorded during auscultation, contains information in low frequencies that are not perceptible by the human auditory system. This problem originated the appearance of digital stethoscopes, devices that allow to visualize the electric waveform of cardiac sounds. With this advance it is possible to analyze the PCG and its time-frequency features as a complement to diagnose heart diseases.

However, the traditional audio coders have good resolution and support to high frequency components of the signal, giving an excess of resources in phonocardiogram coding and making hard to transmit this signals over low data rate networks. For this reason, in this thesis we develop an audio codec adapted to cardiac sounds, representing efficiently its low frequency components, giving a low bit rate, signal compression and the amount of information necessary to reconstruct it with accuracy.

In this framework we have exploited the Matching Pursuit Method with Gabor dictionaries and Linear Predictive Coding to the modelling, representation and parametric coding for phonocardiogram signals. The codifier is evaluated by subjective tests as well as objective measures verifying the quality of the reconstructed signal.

**Keywords: cardiac sounds, phonocardiogram, coding, linear predictive coding, Matching Pursuit.**

## Dedicatoria

*A mi madre y a mis hermanos;  
a mis maestros y a mis amigos.*

## **Agradecimientos**

A mi familia y amigos, de quienes siempre aprendo y disfruto.

A mis instructores y personal del Centro de Investigación Científica y de Educación Superior de Ensenada, institución que ha contribuido en mi formación durante esta etapa de mi vida.

Al Dr. Miguel Ángel Alonso, quien confió en mí para la realización de este trabajo de investigación que ampliamente he disfrutado.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por brindarme el apoyo económico para realizar mis estudios de maestría.

# Tabla de contenido

	Página
<b>Resumen en español</b>	<b>iii</b>
<b>Resumen en inglés</b>	<b>iv</b>
<b>Dedicatoria</b>	<b>v</b>
<b>Agradecimientos</b>	<b>vi</b>
<b>Lista de figuras</b>	<b>x</b>
<b>Lista de tablas</b>	<b>xiii</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Objetivo general de la tesis . . . . .	3
1.1.1. Objetivos particulares . . . . .	3
1.2. Justificación . . . . .	4
1.3. Metodología de la investigación . . . . .	4
1.4. Organización de la tesis . . . . .	5
<b>2. La señal de fonocardiograma</b>	<b>8</b>
2.1. Origen fisiológico y representación de los sonidos cardiacos . . . . .	8
2.2. Características del fonocardiograma . . . . .	10
2.3. Patologías cardiacas . . . . .	11
2.4. Codificación de audio . . . . .	13
2.4.1. La psicoacústica en el modelado y codificación de audio . . . . .	14
2.4.2. El códec de audio . . . . .	16
2.5. Estado del arte en el modelado de sonidos cardiacos . . . . .	21
<b>3. Modelado de la parte determinística mediante Matching Pursuit</b>	<b>25</b>
3.1. Representación dispersa de señales . . . . .	25
3.2. Diccionarios tiempo-frecuencia . . . . .	27
3.2.1. Diccionario y átomos de Gabor . . . . .	28
3.2.2. Descomposiciones atómicas . . . . .	29
3.2.3. Principio de incertidumbre de Heisenberg . . . . .	30
3.3. Descomposición de fonocardiogramas mediante Matching Pursuit (MP)	31
3.3.1. Criterios en la descomposición de fonocardiogramas por MP . . . . .	33
<b>4. Modelado de la parte estocástica mediante Codificación Predictiva Lineal</b>	<b>36</b>
4.1. Filtros digitales y transformada z . . . . .	37
4.1.1. Transformada z . . . . .	37
4.1.2. Filtros digitales . . . . .	38
4.1.2.1. Filtros de respuesta al impulso finita (FIR LCCDE) . . . . .	39
4.1.2.2. Filtros de respuesta al impulso finita (IIR LCCDE) . . . . .	40
4.2. Modelos autorregresivos . . . . .	40
4.3. Codificación mediante predicción lineal . . . . .	42
4.3.1. Modelo de predicción lineal . . . . .	44

## Tabla de contenido (continuación)

4.3.2.	Predicción hacia adelante y hacia atrás . . . . .	45
4.3.3.	Estimación de los coeficientes de predicción . . . . .	46
4.3.3.1.	Recursión de Levinson-Durbin . . . . .	49
4.4.	Estimación espectral . . . . .	51
4.5.	Variantes a LPC: WLPC y VELPC. . . . .	53
4.5.1.	Codificación predictiva lineal con distorsión espectral (WLPC) .	54
4.5.2.	Codificación predictiva lineal con excitación por voz (VELPC) .	56
<b>5.</b>	<b>Conformación del códec propuesto</b>	<b>57</b>
5.1.	Estructura del codificador-decodificador propuesto . . . . .	57
5.2.	Segmentación de los eventos cardiacos . . . . .	59
5.3.	Descomposición atómica de los eventos (MP) . . . . .	60
5.4.	Modelado y tratamiento de la señal residual (LPC) . . . . .	61
5.4.1.	Segmentación del residual . . . . .	62
5.4.2.	Criterios para la determinación del orden $p$ del filtro predictor .	64
5.4.3.	Síntesis/reconstrucción de la señal residual . . . . .	65
5.4.4.	Filtros de pre-énfasis y de-énfasis en LPC . . . . .	66
5.5.	Cuantificación . . . . .	67
5.5.1.	Tipos de cuantificadores . . . . .	67
5.5.2.	Diseño de un cuantificador no uniforme . . . . .	68
5.5.2.1.	Algoritmo de Lloyd . . . . .	68
5.5.3.	Cuantificación de los parámetros extraídos de los modelados .	69
5.5.4.	Cuantificación de los coeficientes de predicción . . . . .	72
5.6.	Representación en bits de los parámetros de codificación . . . . .	74
5.7.	Codificación de los fonocardiogramas y porcentajes de compresión obtenidos . . . . .	76
5.8.	Pruebas en la reconstrucción de los fonocardiogramas . . . . .	76
5.8.1.	Coefficiente de correlación . . . . .	77
5.8.2.	Raíz cuadrada de la diferencia cuadrática media (Percent Root- mean-square Difference, PRD) . . . . .	78
<b>6.</b>	<b>Evaluación subjetiva del códec</b>	<b>80</b>
6.1.	Selección de las señales de prueba . . . . .	80
6.2.	Procedimiento de evaluación MUSHRA . . . . .	81
6.2.1.	Preparación de la interfaz gráfica . . . . .	81
6.2.2.	Fase de adiestramiento . . . . .	83
6.2.3.	Condiciones de audición . . . . .	83
6.2.4.	Selección de los oyentes . . . . .	84
6.3.	Análisis estadístico de los resultados obtenidos en la prueba . . . . .	84
6.3.1.	Cálculo de los histogramas e intervalos de confianza . . . . .	85
<b>7.</b>	<b>Conclusiones</b>	<b>90</b>
7.1.	Sobre los objetivos de la tesis . . . . .	90
7.2.	Trabajo futuro de investigación . . . . .	92

## **Tabla de contenido (continuación)**

**Lista de referencias**

**94**

## Lista de figuras

Figura	Página
1. Localización de las válvulas cardiacas. Tomado de <a href="http://www.nlm.nih.gov/medlineplus/spanish/ency/esp_imagepages/9380.htm">http://www.nlm.nih.gov/medlineplus/spanish/ency/esp_imagepages/9380.htm</a> . . . . .	9
2. Forma de onda de un ciclo cardiaco. . . . .	10
3. Espectrograma de un ciclo cardiaco. . . . .	11
4. Forma de onda de un soplo diastólico. . . . .	12
5. Esquema general del codificador-decodificador diseñado para audio cardiaco. . . . .	13
6. Cuatro tipos de señales: a) Continua en amplitud y tiempo (analógica); b) discreta en amplitud y continua en tiempo; c) continua en amplitud y discreta en tiempo; d) discreta en tiempo y amplitud. . . . .	15
7. Umbral absoluto de audición. Tomado de Painter y Spanias (2000). . . . .	16
8. Diagrama a bloques de un codificador de audio perceptual. . . . .	19
9. Diagrama a bloques de un codificador de audio paramétrico. . . . .	20
10. Representación gráfica de una representación dispersa para una señal $v$ desde un diccionario $D$ . . . . .	26
11. Forma de onda de un átomo de Gabor. . . . .	29
12. Representación de dos átomos tiempo-frecuencia en cajas de Heisenberg. . . . .	31
13. Porcentaje de energía recobrada/extraída tras cada iteración realizada en Matching Pursuit. Se observa en el acercamiento a la gráfica que tras $M = 30$ iteraciones realmente no se retiene el total de la energía. . . . .	34
14. Forma de onda temporal y gráfica en el plano tiempo-frecuencia en la reconstrucción de un evento cardiaco $s_1$ . . . . .	35
15. Diagrama a bloques de un filtro FIR. . . . .	39
16. Diagrama a bloques de un filtro IIR. . . . .	41
17. Modelo AR como salida de un filtro IIR en respuesta a una entrada de un proceso blanco. . . . .	41
18. Modelo de predicción lineal. . . . .	44
19. Modelado mediante LPC. En la parte a) se produce una secuencia $s(n)$ por el modelo todo polos. La parte b) muestra la generación de $e(n)$ como salida del filtro $A(z)$ , cuya inversión en c) reproduce idealmente a $s(n)$ si tiene como entrada al error de predicción $e(n)$ . . . . .	44
20. Simetría entre la predicción hacia adelante y hacia atrás. . . . .	46
21. Aproximación por codificación predictiva lineal para una señal de audio cardiaco con un orden de filtro $p = 30$ . . . . .	53

## Lista de figuras (continuación)

Figura	Página
22. Comparación de la envolvente espectral de una señal proveniente de un sonido de piano muestreada con $f_s = 44kHz$ y un orden de predicción $p = 37$ por los métodos LPC y WLPC. . . . .	55
23. Estructura del codificador para audio cardiaco propuesto. . . . .	58
24. Estructura del decodificador para audio cardiaco propuesto. . . . .	59
25. Segmentación del evento cardiaco normal S1 en una señal de audio cardiaco.	59
26. Descomposición MP de un evento cardiaco y muestra de los primeros 15 átomos mayormente correlacionados. . . . .	61
27. Señal de residual y silencios obtenida de un audio cardiaco soplo diastólico.	63
28. Ventaneo (segmentación) de la señal residual más silencios. . . . .	64
29. Medida de planitud espectral promedio de la señal error $e(n)$ para órdenes del filtro predictor desde 1 hasta 100. . . . .	65
30. Representación gráfica del procedimiento de síntesis “overlap and add (OLA)”. . . . .	66
31. Filtros de pre-énfasis y de-énfasis usados en la codificación LPC. . . . .	67
32. Regiones de decisión y niveles de reconstrucción de un cuantificador. . . . .	69
33. Parámetros extraídos de los modelados MP y LPC para el códec. Se indican los parámetros necesarios de cuantificar. . . . .	70
34. Histogramas normalizados (pdf) para las amplitudes y fases de los átomos extraídos en las descomposiciones. . . . .	71
35. Histogramas normalizados (pdf) para los coeficientes LSFD y ganancias del filtro predictor. . . . .	71
36. Distorsión espectral común por el ruido de cuantificación aplicada a diferentes representaciones equivalentes LPC. . . . .	74
37. Formas de onda original y reconstruida por el códec para un ciclo de un murmullo sistólico tardío. . . . .	78
38. Visualización en la pantalla de la interfaz MUSHRA. . . . .	82
39. Histograma de las calificaciones medias e intervalos de confianza para el estímulo Apertura Normal S1. . . . .	87
40. Histograma de las calificaciones medias e intervalos de confianza para el estímulo Apertura Normal S2. . . . .	87

## Lista de figuras (continuación)

Figura		Página
41.	Histograma de las calificaciones medias e intervalos de confianza para el estímulo S3. . . . .	88
42.	Histograma de las calificaciones medias e intervalos de confianza para el estímulo S4. . . . .	88
43.	Histograma de las calificaciones medias e intervalos de confianza para el estímulo Murmullo Pansistólico. . . . .	89

## Lista de tablas

Tabla		Página
1.	Duración y contenido frecuencial de los principales ruidos cardiacos. Tomado de Castorena (2012) . . . . .	10
2.	Sonidos de la base del sitio web Littman © analizados en este trabajo de tesis. Tomados de (Litmann, 2013). . . . .	58
3.	Características de los diccionarios empleados en la descomposición de los eventos cardiacos del codificador. . . . .	60
4.	Cantidad de eventos y número de átomos requeridos en las descomposiciones MP para las señales de la base de datos. . . . .	62
5.	Tipos de cuantificadores de acuerdo a las características de la fuente de datos de entrada. . . . .	68
6.	Número de bits requeridos por parámetro y tipo de cuantificación realizada. . . . .	75
7.	Porcentajes de compresión y bits requeridos para el codificador diseñado en las señales de prueba. . . . .	77
8.	Coeficientes de correlación de las señales reconstruidas comparadas con los audios originales. . . . .	79
9.	Cálculo de la raíz cuadrada de la diferencia cuadrática media (PRD) para las señales de la base de datos. . . . .	79

## Capítulo 1. Introducción

---

El proveer servicios de salud de calidad para la población se ha convertido en una necesidad prioritaria para cualquier sociedad; no obstante las limitantes geográficas y de infraestructura hacen que esta tarea sea más complicada de llevar a cabo en las zonas marginales. En efecto, las áreas metropolitanas concentran la mayor cantidad de recursos económicos así como especialistas de la salud.

Las tecnologías de la información y comunicaciones (TIC's) han brindado alternativas al problema anterior teniendo herramientas como la telemedicina y la telesalud, las cuales ayudan al intercambio de recursos clínicos y consultas a distancia entre el especialista y algún paciente de alguna zona alejada a la metrópoli. Sin embargo, otra cuestión que limita la labor de las TIC's es que los recursos económicos y tecnológicos necesarios para la comunicación hacia las localidades marginadas son limitados. Las velocidades de transmisión y ancho de banda pueden no tener el soporte necesario ni la fidelidad requerida para transportar datos de tipo clínico, los cuales deben ser cuidadosamente blindados y consecuentemente no tan susceptibles a errores para que el especialista pueda brindar el diagnóstico adecuado.

Por otra parte las enfermedades cardiovasculares (CVD's por sus siglas en Inglés) son la principal causa de mortalidad entre la población a nivel mundial según la OMS (o bien, WHO por sus siglas en Inglés), y es que tan sólo en el 2008 alrededor de 17.3 millones de personas fallecieron por alguna CVD representando el 30 % del total de muertes registradas en ese periodo (WHO, 2012).

Nuestro país, México, no es la excepción. Las enfermedades cardiovasculares son la segunda causa de muerte dado el alto índice de obesidad en la población. Según cifras registradas en el 2010 el Sistema Nacional de Información en la Salud (SINAIS) <sup>1</sup>.

Actualmente existen diversas técnicas para detectar alguna enfermedad cardiovascular; algunas de ellas son bastante sofisticadas como la imagen por resonancia magnética

---

<sup>1</sup>Sistema Nacional de Información en Salud. Principales causas de mortalidad general: Disponible en <http://sinais.salud.gob.mx/mortalidad/index.html>. Al mes de octubre de 2014 no existen datos recientes.

(IMR), la ecocardiografía o ultrasonido cardíaco y la cateterización cardíaca. Todos estos métodos proporcionan una gran cantidad de información importante para realizar algún diagnóstico, sin embargo se vuelven inaccesibles para los habitantes de zonas marginadas por su alto costo y complejidad.

En contraste, un método sencillo y de bajo coste para brindar un primer acercamiento hacia alguna anomalía cardíaca es la *auscultación*, técnica que consiste en escuchar los ruidos internos del corazón desde la pared torácica con un dispositivo llamado *estetoscopio*. En realidad los sonidos que se están escuchando son los producidos por la actividad de las válvulas cardíacas en los procesos de intercambio de oxígeno en la sangre (sístole y diástole).

Los primeros estetoscopios fueron desarrollados con base en principios mecánicos, con lo cual los estudiantes de medicina debían adquirir ciertas habilidades auditivas al auscultar para identificar diferentes patologías (Roguin, 2006). Algunos estetoscopios existentes en la actualidad son digitales, con lo que es posible el registrar y visualizar la forma de onda de audio cardíaco digitalizada, además de poseer una mejor respuesta en frecuencia y controles de volumen (Abbas y Bassam, 2009). El dispositivo puede estar conectado de forma cableada o inalámbrica al PC cuando se registra algún fonocardiograma (PCG)<sup>2</sup>, en este último caso se han desarrollado estetoscopios donde el estándar Bluetooth brinda la conexión al dispositivo terminal (Como el 3200 de la serie 3M Litmann ©).

La digitalización de los sonidos cardíacos hace posible implementar procedimientos que serían muy difíciles de realizar en el dominio analógico; tales como un análisis más completo de la señal en el plano tiempo frecuencia, técnicas para eliminar ruidos existentes *denoising*, filtrado, entre otros. Estos procesos proveerán el acondicionamiento adecuado para la transmisión de estas señales sobre redes de comunicaciones.

Un fonocardiograma digitalizado también representa una ventaja conociendo que en el dominio de la frecuencia existen límites en la audición humana, otorgando a la auscultación mayor precisión para realizar un diagnóstico <sup>3</sup>.

---

<sup>2</sup>Así se le denomina a la señal de audio cardíaco.

<sup>3</sup>La técnica de diagnóstico por medio del análisis digital de señales fisiológicas en biomedicina es cono-

Para representar la mayor cantidad de información de una señal de audio de manera comprimida, se hace el uso de un códec (codificador-decodificador); el cual es una secuencia ordenada de bits que describirán algunos atributos del audio para ser procesados en cualquier sistema que esté digitalizado.

Existen también otros codificadores dedicados a señales de imágenes, video o texto que tienen también el mismo objetivo. Sin embargo, los códecs de audio que existen en la actualidad son dedicados y adaptados a las señales musicales y a la voz humana (vocoders), aunque otros codificadores son de propósito general, por lo regular se basan en las características proporcionadas por estos tipos de sonidos.

### **1.1. Objetivo general de la tesis**

Diseñar un códec que comprima la señal de audio cardíaco para su transmisión sobre redes inalámbricas con bajas tasas de datos. El codificador desarrollado se realizará tomando como base la técnica de representación escasa de señales por medio del algoritmo de Matching Pursuit para la compresión de fonocardiogramas (Nieblas, 2014).

Se tomarán los parámetros adecuados en la descomposición realizada por MP (a dicho modelado se le considerará como *Parte determinística*) así como el modelado de la señal residual y silencios (*Parte estocástica* del códec diseñado) para su cuantificación y posteriormente realizar la conformación de la trama de bits que represente adecuadamente al PCG.

#### **1.1.1. Objetivos particulares**

- Comprimir los sonidos cardíacos de manera eficiente, empleando el menor número de iteraciones/átomos posible y extrayendo la mayor cantidad de energía al emplear el método Matching Pursuit y diccionarios de Gabor.
- Modelar de manera adecuada la señal residual tras ejecutar la descomposición MP así como los silencios entre eventos cardíacos.
- Disminuir la tasa de datos con el códec propuesto con respecto a si se transmitiera un fonocardiograma con un codificador de audio convencional.

---

cida como *diagnóstico asistido por computadora (computer-aided diagnosis)*(Doi, 2008).

## 1.2. Justificación

El diagnóstico de patologías cardíacas que da origen a algunas de las enfermedades cardiovasculares se puede otorgar de manera sencilla y económica por medio de la auscultación, donde se obtendrán señales de audio cardíaco que proporcionarán información sobre la actividad mecánica del corazón y el estado de sus válvulas.

Más del 99% de la energía del fonocardiograma se concentra en la banda de 5-1000Hz (Djebbari y Reguig, 2000). En algunos trabajos de la literatura se ha muestreado la señal de PCG empleando tasas en el rango de 8,000-22,500Hz y con 8-16 bits de resolución (Castorena, 2012). Con lo anterior se requieren tasas de datos efectivas en el rango de 64 kbps a 352.8 kbps. Estas medidas tienen como consecuencia un sistema de tele-auscultación costoso y complicado de implementar en nuestro país (Nieblas, 2014).

Además de las anteriores consecuencias, si se tiene un escenario real de aplicación y en el mejor de los casos la tasa de datos neta será del orden del 50% de la tasa bruta de transferencia, el resultado será la degradación en la calidad de la señal recibida.

Aprovechando las técnicas por representación escasa de señales se pretende lograr comprimir fonocardiogramas extrayendo en la descomposición más del 90% de la energía de la señal de audio. Se pretende crear un códec de propósito específico para señales de audio cardíaco, con lo que se logrará fidelidad al intercambio de este tipo de información por redes de bajas tasas de datos, se proporcionará un primer medio para que a distancia se puedan diagnosticar eficazmente patologías o anomalías cardíacas.

## 1.3. Metodología de la investigación

Para alcanzar los objetivos planteados en el presente trabajo se han presentado varias etapas. Los siguientes enunciados resumen los puntos más importantes de cada una de ellas:

- Se revisó el estado del arte concerniente al modelado de sonidos cardíacos.
- Se reunió/generó una base de datos de audio cardíaco para ser analizada (sonidos normales y con patologías).

- Se realizó la segmentación de eventos cardíacos como primer paso en el análisis.
- Se analizó el modelado y compresión de fonocardiogramas utilizando la representación escasa de señales del algoritmo Matching Pursuit (MP).
- Se comprimió de manera eficiente el PCG por medio de MP empleando el menor número de átomos posible.
- Se modeló adecuadamente el residual de la señal producto de la representación escasa además de los silencios entre los eventos cardíacos por medio de predicción lineal (LPC).
- Se extrajeron y analizaron los parámetros más significativos producto de la descomposición MP y la codificación LPC.
- Se cuantificaron de manera adecuada los parámetros extraídos evaluando el desempeño de los cuantificadores existentes en la literatura.
- Se verificó la cantidad de bits requerida por cada parámetro necesario en la compresión del PCG por el códec propuesto, con ello se calcularon también los porcentajes de compresión para velocidades de 8 kbps y 16 kbps.
- Se evaluó de manera objetiva y subjetiva por medio del test MUSHRA, la calidad en la compresión del códec propuesto así como su desempeño al compararlo con MPEG-1 capa 3 (MP3) y OPUS (Codificador de acceso libre).

#### **1.4. Organización de la tesis**

La presente tesis se expondrá conforme a los puntos mostrados en la metodología anterior en cuanto a secuencia y objetivos específicos para cada capítulo. A continuación se describe el contenido de cada una de dichas secciones.

El *Capítulo 2* describe de manera fisiológica al audio cardíaco, es decir, qué origina esta señal físicamente y qué representa; cómo detectar anomalías cardíacas por medio de un fonocardiograma además de sus características como señal acústica y eléctrica: forma de onda temporal, duración, respuesta en frecuencia y ancho de banda.

También en esta sección se ha dado una revisión del estado de arte en codificación de audio y en el modelado del PCG, algunas técnicas que han logrado su compresión y procesamiento de manera adecuada, concluyendo que MP es un método muy adecuado por las características de esta señal.

El códec para audio cardíaco diseñado es dividido en una parte determinística y una parte estocástica. En el primer caso se le conoce como señal determinística a aquella que pueda ser descrita por una expresión matemática cerrada o bien puede ser reproducida en repetidas ocasiones.

Por otra parte, un proceso aleatorio en tiempo discreto es un *ensemble* o colección de señales que es definido en términos de las propiedades estadísticas del proceso sufrido por este conjunto. Se trata de una secuencia *indizada* de valores aleatorios, una estructura que describe estos valores y las relaciones existentes entre ellos. Un proceso aleatorio en tiempo discreto puede emplear conjuntos de señales determinísticas (Papoulis, 1984).

En el *Capítulo 3* se expone el modelado Matching Pursuit como una parte determinística del códec, ya que se ha encontrado en algunos trabajos recientes que se obtiene una gran aproximación dada la alta correlación presentada entre las ondas denominadas átomos de Gabor y los eventos de audio cardíaco.

Se expone en el *Capítulo 4* una parte del fonocardiograma que no guarda una buena correlación con los átomos de Gabor y por lo tanto no es conveniente modelar por MP, ya que la descomposición emplea un elevado número de átomos. Este *residual* o parte estocástica del PCG (parecida a ruido o *noise-like*) es modelada por medio de codificación por predicción lineal (LPC).

Ya teniendo completa la representación del audio cardíaco se procede a conformar el códec juntando las partes expuestas en los capítulos 3 y 4. Los parámetros más importantes de estos modelados, es decir los que preservan la mayor información de la señal, se extraerán para ser cuantificados.

Posteriormente se calcularán las cantidades de bits requeridas por parámetro y consecuentemente los porcentajes de compresión en los fonocardiogramas analizados. Este

procedimiento es descrito en el *Capítulo 5*.

Siguiendo un protocolo científico se deberá dar una validación para el producto obtenido en el presente trabajo. En el caso de codificación de audio se realizan pruebas subjetivas y un test muy común para evaluación es el MUSHRA, avalado por la Unión Internacional de Telecomunicaciones (ITU).

La realización del método MUSHRA aplicando el codificador de audio cardiaco diseñado y el análisis de los resultados arrojados por estas pruebas son expuestos en el *Capítulo 6*.

Para la finalización del presente trabajo se exponen las conclusiones respectivas en el *Capítulo 7*, además se plantean algunas futuras líneas de investigación para el desarrollo de futuras tesis enfocadas en la codificación de audio cardiaco.

## Capítulo 2. La señal de fonocardiograma

---

En este capítulo se conocerán las características fisiológicas del audio cardíaco, comprendiendo lo que representan cada uno de los sonidos generados así como su clasificación. De igual manera se expondrá una descripción del fonocardiograma como señal eléctrica y algunos parámetros importantes de la misma que servirán en su procesamiento digital, como la respuesta en frecuencia y forma de onda temporal. Se definirán también en esta sección las características de los sonidos que representen algún mal o anomalía en el funcionamiento del corazón, qué parámetros las caracterizan así como algunas consideraciones en su tratamiento como señal eléctrica y acústica.

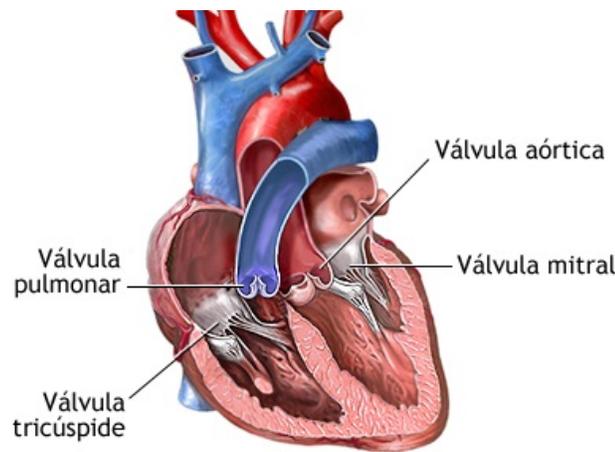
El codificador de fonocardiograma (al cual denotaremos como PCG por sus siglas en inglés) involucra diversos procedimientos en su realización. Es importante por lo tanto, conocer las características y parámetros de algunos códecs existentes; debido a esto se ha incluido en este capítulo una breve introducción a la codificación de audio en general.

De igual manera se mostrará en esta sección un breve panorama o *estado del arte* en codificación de sonidos cardíacos, basado en los trabajos realizados en la literatura para el modelado y procesamiento de fonocardiogramas. Se tomarán algunas consideraciones importantes sobre estos trabajos y se definirá cuáles se han empleado en esta tesis para el desarrollo del códec.

### 2.1. Origen fisiológico y representación de los sonidos cardíacos

Como se ha definido anteriormente un fonocardiograma representa al audio cardíaco; se trata de una señal mecánica sonora tomada con un dispositivo médico llamado estetoscopio. Los sonidos generados por esta señal describen el comportamiento o acción mecánica de las válvulas o valvas del corazón, las cuales son compuertas que abren o cierran sistemáticamente para dar paso a la regulación de la sangre a través de las venas y arterias (Abbas y Bassam, 2009).

Existen cuatro válvulas, cada una asociada a las principales venas que están conectadas al órgano cardíaco: válvula aórtica, válvula pulmonar, válvula mitral y válvula tricúspide. Estos componentes se muestran en la Figura 1.



**Figura 1: Localización de las válvulas cardíacas.** Tomado de [http://www.nlm.nih.gov/medlineplus/spanish/ency/esp\\_imagepages/9380.htm](http://www.nlm.nih.gov/medlineplus/spanish/ency/esp_imagepages/9380.htm)

Son dos eventos importantes los que constituyen un PCG ordinario, denominados como S1 y S2. En el caso del primer evento o S1 se tiene un sonido que describe la actividad dada por la apertura de las válvulas aórtica y pulmonar y el cierre de las válvulas mitral y tricúspide (esto evita que la sangre circule hacia esta sección). En este evento se completa una **sístole ventricular**, donde las ventrículos se contraen para expulsar la sangre hacia las arterias (Djebbari y Reguig, 2000).

El segundo evento cardíaco S2 describe el complemento de los acontecimientos dados en S1, esto es, las válvulas mitral y tricúspide ahora se abren para recibir la sangre mientras que los conductos aórtico y pulmonar se han cerrado. Este proceso se denomina **sístole auricular** y en efecto, ahora son las aurículas quienes se han contraído para expulsar la sangre hacia los ventrículos. Se denomina *ciclo cardíaco* cuando secuencialmente han ocurrido los sonidos S1 y S2. La **diástole** ocurre después de completarse un ciclo cardíaco y se trata de un periodo donde se relajan las aurículas y ventrículos. Estos dos sonidos son identificados en un fonocardiograma regular durante la actividad de un ciclo (Abbas y Bassam, 2009).

Existen otros dos sonidos cardíacos: S3 y S4, los cuales son asociados a otros eventos y generalmente son despreciados en el análisis del PCG. En el caso del llenado o asimilación de sangre hacia los ventrículos se origina S3, mientras que S4 ocurre al final de la diástole surgido por la contracción de las aurículas cuando han desplazado el flujo

de los ventrículos (Köymen *et al.*, 1987).

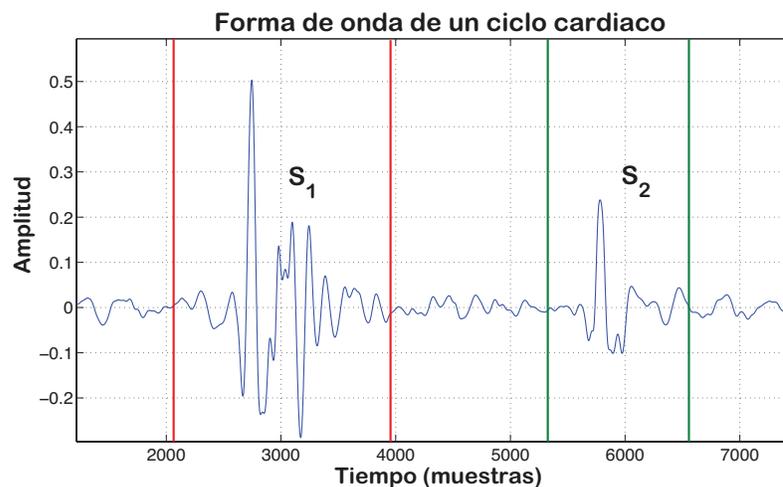
Al espacio donde no ocurre algún evento cardíaco se le denomina *silencio*.

## 2.2. Características del fonocardiograma

Como señal eléctrica un fonocardiograma consta de formas de onda específicas asociadas a los sonidos cardíacos que se mencionaron anteriormente (S1 y S2). Los eventos cardíacos preservan cierta duración en tiempo y por lo tanto también contienen cierto intervalo de frecuencias, características que se muestran en la Tabla 1. En la Figura 2 puede observarse la forma de onda de los ruidos o eventos cardíacos normales que constituyen un ciclo cardíaco.

**Tabla 1: Duración y contenido frecuencial de los principales ruidos cardíacos. Tomado de Castorena (2012)**

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

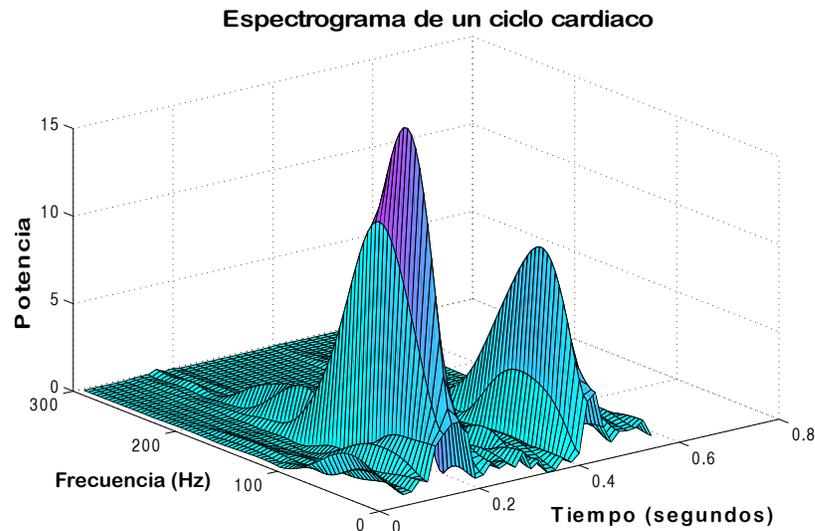


**Figura 2: Forma de onda de un ciclo cardíaco.**

Los sonidos cardíacos presentan cambios rápidos con respecto al tiempo y a la frecuencia debido al comportamiento no estacionario de la señal PCG. Es por esto que el

estudio y entendimiento de sus características se realiza en un plano conjunto tiempo-frecuencia. La STFT<sup>1</sup> es una herramienta propicia para el análisis en los cambios del comportamiento del fonocardiograma (Djebbari y Reguig, 2000).

En la Figura 3 se muestra el espectrograma (gráfica de la STFT) de un ciclo cardíaco, con el cual se observa la evolución tiempo-frecuencia de los eventos S1 y S2 además de su contenido frecuencial ya antes definido.



**Figura 3: Espectrograma de un ciclo cardíaco.**

La representación temporal de la envolvente de los sonidos cardiacos así como su contenido frecuencial permite identificar correctamente si se trata de eventos normales, como los que hemos mencionado en esta sección. Estos sonidos también garantizarán un funcionamiento mecánico adecuado del corazón, esto es, válvulas cardíacas trabajando adecuadamente.

### **2.3. Patologías cardiacas**

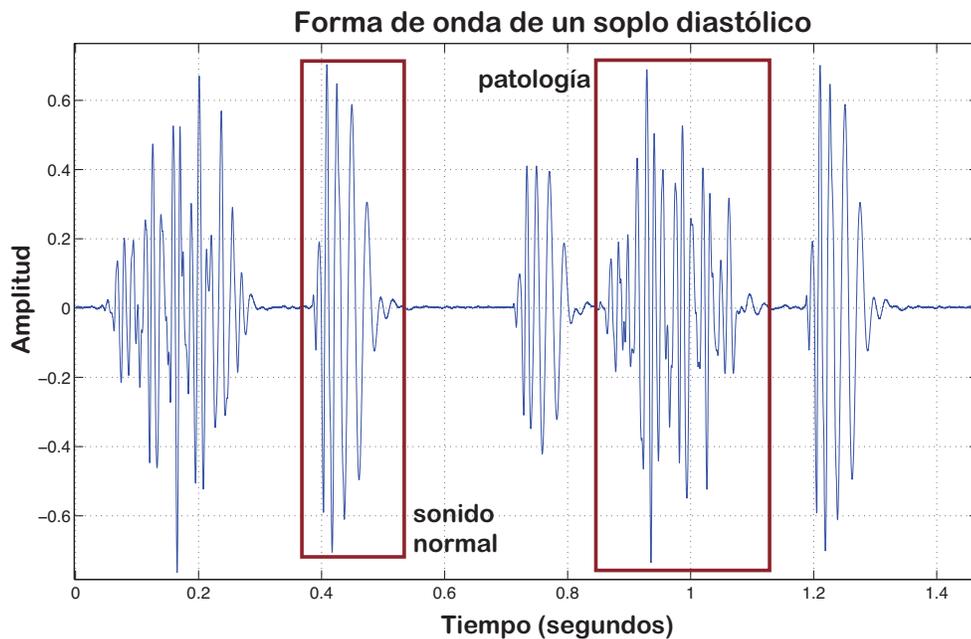
Las válvulas o valvas cardiacas son compuertas encargadas de regular el flujo sanguíneo a través de las aurículas y ventrículos del corazón, es por ello que la principal fuente de sonidos cardiacos anormales o patologías detectadas son daños en alguna(s)

<sup>1</sup>Transformada de Fourier de Tiempo Corto (*Short Time Fourier Transform*).

de éstas. Se distinguen principalmente dos tipos de daños valvulares en cuanto a su funcionamiento Leatham (1987).

Particularmente son las válvulas cardiacas izquierdas las que pueden fallar al no abrir apropiadamente, a esta anomalía se le conoce como *estenosis*. Por otra parte, cuando las valvas no puedan cerrar de manera adecuada se les denominará *incompetentes*, provocando un bajo flujo sanguíneo o *regurgitación* (Abbas y Bassam, 2009).

Los daños valvulares mencionados anteriormente producen un flujo turbulento del torrente sanguíneo a través del corazón, el cual presenta un comportamiento más brusco y puede ser percibido de manera auditiva (Köymen *et al.*, 1987). En efecto, esto se traduce a que el PCG de una patología o anomalía cardiaca presenta cambios a frecuencias más elevadas y tiempos más cortos, teniendo alteraciones en la forma de onda y desde luego en el espectrograma. La Figura 4 muestra la forma de onda de una señal patológica producto de un soplo diastólico, evento que se observa entre S1 y S2 presentando a diferencia de éstos una frecuencia no tan consistente y una duración temporal mayor.



**Figura 4: Forma de onda de un soplo diastólico.**

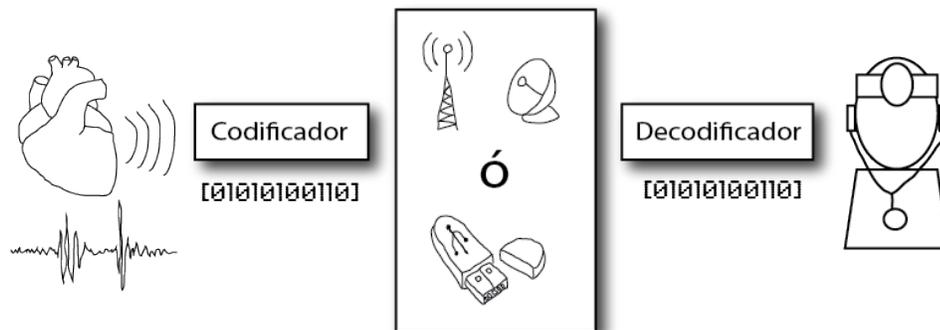
## 2.4. Codificación de audio

El sonido es un fenómeno dado por la propagación de ondas mecánicas a través de un medio, que generalmente es el aire en nuestro caso. Cualquier sonido tiene características analógicas por naturaleza, esto dado que por medio de nuestro sistema auditivo tenemos la capacidad en efecto de escuchar un sonido de manera continua (Painter y Spanias, 2000; Bosi y Goldberg, 2003).

Sin embargo, el empleo y demanda de sistemas digitales en la actualidad crea la necesidad de reproducir, transmitir, procesar y grabar al sonido digitalmente. Un *códec* (codificador-decodificador) es un dispositivo que toma las señales analógicas de audio como entrada y las transforma temporalmente en una representación digital<sup>2</sup> conveniente para realizar las tareas antes mencionadas (Bosi y Goldberg, 2003).

En el caso de este trabajo de tesis se requiere transmitir y almacenar señales de audio cardiaco, las cuales son producto de un fenómeno continuo y por lo tanto analógicas por naturaleza. El canal de almacenamiento y/o transmisión es limitado, la baja tasa de datos disponible hace necesaria una compresión y codificación eficiente de la señal del fonocardiograma.

La Figura 5 muestra la funcionalidad que tendrá el códec adaptado a audio cardiaco que se presenta en este documento.



**Figura 5: Esquema general del codificador-decodificador diseñado para audio cardiaco.**

La codificación de audio es un proceso que requiere la digitalización de una forma de

<sup>2</sup>Es decir, una cantidad adecuada (limitada) de bits para poder almacenar y representar una señal de audio.

onda completamente analógica, para ello la señal deberá de sufrir diferentes procesos (Jayant, 1974).

En la Figura 6 se muestra una ilustración cualitativa de cómo una forma de onda análoga, la cual es *continua* en tiempo y en amplitud es transformada al dominio digital.

El proceso comienza con la discretización en tiempo, lo cual es conocido como *muestreo*. Posteriormente se realiza la discretización en amplitud, a lo cual se le denomina *cuantificación*.

La forma de onda mostrada en la Figura 6 de la parte a) es análoga, y hasta completar el proceso en la parte d) puede apreciarse la versión digitalizada de la misma. Cada una de las muestras de la forma de onda en tiempo mostradas en d) tiene un número finito de valores de amplitud y tiempo permisibles.

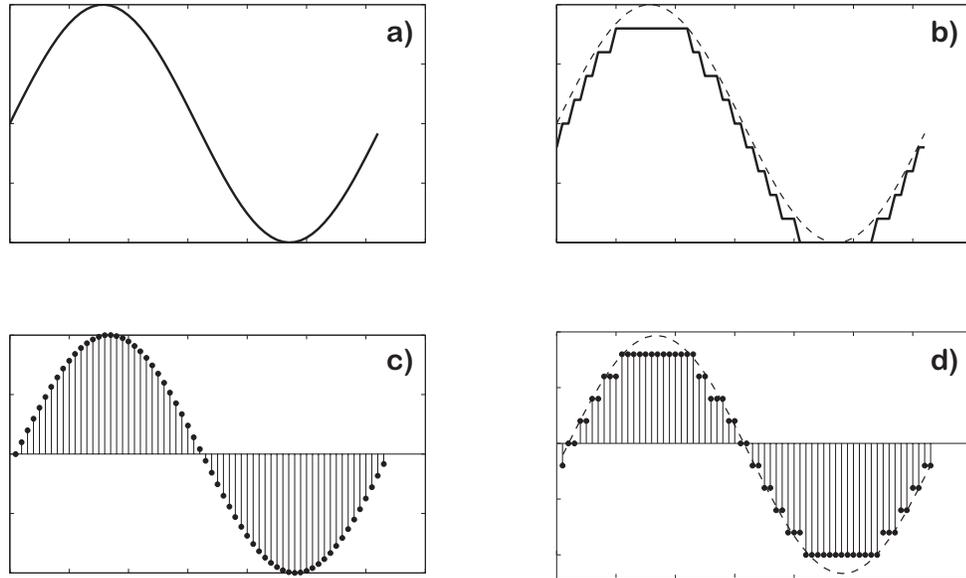
La transformación de a) hacia d) puede ser por medio del paso c) y posteriormente el d) generalmente, es decir, primero se muestrea a la señal y posteriormente se cuantifica. Es importante verificar la resolución en la realización de los procedimientos anteriores, pues en efecto la señal perderá calidad y en especial durante la cuantificación, que en efecto es un proceso de compresión con pérdidas. Durante el muestreo es importante respetar el parámetro de *frecuencia de muestreo* para una adecuada representación de la señal.

#### **2.4.1. La psicoacústica en el modelado y codificación de audio**

El diseño de los algoritmos de codificación está basado en las características del receptor, con lo cual se garantizará una optimización en la eficiencia y calidad del código. En el caso del audio el receptor es el oído humano, cuya percepción se ve afectada por sus propiedades de *enmascaramiento sonoro*<sup>3</sup>. Por medio de la *psicoacústica* se realiza la caracterización adecuada del sistema de percepción auditiva humana y particularmente el análisis de las capacidades tiempo-frecuencia del oído interno (Painter y Spanias, 2000).

---

<sup>3</sup>Fenómeno que ocurre cuando dos o más sonidos son reproducidos simultáneamente. El más débil de ellos resultará inaudible.



**Figura 6: Cuatro tipos de señales: a) Continua en amplitud y tiempo (analógica); b) discreta en amplitud y continua en tiempo; c) continua en amplitud y discreta en tiempo; d) discreta en tiempo y amplitud.**

Aplicar principios de percepción del oído para la codificación de audio es realmente relevante para diversos codificadores, algunos explotan el hecho de descartar información que no es detectable para cualquier oyente bien entrenado y así comprimir la señal (Schroeder *et al.*, 1979). Esta información irrelevante se identifica durante el análisis de señal incorporando en el codificador principios psicoacústicos como umbrales absolutos de audición, análisis de bandas de frecuencia críticas, enmascaramiento simultáneo y temporal, entre otros.

El umbral absoluto de audición caracteriza la cantidad de energía necesaria en un tono puro tal que pueda ser detectado por un oyente en un ambiente sin ruido. La Figura 7 muestra el comportamiento del umbral de audición absoluto en el espectro frecuencial. Dicha gráfica es el resultado de cuantificar el nivel de presión de sonido (SPL) requerido para cada frecuencia detectada por el oyente (tono puro).

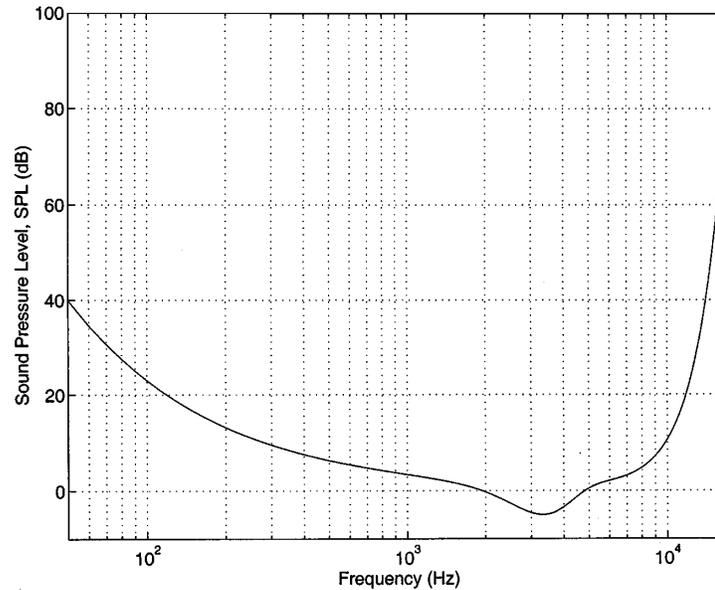


Figura 7: Umbral absoluto de audición. Tomado de Painter y Spanias (2000).

#### 2.4.2. El códec de audio

Como mencionamos anteriormente un códec de audio es el sistema que incluye el conjunto de algoritmos para permitir codificar y decodificar datos auditivos en una cantidad limitada de bits (Bosi y Goldberg, 2003; Painter y Spanias, 2000). Es importante que un códec pueda comprimir el archivo de audio para ocupar el menor espacio posible y una calidad aceptable para reproducirlo adecuadamente (Painter y Spanias, 2000). Los siguientes parámetros son importantes en la definición de un códec de audio:

- **Número de canales.** Esto se refiere al número de señales simultáneas de audio que contiene el flujo de datos. De acuerdo a ello la señal puede ser mono (un canal), estéreo (dos canales) o multicanal 5.1 (6 canales) ó 7.1 (8 canales).
- **Frecuencia de muestreo.** Una señal analógica de audio necesita ser discretizada primeramente en el tiempo, para ello se emplea el muestreo. El teorema de Nyquist establece que la frecuencia de muestreo adecuada para representar una señal debe ser el doble de la frecuencia máxima contenida en el ancho de banda de la señal. Ejemplos de ello son los estándares MPEG-3 (MP3), el cual muestrea con calidad de CD a diferentes frecuencias de muestreo; la más común es de 44,100 Hz teniendo

en cuenta que el oído humano no es capaz de escuchar frecuencias superiores a 22 kHz.

- **Número de bits por muestra.** Con ello se determina qué precisión existe al reproducir la señal original así como el rango dinámico de la misma. Suelen utilizarse en los codificadores convencionales 8, 16 ó 24 bits por muestra aunque en calidad CD comúnmente se emplean 16.
- **Tipo de compresión.** Se distingue entre compresión con pérdidas (*lossy compression*) y sin pérdidas (*lossless compression*).
  - *Compresión con pérdidas:* La información que se considera irrelevante es despreciada, aunque con ello es evidente tener una pérdida en la calidad del resultado al ejecutar la decodificación final.
  - *Compresión sin pérdidas:* Toda la información es transmitida, sólo se elimina la información repetida o redundante y la información que no se repite se almacena ocupando un menor espacio.
- **Tasa de bits (datos).** Esta cantidad determina el número de bits necesarios por unidad de tiempo y no puede deducirse a través de los parámetros anteriores, ya que pudo aplicarse una compresión con o sin pérdidas. De acuerdo a su estabilidad puede ser *constante (CBR)* o *variable (VBR)*. VBR es de uso más frecuente en audio y además más eficiente que CBR ya que existen tramos de silencios aleatorios donde deberá bajar la cantidad de bits a transmitir y por lo tanto variar la tasa.
  - *Tasa de bits constante (CBR):* La tasa de salida del codificador es constante. CBR es útil en flujo de datos multimedia para canales de capacidad limitada, sin embargo no es la mejor opción para almacenamiento, ya que no asigna suficientes bits para las secciones complejas (por lo tanto se degrada la calidad) además de usar bits innecesarios en secciones de menor complejidad.
  - *Tasa de bits variable (VBR):* En este método de compresión se consigue una mayor calidad de sonido para un tamaño de archivo determinado. En contraste a CBR se otorga la tasa de bits necesaria a cada parte del archivo que se pretende codificar consiguiendo una calidad mayor en archivos de tamaño

reducido. Al codificar con VBR, el codificador asignará las tasas de bits que variarán según la complejidad de la onda de audio a lo largo del archivo. Este formato es muy apropiado para la transmisión de audio en tiempo real.

En la actualidad existen diversos códecs de audio, aunque cada vez son más complejos y presentan características adicionales a las mencionadas anteriormente. De acuerdo a los parámetros existentes en algunos codificadores de audio expuestos por Bosi y Goldberg (2003), se ha realizado en esta tesis una clasificación en cuatro grandes grupos:

- **Codificadores de forma de onda.** Estos codificadores están basados en el estudio general de la señal, de tal manera que intentan reproducir la forma de la señal de entrada (O'Reilly, 1984). Son diseñados en general para ser independientes de la señal, de manera que puedan ser útiles para codificar una variedad extensa de señales; la redundancia de la señal es aprovechada para lograr una compresión.

Algunos codificadores de forma de onda son:

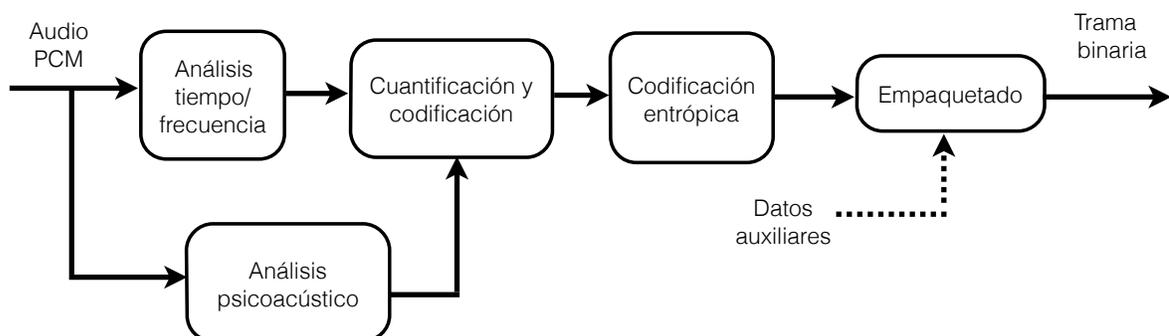
- Modulación por codificación de pulsos (*Pulse Code Modulation* ó PCM). Esta codificación es la base para la mayoría de los codificadores y a la vez se incluye en el procedimiento para digitalizar una señal. Basta con representar mediante impulsos las variaciones en amplitud de las muestras de la señal de entrada.
- Modulación diferencial por pulsos codificados (DPCM). Basándose en el principio de PCM, DPCM codifica solamente la diferencia entre la muestra presente y la muestra anterior de la señal de entrada. Se envía una cantidad mayor de información y un número más reducido de bits si se compara con PCM, sin embargo si existe un error en alguna trama codificada con PCM éste afectará tanto a la muestra actual y las muestras subsecuentes.
- Modulación diferencial adaptativa por pulsos codificados (ADPCM). Se trata de una variante de DPCM donde se varía la escala de codificación según las diferencias entre las muestras actual y subsecuente, es decir se asignan cantidades variables de bits a éstas según corresponda.

Algunos codificadores de forma de onda basados en PCM u otras de sus variantes

han tenido éxito para transmitir señales de voz sobre canales telefónicos. Ejemplos de éstos son el codificador  $\mu - 255PCM$  (Dammann *et al.*, 1972) y el  $G722$  (Mermelstein, 1988).

- Codificadores perceptuales.** Se aprovechan las limitaciones en la percepción del sistema auditivo humano para codificar el flujo de datos. Las muestras son codificadas en formato PCM y son convertidas al dominio de la frecuencia. Dada esta conversión la señal puede descomponerse en sub-bandas y proporcionar una resolución mayor a las bandas de frecuencia donde el oído humano sea más sensible. La codificación puede también llevarse a cabo por transformada, codificando las componentes de una transformada unitaria de la señal realizando la operación inversa (antitransformada o transformada inversa) en el decodificador. En este caso la compresión se presenta debido a lo no correlacionadas que tienden a ser las componentes generadas por las transformadas unitarias de manera que pueden codificarse independientemente. Otra manera útil de desechar información irrelevante al codificar audio de manera perceptual es la codificación en *sub-bandas*, pasando la señal de audio por diversos filtros pasa-banda a diferentes frecuencias para codificarla de manera independiente y dar diferente resolución a las bandas más sensibles.

Se muestra en la Figura 8 el diagrama a bloques que ilustra el procedimiento de codificación de audio de manera perceptual.



**Figura 8: Diagrama a bloques de un codificador de audio perceptual.**

El grupo de trabajo **Moving Picture Experts Group (MPEG)** ha establecido diver-

Estos estándares para la transmisión de audio, aprovechando en ellos la codificación perceptual para diseñar codificadores como MP3 (Musmann, 2006; Noll, 1997). Otro codificador perceptual de audio que ha tenido reconocimiento en el mercado es el *Advanced Audio Coding (AAC)* (Bosi et al., 1997).

- Codificadores paramétricos.** Estos codificadores se basan en el hecho de que el audio y la voz pueden representarse y sintetizarse como tonos aislados o patrones armónicos, esto es, representarlos con ondas senoidales y componentes ruidosas (*sinusoid plus noise modeling*). Los parámetros de las sinusoides son los que se codifican: amplitud, frecuencia fundamental o componentes espectrales. Estos parámetros requieren por lo general pocos bits para ser representados (Quatieri y McAulay, 1986; McAulay y Quatieri, 1986; Ruiz Reyes y Candeas, 2010).

En la Figura se muestra el diagrama a bloques de un codificador de audio tipo paramétrico.

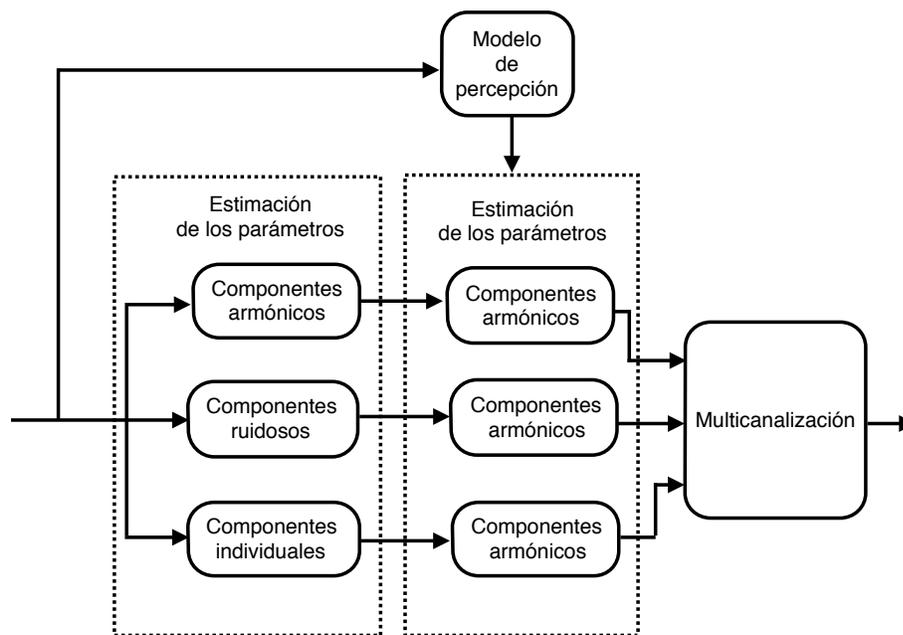


Figura 9: Diagrama a bloques de un codificador de audio paramétrico.

- Codificadores híbridos.** Combinan las técnicas de los codificadores de forma de onda para obtener la señal de audio de más alta calidad con tasas bajas de bits. Un conjunto de muestras es analizado como si se tratase de una sola para obtener los

parámetros de la señal. Cuando la trama se decodifica los parámetros son sintetizados para conseguir que se parezca al original. Se conocen algunos codificadores híbridos de la literatura:

- El vocoder (codificador de voz) basado en la técnica *Residual-Excited Linear Prediction (RELP)* desarrollado por Un y Magill (1975).
- Algunos codificadores de voz basados en el algoritmo *Code Excited Linear Prediction (CELP)*, donde se emplean diccionarios o bloques de código adaptativos (Mano *et al.*, 1995; Kipper *et al.*, 1991; Miki *et al.*, 1994; Koishida *et al.*, 1996).
- Codificadores que emplean predicción lineal mediante suma de vectores, conocidos como *Vector Sum Excited Linear Prediction (VSELP)*. Algunos de ellos han sido desarrollados por Gu *et al.* (1995); Sunwoo y Park (1991); Choi *et al.* (1996); Gerson y Jasiuk (1990).
- Codificación mediante el método *Regular Pulse Excitation-Long Term Prediction (RPE-LTP)*. Los codificadores diseñados por Huerta y Stern (2001); McLoughlin y Zhong-Qiang (2000); Coleman *et al.* (1989) se basan en esta técnica.

En este trabajo de tesis se desarrollará un codificador tipo *paramétrico* con pérdidas aprovechando la alta correlación existente entre los eventos del audio cardiaco y las formas de onda conocidas como átomos de Gabor. La siguiente sección describe las técnicas que se han usado en la literatura para el modelado del fonocardiograma.

## 2.5. Estado del arte en el modelado de sonidos cardiacos

La primer parte para el diseño de un codificador de audio cardiaco paramétrico consiste en el modelado matemático, el cual dotará de características propias y codificables al PCG además de encontrar la representación más adecuada y que represente mayor compresión. Robert Hooke, conocido por sus trabajos en elasticidad de materiales, otorgó potencialidad en el diagnóstico de patologías a la auscultación cardiaca (Wood y Barry, 1995). Y es que Hooke fue capaz de escuchar claramente los sonidos pertenecientes a los latidos del corazón y planteaba que tal vez pudiese indagar sobre el funcio-

namiento de otros órganos del cuerpo humano mediante el sonido que éstos generaban (Leatham, 1987).

El estetoscopio fue creado por Laënnec en 1816 (Roguin, 2006) e hizo del fonocardiograma una herramienta fundamental en la diagnosis clínica hasta nuestros días. Sin embargo, la escasez de estándares en la nomenclatura, los transductores empleados en el registro y las locaciones para llevarlo a cabo mostraron un progreso lento; como resultado se tuvo un pobre entendimiento de los mecanismos del sonido cardiaco y la inherente complejidad de los sonidos cardiacos.

Sin embargo, algunos factores han hecho del PCG una herramienta de diagnóstico limitada. La percepción auditiva humana es la primer limitante; aunque los especialistas y cardiólogos muestren singulares habilidades en el reconocimiento de diversas patologías no se garantiza que la auscultación tenga una precisión para diagnosticar. En efecto, toda señal de audio es pasa-bajas, de baja intensidad para el oído humano y muy corta duración. La banda baja de frecuencias puede concentrar también información importante del fonocardiograma.

Los avances en el procesamiento digital de señales así como la aparición del diagnóstico asistido por computadora (*Computer-Aided Diagnosis*) han logrado reforzar al PCG para ser una herramienta importante en el estudio de anomalías cardiacas. Los estetoscopios actuales muestran ahora las formas de onda y el análisis de la señal es más extenso (Mahnke, 2009; Watrous *et al.*, 2008).

Algunos trabajos en la literatura han realizado un análisis extenso en el plano tiempo-frecuencia del PCG, mostrando una caracterización espectral del mismo ya sea por espectrogramas calculados por la STFT o distribuciones Wigner-Ville (Ghofrani *et al.*, 2002), tal es el caso de Djebbari y Reguig (2000); Debbal y Bereksi-Reguig (2007, 2008) y Jasper y Othman (2010). En otros trabajos como los realizados por Boutana *et al.* (2011) y Debbal y Bereksi-Reguig (2006) se realiza la identificación de cardiopatologías usando como herramientas las distribuciones tiempo-frecuencia antes mencionadas.

De acuerdo a lo mencionado en las secciones anteriores de este trabajo con respecto al PCG se trata de una señal no estacionaria, esto es, los ciclos cardiacos no son consis-

tentes en posición, duración temporal ni contenido frecuencial. Dada esta característica los autores se han dedicado a analizar las componentes principales por separado: los primeros eventos cardiacos (S1) son estudiados en los trabajos de Chen *et al.* (1997); Wood y Barry (1995) y Wood *et al.* (1992), mientras que los trabajos de Hedayioglu *et al.* (2012); Wang y Pan (2004) han estudiado de manera específica el comportamiento del evento S2.

Se ha propuesto en el trabajo de Köymen *et al.* (1987) un modelado del PCG a partir de sinusoides amortiguadas, dadas que las estructuras en el pecho (lugar donde se realiza la auscultación) y venas de mayor contribución constituyen un sistema selectivo en frecuencia excitado por la aceleración y desaceleración de las válvulas.

El modelado del fonocardiograma mediante senoidales amortiguadas se ha extendido en la literatura para realizar diversos procesos en el PCG tales como eliminación de ruido (*denoising*), segmentación de eventos, separación de componentes (diagnóstico de cardiopatologías así como la división de los eventos S1 y S2) y escalamiento en el plano tiempo-frecuencia (Nieblas *et al.*, 2013; Zhang *et al.*, 1998b,a; Sava *et al.*, 1998).

Los trabajos anteriormente citados emplean el algoritmo **Matching Pursuit** (Mallat y Zhang, 1993) con éxito en el análisis y síntesis del fonocardiograma, tal método selecciona un conjunto de formas de onda cosenoidales (cuyo *amortiguamiento* viene dado por la modulación mediante una ventana gaussiana) denominadas *átomos de Gabor* pertenecientes a un bloque conocido como diccionario.

La compresión de fonocardiogramas es el paso donde se extraen los parámetros más importantes del modelado de señal para representarlos de manera digital. Otros autores de la literatura han realizado este procedimiento efectuando la representación del audio cardiaco mediante ondeletas (en Inglés *wavelets*) como Alajarín y Merino (2004, 2005); Martínez-Alajarín *et al.* (2006), además de realizar también por el mismo método el proceso de *de-noising* (Messer *et al.*, 2001).

Por otra parte, en la tesis desarrollada por Castorena (2012) se propone adaptar códecs ya existentes para transmitir el audio cardiaco sobre diversos escenarios que involucran redes con bajas tasas de datos. Posteriormente Nieblas (2014) analizó diversos

bloques de ondeletas mediante Matching Pursuit para elegir el que represente a la señal PCG de la manera más adecuada.

No obstante, hasta el momento no se conoce que se haya diseñado algún codificador de audio cardíaco, se ha revisado en esta sección que el modelado y la compresión del PCG han sido explorados quedando pendiente la realización de un códec que permita describir en bajas tasas de datos a esta señal para su transmisión en redes de bajos recursos.

En el siguiente capítulo se aplicará la compresión de fonocardiogramas mediante **Matching Pursuit**, ya que conforma una parte medular del codificador a diseñar en esta tesis.

## Capítulo 3. Modelado de la parte determinística mediante Matching Pursuit

---

Este capítulo presenta una breve descripción del algoritmo iterativo *Matching Pursuit* (MP), que descompone una señal en una combinación lineal de formas de onda elementales pertenecientes a un diccionario redundante de funciones. La compresión se realiza debido a que esta técnica es del tipo *representación dispersa* de señales, es decir, mediante un número reducido de ondas se representan eventos cardiacos extrayendo una gran cantidad de su energía. De igual manera MP es conocido como un método de *representación atómica*, dado que los bloques o diccionarios redundantes de funciones que se les conoce también como *átomos*.

Se ha seleccionado un diccionario multi-bloques para la descomposición cuyos elementos son conocidos como *átomos de Gabor* dada su alta correlación con los eventos cardiacos.

### 3.1. Representación dispersa de señales

Una herramienta que ha demostrado ser eficiente en el análisis y representación de señales es la representación dispersa (Gribonval y Nielsen, 2003, 2007; Elad, 2010; Mallat y Zhang, 1993). Varios de los procedimientos efectuados en el tratamiento de señales de audio e imágenes como la transformación, compresión, *de-noising*, separación de fuentes y otros métodos de codificación se han efectuado con eficiencia mediante esta técnica. A continuación se da una breve definición sobre la representación dispersa<sup>1</sup> de señales.

El objetivo de la representación dispersa es encontrar una representación eficiente de un vector o señal  $\mathbf{v} \in \mathcal{H}$ , donde  $\mathcal{H} = \mathbb{R}^N$

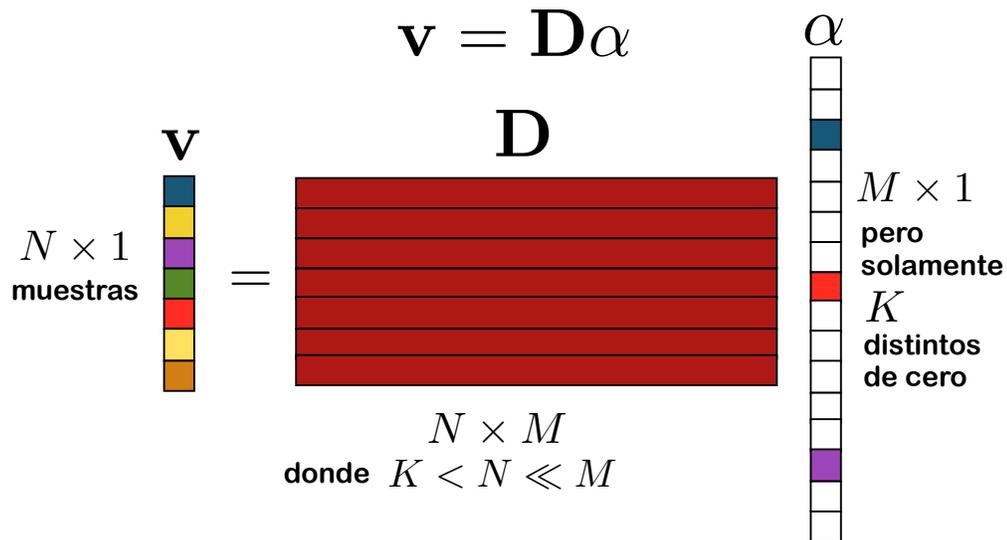
Una solución sencilla consistiría en tomar una base ortonormal  $\Phi = \{\phi_1, \dots, \phi_N\}$  para  $\mathcal{H}$  y usar los coeficientes de Fourier  $\{\langle \mathbf{v}, \phi_k \rangle\}_{k=1}^N$  para representar a  $\mathbf{v}$ . Esta respuesta es simple y funciona adecuadamente en muchos casos, sin embargo, también puede

<sup>1</sup>Es conocida en Inglés como *sparse representation*.

considerarse un tipo de expansión más general donde la base ortonormal es reemplazada por un diccionario adaptado formado desde  $\mathcal{H}$ .

Un **diccionario** en  $\mathcal{H} = \mathbb{R}^N$  es una familia de  $K > N$  vectores columna unitarios  $\{\mathbf{g}_k\}$  que expande o genera a  $\mathcal{H}$ . En notación matricial se puede expresar como  $\mathbf{D} = [\mathbf{g}_1, \dots, \mathbf{g}_K]$ .

Se entiende como una representación de  $\mathbf{v}$  en  $\mathbf{D}$  un vector columna  $\boldsymbol{\alpha} = (\alpha_k) \in \mathbb{R}^K$  tal que  $\mathbf{v} = \mathbf{D}\boldsymbol{\alpha}$ . Es importante verificar que  $K > N$ , esto es, los vectores en  $\mathcal{D}$  no son lo suficientemente independientes linealmente y la representación de  $\mathbf{v}$  no es única. Se espera que para todas las representaciones posibles de  $\mathbf{v}$  exista una *muy reducida*, esto es, la representación con el menor número de coeficientes diferentes de cero. Este procedimiento es representado de manera sencilla en la Figura 10, donde se verifica que los pocos elementos distintos de cero del diccionario conforman la representación dispersa ideal de  $\mathbf{v}$ .



**Figura 10: Representación gráfica de una representación dispersa para una señal  $\mathbf{v}$  desde un diccionario  $\mathbf{D}$ .**

Un compromiso que representa un problema en este proceso es el de encontrar todas las representaciones de  $\mathbf{v}$  para encontrar las representaciones dispersas (procedimiento que lleva un largo tiempo) y entonces determinar si existe un conjunto único de representaciones dispersas. Varios tipos de diccionarios así como diversos procedimientos y

modificaciones a los mismos se han analizado en la literatura para resolver esta encomienda en el menor tiempo posible (Gribonval y Nielsen, 2007, 2003; Elad, 2010; Ruiz Reyes y Candeas, 2010).

### 3.2. Diccionarios tiempo-frecuencia

Es a través de la transformada de Fourier como se describe una señal en términos de su contenido en frecuencia, lo cual se trata de una combinación de formas de onda ortogonales. Esta herramienta es la de uso más común en el análisis de señales en el dominio frecuencial debido a la facilidad que presentan las funciones senoidales siendo su base.

Mediante una transformación por Fourier se lleva a cabo una proyección de una señal  $x(t)$  sobre funciones base exponenciales complejas de frecuencia  $\omega = 2\pi f$  :

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t} dt, \quad (1)$$

donde  $i = \sqrt{-1}$  y  $X(\omega)$  es la *transformada de Fourier* de la señal  $x(t)$ .

Sin embargo, no todas las señales son idealmente representadas por esta transformación, por ejemplo aquellas que presentan transitorios (cambios abruptos en frecuencia para instantes cortos de tiempo), cuyas irregularidades requieren un análisis en tramos más consistentes. De esta manera surge la Transformada de Fourier de Tiempo Corto (STFT), la cual consiste en simplemente aplicar la transformada de Fourier a estructuras locales (segmentos con traslape) de la señal.

En la STFT se incluye el uso de una *ventana deslizante*  $w(\tau - t)$  cuya localización es desplazada en el eje temporal, con lo que se proveerá un análisis en el plano tiempo frecuencia dado por la expresión 2:

$$S(\omega, t) = \int_{-\infty}^{\infty} x(t)w(\tau - t)e^{-j\omega t} dt. \quad (2)$$

De aquí se deduce que el empleo de ondas bien concentradas<sup>2</sup> en el plano tiempo-

<sup>2</sup>Que se conozcan completamente tanto en tiempo como frecuencia (finitas en ambos planos).

frecuencia otorgarán una representación más adecuada en caso que la señal presente transitorios. El desarrollo de un modelo para el análisis o descomposición de señales se verá favorecido mediante la selección de los átomos óptimos, esto es que coincidan con las características básicas de la señal tanto en tiempo como en frecuencia (Mallat, 1999).

### 3.2.1. Diccionario y átomos de Gabor

El seleccionar un diccionario adecuado es una tarea fundamental en la descomposición y representación escasa de una señal. En la literatura se han propuesto diversos paquetes de ondas tiempo-frecuencia tales como ondeletas (*wavelets*), paquetes de cosenos, transformadas de coseno discreto modificada y diccionarios de Gabor (Wolfe *et al.*, 2001), etc.

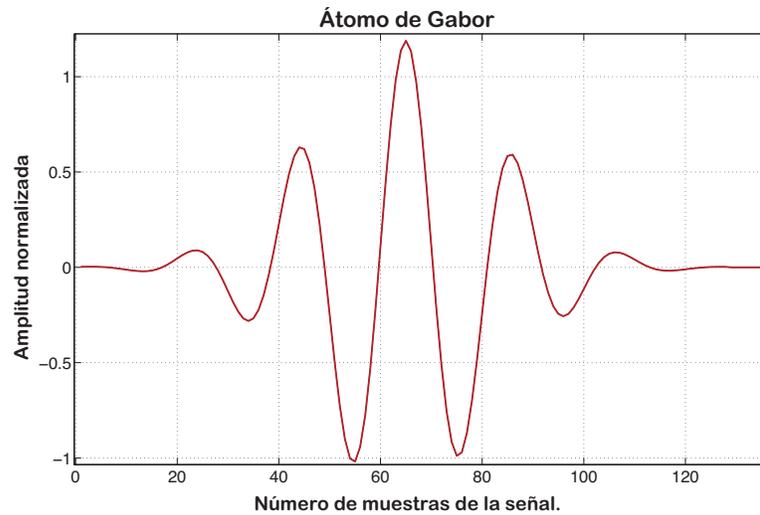
Estos bloques de ondas elementales han sido probados en la reconstrucción de fonocardiogramas (Nieblas, 2014), obteniendo que los diccionarios de Gabor son los más adecuados para este propósito.

Los átomos de Gabor son ondas cosenoidales que se obtienen por la dilatación, traslación y modulación de una ventana madre  $w(t)$ , la cual generalmente toma valores reales y positivos además de tener una norma unitaria. La ecuación (3) define esta forma de onda:

$$g_\gamma(t) = \frac{1}{\sqrt{s}} w\left(\frac{t-u}{s}\right) e^{i2\pi\xi(t-u)}, \quad (3)$$

donde  $w$  es una ventana gaussiana deslizante:  $w(t) = \sqrt[4]{2}e^{-\pi t^2}$ , la longitud de esta envolvente es controlada mediante la variable  $s$ , mientras que  $u$  es un desplazamiento para especificar la localización temporal del átomo y  $\xi$  es la frecuencia de modulación de la senoide.

Se define entonces como diccionario o bloque de Gabor al conjunto constituido por las ondas previamente descritas:  $\Gamma = \{\gamma_m\}_{m=1}^K$ , donde cada uno de sus elementos se definen como  $\gamma_m = (s_m, u_m, \xi_m)$ . La Figura 11 muestra la forma de onda de un átomo de Gabor.



**Figura 11: Forma de onda de un átomo de Gabor.**

### 3.2.2. Descomposiciones atómicas

Las bases de ondeletas fueron las predecesoras a las bases de Fourier en la reconstrucción de señales. Pocos coeficientes son necesarios para representar las estructuras correspondientes a los transitorios locales de la señal mediante ondeletas (Mallat, 1999), constituyendo mediante este método un esquema de compresión.

La historia de la reconstrucción por ondeletas comienza con Haar (1910), quien propuso una función constante cuyas traslaciones y dilataciones constituyen una base ortonormal para la representación de señales. Esta ondeleta lleva su mismo nombre.

Posteriormente Gabor (1946) propuso la descomposición de una señal sobre diccionarios de formas de onda elementales llamados átomos, los cuales tienen un esparcimiento mínimo en el plano tiempo-frecuencia. En efecto, a este método se le denomina descomposición atómica de señales.

Para la reconstrucción de la señal se lleva a cabo la suma de los átomos tiempo-frecuencia seleccionados del diccionario. Es fundamental el hecho de entender que los diccionarios son bloques de formas de onda atómicas los cuales presenten elementos de características tiempo-frecuencia adaptados a las propiedades de la señal; la mejor representación de la señal es un problema complejo, dado que existen un sinnúmero de des-

composiciones posibles (Gribonval y Nielsen, 2007).

Sea una señal continua  $x(t)$ , cuya representación atómica en un conjunto de  $K$  funciones  $g_m(t)$  tiene la siguiente estructura:

$$x(t) = \sum_{i=1}^K \alpha_i g_i(t), \quad (4)$$

donde los coeficientes  $\alpha_i$  denotan la ponderación asociada a cada función átomo  $g_i(t)$ . También deberá asimilarse que la señal  $x(t)$  es finita, para que su energía sea adecuadamente representada tanto en tiempo como en frecuencia (esta representación también puede realizarse con una señal discreta  $x(n)$  finita de  $N$  elementos).

La selección de funciones elementales o átomos que mejor aproximan a  $x(t)$  así como el cálculo de los coeficientes  $\alpha_i$  puede realizarse mediante diversos métodos (Nieblas, 2014), en este caso se ha seleccionado el algoritmo iterativo *Matching Pursuit* (Mallat y Zhang, 1993).

### 3.2.3. Principio de incertidumbre de Heisenberg

Es importante en una descomposición atómica el conocer la concentración de energía de las ondas tiempo-frecuencia empleadas, lo cual garantizará una buena reconstrucción de la señal tratada (Mallat, 1999). En este caso, los átomos de Gabor son construidos por la traslación en tiempo y frecuencia de una ventana  $g$ :

$$g_{u,\xi}(t) = g(t - u)e^{i\xi t}, \quad (5)$$

donde  $u$  y  $\xi$  son las traslaciones o desplazamientos en tiempo y frecuencia correspondientemente.

La energía de  $g_{u,\xi}$  se concentra en la vecindad de  $u$  sobre un intervalo de tiempo de  $\sigma_t$ . La transformada de Fourier  $\widehat{G}$  se traslada mediante  $\xi$  por medio de:

$$g_{u,\xi}(\omega) = g(\omega - \xi)e^{-iu(\omega - \xi)}. \quad (6)$$

La energía de  $\widehat{g}_{u,\xi}$  se localiza sobre la frecuencia  $\xi$  en un intervalo de tamaño  $\sigma_\omega$ . Para el plano tiempo-frecuencia  $(t, \omega)$  la energía del átomo  $g_{u,\xi}$  se representa simbólicamente por un *rectángulo de Heisenberg*, el cual está centrado en  $(u, \xi)$  y tiene un largo de duración de tiempo  $\sigma_t$  y un ancho en frecuencia  $\sigma_\omega$ . El principio de incertidumbre de Heisenberg satisface la desigualdad siguiente:

$$\sigma_t \sigma_\omega \geq \frac{1}{2}. \quad (7)$$

El área del rectángulo de Heisenberg  $\sigma_t \sigma_\omega$  es mínima para una  $g$  tipo gaussiana, en cuyo caso hemos visto que los átomos  $g_{u,\xi}$  son llamados funciones de Gabor. La Figura 12 otorga una descripción gráfica en cajas de Heisenberg de dos átomos tiempo-frecuencia, cuya estructura gaussiana favorece el soporte en ambos planos.

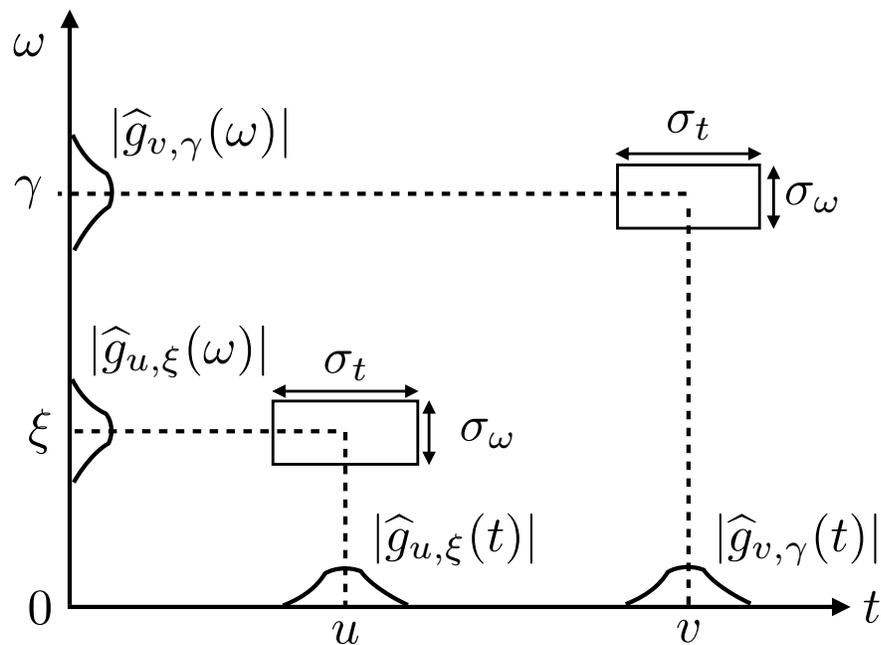


Figura 12: Representación de dos átomos tiempo-frecuencia en cajas de Heisenberg.

### 3.3. Descomposición de fonocardiogramas mediante Matching Pursuit (MP)

Tras con anterioridad haber descrito la importancia de las descomposiciones atómicas y la representación escasa de señales por medio de diccionarios tiempo-frecuencia (en específico los definidos por Gabor) es conveniente mencionar un algoritmo que conjunta estas técnicas como lo es Matching Pursuit (Mallat y Zhang, 1993).

Se ha presentado también MP en la literatura como un auxiliar muy completo en el análisis de fonocardiogramas, debido a que la mayor cantidad de la energía de éstos puede concentrarse por medio de este método en una representación de un número iterativo deseado de formas de onda elementales (Zhang *et al.*, 1998a,b; Wang y Pan, 2004; Sava *et al.*, 1998). Esto es, la cantidad de iteraciones deseada es proporcional al número de átomos de la representación y cuan mayor sea se preservará en efecto mayor energía en la reconstrucción.

Matching Pursuit es un algoritmo codicioso (*greedy*) debido a que en cada iteración sigue extrayendo energía a la señal analizada, resolviendo el problema de la mejor representación analizado al inicio de este capítulo mediante la realización del máximo absoluto del producto interno (también considerado como *correlación* en este caso). Esta operación garantiza el encontrar los átomos que mejor coincidan (*best-match*) con la estructura local de la señal en la iteración correspondiente.

La combinación lineal de formas de onda elementales, representación escasa o descomposición atómica de una señal  $x(t)$  mediante MP está dada por la expresión (8):

$$x(t) = \sum_{m=1}^M \alpha_m \cdot g_{\gamma_m}(t) + R_M(t), \quad (8)$$

donde  $M$  es el número de iteraciones ó átomos deseado en la descomposición,  $g_{\gamma_m}(t)$  es el átomo  $m$ -ésimo óptimo del bloque o diccionario (esto es,  $\gamma \in \Gamma$ ),  $\alpha_m$  el factor de ponderación y  $R_M(t)$  un término residual surgido de la diferencia entre la señal y el átomo ponderado después de la  $m$ -ésima iteración.

Por medio del Algoritmo 1 se observan los procedimientos correspondientes a la obtención de dichos términos, donde la relación señal a ruido (Signal to Noise Ratio, SNR) es obtenida como el cociente de la energía de la señal reconstruida por la suma de los átomos ponderados entre la energía de la señal residual.

Observando la forma de onda de los átomos de Gabor puede deducirse su alta correlación con los eventos cardíacos, por lo cual solamente estas partes de la señal son adecuadamente modeladas y reconstruidas mediante Matching Pursuit y diccionarios de

---

**Algoritmo 1 Matching Pursuit**


---

**Entrada:**  $x(t)$ ,  $\mathcal{D} = \{g_\gamma(t), \gamma \in \Gamma\}$ 
**Salida:**  $\alpha_m, g_{\gamma_m}(t)$ 

$$R = x(t)$$

$$\alpha_m = 0$$

**repetir**

$$g_{\gamma_m} = \arg \max_{\gamma \in \Gamma} |\langle R, g_\gamma \rangle|$$

$$\alpha_m = \langle R, g_{\gamma_m} \rangle$$

$$R = R - \alpha_m \cdot g_{\gamma_m}$$

**hasta que** se alcance la cantidad de iteraciones o SNR deseadas.

---

este tipo logrando una compresión.

La *segmentación* es el primer paso para el análisis del fonocardiograma, de esta manera se extraerán solamente los sonidos cardiacos principales y patologías para realizar su descomposición. Este proceso puede realizarse de manera manual o automática por medio de algún algoritmo (Nieblas *et al.*, 2013).

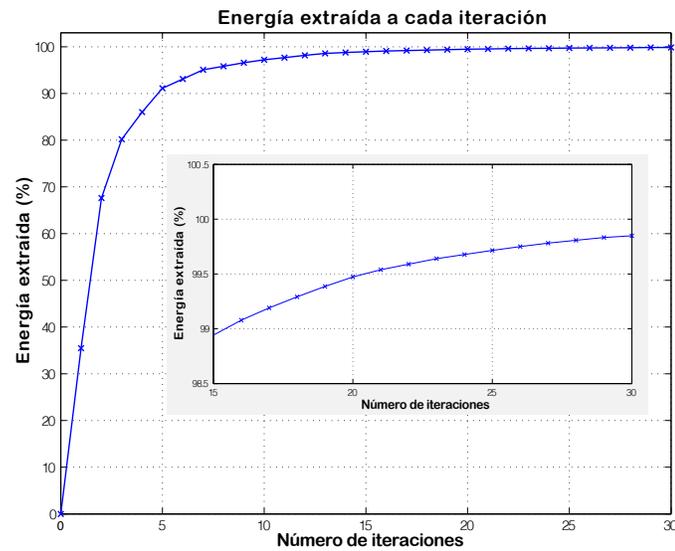
Durante la segmentación se detectan los tiempos de inicio (*on-set*) y final (*off-set*) de algún sonido normal o patología cardiaca. Estos instantes de tiempo ayudan a diferenciar las partes del fonocardiograma que son silencio y las que se consideran evento.

### 3.3.1. Criterios en la descomposición de fonocardiogramas por MP

A pesar de tener una reconstrucción aceptable en las señales de fonocardiograma tras representarlas en los átomos seleccionados desde Matching Pursuit el 100% de la energía realmente nunca es alcanzado, se tiene un comportamiento asintótico entre ésta variable y el número de iteraciones.

Esto se observa con detenimiento en la Figura 13, para lo cual se ha hecho un acercamiento en la parte donde parece haberse restablecido el total de la energía (de las 21 a las 30 iteraciones).

Dado este comportamiento, en este trabajo se ha decidido representar mediante MP un 99% de la energía de la señal, preservando el 1% restante en la señal residual  $R_M(t)$ . Este residuo tiene correlación mucho más baja con respecto a las ondas de Gabor que el evento cardiaco inicial, que al igual que los silencios aún son perceptibles al oído hu-

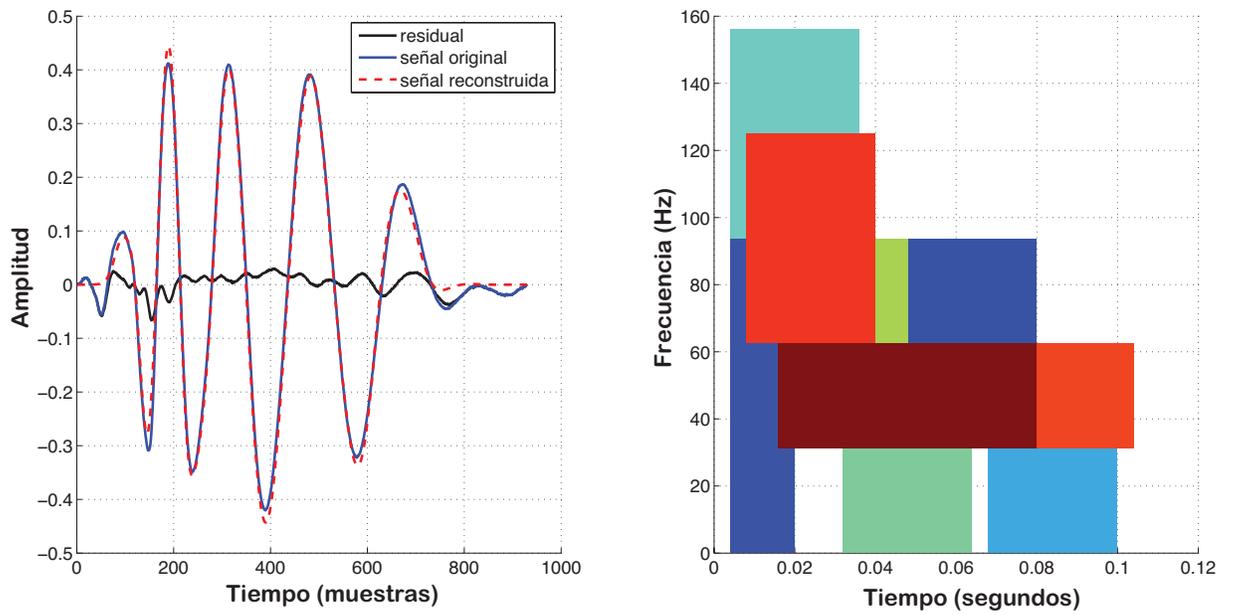


**Figura 13: Porcentaje de energía recobrada/extraída tras cada iteración realizada en Matching Pursuit. Se observa en el acercamiento a la gráfica que tras  $M = 30$  iteraciones realmente no se retiene el total de la energía.**

mano debido al comportamiento logarítmico de este órgano natural receptor (Smith y Abel, 1999).

En la Figura 14 se ha realizado la reconstrucción de un evento cardíaco  $s_1$  con 10 iteraciones, cantidad necesaria para representar en un 99% su energía. Se muestra la forma de onda temporal de la señal original, el residual y la señal reconstruida, así como la representación en el plano tiempo-frecuencia en cajas de Heisenberg de los átomos necesarios para la descomposición.

La señal residual se modelará con técnicas de predicción lineal (LPC), presentadas en el siguiente capítulo. Esta señal se considera como la parte *estocástica* del códec.



**Figura 14: Forma de onda temporal y gráfica en el plano tiempo-frecuencia en la reconstrucción de un evento cardiaco s1.**

## Capítulo 4. Modelado de la parte estocástica mediante Codificación Predictiva Lineal

---

Las señales tratadas en tiempo discreto pueden ser de naturaleza determinística o aleatoria, esto dependerá del tipo de expresiones matemáticas con las que puedan ser caracterizados (Hayes, 2009).

Los procesos aleatorios discretos en tiempo resultan más complicados de modelar matemáticamente, por ello deben emplearse herramientas de tipo estadístico, por ejemplo la **correlación**, que es una caracterización de segundo orden estadístico que provee información acerca de cuánto dependen linealmente dos señales. Esta propiedad puede ser aprovechada para diseñar sistemas de predicción de señal.

En el presente capítulo se describirá de manera resumida la técnica de codificación predictiva lineal LPC por sus siglas en Inglés, la cual caracteriza a una señal aleatoria discreta. Es una técnica de compresión y representación de secuencias discretas que en este caso se empleará como modelo de una fuente de ruido (*noise-like*) del codificador de audio a diseñar. Particularmente LPC ha tenido éxito en el caso de la codificación de voz.

Desde su introducción por Makhoul (1973), los modelos de predicción lineal han tenido éxito en la codificación de audio debido a su baja complejidad, buena eficiencia y escaso costo computacional. Basándose en técnicas de regresión por mínimos cuadrados se obtiene la envolvente espectral de una señal, la cual proporciona la información suficiente para ser percibida de manera auditiva (Makhoul, 1975a).

Las ecuaciones en diferencias y los filtros digitales son elementos con los que se tratan a las señales en tiempo discreto, en especial de manera estadística por tratarse principalmente de conjuntos de valores aleatorios (Hayes, 2009), por consiguiente se definirán en el presente capítulo. De igual forma es necesario introducir a los modelos autorregresivos (AR) como técnicas de recursividad para la caracterización de algunos de estos procesos. LPC, en efecto, es un modelado autorregresivo.

En esta sección describen los principales parámetros que emplea LPC al abordar el

problema de la predicción lineal tal y como es planteado por Vaidyanathan (2008). De igual manera, en este capítulo se ahondará el problema de encontrar los coeficientes que conforman el predictor lineal desde la perspectiva de Rabiner *et al.* (1978); Rabiner y Juang (1993), quien mediante el método de la autocorrelación otorga una respuesta favorable y sencilla que al problema que Makhoul (1975a) había introducido de manera determinística.

Por último se extrapolará la idea de predicción lineal desde el dominio temporal hacia el frecuencial (Makhoul, 1975b, 1973), donde la autocorrelación actúa como vínculo para lograr una coincidencia en la representación para ambos dominios.

Es necesario resaltar la importancia de LPC como técnica de compresión, tal y como lo manifiesta Gersho y Gray (1992), ya que se lograrán representar segmentos de señal con grandes cantidades de muestras con un escaso número de coeficientes de una representación puramente lineal.

## 4.1. Filtros digitales y transformada $z$

### 4.1.1. Transformada $z$

No todas las señales pueden ser descritas por la transformada de Fourier de tiempo discreto (DTFT) para su representación en el dominio frecuencial, por ello se emplea la *transformada  $z$*  como generalización para incluir a estas secuencias y representar su caracterización espectral.

La transformada  $z$  de una señal en tiempo discreto se define como:

$$S(z) = \sum_{n=-\infty}^{\infty} s(n)z^{-n}, \quad (9)$$

donde  $z = re^{j\omega}$  y teniendo el caso de que cuando  $r = 1$ :

$$S(e^{j\omega}) = S(z)|_{z=e^{j\omega}} = \sum_{n=-\infty}^{\infty} s(n)e^{-jn\omega}, \quad (10)$$

siendo entonces lo definido en la ecuación (10) la misma transformada de Fourier en

tiempo discreto (DTFT) de  $s(n)$ .

Se puede también definir como *función del sistema* a la transformada  $z$  de la respuesta al impulso  $h(n)$  de un sistema. Esta función servirá para conocer la salida de un sistema a una entrada específica y será de utilidad en las siguientes definiciones:

$$H(z) = \sum_{n=-\infty}^{\infty} h(n)z^{-n}. \quad (11)$$

#### 4.1.2. Filtros digitales

Un filtro digital es definido como un sistema lineal e invariante en tiempo (LTI) que modifica bajo circunstancias controladas a una señal discreta tanto en frecuencia como en tiempo (Hayes, 2009). La función que define a un filtro digital consiste en una ecuación en diferencias, en este caso y por simplicidad de constantes lineales (LCCDE):

$$y(n) + \sum_{k=1}^p a(k)y(n-k) = \sum_{k=0}^q b(k)s(n-k), \quad (12)$$

donde  $s(n)$  es la entrada y  $y(n)$  la salida correspondiente del filtro digital, los enteros  $p$  y  $q$  determinan su orden mientras que  $a(1), a(2), \dots, a(p)$  y  $b(0), b(1), \dots, b(q)$  lo definen o caracterizan <sup>1</sup>.

La salida del sistema  $y(n)$  dada por el modelo matemático surge al reescribir la expresión (12):

$$y(n) = \sum_{k=0}^q b(k)s(n-k) - \sum_{k=1}^p a(k)y(n-k). \quad (13)$$

En efecto,  $y(n)$  es una combinación lineal de los valores anteriores  $y(n-k)$ . De acuerdo a la existencia o no de los coeficientes  $a's$  y  $b's$  se definen los filtros de respuesta al impulso finita o infinita (FIR o IIR) a continuación.

---

<sup>1</sup>Más adelante se analizarán las características de acuerdo a la magnitud de estos coeficientes.

#### 4.1.2.1. Filtros de respuesta al impulso finita (FIR LCCDE)

De la ecuación (13) en el caso de que  $p = 0$  la salida  $y(n)$  del sistema depende solamente de los valores de entradas anteriores  $x(n - k)$  teniendo un *Filtro de Respuesta al Impulso Finita*.

La función del sistema si se conoce la respuesta al impulso  $h(n)$  se define por el siguiente polinomio en  $z^{-1}$ :

$$H(z) = \sum_{k=0}^q b(k)z^{-k} = b(0) \prod_{k=1}^q (1 - z_k z^{-1}). \quad (14)$$

El polinomio determinado por (14) genera raíces  $z_k$  denominadas “ceros”. Debido a esta estructura el filtro FIR se define como *todo ceros*.

La figura 15 muestra un diagrama a bloques del filtro FIR, consistente en una serie de retardos  $z^{-1}$  conectados en serie y sumados a la salida. Es evidente que el número de bloques deseado es el mismo que indica el orden del sistema.

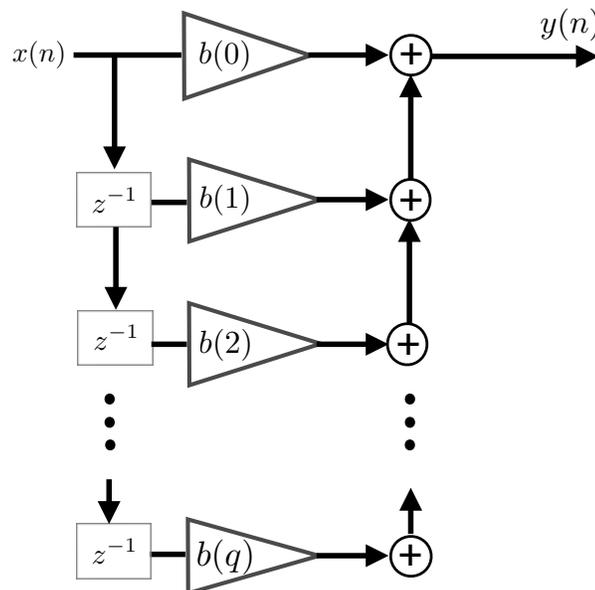


Figura 15: Diagrama a bloques de un filtro FIR.

#### 4.1.2.2. Filtros de respuesta al impulso finita (IIR LCCDE)

Cuando en la expresión (13) se tiene que  $p \neq 0$  la relación entrada-salida o función del sistema está dada por una operación fraccionaria, las raíces de los polinomios del denominador y numerador producen los ceros y polos del sistema respectivamente, los cuales son capaces de dar características sobre la estabilidad del sistema lineal.

La ecuación (15) denota la función de sistema de un filtro IIR:

$$H(z) = \frac{\sum_{k=0}^q b(k)z^{-k}}{1 + \sum_{k=1}^p a(k)z^{-k}} = b(0) \frac{\prod_{k=1}^q (1 - z_k z^{-1})}{\prod_{k=1}^p (1 - p_k z^{-1})}, \quad (15)$$

donde las raíces del denominador  $p_k$  son conocidas como los *polos* del filtro.

Si el orden del polinomio del numerador es igual a cero entonces:

$$H(z) = \frac{b(0)}{1 + \sum_{k=1}^p a(k)z^{-k}} = \frac{b(0)}{\prod_{k=1}^p (1 - p_k z^{-1})} \quad (16)$$

y se le conoce como un filtro *todo polos* al no contener ceros.

Teniendo que  $a(k), b(k) \in \mathbb{R}$  es posible que  $H(z) = H^*(z^*)$ , esto es, si  $H(z)$  tiene un polo en  $z = a$  también tendrá un cero en  $z = a^*$ .

Se observa en la figura 16 la estructura de un filtro IIR, observando que se requieren las muestras de los instantes anteriores de tiempo tanto de la entrada  $x(n)$  como de la salida  $y(n)$  del filtro.

## 4.2. Modelos autorregresivos

Un proceso estacionario en sentido amplio  $w(n)$  (WSS) <sup>2</sup> (Papoulis, 1984) se dice que es autorregresivo *AR* si puede ser generado empleando una ecuación en diferencias

<sup>2</sup>Proceso estacionario en sentido amplio: señal aleatoria o proceso estocástico que en sus valores no guarda correlación estadística en dos instantes de tiempo diferentes.

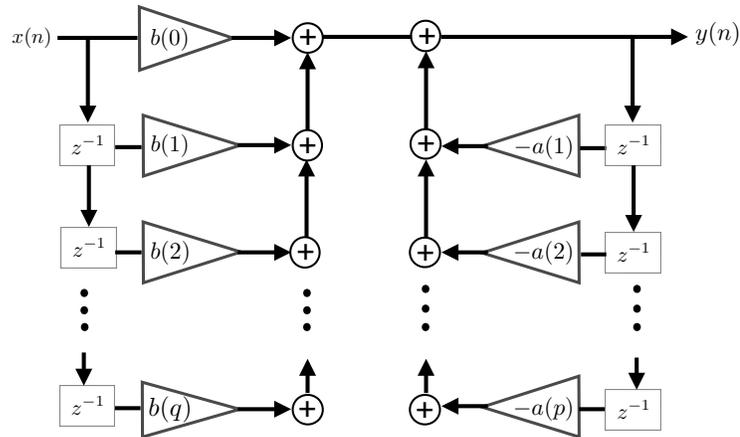


Figura 16: Diagrama a bloques de un filtro IIR.

recursiva como la siguiente:

$$w(n) = - \sum_{i=1}^N d_i^* w(n-i) + e(n) \quad (17)$$

Donde  $e(n)$  es también un proceso WSS con media cero.

Suponiendo que el siguiente polinomio es *todos ceros*:

$$D(z) = 1 + \sum_{i=1}^N d_i^* z^{-i} \quad (18)$$

se tiene ahora que  $w(n)$  es la salida de un filtro IIR *todo polos* definido por  $1/D(z)$  en respuesta a una entrada  $e(n)$ . Si los coeficientes  $d_i \neq 0$  el proceso es  $AR(n)$  ( $AR$  de orden  $n$ ). El proceso  $e(n)$  es WSS, tiene media cero y es de ruido blanco.

En la Figura 17 se muestra el modelo de un proceso  $AR$  como salida de un filtro de respuesta al impulso infinita cuya entrada es un proceso de ruido blanco.

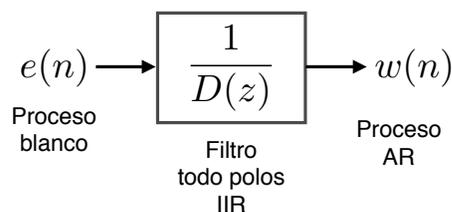


Figura 17: Modelo  $AR$  como salida de un filtro IIR en respuesta a una entrada de un proceso blanco.

Si se supone que para un proceso WSS  $s(n)$  se ha encontrado el polinomio predictor de coeficientes  $A_N(z)$  de  $n$ -ésimo orden, y además como se mencionó anteriormente, se puede representar como la salida de un filtro IIR, se tiene que la entrada a este sistema es el error de predicción  $e_N(n)$  y en dominio temporal se puede expresar como:

$$s(n) = - \sum_{i=1}^N a_i^* s(n-i) + e_N(n). \quad (19)$$

Prácticamente  $e_N(n)$  tiende a ser cercanamente blanco si  $N$  es razonablemente grande. Si se reemplaza a este proceso por ruido blanco con media cero,  $e(n)$ , y conociendo el error medio cuadrático  $\epsilon_N$  se obtendrá el modelo  $AR$  al cual denominaremos  $y(n)$  del proceso  $s(n)$ .

### 4.3. Codificación mediante predicción lineal

La idea básica de codificación predictiva lineal (LPC) consiste en que una muestra de un proceso aleatorio puede ser representado como una combinación lineal de las muestras anteriores, en la codificación se representan las envolventes en energía y en frecuencia como parámetros del modelo por medio de *coeficientes de predicción* de un filtro digital.

El conjunto de coeficientes predictores es único y puede determinarse minimizando la suma de las diferencias al cuadrado (la cual se denominará como *error cuadrático*) sobre un intervalo finito entre la muestra actual y las predichas linealmente.

Las muestras de una señal aleatoria discreta  $s(n)$  pueden ser producidas por una señal de excitación  $u(n)$  como se describe en la siguiente ecuación:

$$s(n) = \sum_{k=1}^p a_k \cdot s(n-k) + Gu(n), \quad (20)$$

donde  $a_k$  son los coeficientes predictores y  $G$  es la ganancia del filtro predictor;  $u(n)$  es la señal de excitación a la entrada del filtro.

La función  $H(z)$  que relaciona la entrada y la salida del sistema del filtro es la siguiente:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}}, \quad (21)$$

esto es un modelo denominado *todo polos* con el cual se representan secuencias de voz de manera natural (Rabiner y Juang, 1993; Rabiner *et al.*, 1978) y será utilizado en el codificador diseñado en esta tesis para el modelado de la parte estocástica.

Un predictor lineal de  $p$ -ésimo orden es un sistema de la forma:

$$\hat{s}(n) = \sum_{k=1}^p \alpha_k s(n-k) \Leftrightarrow P(z) = \sum_{k=1}^p \alpha_k z^{-k}. \quad (22)$$

A la diferencia entre las muestras de señal originales y las muestras estimadas por el predictor lineal se le conoce como *error de predicción*  $e(n)$ , el cual está dado por:

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k). \quad (23)$$

El error de predicción es también la salida de un sistema con la siguiente función de transferencia:

$$A(z) = \frac{E(z)}{S(z)} = 1 - P(z) = 1 - \sum_{k=1}^p \alpha_k z^{-k}, \quad (24)$$

si la secuencia original  $s(n)$  cumple exactamente con el modelo de predicción se tendrá una reconstrucción original. Esto se traduce a que  $\alpha_k = a_k$ , para  $1 \leq k \leq p$ ,  $e(n) = Gu(n)$  y además  $A(z)$  es un filtro inverso para  $H(z)$ , esto es:

$$H(z) = \frac{1}{A(z)}. \quad (25)$$

En la Figura 18 se muestra un diagrama del modelo de predicción lineal LPC, conjuntando las etapas de obtención de  $s(n)$  en cascada por medio del sistema *todo polos* (parte izquierda) y de reconstrucción por medio del filtro  $A(z)$  (parte derecha).

Por otra parte, la Figura 19 muestra la generación de una secuencia  $s(n)$  mediante

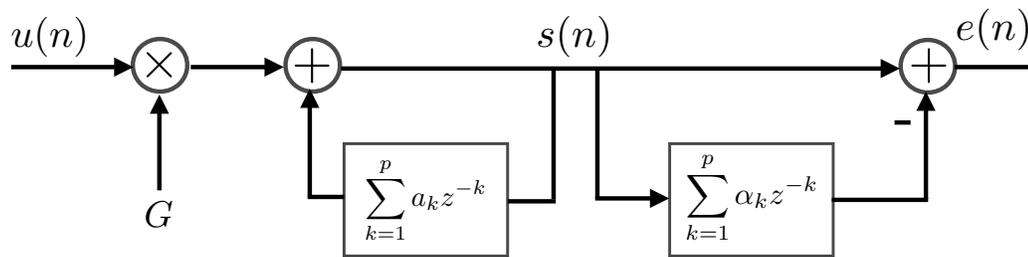


Figura 18: Modelo de predicción lineal.

LPC, la creación de la señal del error de predicción  $e(n)$  y mediante ésta la reconstrucción ideal de  $s(n)$ .

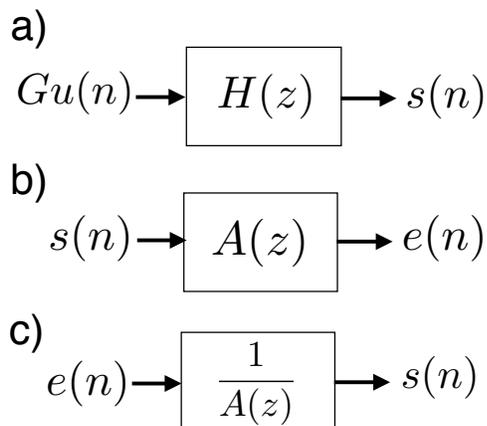


Figura 19: Modelado mediante LPC. En la parte a) se produce una secuencia  $s(n)$  por el modelo todo polos. La parte b) muestra la generación de  $e(n)$  como salida del filtro  $A(z)$ , cuya inversión en c) reproduce idealmente a  $s(n)$  si tiene como entrada al error de predicción  $e(n)$ .

#### 4.3.1. Modelo de predicción lineal

La naturaleza aleatoria de una secuencia de entrada  $s(n)$  de audio ha forzado el calcular los coeficientes de un filtro de predicción en segmentos pequeños de la señal (comúnmente de 10-20 ms para el caso de la voz)<sup>3</sup>.

Las características del error de predicción  $e(n)$  son importantes para definir qué clase de secuencia elegir como la entrada  $u(n)$ .

Según los trabajos realizados por Rabiner *et al.* (1978), se tienen dos casos para esta elección :

<sup>3</sup>En nuestro caso, el audio cardíaco posee baja correlación con los bloques de reconstrucción del modelo determinista elegido, por lo tanto el residual se considerará de esta naturaleza y se analizará también en segmentos o tramos de la misma longitud.

- Trama de audio vocalizada (*voiced sound*): El segmento analizado de audio provee a la salida una señal de  $e(n)$  con características **periódicas**, donde se ha encontrado una frecuencia fundamental  $f_0$  (llamada también tono o *pitch*). Dada esta periodicidad, se elige a  $u(n)$  como un tren de pulsos con periodo  $T_0$  (*pitch period*)<sup>4</sup>.
- Trama de audio no vocalizada (*unvoiced sound*): Se tiene un segmento de audio que arrojó a la salida una secuencia  $e(n)$  que **no presenta periodicidad alguna**, sus características son similares a las de una fuente de ruido blanco gaussiano y se elige a  $u(n)$  como una secuencia aleatoria con este comportamiento<sup>5</sup>.

Las características anteriores permiten producir a  $s(n)$  a través de una señal de excitación  $u(n)$  que será la entrada de un filtro inverso predictor lineal <sup>6</sup>.

En las siguientes secciones se abordará el problema de encontrar los coeficientes de predicción  $\alpha_k$  y la ganancia  $G$  que acompaña o complementa a  $u(n)$ .

#### 4.3.2. Predicción hacia adelante y hacia atrás

Como anteriormente se había considerado, la predicción lineal consiste en representar una muestra  $s(n)$  de una secuencia discreta a partir de la combinación lineal de los  $p$  valores anteriores

$$s(n-1), s(n-2), \dots, s(n-p).$$

El predictor lineal diseñado a partir de estos valores fue definido en la ecuación (22). A este procedimiento se le conoce como *predicción hacia adelante (forward prediction)* ya que se está prediciendo a partir de las muestras consecuentes en el tiempo.

En contraste, también es posible predecir *hacia atrás (backward prediction)*, usando las mismas  $p$  muestras u observaciones de la señal en una combinación lineal pero ahora para estimar el valor  $s(n-p-1)$  en un instante o unidad de tiempo anterior al valor observado.

<sup>4</sup>En el caso de las señales de voz ejemplos de sonidos vocalizados son los producidos al pronunciar fonemas como los de las letras “a”, “o”, entre otras.

<sup>5</sup>Sonidos como el de la “s” presentan este comportamiento.

<sup>6</sup>Se trata de un filtro con las mismas características que el filtro de predicción FIR lineal directo, sólo que la operación de filtrado se realiza de manera inversa.

Dadas estas condiciones el predictor *hacia atrás* se define como:

$$\hat{s}(n-p-1) = -\sum_{k=1}^p b_k \cdot s(n-k). \quad (26)$$

Ahora las mismas muestras relativas al instante de tiempo  $n-p-1$  son vistas como *valores futuros* y los coeficientes que denotan la predicción hacia atrás se denotan como  $b_k$ . La Figura 20 denota la simetría existente entre la predicción hacia adelante y hacia atrás, esto es, se emplean las mismas  $i$  muestras.

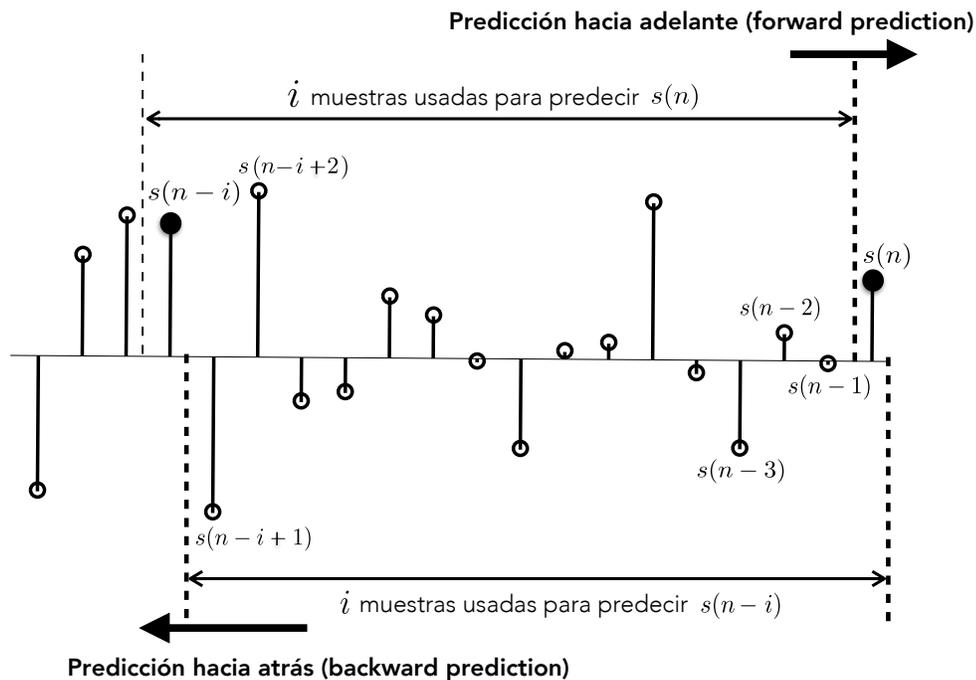


Figura 20: Simetría entre la predicción hacia adelante y hacia atrás.

#### 4.3.3. Estimación de los coeficientes de predicción

En el procedimiento de LPC, una tarea muy importante es la de encontrar los coeficientes predictores  $\alpha_k$  que minimicen el *error cuadrático medio*  $E$ , definido como:

$$E = \sum_{n=-\infty}^{\infty} e(n)^2. \quad (27)$$

Para esto existen los métodos basados en la autocorrelación y autocovarianza, se ha elegido el primero de ellos a manera de ilustrar con simplicidad la determinación de los  $\alpha_k$ 's.

De acuerdo a la expresión (27) para calcular  $E$  podemos reemplazar a  $e(n)$  de acuerdo a como se definió en (23) y tendremos que:

$$E = \sum_{n=-\infty}^{\infty} (s(n) - \hat{s}(n))^2 = \sum_{n=-\infty}^{\infty} \left[ s(n) - \sum_{k=1}^p \alpha_k s(n-k) \right]^2. \quad (28)$$

El predictor lineal óptimo es aquel diseñado para el mínimo valor de  $E$ . De acuerdo a lo mencionado en Makhoul (1975a), la minimización ocurrirá cuando:

$$\frac{\partial E}{\partial \alpha_i} = 0, \quad (29)$$

para  $i = 1, 2, 3, \dots, p$ , logrando diseñar un predictor de resolución variable de acuerdo al número de coeficientes.

Realizando la derivación parcial de  $E$  con respecto a  $\alpha_i$  como lo indica (29) desde (28) se tendrá que:

$$\begin{aligned} \frac{\partial E}{\partial \alpha_i} &= \sum_{n=-\infty}^{\infty} 2 \left[ s(n) - \sum_{k=1}^p \alpha_k s(n-k) \right] [-s(n-i)] = 0, \\ \sum_{n=-\infty}^{\infty} s(n)s(n-i) - \sum_{n=-\infty}^{\infty} \sum_{k=1}^p \alpha_k s(n-k)s(n-i) &= 0 \\ \sum_{n=-\infty}^{\infty} s(n)s(n-i) - \sum_{k=1}^p \alpha_k \sum_{n=-\infty}^{\infty} s(n-k)s(n-i) &= 0, \end{aligned}$$

de donde se puede obtener un sistema de  $p$  ecuaciones lineales simultáneas:

$$\sum_{k=1}^p \alpha_k \sum_{n=-\infty}^{\infty} s(n-k)s(n-i) = \sum_{n=-\infty}^{\infty} s(n)s(n-i), \quad (30)$$

el cual puede resolverse empleando la autocorrelación, definida como:

$$R(k) = \sum_{n=-\infty}^{\infty} s(n)s(n-k) = \sum_{n=-\infty}^{\infty} s(n)s(n+k), \quad (31)$$

esta función es simétrica, es decir, se cumple que  $R(k) = R(-k)$ .

Suponiendo que la señal  $s(n)$  ha sido segmentada o *ventaneada*, es decir  $s(l) = 0$  para  $l < 1$  ó  $l > N$ , puede realizarse un cambio de variable  $n - k$  por  $l$  en (30) para compactar en (31) al término que acompaña a la sumatoria de  $a_k$ , esto es:

$$\sum_{n=0}^{\infty} s(n-k)s(n-i) \rightarrow \sum_{l+k=-\infty}^{l+k=\infty} s(l)s(l+k-i) = R(i-k),$$

de la misma forma en (30) se puede denotar una expresión cerrada como la (31) en el término de la derecha, con lo que se tendrá:

$$\sum_{n=-\infty}^{\infty} s(n)s(n-i) = R(i).$$

Es importante observar que los límites de la sumatoria se conservan dada la suposición de que los valores de  $s(n)$  son cero fuera del intervalo  $0 \leq l \leq N - 1$ , lo cual puede realizarse multiplicando esta secuencia por una ventana de tiempo.

El procedimiento completo de realizar las sustituciones en (30) es el siguiente:

$$\sum_{k=1}^p \alpha_k \sum_{n=-\infty}^{\infty} s(n-k)s(n-i) = \sum_{n=-\infty}^{\infty} s(n)s(n-i) \rightarrow \sum_{k=1}^p \alpha_k R(i-k) = R(i). \quad (32)$$

Si se continúa desarrollando para  $i = 1, 2, \dots, p$  lo definido en (32) se produce el siguiente conjunto de ecuaciones lineales:

$$\begin{aligned} \alpha_1 R(0) + \alpha_2 R(1) + \alpha_3 R(2) + \dots + \alpha_p R(p-1) &= R(1) \\ \alpha_1 R(1) + \alpha_2 R(0) + \alpha_3 R(1) + \dots + \alpha_p R(p-2) &= R(2) \\ \alpha_1 R(2) + \alpha_2 R(1) + \alpha_3 R(0) + \dots + \alpha_p R(p-3) &= R(3) \\ &\dots \\ \alpha_1 R(p-1) + \alpha_2 R(p-2) + \alpha_3 R(p-3) + \dots + \alpha_p R(0) &= R(0). \end{aligned}$$

Reescribiendo de manera matricial el conjunto de ecuaciones de (32) se tiene:

$$\begin{bmatrix} R(0) & R(1) & \cdots & R(p-2) & R(p-1) \\ R(1) & R(0) & \cdots & \cdots & R(p-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R(p-2) & \cdots & \cdots & R(0) & R(1) \\ R(p-1) & R(p-2) & \cdots & R(1) & R(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ \vdots \\ R(p) \end{bmatrix}, \quad (33)$$

que en forma compacta corresponde a:

$$\mathbf{R}\boldsymbol{\alpha} = \mathbf{r}. \quad (34)$$

El conjunto de soluciones  $\alpha_k$  corresponde al vector  $\boldsymbol{\alpha}$  que se obtiene algebraicamente por:

$$\boldsymbol{\alpha} = \mathbf{R}^{-1}\mathbf{r}.$$

La matriz  $\mathbf{R}$  del sistema (34) es de tipo *Toeplitz*<sup>7</sup>, simétrica<sup>8</sup> y definida positiva (Gray, 2006). Con ello, el cálculo de la energía de la señal  $R(0)$  dará los valores de la diagonal principal en esta matriz.

El método denominado recursión de Levinson-Durbin es adecuado para resolver este sistema de ecuaciones y se explicará en la siguiente sección.

#### 4.3.3.1. Recursión de Levinson-Durbin

La estructura (33) presenta propiedades para derivar un algoritmo recursivo que permita invertir la matriz  $\mathbf{R}$ , además de observar en esta misma expresión que el vector  $\mathbf{r}$  es una versión sintetizada<sup>9</sup> del contenido existente en  $\mathbf{R}$ .

El algoritmo de Levinson-Durbin se ha formulado para diseñar un filtro predictor hacia adelante de tamaño  $p$  de memoria, comenzando con cero y realizando iteraciones en una

<sup>7</sup>Matriz cuadrada que se caracteriza por tener los mismos elementos en las diagonales de izquierda a derecha (son constantes).

<sup>8</sup>Obsérvese su simetría en la igualdad de los valores triangular inferior y superiores, que los elementos en la diagonal principal son los mismos y que es cuadrada en efecto.

<sup>9</sup>En efecto,  $\mathbf{r}$  presenta sólo una vez los valores de  $R(1)$  a  $R(p)$  mientras que  $\mathbf{R}$  repite a éstos y contiene además a  $R(0)$ .

cantidad deseada actualizando su tamaño (Rabiner *et al.*, 1978; Hayes, 2009; Vaidyanathan, 2008). Se trata de un procedimiento que por medio de una *recursión* determina el óptimo predictor de  $i$ -ésimo orden a partir del  $(i - 1)$ -ésimo predictor (filtro anterior).

Los valores de autocorrelación de  $\mathbf{R}$  determinan los coeficientes de reflexión  $k_i$  y los coeficientes de predicción  $\alpha_i$  calculándose de manera consecutiva.

La función del sistema del error de predicción  $A(z)$  definida en (24) satisface que:

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{(-i)} A^{(i-1)}(z^{(-1)}).$$

Esto es, el  $i$ -ésimo error de predicción hacia adelante  $e^{(i)}(n)$  es la salida de la  $i$ -ésima función de sistema del filtro de error de predicción  $A^{(i)}(z)$ . Todos los ceros de este sistema estrictamente se encuentran dentro del círculo unitario del plano  $z$  (Makhoul, 1975a).

El Algoritmo 2 es el procedimiento de recursión de Levinson-Durbin, donde es importante verificar que el índice superior marcado entre paréntesis denota el orden  $p$ -ésimo del predictor previamente seleccionado y el índice inferior el  $i$ -ésimo coeficiente a calcular. Los coeficientes de reflexión  $k_i$ , obtenidos a la par que los coeficientes  $\alpha_i$  de predicción

---

#### Algoritmo 2 Recursión de Levinson-Durbin

---

```

 $E^0 = R(0);$ 
para  $i = 1, 2, \dots, p$  hacer
    
$$k_i = \frac{\left( R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j) \right)}{E^{(i-1)}}$$

     $\alpha_i^{(i)} = k_i$ 
    si  $i > 1$  entonces
        para  $j = 1, 2, \dots, i - 1$  hacer
             $\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}$ 
        fin para
    fin si
     $E^{(i)} = (1 - k_i^2) E^{(i-1)}$ 
fin para
 $\alpha_j = \alpha_j^{(p)}$ 
 $j = 1, 2, \dots, p$ 

```

---

a la salida del algoritmo de Levinson, también son denominados *PARCOR*. Esto se debe a que son obtenidos por medio de correlaciones parciales (del Inglés *partial correlation*).

Una propiedad muy importante es que para todos los coeficientes reflectores se cumple que  $|k_i| \leq 1$ , lo cual sirve para garantizar la estabilidad del filtro predictor<sup>10</sup>.

Itakura (1975); Itakura y Saito (1968) lograron establecer una forma para calcular directamente estos coeficientes mediante una expresión que relaciona los valores de la correlación cruzada con los errores de predicción hacia adelante y hacia atrás, esto es:

$$k_i = \frac{\sum_{m=-\infty}^{\infty} e^{(i-1)}(m)b^{(i-1)}(m-1)}{\left( \sum_{m=-\infty}^{\infty} (e^{(i-1)}(m))^2 \sum_{m=-\infty}^{\infty} (b^{(i-1)}(m-1))^2 \right)}, \quad (35)$$

Donde  $b^{(i)}(m)$  es la salida de error de predicción del filtro hacia atrás.

#### 4.4. Estimación espectral

La interpretación en dominio frecuencial del análisis de predicción lineal por medio de la autocorrelación, provee información necesaria para representar el espectro de potencia de una señal dado por la STFT (Rabiner y Juang, 1993).

Todo parte del principio fundamental de que la función de autocorrelación  $R(n)$  es un proceso inverso al cálculo de magnitud al cuadrado del valor absoluto de la transformada de Fourier  $|S(e^{j\omega})|^2$  (conocido también como *espectro de potencia de la señal*).

El espectro de potencia estimado por LPC  $\hat{P}(\omega)$  ó  $|\hat{S}(\omega)|^2$ , corresponde al cálculo de la magnitud al cuadrado de la transformada de Fourier de la función del sistema  $H(z)$  que se expresó en (21):

$$\hat{P}(\omega) = |\hat{S}(\omega)|^2 = |H(e^{j\omega})|^2 = \frac{G^2}{|A(e^{j\omega})|^2}, \quad (36)$$

donde se ha sustituido  $z$  por  $e^{j\omega}$ .

Por otra parte, el espectro de potencia de la señal original  $|S(e^{j\omega})|^2$  ó  $P(\omega)$  es obtenido a partir de la función de sistema del error de predicción. Este espectro se calcula

<sup>10</sup>Esta propiedad será importante al cuantificar.

despejando  $S(z)$  de (24), y de igual manera, sustituyendo  $z$  por  $e^{j\omega}$  se obtiene:

$$P(\omega) = |S(e^{j\omega})|^2 = \frac{|E(e^{j\omega})|^2}{|A(e^{j\omega})|^2}. \quad (37)$$

Comparando el espectro estimado por LPC calculado en (36) y el espectro original definido en (37) se ha demostrado que  $|E(e^{j\omega})|^2$  es modelado por una señal de espectro plano (ruido blanco, por ejemplo)  $G^2$ , dado este efecto al filtro  $A(z)$  también se le conoce como *blanqueador* ya que produce una señal de este tipo (Makhoul, 1973).

Los transitorios que determinen la estacionariedad de la señal afectarán el cálculo de  $P(\omega)$  y por consiguiente su autocorrelación (que ahora se denotará como  $R_{\hat{n}}(i)$ ). Por lo tanto, la señal variará muy rápidamente en el tiempo y es preciso que sea *ventaneada*:

$$s_{\hat{n}}(m) = s(n + m) \cdot w(m),$$

donde la señal  $w(m)$  es una *ventana* de tiempo que idealmente sería rectangular y cuya función es hacer que se tenga estacionariedad por un gran periodo de tiempo.

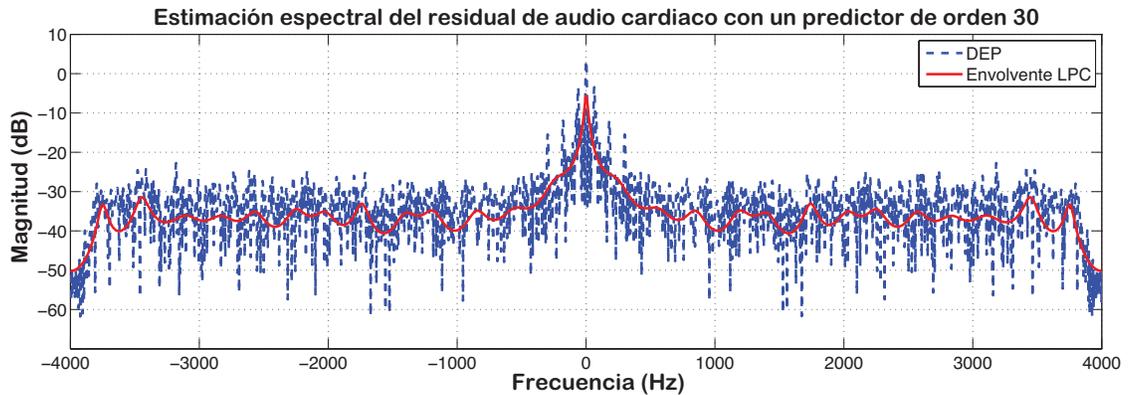
En este caso, la transformada de Fourier de tiempo corto *STFT* es la herramienta auxiliar para calcular el espectro de la señal  $P(\omega) = |s_{\hat{n}}(e^{j\omega})|^2$  y el espectro estimado, dado por:

$$|H(e^{j2\pi f/f_s})|^2 = \left| \frac{G}{1 - \sum_{k=1}^p \alpha_k H(e^{-j2\pi f/f_s})} \right|^2$$

siendo  $f_s$  la frecuencia de muestreo de  $s(n)$ .

En la Figura 21 se observa lo realizado por LPC en términos de la aproximación de la envolvente espectral de energía para una señal residual de audio cardiaco. La resolución en esta envolvente radica en incrementar el orden del filtro predictor.

La predicción del *futuro* mediante procedimientos estadísticos desde luego que no es con certeza, pero es posible determinar una estructura similar a través de la inercia presentada en las observaciones del pasado y el presente, premisa retomada por las



**Figura 21:** Aproximación por codificación predictiva lineal para una señal de audio cardíaco con un orden de filtro  $p = 30$ .

técnicas matemáticas de predicción en secuencias discretas que son de tipo aleatorio (Gersho y Gray, 1992).

En el caso de la predicción lineal se tiene una versión especializada de la teoría de estimación por mínimos cuadrados o regresión lineal, capaz de estimar modelos en tiempo y frecuencia. En efecto, se trata de un criterio de regresión por suma de diferencias al cuadrado en dominio temporal capaz de trasladar a un criterio de coincidencia espectral para el dominio frecuencial, lo cual es posible gracias al análisis desde la autocorrelación.

De igual manera, la predicción lineal es una técnica de compresión de señal ya que encuentra una representación eficiente y compacta de un proceso aleatorio.

#### 4.5. Variantes a LPC: WLPC y VELPC.

Las formas de excitar los filtros de reconstrucción son diversas en la codificación predictiva lineal. En ocasiones se desea prescindir de un algoritmo para la estimación del periodo tonal y se reconstruye al segmento de señal mediante un filtro excitado con alguna secuencia de características conocidas. Lo importante es preservar la compresión.

De igual forma los filtros digitales pueden ser modificados con el objeto de cambiar la distribución de los polos. Esto dará una mejor resolución al reconstruir una determinada banda de frecuencias.

A partir de estas modificaciones, la predicción lineal presenta otras dos versiones: co-

dificación predictiva lineal con distorsión espectral (*Warped LPC*) y codificación predictiva lineal con excitación por voz (*Voice-excited LPC*), las cuales son presentadas a continuación.

#### 4.5.1. Codificación predictiva lineal con distorsión espectral (WLPC)

El comportamiento logarítmico del sistema auditivo humano y la naturaleza pasa-bajas en dominio frecuencial de las señales de audio hace posible el modelado por predicción lineal mediante escalas deformadas (*warped*), donde la aproximación a la envolvente espectral posee mejor resolución en el intervalo de frecuencias audibles de interés. Esto origina las técnicas de predicción lineal a escalas compactas distorsionadas (Sturbe, 1980), donde la mejor resolución se da al tener una mayor concentración de polos en la banda frecuencial de interés para representar mejor la región espectral deseada y poderla escuchar mejor.

Una deformación (*warping*) puede aplicarse ya sea en el análisis o síntesis por predicción lineal a una señal de audio, con esto la envolvente espectral es arbitrariamente distorsionada sin afectar su estructura interna.

Existen diversos métodos de deformación de la escala frecuencial, el más común de ellos consiste en reemplazar los elementos de retardo unitarios de una estructura convencional por elementos pasa-todo de primer orden. El ajuste en la resolución espectral puede aproximar la resolución frecuencial audible humana. Autores como Härmä y Laine (2001) han empleado técnicas de predicción lineal por deformación espectral (WLPC) en el análisis de voz y aplicaciones de codificación.

La función de transferencia paso-todo de primer orden de un filtro todos polos está dada por:

$$D(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}, \quad (38)$$

lo cual reduce a una simple unidad de retardo con fase lineal y retardo de grupo constante.

La frecuencia de punto de inflexión  $f_{tp}$  (*turning-point frequency*) define la frecuencia a la cual la distorsión no afecta la resolución frecuencial, esto es, cuando el retardo de grupo es uno. La expresión (39) define a  $f_{tp}$  como una función del factor de distorsión  $\lambda$  y

la frecuencia de muestreo y es dada por:

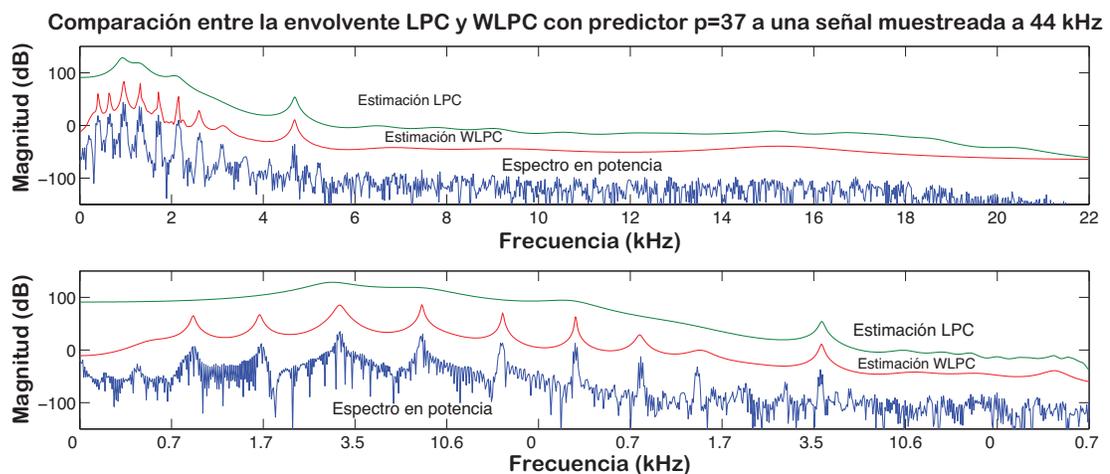
$$f_{tp} = \pm \frac{f_s}{2\pi} \arccos \lambda. \quad (39)$$

La resolución frecuencial de un sistema distorsionado con  $\lambda \geq 0$  es altamente por debajo y altamente por arriba de  $f_{tp}$  que en un sistema convencional con resolución frecuencial uniforme.

El cálculo del valor de distorsión espectral  $\lambda$  es un factor fundamental en WLPC que puede ser calculado en función de la frecuencia de muestreo y coincide con la transición hacia la escala psicoacústica Bark (Smith y Abel, 1999), la cual es modelada para coincidir con la percepción logarítmica humana:

$$\lambda_{f_s} = 1.0674 \left( \frac{2}{\pi} \arctan(0.06583 f_s / 1000) \right)^{1/2} - 0.1916. \quad (40)$$

En la Figura 22 se muestra una comparación entre las envolventes espectrales estimadas por predicción lineal convencional (LPC) y predicción lineal con distorsión (WLPC) para una señal de audio que fue muestreada a 44kHz. Puede observarse una resolución mejor en la banda baja de frecuencias por WLPC (por debajo de los 7kHz) en la definición de la envolvente.



**Figura 22: Comparación de la envolvente espectral de una señal proveniente de un sonido de piano muestreada con  $f_s = 44kHz$  y un orden de predicción  $p = 37$  por los métodos LPC y WLPC.**

#### 4.5.2. Codificación predictiva lineal con excitación por voz (VELPC)

La señal de error  $e(n)$  proveniente de la salida del filtro FIR predictor reconstruye perfectamente la entrada  $x(n)$  a la salida del filtro reconstructor inverso IIR. Si se toma un número de coeficientes  $M$  de la transformada discreta del coseno (DCT) de  $e(n)$  tal que  $M < N$  donde  $N$  es la longitud de esta señal se habrá logrado una compresión. En efecto, la excitación al filtro IIR de reconstrucción estará dado por la transformada inversa del coseno (IDCT) correspondiente a estos  $M$  coeficientes. Esta técnica es conocida como predicción lineal con excitación por voz (*Voice-excited LPC*).

La tarea de determinar el número  $M$  de coeficientes ideal para reconstruir con precisión  $x(n)$  puede realizarse con un *análisis de componentes principales* dado por la *transformada de Karhunen Loève* (Jayant, 1974). Este procedimiento trata de descomponer en valores singulares la matriz de autocorrelación de la entrada  $x(n)$  y analizar su comportamiento gráfico.

Para el codificador diseñado por este trabajo de tesis se usará la técnica convencional de LPC para el modelado de la parte estocástica, sin embargo, es importante para el lector conocer las variantes WLPC y VELPC que pueden emplearse en esta codificación para trabajos posteriores.

## Capítulo 5. Conformación del códec propuesto

---

Una vez definidos los conceptos necesarios para el modelado de la parte determinística (Matching Pursuit) y la parte estocástica (Codificación predictiva lineal) del codificador de audio propuesto para este trabajo de tesis, es conveniente mostrar los elementos necesarios para diseñarlo. Para ello, se tomarán los parámetros necesarios de ambos modelados que reconstruyan con precisión la señal de audio cardíaco.

Una vez obtenidos los parámetros necesarios para conformar el códec es necesario cuantificarlos, este requerimiento es necesario en el sentido de representarlos por medio de palabras código para ser almacenados. La cuantificación es un proceso irreversible de compresión con pérdidas en la codificación de toda señal de audio, el cual por lo tanto debe ser cuidadosamente analizado para que no se distorsionen en gran medida los parámetros que se han cuantificado.

En este capítulo se muestran los pasos necesarios para la conformación del codificador de audio cardíaco; la extracción de los parámetros y su cuantificación. Por último se definirán las tasas de compresión y medidas de distorsión generadas por las señales codificadas para comparar el códec de manera objetiva.

### 5.1. Estructura del codificador-decodificador propuesto

Para el desarrollo, modelado y pruebas realizadas al codificador y otros procesos durante este trabajo de tesis se han seleccionado 10 señales de audio cardíaco con duración de 5 segundos desde una base de datos disponible en el sitio web oficial de Litman ©, desarrolladores de estetoscopios y otras herramientas de diagnóstico clínico (Litmann, 2013).

Se han obtenido las señales indicadas en la Tabla 2 en formato .wav muestreadas a 11,025 Hz. Por cuestiones de compatibilidad con los diccionarios en la descomposición las señales han sido remuestreadas a 8,000 Hz.

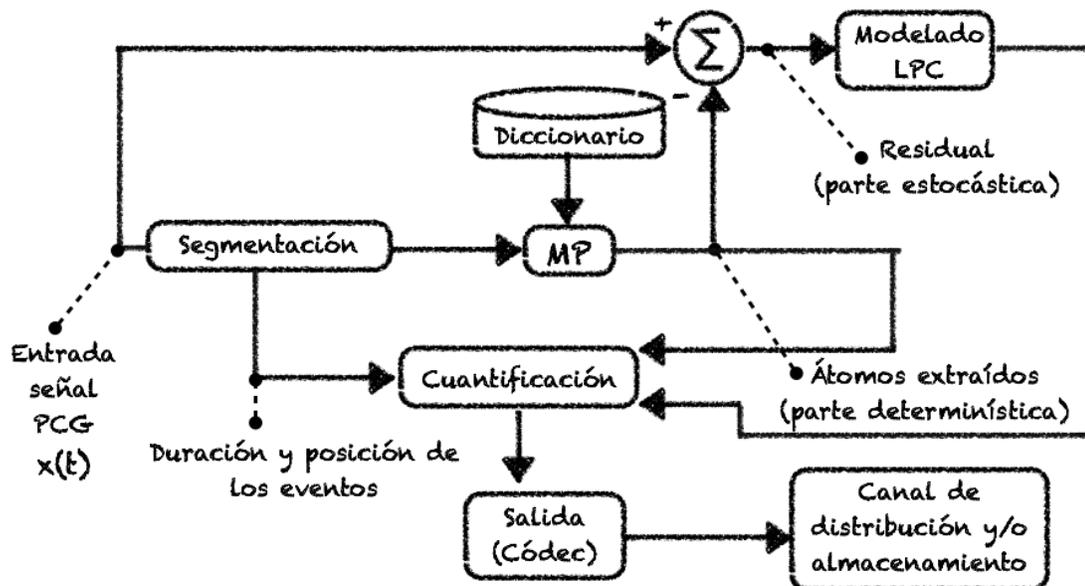
El codificador de audio propuesto para este trabajo de tesis consta de los dos modelados matemáticos básicos mostrados en los capítulos anteriores (MP y LPC).

**Tabla 2: Sonidos de la base del sitio web Littman © analizados en este trabajo de tesis. Tomados de (Litmann, 2013).**

Nombre de la señal
División normal del primer ruido (Normal Split S1)
División normal del segundo ruido (Normal Split S2)
S3
S4
Murmulo sistólico temprano (Early Systolic Murmur)
Murmulo sistólico tardío (Late Systolic Murmur)
Clic de eyección (Ejection Click)
Chasquido de apertura (Opening Snap)
Murmulo pansistólico (Pansystolic Murmur)
Soplo diastólico (Diastolic Rumble)

La Figura 23 muestra la estructura del codificador diseñado, cuyas etapas serán brevemente explicadas en esta sección. El canal de distribución y/o almacenamiento no será referido durante este trabajo.

Por otra parte, en la Figura 24 se muestra el procedimiento de decodificación, cuyas etapas son las necesarias para la conformación de la señal de audio reconstruida, la cual será analizada en términos de porcentaje de compresión, distorsión y niveles de error cuadrático medidos desde la señal de audio cardiaco original.



**Figura 23: Estructura del codificador para audio cardiaco propuesto.**

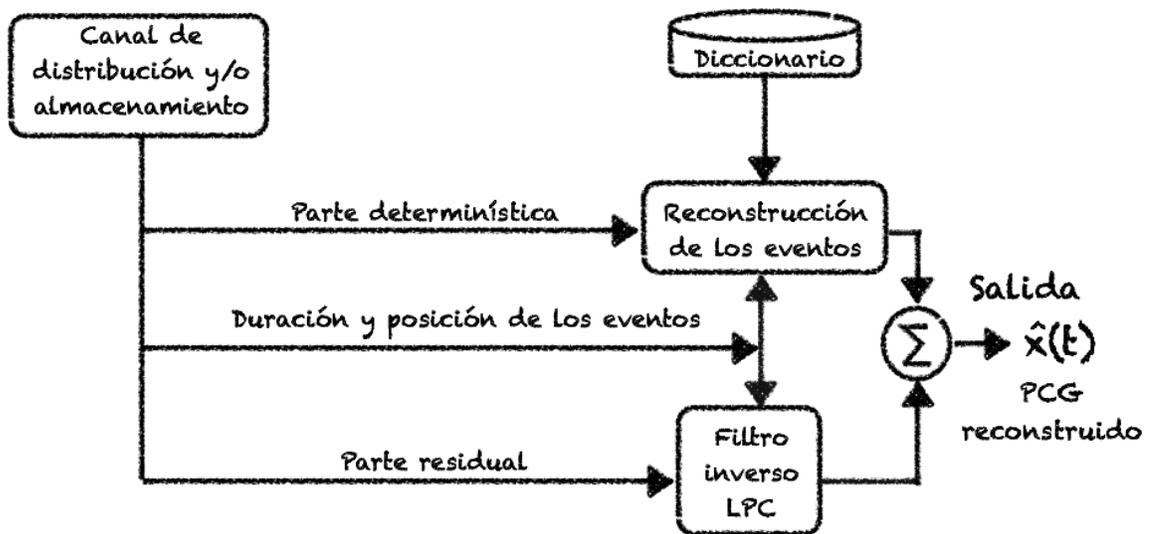


Figura 24: Estructura del decodificador para audio cardiaco propuesto.

## 5.2. Segmentación de los eventos cardiacos

El proceso de codificación de audio cardiaco inicia con la *segmentación*, tarea que consiste en extraer los instantes de tiempo (ya sea en segundos o en muestras) donde comienza (*onset*) y termina (*offset*) un sonido o evento cardiaco. La precisión en la localización de estos intervalos es importante<sup>1</sup>. El diseño del algoritmo automático de segmentación no es parte de los objetivos de este trabajo de tesis, por lo tanto se ha realizado de manera manual.

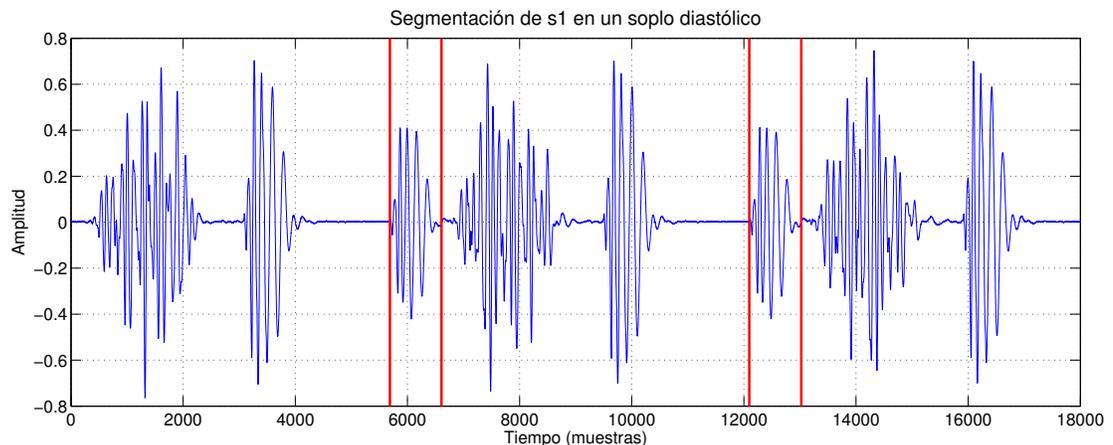


Figura 25: Segmentación del evento cardiaco normal S1 en una señal de audio cardiaco.

<sup>1</sup>Esto es, que los valores obtenidos difieran en pocas muestras.

### 5.3. Descomposición atómica de los eventos (MP)

Como se ha especificado en el capítulo 3, el proceso de descomposición atómica dado por Matching Pursuit otorga una reconstrucción y compresión aceptables para los eventos cardiacos. La selección de un diccionario *multi-bloques* de Gabor en la descomposición otorga una diversidad atómica en cuanto a frecuencias, posiciones y tamaños de los átomos para la regeneración del evento (Nieblas, 2014).

**The Matching Pursuit Toolkit (MPTK)** es una herramienta eficiente en el análisis y síntesis de señales de audio por la técnica Matching Pursuit (Gribonval y Krstulovic, 2006). La gran ventaja de la herramienta MPTK se basa en la rapidez para calcular las correlaciones (productos internos) entre los átomos del diccionario y las estructuras locales de la señal. Los productos internos son calculados auxiliándose en la FFT y además son actualizados para no re-calcularse (se han preservado en instantes de tiempo anteriores).

En el codificador diseñado para este trabajo de tesis se han realizado las descomposiciones de los eventos cardiacos mediante MPTK y se han seleccionado diccionarios multi-bloques de Gabor como sugiere el trabajo de tesis de Nieblas (2014). Se trata de dos diccionarios con 4 y 5 bloques respectivamente con las siguientes características que se muestran en la Tabla 3.

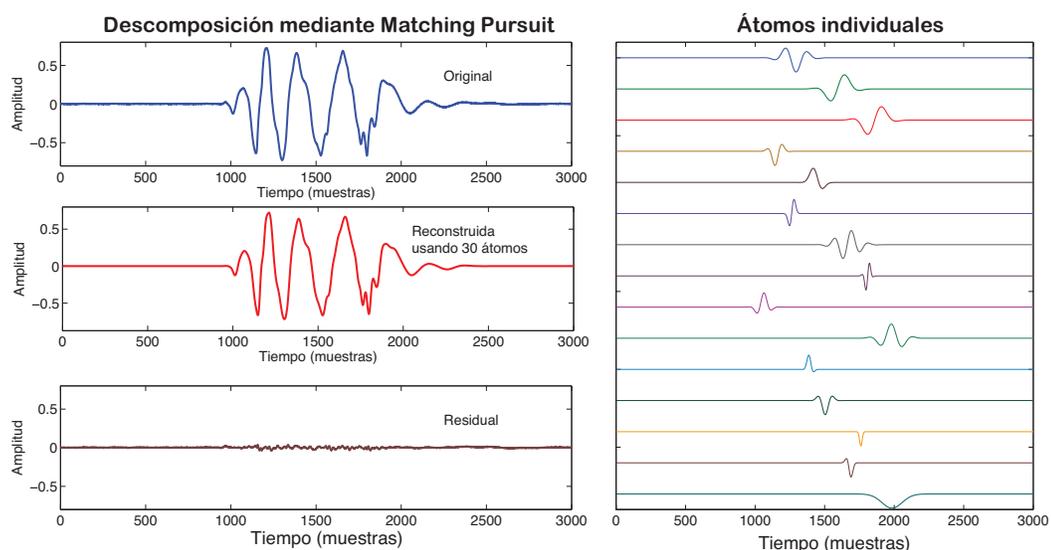
**Tabla 3: Características de los diccionarios empleados en la descomposición de los eventos cardiacos del codificador.**

<b>Parámetros del diccionario</b>	Diccionario 1	Diccionario 2
Longitudes de las ventanas y átomos (muestras)	64, 128, 256, 512	64, 128, 256, 512, 1024
Tamaños de la FFT (muestras)	64, 128, 256, 512	64, 128, 256, 512, 1024
Tipo de ventana	gaussiana	gaussiana
Desplazamiento de la ventana (muestras)	8, 16, 24, 32	8, 16, 24, 32, 64

Además, en la Tabla 3, se observa que es conveniente fijar como iguales las longitudes de los átomos y los tamaños de las transformadas rápidas de Fourier (FFT). Esto dará consistencia en las frecuencias atómicas encontradas tras la realización del máximo del producto interno.

Otro aspecto importante a verificar de los diccionarios es el empalme u *overlap* dado por el corrimiento de la ventana. En este caso las ventanas se irán desplazando por la estructura de la señal en  $1/8$  de muestras con relación a su tamaño para calcular las correlaciones correspondientes.

La Figura 26 muestra la descomposición mediante MP de un evento cardiaco normal s1 tras realizar 30 iteraciones. Se muestran los primeros 15 átomos seleccionados del análisis ya que presentan mayor amplitud. Se observan en la Figura las diferentes posiciones y frecuencias atómicas (frecuencias de modulación del coseno).



**Figura 26: Descomposición MP de un evento cardiaco y muestra de los primeros 15 átomos mayormente correlacionados.**

La Tabla 4 muestra el número de eventos y la cantidad de átomos requerida para las descomposiciones Matching Pursuit de las señales de la base de datos. Se observa que la señal del murmullo pansistólico es la que más iteraciones requiere para su representación (392 átomos). Sin duda alguna esta señal presentará menor índice de compresión, dado que la cantidad de átomos es proporcional a la cantidad de bits requeridos para su codificación.

#### 5.4. Modelado y tratamiento de la señal residual (LPC)

Matching Pursuit, a pesar de ser un algoritmo codicioso (*greedy*) presenta limitantes en cuanto a criterios de selección del número de átomos (iteraciones) o SNR deseada,

**Tabla 4: Cantidad de eventos y número de átomos requeridos en las descomposiciones MP para las señales de la base de datos.**

Nombre de la señal	Cantidad de eventos	Total de átomos
Soplo diastólico	17	264
Clic de eyección	19	275
Murmullo sistólico temprano	20	365
Murmullo sistólico tardío	20	340
Chasquido de apertura	19	223
S3	19	197
S4	20	208
Murmullo pansistólico	21	392
Apertura normal S1	12	130
Apertura normal S2	12	172

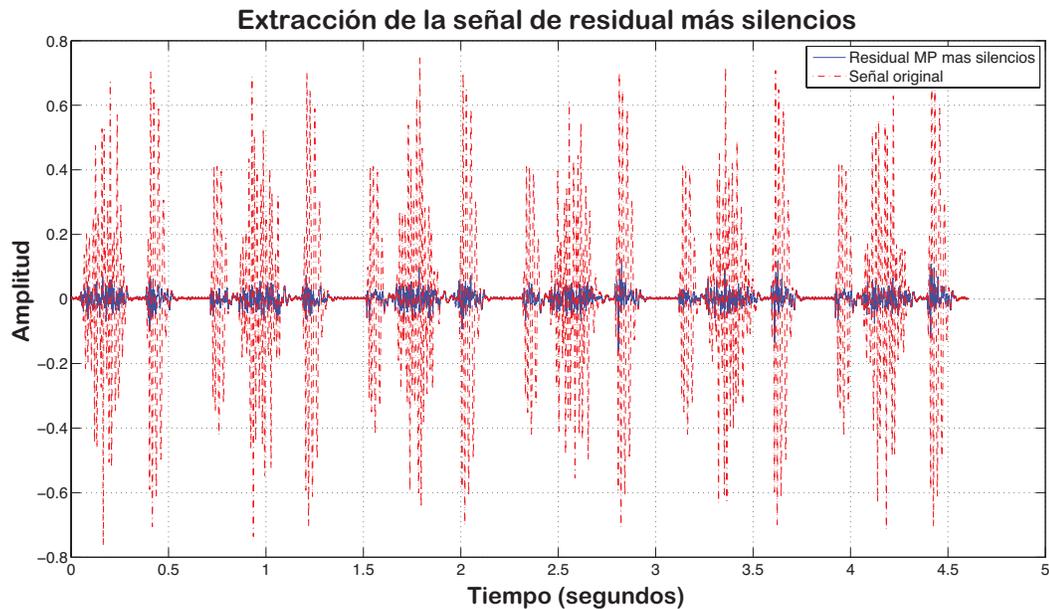
como se ha mencionado en el capítulo 3. También se ha mencionado que en la práctica el número  $N$  de iteraciones nunca es suficiente para representar o retener el 100 % de la energía de la señal analizada, teniendo en la señal residual  $R_M(t)$  energía aún perceptible para el oído humano dado su comportamiento logarítmico.

En efecto  $R_M(t)$  presenta correlación muy baja con los átomos de Gabor del diccionario seleccionado, lo cual puede observarse en requerir un gran número de iteraciones sin modelar con éxito esta señal, la cual es considerada por lo tanto como la *parte estocástica* de este codificador de audio.

Es a través de las técnicas de codificación predictiva lineal (LPC) como se ha modelado a la señal residual más silencios del fonocardiograma para el codificador diseñado en esta tesis, parte que se ha establecido que aún contendrá el 1 % de la energía de cada evento del PCG. El criterio de selección de  $N$  iteraciones nos da como promedio 10 átomos seleccionados por evento normal y un máximo de 30 átomos por patología o evento de anomalía cardíaca. La Figura 27 muestra la forma de onda de una señal de soplo diastólico y su señal residual más silencios correspondiente tras realizar el criterio de descomposición antes mencionado.

#### 5.4.1. Segmentación del residual

El análisis LPC de una señal de audio debe realizarse sobre tramos relativamente estacionarios, lo cual reduce la posibilidad de tener cambios abruptos en la autocorrela-



**Figura 27: Señal de residual y silencios obtenida de un audio cardiaco soplo diastólico.**

ción y por lo tanto menor cantidad de error en los coeficientes predictores. Para la señal residual más silencios de audio cardiaco fue necesario tomar tramos de señal de 256 muestras (equivalentes a 32 ms de señal si se considera una frecuencia de muestreo de 8kHz) y posteriormente multiplicarlos por ventanas de análisis tipo *Blackman* para evitar fugas espectrales.

El empalme u *overlap* de los segmentos analizados fue de 128 muestras (50 %) para asegurar consistencia en la reconstrucción. Se muestra en la Figura 28 una descripción gráfica del proceso de segmentación (ventaneo) de la señal residual más silencios cardiacos.

Las tramas de señal obtenidas en el ventaneo son sometidas al análisis LPC, donde será importante extraer los siguientes parámetros para el filtro de síntesis o reconstrucción:

- Coeficientes predictores  $a_k$ : Calculados mediante la recursión de Levinson por el método de autocorrelación como se mostró en el capítulo anterior. El criterio para determinar el número de coeficientes (orden del filtro) se especificará en el siguiente apartado.

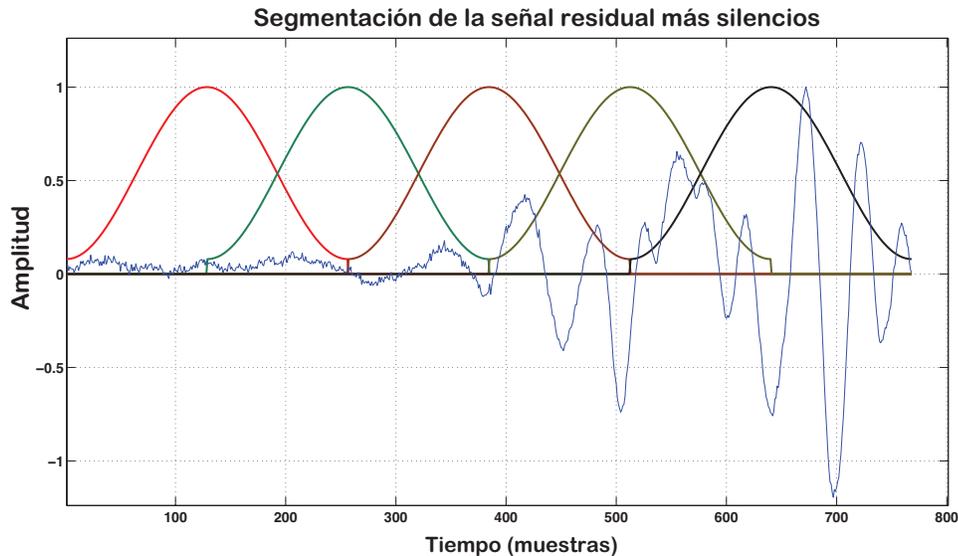


Figura 28: Ventaneo (segmentación) de la señal residual más silencios.

- Ganancia del filtro predictor: Obtenida para cada trama por la autocorrelación en el instante  $t = 0$ , esto es  $\mathbf{R}(0)$ .
- El periodo tonal o pitch: Que se ha obtenido en este trabajo por medio de estimación a través de la función de autocorrelación (ACF) de la trama de residual de entrada (máximo de la ACF no ubicado en cero).

#### 5.4.2. Criterios para la determinación del orden $p$ del filtro predictor

El filtro predictor lineal es un bloque que se dice *blanquea* o *aplana* el espectro en potencia de la señal de entrada, esto debido a que remueve las correlaciones a corto término. La calidad de un filtro predictor se basará en calcular qué tanto es capaz de *aplanar* un espectro en potencia, con lo cual, determinará el grado de tonalidad de un sonido o bien cuánto éste tiende a parecerse a ruido blanco (el cual tiene un espectro plano). Esta medida puede ser obtenida por medio de la **planitud** o **aplanamiento espectral**, conocida como *spectral flatness measure* en Inglés. Se denotará en este caso por sus siglas SFM (Jayant, 1974).

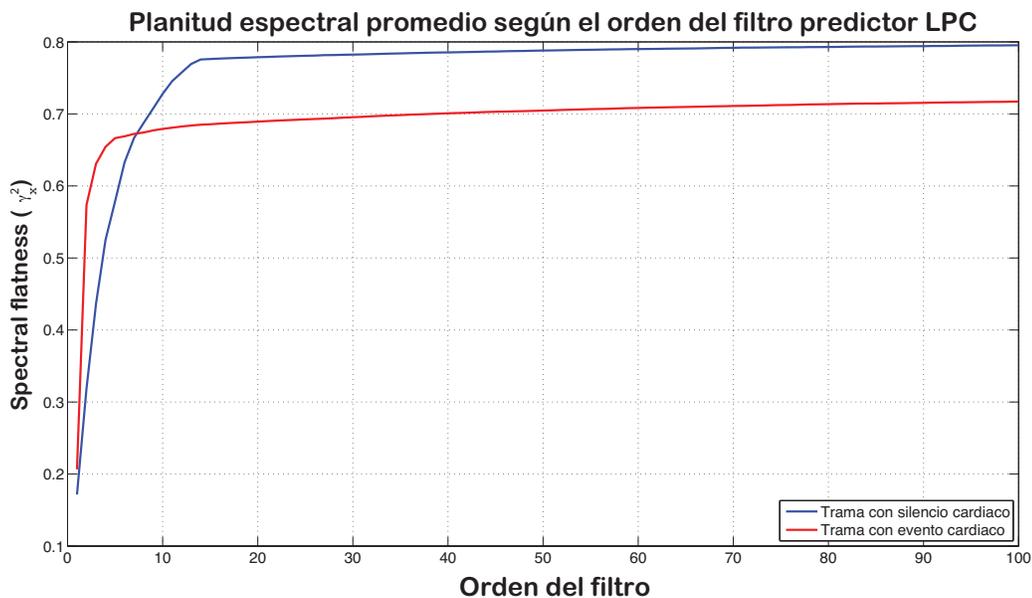
La SFM,  $\gamma_x^2$  se calcula como la razón entre la media geométrica y la media aritmética

de la densidad espectral de energía  $S_x x(e^{j\omega})$  de una señal de audio:

$$\gamma_x^2 = \frac{\exp \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log_e S_{xx}(e^{j\omega}) d\omega \right]}{\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xx}(e^{j\omega}) d\omega}. \quad (41)$$

Para la determinación del orden del filtro predictor en la señal residual de audio cardíaco se calculó la SFM promedio de la señal error de cada trama (esta es la salida del predictor lineal) para órdenes predictores  $p = 1, 2, \dots, 100$ . Las tramas de audio fueron separadas de acuerdo a si contenían evento cardíaco o presentaban sólo silencio.

De acuerdo a la Figura 29, se ha considerado un aplanamiento aceptable (próximo a volverse constante) a partir de un orden  $p = 15$  en tramas de silencio cardíaco y tramas con evento respectivamente.



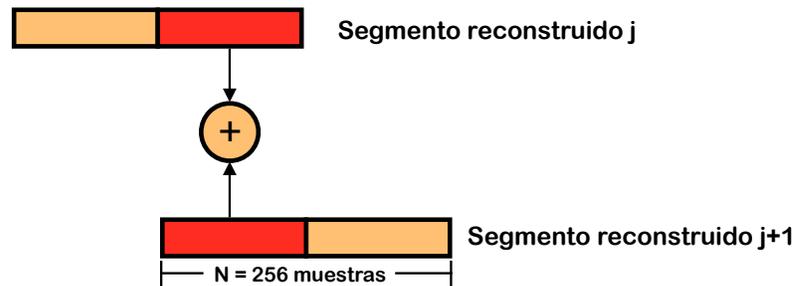
**Figura 29: Medida de planitud espectral promedio de la señal error  $e(n)$  para órdenes del filtro predictor desde 1 hasta 100.**

### 5.4.3. Síntesis/reconstrucción de la señal residual

El filtro inverso de reconstrucción IIR por medio de los coeficientes de predicción y la ganancia del filtro predictor reconstruye cada uno de los tramos de señal. Tal como se indicó en el capítulo 4, se presentarán los casos de tener una trama de audio *vocalizada* o *no vocalizada*, excitando al filtro reconstructor con un tren de pulsos separados por el

periodo tonal (pitch) en el primer caso y en el segundo mediante ruido blanco gaussiano.

Una vez regenerados todos los tramos de señal se procede a sintetizar toda la señal de audio cardiaco (contraparte del ventaneo), sumando las partes de los segmentos donde existió el traslape en el ventaneo como se indica en la Figura 30. Dicho procedimiento es conocido como *overlap and add (OLA)* en Inglés.



Para  $j=1,2,\dots,M$  segmentos reconstruidos de señal

Figura 30: Representación gráfica del procedimiento de síntesis “overlap and add (OLA)”.

#### 5.4.4. Filtros de pre-énfasis y de-énfasis en LPC

La señal de audio cardiaco es pasa-bajas, lo cual indica que en su espectro en potencia a medida que incrementa el rango frecuencial decrece la energía presentada.

Previo al análisis LPC es conveniente realizar un filtrado de pre-énfasis, cuyo objetivo será realzar las altas frecuencias de la señal y representarlas con una mayor precisión. En la reconstrucción de los segmentos deberá ejecutarse la contraparte de este proceso, el de-énfasis correspondiente que habrá de contrarrestar el efecto.

Los filtros de pre-énfasis y de-énfasis empleados comúnmente en LPC para los codificadores de voz existentes tienen la siguiente estructura:

$$y(n) = x(n) - a \cdot x(n - 1) \quad (42)$$

$$y(n) = x(n) + a \cdot x(n - 1), \quad (43)$$

donde generalmente se emplea  $a = 0.9$  en los métodos de predicción lineal.

En la Figura 31 se muestra la gráfica de la respuesta en frecuencia de los filtros mostrados en (42) y (43), mostrando la parte pasa-altas (pre-énfasis) y pasa-bajas (de-énfasis).

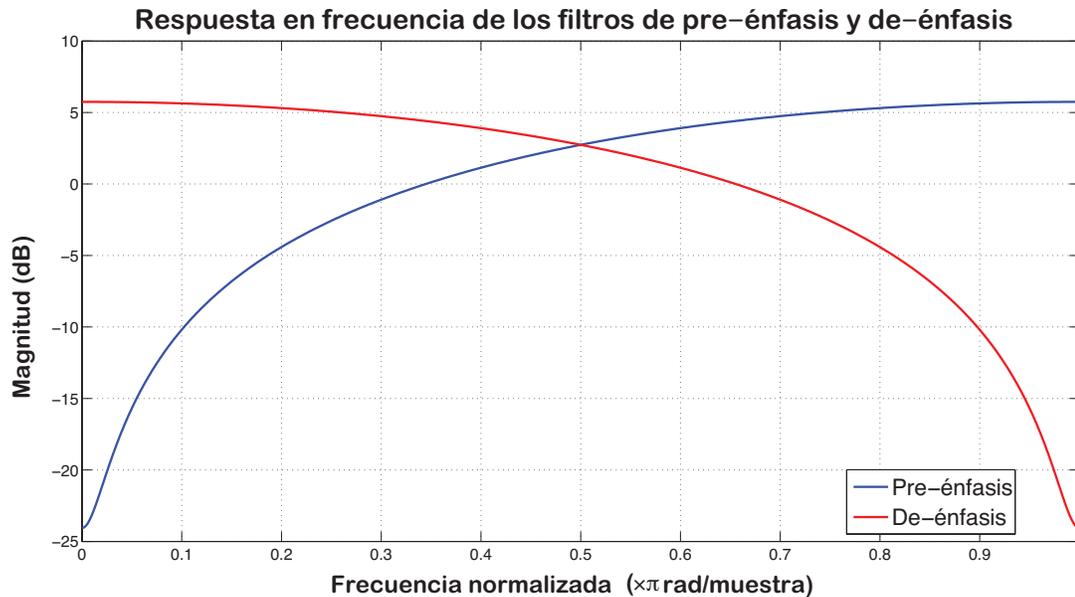


Figura 31: Filtros de pre-énfasis y de-énfasis usados en la codificación LPC.

## 5.5. Cuantificación

### 5.5.1. Tipos de cuantificadores

En la digitalización de una señal de audio está presente un proceso irreversible de compresión con pérdidas como la cuantificación, el cual tiene como objetivo *nivelizar* los parámetros obtenidos para su representación binaria. De manera matemática la cuantificación equivale a sumarle ruido y distorsionar consecuentemente la señal de entrada.

El cuantificador más sencillo de diseñar corresponde a aquella fuente de datos cuya distribución de probabilidad (pdf) es uniforme, sin embargo en la realidad esto no siempre ocurre y es conveniente adaptarlo a una función de distribución de probabilidad no uniforme.

La Tabla 5 define los tipos de cuantificadores con respecto a la dimensión de la fuente de datos de entrada así como a su pdf.

**Tabla 5: Tipos de cuantificadores de acuerdo a las características de la fuente de datos de entrada.**

Parámetro de la fuente de datos	Tipo de cuantificador	Características
Función de distribución de probabilidad	Cuantificador uniforme	Niveles equidistantes. Implementación sencilla.
	Cuantificador no uniforme	Niveles variantes en longitud. Menor distorsión introducida.
Dimensión	Cuantificador escalar	Vector unidimensional como fuente de entrada.
	Cuantificador vectorial	Cuantificación en bloques (vectores de bits) al agrupar diversas fuentes en una sola entidad.

### 5.5.2. Diseño de un cuantificador no uniforme

El cuantificador que introducirá la menor distorsión será aquel que sea diseñado proporcionalmente a la función de distribución de probabilidad de la fuente de entrada.

El error cuadrado promedio de cuantificación  $\sigma_q^2$  (MSQE) debe ser el mínimo para presentar la menor distorsión posible tras cuantificar y se define como:

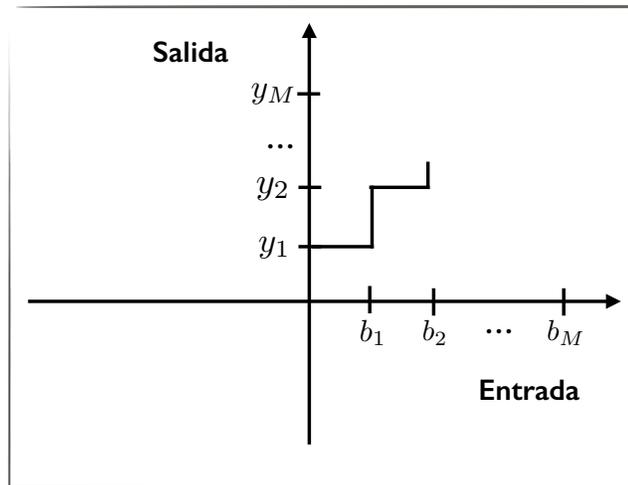
$$\sigma_q^2 = \int_{-\infty}^{\infty} (x - Q(x))^2 f_X(x) dx = \sum_{i=1}^M \int_{b_{i-1}}^{b_i} (x - y_i)^2 f_X(x) dx, \quad (44)$$

donde  $x$  es la fuente de entrada,  $Q(x)$  la operación de cuantificación,  $f_X(x)$  la función de distribución de probabilidad de la fuente,  $\{b_i\}_{i=1}^M$  las regiones de decisión y  $\{y_i\}_{i=1}^M$  los niveles de reconstrucción (mostrados gráficamente en la Figura 32).

#### 5.5.2.1. Algoritmo de Lloyd

Según lo discutido anteriormente el mejor cuantificador no uniforme surge al tener un modelo probabilístico para la fuente, encontrando los valores  $b_i$  y  $y_i$  que minimicen lo definido en la ecuación (44). La derivación parcial de esta ecuación es un camino para lograr esta minimización, con la cual al resolver para  $y_j$  se tiene:

$$y_j = \frac{\int_{b_{j-1}}^{b_j} x f_X(x) dx}{\int_{b_{j-1}}^{b_j} f_X(x) dx}. \quad (45)$$



**Figura 32: Regiones de decisión y niveles de reconstrucción de un cuantificador.**

Teniendo que  $\int_{-\infty}^{\infty} x f_X(x) dx$  es el valor esperado de la fuente de datos  $x$ , se define que cada punto a la salida  $y_j$  para cada intervalo de cuantificación es el centroide de la función de probabilidad en ese intervalo.

Si se realiza el mismo procedimiento en (44), ahora para  $b_j$  (igualando a cero la derivada parcial con respecto a este término) se tiene:

$$b_j = \frac{y_{j+1} + y_j}{2}. \quad (46)$$

Con esto se observa que la región de decisión es simplemente el punto medio de dos niveles de reconstrucción vecinos. Al resolver las ecuaciones (45) y (46) se obtendrán los valores que minimicen el MSQE definido en (44), desafortunadamente para resolver  $y_j$  se requieren los valores  $b_j$  y  $b_{j-1}$  y viceversa.

El algoritmo de Lloyd, desarrollado por Stuart P. Lloyd describe cómo resolver las ecuaciones (45) y (46) de manera iterativa, obteniendo los valores  $b_j$  y  $y_j$  a base de simetría y suponiendo los valores para la primer región de decisión  $b_0$  como cero, así como suponiendo que la última región  $b_{\frac{M}{2}}$  es el valor más grande que puede tener la entrada.

### 5.5.3. Cuantificación de los parámetros extraídos de los modelados

Para el desarrollo del códec de audio propuesto se han cuantificado aquellos parámetros que estén en formato de punto flotante para convertirse en enteros con signo. Es-

to garantizará la correcta representación de los mismos con el menor número de bits posibles. Es necesario determinar ahora de los parámetros extraídos del PCG de los modelados MP y LPC, cuáles es necesario cuantificar y cuáles será necesario codificar directamente (ya que son enteros con signo).

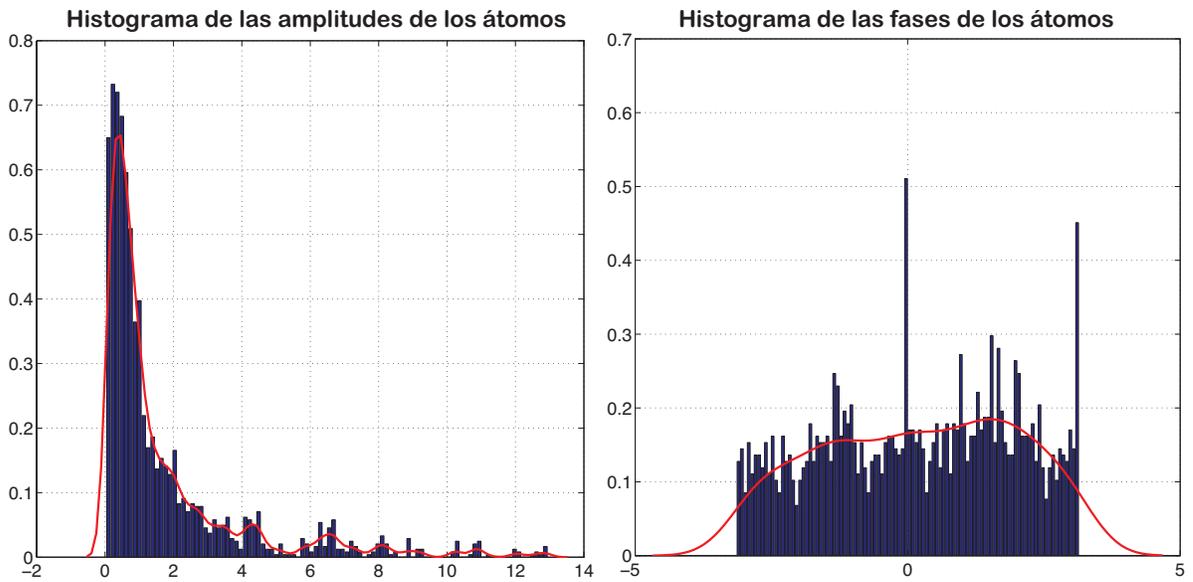


**Figura 33: Parámetros extraídos de los modelados MP y LPC para el códec. Se indican los parámetros necesarios de cuantificar.**

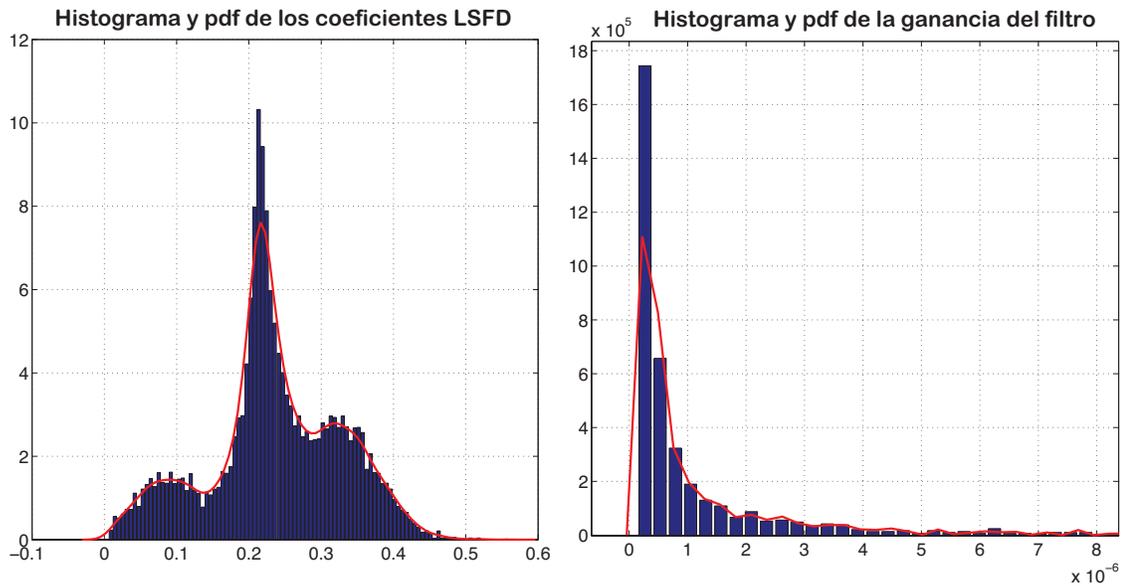
De acuerdo a lo observado en la Figura 33, habrá que diseñar 4 cuantificadores para el códec, correspondientes a las amplitudes y fases de los átomos seleccionados en MP, los coeficientes de líneas espectrales diferenciales (LSFD, obtenidos a través de los coeficientes predictores) y las ganancias de los filtros de predicción de LPC. El primer paso corresponderá a observar las funciones de distribución de probabilidad (pdf) de las variables para posteriormente determinar sus cuantificadores optimizados mediante el algoritmo de Lloyd.

Se han realizado los histogramas normalizados para determinar las funciones de probabilidad de las amplitudes y fases atómicas extraídas desde MPTK para las señales de la base de datos.

En la Figura 34 se observa que la distribución de probabilidad de las fases de los átomos es de manera uniforme en el intervalo  $[-\pi, \pi]$ , siendo necesario diseñar un cuan-



**Figura 34:** Histogramas normalizados (pdf) para las amplitudes y fases de los átomos extraídos en las descomposiciones.



**Figura 35:** Histogramas normalizados (pdf) para los coeficientes LSFD y ganancias del filtro predictor.

tificador sencillo para esta variable. En el caso de la amplitud se procederá a diseñar con el algoritmo de Lloyd un cuantificador no uniforme basado en su pdf.

Se observa un comportamiento no uniforme en las funciones de probabilidad obtenidas a través de los histogramas de los coeficientes LSFD y ganancia del filtro predictor,

esto según la Figura 35. De igual manera que con la amplitud, mediante el método de Lloyd se realizará el diseño de estos cuantificadores no uniformes optimizados.

#### 5.5.4. Cuantificación de los coeficientes de predicción

Como se ha manejado anteriormente, los coeficientes obtenidos para la predicción lineal son bastante sensibles a la inestabilidad, por lo tanto su cuantificación no es realizable de manera directa para su transmisión, ya que la distorsión introducida afectaría en gran medida en la síntesis de la señal (Noah, 1984).

Sin embargo, existen otras representaciones equivalentes de los coeficientes predictores LPC empleadas en algunos vocoders diseñados en la literatura, con las cuales puede realizarse la cuantificación presentando la menor distorsión posible:

- **Coefficientes reflectores o PARCORS:** Se obtienen directamente a partir de la recursión de Levinson a la par de los coeficientes predictores.
- **Coefficientes de relación log-área (log-area ratios o LAR):** Los coeficientes reflectores  $k_i$  pueden sufrir transformaciones para expandir su región en el plano polar (Viswanathan y Makhoul, 1975).

Una de estas transformaciones es calculada como el logaritmo de la razón de la suma y resta del  $i$ -ésimo coeficiente reflector, en una representación conocida como relación log-área (LAR), esto es:

$$LAR_i = \log \frac{1 + k_i}{1 - k_i}.$$

- **Pares o frecuencias de líneas espectrales (line spectrum pairs/frequencies, LSF):** Itakura (1975), encontró que el polinomio del filtro predictor  $A(z)$  tiene una representación equivalente a partir de un par de polinomios:

$$A(z) = 0.5[P(z) + Q(z)]$$

$$P(z) = A(z) + z^{-(p+1)}A(z^{-1})$$

$$Q(z) = A(z) - z^{-(p+1)}A(z^{-1}).$$

Las raíces de  $P(z)$  y  $Q(z)$  ocurren en pares conjugados y se les denomina *frecuencias* o *pares* de líneas espectrales.

- **Pares o frecuencias de líneas espectrales diferenciales (LSFD's):** Soong y Juang (1984) explotaron la correlación entre LSF's sucesivos, para ello cuantificaron las diferencias entre éstos. A tal representación equivalente se le conoce como LSF diferenciales ó LSFD's:

$$LSFD_i = LSF_{i+1} - LSF_i.$$

Seleccionar la representación LPC adecuada en el diseño del codificador será una medida que intervendrá en la calidad de la reconstrucción de la señal. La distorsión espectral (medida en decibeles) es una medida que indica la cantidad de degradación del espectro en potencia de la señal analizada tras seleccionar alguna representación equivalente de las antes mencionadas y haber cuantificado.

Se define como la distorsión espectral  $D_i$  común entre dos tramas de audio a la siguiente expresión:

$$D_i = \sqrt{\frac{1}{f_s} \int_0^{f_s} [10 \log_{10}(P_i(f)) - 10 \log_{10}(\hat{P}_i(f))]^2 df}, \quad (47)$$

donde  $f_s$  es la frecuencia de muestreo en Hz,  $P_i(f)$  y  $\hat{P}_i(f)$  son los espectros de potencia LPC de las  $i$ -ésimas tramas, dados por:

$$P_i(f) = \frac{1}{|A_i(\exp(j2\pi f/f_s))|^2}$$

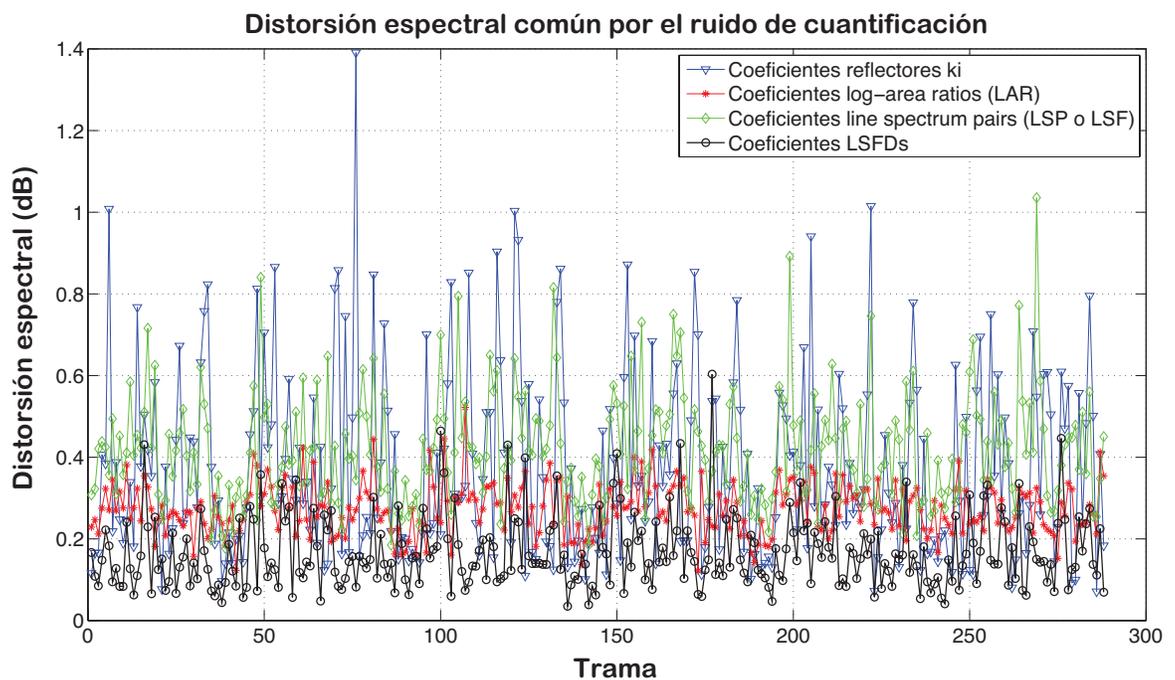
$$\hat{P}_i(f) = \frac{1}{|\hat{A}_i(\exp(j2\pi f/f_s))|^2},$$

donde  $A_i(z)$  y  $\hat{A}_i(z)$  son los polinomios LPC originales y cuantificados de la  $i$ -ésima trama respectivamente.

La distorsión espectral es evaluada para todas las tramas en los datos experimentales

y su valor promedio es calculado. Este valor es la distorsión asociada a un cuantificador en particular.

Los coeficientes seleccionados son los que introduzcan menores cantidades de distorsión espectral tras reconstruir la señal después de su cuantificación, para ello se ha calculado  $D_i$  en las 288 tramas resultantes de 32ms (256 muestras) de la señal soplo diastólico para los coeficientes LPC reflectores ( $k_i$ ), log-area (LAR), LSF y LSFD realizando la cuantificación no uniforme con un cuantificador optimizado de Lloyds para cada caso. Los resultados de este procedimiento se muestran en la Figura 36.



**Figura 36: Distorsión espectral común por el ruido de cuantificación aplicada a diferentes representaciones equivalentes LPC.**

Tras observar la Figura 36, se ha decidido el emplear coeficientes del tipo LSFD ya que introducen menores niveles de distorsión espectral tras ser cuantificados. De igual forma, disminuyen el rango dinámico empleado para las regiones de decisión del cuantificador, permitiendo representar con menos bits a cada uno de estos coeficientes.

## 5.6. Representación en bits de los parámetros de codificación

Una vez diseñados los cuantificadores optimizados de los parámetros requeridos para codificar es conveniente determinar el número de bits para representar cada uno de ellos.

De igual manera, se requiere conocer dicha cantidad para aquellos parámetros que son enteros con signo y no requerirán cuantificación. Estos conceptos son mostrados en la Tabla 6.

**Tabla 6: Número de bits requeridos por parámetro y tipo de cuantificación realizada.**

Parámetros de la parte determinística (MP)		
Parámetro extraído	Núm. bits requeridos/átomo	Tipo de cuantificador
Amplitud atómica	8	No uniforme
Longitud atómica	3	Sin cuantificador
Fase atómica	8	Uniforme
Posición atómica	8	Sin cuantificador
Frecuencia atómica	6	Sin cuantificador
Parámetros de la parte estocástica (LPC)		
Parámetro extraído	Núm. bits requeridos/segmento	Tipo de cuantificador
Ganancia del filtro	4	No uniforme
Periodo tonal (Pitch)	8	Sin cuantificador
LSFD	7	No uniforme

A continuación se enlistan las razones por las cuales algunos de los parámetros enlistados de la Tabla 6 no requirieron el diseño de un cuantificador:

- Las *longitudes* de los átomos (que son equivalentes a las longitudes de las ventanas) fueron pre-seleccionadas con el tipo de diccionario empleado para la descomposición y corresponden a 64, 128, 256, 512 y 1024 muestras. Se trata de 5 cantidades que pueden representarse con sólo 3 bits.
- Con la elección del diccionario también se permitió preseleccionar el desplazamiento de las ventanas gaussianas, el cual fue a 1/8 de sus longitudes. Por ello las *posiciones* de los átomos son múltiplos de 8 que pueden representarse con enteros de 8 bits sin signo.
- Las *frecuencias* atómicas pudieron convertirse a enteros tras multiplicar su valor decimal por el tamaño de la FFT, cantidad que es equivalente al tamaño de las ven-

tanías en el diccionario. De acuerdo a su rango dinámico se consideraron necesarios 6 bits para su representación.

- El periodo tonal o *pitch* fue observado en cuanto a rango dinámico. Al ser necesario su valor en muestras para la implementación éste debe ser un entero. Se concluyó que 7 bits son necesarios para representar cada uno de sus valores.

### 5.7. Codificación de los fonocardiogramas y porcentajes de compresión obtenidos

Para la evaluación del desempeño del codificador diseñado es necesario calcular las tasas de compresión al implementarlo en el catálogo de señales de prueba antes mencionado. Se ha decidido calcular el porcentaje de compresión  $C_r$  como una relación entre bits utilizados por el codificador y bits empleados para codificar a velocidades de transmisión de  $R_b = 64$  y  $R_b = 128$  kilobits por segundo (kbps), esto es:

$$C_r = \frac{\text{cantidad de bits codificador}}{\text{cantidad de bits } R_b}.$$

La Tabla 7 muestra los porcentajes de compresión  $C_r$  calculados para las señales de la base Litman por el codificador, de igual manera se indican los bits que se requieren para codificar cada una de las muestras de audio cardiaco. Estos porcentajes son en proporción a las cantidades de átomos obtenidas en la Tabla 4.

### 5.8. Pruebas en la reconstrucción de los fonocardiogramas

Tal y como se ha definido en el diagrama de la Figura 24, los procedimientos de decodificación dan lugar a la reconstrucción de la señal  $\hat{x}(t)$ , la cual necesita ser comparada con el fonocardiograma de entrada  $x(t)$ .

Para esto la Figura 37 muestra las formas de onda de la señal original y la señal reconstruida tras aplicar el procedimiento de codificación planteado en el caso del murmullo sistólico tardío, las cuales se observan muy similares.

**Tabla 7: Porcentajes de compresión y bits requeridos para el codificador diseñado en las señales de prueba.**

Nombre de la señal	Porcentaje de compresión (%)		Bits requeridos por el codificador
	@128 kbps	@ 64 kbps	
Soplo diastólico	93.24	86.47	39,818
Clic de eyección	93.88	87.68	39,115
Murmullo sistólico temprano	93.35	86.70	43,301
Murmullo sistólico tardío	93.30	86.61	44,333
Chasquido de apertura	94.06	88.13	38,077
S3	93.99	87.98	38,385
S4	94.06	88.16	39,031
Murmullo pansistólico	93.00	86.00	47,358
División normal de S1	94.49	88.98	31,602
División normal de S2	94.21	88.43	34,005

### 5.8.1. Coeficiente de correlación

A pesar de que el codificador diseñado ha mostrado buenos resultados en cuanto a replicar la forma de onda es necesario medir el grado de intensidad que guardan las variables  $x(t)$  y  $\hat{x}(t)$  matemáticamente (Senhadji *et al.*, 1995).

El coeficiente de correlación  $\rho_{x,\hat{x}}$  es una medida que cuantifica y describe esta relación por medio de la razón de las covarianzas entre éstas dos variables, esto es:

$$\rho_{x,\hat{x}} = \frac{\sigma_{x\hat{x}}}{\sigma_x \sigma_{\hat{x}}} = \frac{E[(x - \mu_x)(\hat{x} - \mu_{\hat{x}})]}{\sigma_x \sigma_{\hat{x}}}, \quad (48)$$

donde  $\sigma_{x\hat{x}}$  es la covarianza de las variables  $x$  y  $\hat{x}$ ;  $\sigma_x$  es la desviación estándar de la variable  $x$  y  $\sigma_{\hat{x}}$  es la desviación estándar de la variable  $\hat{x}$  respectivamente.

Se calculó el coeficiente de correlación después de codificar y reconstruir cada una de las señales analizadas en la base de datos de la Tabla 2 con respecto a las señales reconstruidas. Los resultados se muestran en la Tabla 8.

Según lo observado en la Tabla 8, los audios reconstruidos han mostrado fuerte correlación (cercana a 1) en comparación a los fonocardiogramas originales, lo cual indica dependencia lineal y buenos resultados para el codificador.

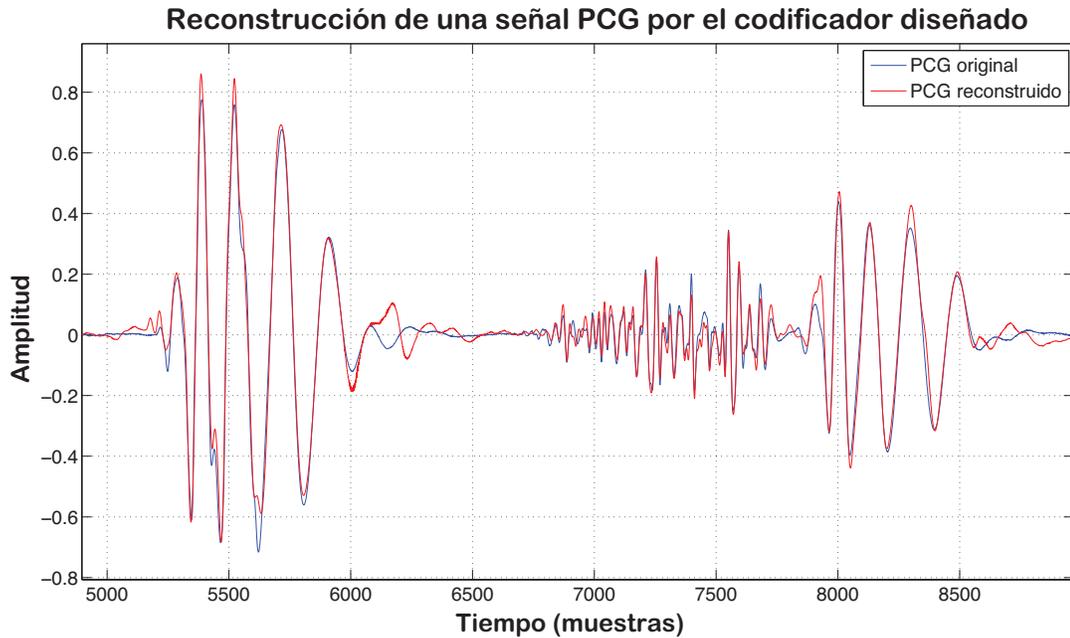


Figura 37: Formas de onda original y reconstruida por el códec para un ciclo de un murmullo sistólico tardío.

### 5.8.2. Raíz cuadrada de la diferencia cuadrática media (Percent Root-mean-square Difference, PRD)

La calidad de reconstrucción en un esquema de compresión de señales PCG se ha obtenido comúnmente por medio de una cantidad denominada *Raíz cuadrada de la diferencia cuadrática media*, lo cual es realmente una medida del porcentaje de degradación de una señal tras sufrir cambios en el proceso de compresión (Blanco-Velasco *et al.*, 2005). Suponiendo las mismas variables,  $x$  como el fonocardiograma original y  $\hat{x}$  como el PCG reconstruido posterior a la codificación se define su PRD como:

$$PRD = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{\sum_{i=1}^N (x_i - \mu_x)^2}} \times 100, \quad (49)$$

donde  $\mu_x$  es la media o valor esperado de la variable  $x$ .

Es evidente que la señal  $x$  habrá sido exitosamente reconstruida tras aplicar el esquema de compresión si el PRD calculado es lo más reducido posible.

**Tabla 8: Coeficientes de correlación de las señales reconstruidas comparadas con los audios originales.**

Nombre de la señal	Coeficiente de correlación
Soplo diastólico	0.9697
Clic de eyección	0.9724
Murmullo sistólico temprano	0.9335
Murmullo sistólico tardío	0.9330
Chasquido de apertura	0.9406
S3	0.9614
S4	0.9638
Murmullo pansistólico	0.9648
División normal de S1	0.9634
División normal de S2	0.9675

La Tabla 9 muestra los valores de la raíz cuadrada de la diferencia cuadrática media calculados para las señales reconstruidas de la base de datos tomada comparándolas con las señales reconstruidas por el codificador propuesto, observando muy buenos resultados debido a su bajo valor de degradación (no rebasa el 3% en ninguno de los casos).

**Tabla 9: Cálculo de la raíz cuadrada de la diferencia cuadrática media (PRD) para las señales de la base de datos.**

Nombre de la señal	PRD
Soplo diastólico	2.50
Clic de eyección	2.38
Murmullo sistólico temprano	2.33
Murmullo sistólico tardío	2.41
Chasquido de apertura	2.15
S3	2.87
S4	2.77
Murmullo pansistólico	2.74
División normal de S1	2.77
División normal de S2	2.61

## Capítulo 6. Evaluación subjetiva del códec

---

Las pruebas matemáticas presentadas al final del capítulo anterior no son las más relevantes para evaluar completamente la calidad de un codificador de audio. Es importante que el público auditor logre valorar su percepción con respecto al funcionamiento del producto realizado. En efecto, la señal ha sufrido diversos procedimientos de compresión con pérdidas y se vuelve necesario describir en qué medida éstas la han afectado.

Un test subjetivo que ha trascendido entre expertos de audio y diseñadores de códecs es el descrito por la recomendación BS.1534-11 del ITU-R (Series, 2014), conocido también como MUSHRA (acrónimo del Inglés *Multiple stimuli with hidden reference anchor*).

MUSHRA es una prueba que, como su nombre lo indica, emplea múltiples *estímulos*, los cuales son diferentes señales que han sido previamente codificadas y por lo tanto han sido distorsionadas por el codificador de audio diseñado. Se le presenta al oyente una referencia (señal original) y una versión oculta de ésta (es decir, no se le indica como tal), la cual se muestra en forma aleatoria en conjunto con el resto de las señales distorsionadas (anclas).

Las anclas son otras versiones codificadas de la señal con niveles perceptibles de degradación, las cuales tienen como propósito ajustar la escala de valores del oyente lo más posible a la escala absoluta, asegurándose que éste ha marcado con una puntuación más baja las muestras claramente más pobres y con una puntuación alta a las muestras claramente mejores. Dicha distorsión es realizada de manera intencional.

En esta sección se describirán los pasos para realizar el test MUSHRA como valoración subjetiva a las señales de audio codificadas, se valorarán los resultados obtenidos enunciando diversas conclusiones sobre los mismos.

### 6.1. Selección de las señales de prueba

Para la realización de las pruebas subjetivas del códec, MUSHRA recomienda seleccionar estímulos y/o señales de audio bajo los siguientes requerimientos:

- Duración no mayor a 20 segundos.

- No seleccionar un número mayor a 10 estímulos para comodidad del oyente.
- Una versión de la referencia pasada por un filtro paso bajas a 3.5 kHz, situada en alguna de las anclas.

La primera recomendación se cumple, ya que las señales de la base no rebasan los 5 segundos. En el caso del segundo requerimiento se ha decidido seleccionar a los siguientes 5 estímulos debido a considerarse los mejores en calidad desde la codificación:

- Apertura normal del S2.
- S4.
- Apertura normal del S1.
- S3.
- Murmullo pansistólico.

Las anclas son versiones de las señales anteriores codificadas bajo los formatos MP3 (Noll, 1997) y OPUS (Valin *et al.*, 2010), otra versión con 30 dB de AWGN (ruido blanco gaussiano aditivo), la versión cifrada utilizando el códec propuesto y en efecto la referencia oculta (versión WAV). Las condiciones para estos formatos fueron tasas de datos de 8 kilobits por segundo CBR (excepto WAV), 16 bits por muestra y frecuencia de muestreo de 8,000 Hz. Las anclas con versión pasabajas no son realizables en el caso del audio cardiaco, ya que ninguna de estas señales rebasa en ancho de banda los 3 kHz y dicha diferencia no resultará audible.

## **6.2. Procedimiento de evaluación MUSHRA**

### **6.2.1. Preparación de la interfaz gráfica**

Se realizó por medio del software MAX© una interfaz gráfica adaptada al test MUSHRA (Hummersone, 2011). Por medio de esta interfaz el usuario califica la calidad de la señal de audio mostrada en alguno de los 5 canales con respecto a la referencia (canal

R) con notas en una escala de 0 a 100 y relacionando los adjetivos calificativos: 0 a 20 si la calidad es *mala*, 20 a 40 si la calidad es *mediocre*, 40 a 60 si la calidad es *regular*, 60 a 80 si la calidad es *buena* y 80 a 100 si es de una *excelente* calidad.

Se define un estímulo por página, donde se podrán reproducir las anclas y la referencia cuantas veces sea necesario para el oyente. Cada una de las anclas está situada de manera aleatoria con el propósito de establecer el criterio de la *referencia oculta*. La interfaz gráfica está diseñada para que no sea posible calificar algún ancla mientras no se esté reproduciendo. Se permitirá cambiar de página y/o estímulo una vez que todas las anclas en ésta hayan sido calificadas.

La Figura 38 muestra la visualización en pantalla de la interfaz MUSHRA adaptada desde MAX© para la evaluación subjetiva del códec diseñado.



Figura 38: Visualización en la pantalla de la interfaz MUSHRA.

Las calificaciones otorgadas por cada uno de los oyentes son guardadas al terminar

la prueba.

### **6.2.2. Fase de adiestramiento**

Es importante la familiarización de los oyentes con las señales de audio empleadas en la prueba, el entorno (manejo de la interfaz), las herramientas empleadas y el procedimiento a realizar. Con esto los sujetos serán capaces de haber adquirido el conocimiento y las capacidades necesarias para apreciar y valorar con objetividad cada uno de los estímulos aportando información relevante a la prueba.

La condición auditiva de los oyentes se ve fortalecida mediante una fase previa de adiestramiento, la cual se definió para este caso en la pre-evaluación de 3 de los estímulos sin considerar estas calificaciones para la valoración formal final.

### **6.2.3. Condiciones de audición**

Se han empleado para la prueba auriculares de estudio de calidad profesional con las siguientes características:

- Marca: audio technica.
- Modelo: ATH-M50X.
- Respuesta en frecuencia: 15-28,000 Hz.
- Sensibilidad: 99 dB.
- Máxima potencia de entrada: 1,600 mW a 1 kHz.
- Impedancia de 38 ohms

Las pruebas fueron realizadas en un ambiente libre de ruido, con un clima agradable y fresco. Se percató que los oyentes no estuviesen bajo circunstancias no favorables en cuanto a salud o estado anímico.

#### 6.2.4. Selección de los oyentes

Según la experiencia, los datos procedentes de alrededor de 20 sujetos son suficientes para extraer conclusiones adecuadas de la prueba. Sin embargo, aumentando el tamaño de oyentes, y en particular si son experimentados, es más conveniente. La inserción de una técnica de rechazo ya sea anterior (preselección) o posterior (postselección) a la prueba real resulta conveniente en ocasiones, ya que los resultados pueden sesgarse.

Se han seleccionado 20 sujetos para la realización de las pruebas mostradas en este trabajo de tesis, los cuales son estudiantes de posgrado del CICESE. Estas personas participaron de manera voluntaria y no recibieron compensación económica alguna por su participación.

Las edades de los oyentes se encuentran en el intervalo de 22 a 34 años, presentando una edad promedio de 24.95 años y una desviación estándar de 2.84 años entre ellos. Cabe destacar que el 20 % de los participantes son del sexo femenino y el 10 % de ellos cuenta con alguna formación musical. Ninguno de los participantes presentó problemas de audición ni locomoción durante la prueba, cuya duración promedio fue de 15 segundos.

#### 6.3. Análisis estadístico de los resultados obtenidos en la prueba

La calidad del funcionamiento medio de cada uno de los estímulos propuestos en esta investigación es identificada por medio del análisis estadístico de los resultados obtenidos, de igual manera puede corroborarse por medio de ello la fiabilidad de cualquier diferencia entre los valores que han sido obtenidos.

El primer paso en el análisis estadístico consiste en clasificar los datos en una matriz ordenada como la siguiente:

$$\begin{bmatrix} & \text{Sujeto}_1 & \text{Sujeto}_2 & \cdots & \text{Sujeto}_n \\ \text{Audio}_1 & x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ \text{Audio}_2 & \vdots & \vdots & \vdots & x_{2,n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \text{Audio}_m & x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{bmatrix} \quad (50)$$

De esta manera es más sencillo identificar las dispersiones individuales por sujeto, así como las calificaciones promedio de cada señal de audio evaluada.

### 6.3.1. Cálculo de los histogramas e intervalos de confianza

La concentración de los valores medios de cada señal de audio presentan cierto grado de incertidumbre, asociado a los grandes rangos de valoración existentes para la prueba. Este grado de incertidumbre puede ser calculado y medido por medio de los *intervalos de confianza*, los cuales expresan un rango de valores con los que se puede comparar la incertidumbre entre las valoraciones medias de los diversos estímulos empleados.

Para el cálculo de los intervalos de confianza, se requiere la siguiente información:

- Niveles de confianza
- Estadística
- Margen de error

Dadas estas tres entradas, el conjunto de valores del intervalo de confianza se define por la *muestra  $\pm$  margen de error*, y la incertidumbre asociada con el intervalo de confianza se especifica por el nivel de confianza.

El cálculo de los intervalos de confianza puede definirse en los siguientes cuatro pasos:

- Identificar/seleccionar una muestra estadística que se desea emplear para estimar como parámetro (media, desviación estándar).
- Seleccionar un nivel de confianza. Es necesario describir el grado de incertidumbre de algún método de muestreo. Frecuentemente en las investigaciones se elige el 90 %, 95 % ó 99 % de niveles pero cualquier porcentaje puede ser utilizado.
- Encontrar el margen de error, el cual es el producto del valor crítico y la desviación estándar en la estadística. Para una distribución normal con  $n$  elementos el error

estándar puede definirse como:

$$STD_e = \frac{\sigma}{n}, \quad (51)$$

donde

$$\sigma = \sqrt{\frac{1}{N}(x_i - \mu)^2}$$

es la desviación estándar del conjunto de  $n$  elementos  $x_i$  con media  $\mu$ .

- Especificar el intervalo de confianza. La incertidumbre es denotada por el nivel de confianza, y el rango del intervalo de confianza se define por medio de la siguiente expresión:

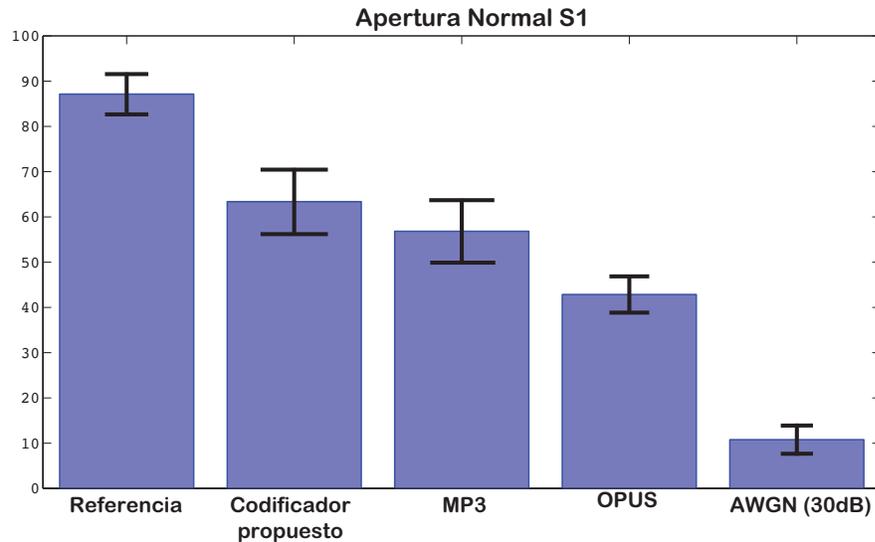
$$IC = muestra \pm margen\ de\ error, \quad (52)$$

lo cual demuestra que una población de mayor número  $n$  de elementos estadísticos evaluados denotará un menor margen de error y proporcionalmente mejores intervalos de confianza calculados.

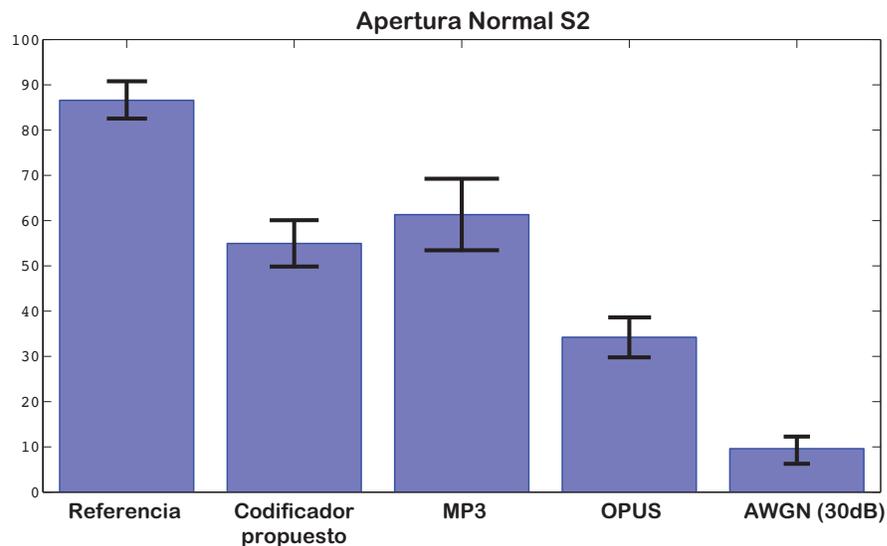
Para los estímulos seleccionados en esta prueba se han calculado las calificaciones medias de los 20 oyentes (sujetos) y se han calculado los histogramas e intervalos de confianza correspondientes a un nivel de confianza del 95 %. Las Figuras 39, 40, 41, 42, y 43 muestran los resultados obtenidos.

Los resultados arrojados por la prueba avalan el buen desempeño del codificador diseñado en este trabajo de tesis, cuya valoración promedio se encuentra en el intervalo de *acceptable* (40 a 60) según la escala MUSHRA. Como se indicó en un principio, la referencia o señal original obtuvo la calificación más alta posible para todos los estímulos. Por otro lado, la versión con ruido aditivo fue calificada con la nota más baja. La prueba fue intencionada para que estos resultados se ocurrieran y existiesen calificativos más cercanos para las otras anclas.

En las Figuras puede observarse que, aunque las versiones de MP3 en la mayoría de los estímulos superan en calificación promedio al códec (excepto en la Apertura normal del S1), de manera estadística los intervalos de confianza son muy cercanos entre estas dos anclas. En contraste, la versión codificada por OPUS presentó calificaciones más



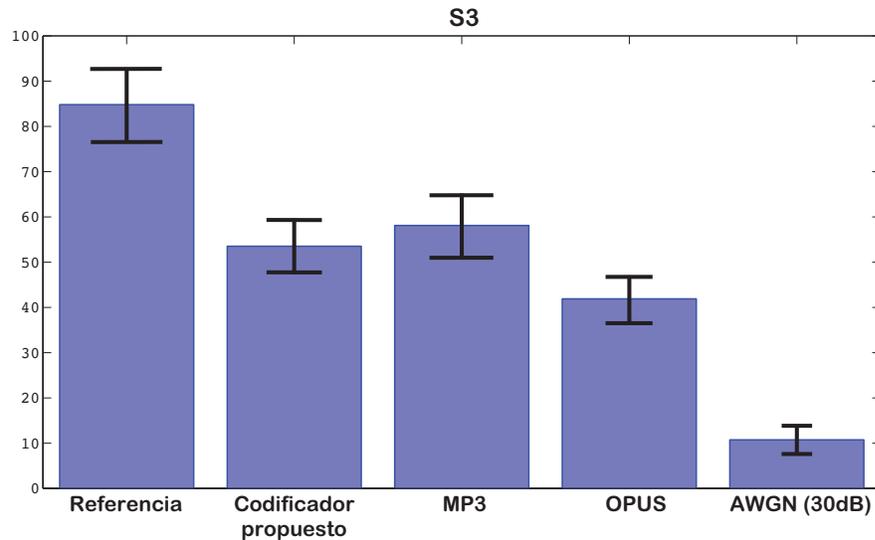
**Figura 39: Histograma de las calificaciones medias e intervalos de confianza para el estímulo Apertura Normal S1.**



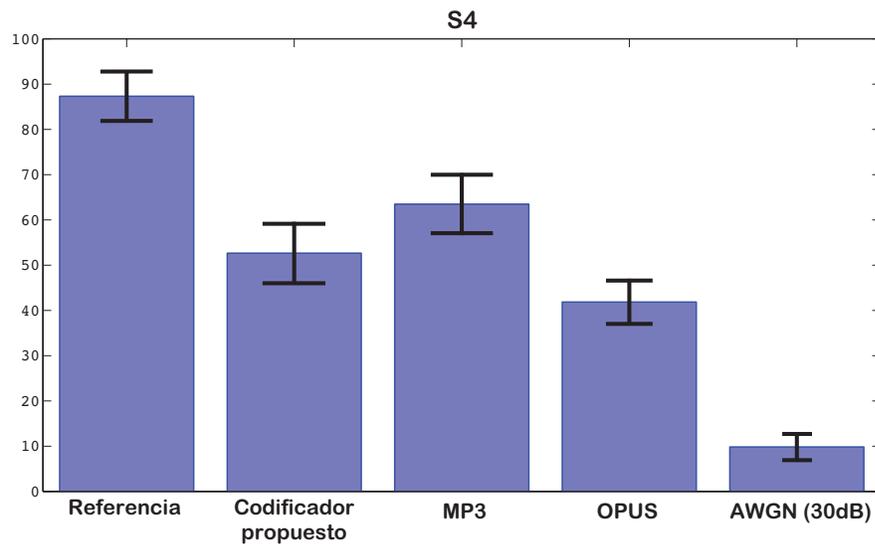
**Figura 40: Histograma de las calificaciones medias e intervalos de confianza para el estímulo Apertura Normal S2.**

bajas, y en este caso los intervalos de confianza no lograron acercarse en ningún estímulo a los del codificador.

Es importante mencionar que los codificadores MP3 y OPUS utilizan técnicas de codificación entrópica (no empleando ninguna en el codificador diseñado) y técnicas de cuantificación más sofisticadas que las empleadas en esta propuesta. A pesar de no haber implementado estos métodos en el códec, su calidad subjetiva fue evaluada como



**Figura 41: Histograma de las calificaciones medias e intervalos de confianza para el estímulo S3.**

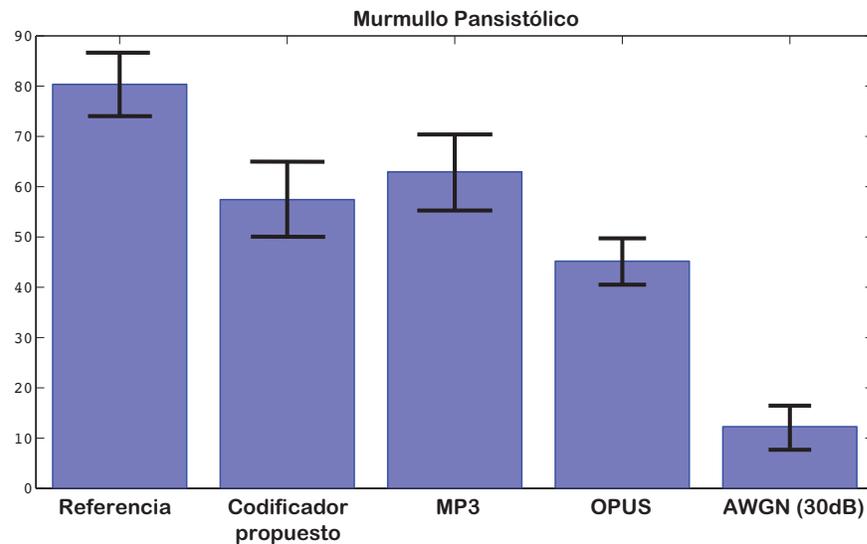


**Figura 42: Histograma de las calificaciones medias e intervalos de confianza para el estímulo S4.**

aceptable y competente<sup>1</sup> en todos los estímulos evaluados en la prueba.

Los resultados arrojados por la prueba MUSHRA para el codificador propuesto deben tomarse con cautela, son los especialistas en la salud quienes en realidad valorarán si la calidad perceptual otorgada por el codificador diseñado es la necesaria para que pueda realizarse la compresión, representación y transmisión de fonocardiogramas por el mismo.

<sup>1</sup>Los intervalos de confianza entre el codificador y MP3 fueron muy cercanos.



**Figura 43: Histograma de las calificaciones medias e intervalos de confianza para el estímulo Murmullo Pansistólico.**

De igual manera, los oyentes expertos en salud podrán determinar si las anomalías cardiacas son aún diagnosticables al realizar estos procedimientos. De igual manera los analistas clínicos podrán catalogar si los artefactos introducidos por el códec pueden detectarse como cardiopatías. Ninguno de los participantes seleccionados para el desarrollo de esta prueba tiene conocimientos especializados de este tipo.

A partir de estos resultados se formularán conclusiones importantes en la siguiente sección, ya que esta investigación ha desarrollado un codificador adaptado a una señal biológica importante para el diagnóstico clínico.

## Capítulo 7. Conclusiones

---

En esta sección se pretenden mostrar las conclusiones generadas durante el desarrollo de esta tesis, determinando los objetivos con los que ha cumplido.

Es importante también señalar durante esta sección algunas propuestas que pudiesen colaborar en la realización de trabajos futuros sobre la misma línea de investigación.

### 7.1. Sobre los objetivos de la tesis

Se ha propuesto un codificador-decodificador para señales de audio cardiaco  $x(t)$ , el cual está basado en la suma de una parte *determinística* y una parte *estocástica* dado por:

$$x(t) = \underbrace{\sum_m^M \alpha_k g_{\gamma_m}(t)}_{\text{parte determinística}} + \underbrace{R_M(t)}_{\text{parte estocástica}},$$

que tiene como objetivo el obtener una forma de onda y señal de audio óptimas, con la mayor cantidad posible de información y con un carácter de compresión de señal. Es importante señalar que no se ha encontrado durante el desarrollo de esta tesis un codificador para audio cardiaco con estas características.

Durante el desarrollo de este trabajo han surgido como conclusiones las descritas en los siguientes puntos:

- Los resultados de la presente investigación se consideran satisfactorios, un argumento se basa en el hecho de que no se ha encontrado en la literatura un codificador para audio cardiaco que conjunte estos modelados.
- La descomposición atómica por *Matching Pursuit* proporciona un esquema de compresión poderoso en la representación del PCG, sin embargo la información puede ser mayormente aprovechada por medio de aplicar este método solamente a los eventos cardiacos segmentados, ya que a diferencia de los silencios son los que presentan mayor correlación con los átomos del diccionario de Gabor.

- Los eventos correspondientes a patologías cardíacas requieren un número mayor a 10 átomos para su representación en MP. Es importante establecer un mecanismo para la extracción de energía en estas partes de la señal, que preserve la información contenida y logre el menor número de iteraciones posible.
- La señal residual  $R_M(t)$  es muy importante de modelar, a pesar de representar en este caso sólo el 1 % de la energía de cada evento cardíaco aún es bastante perceptible debido al carácter logarítmico del sistema auditivo humano.
- El modelado de la señal residual y los silencios cardíacos es posible mediante la predicción lineal (LPC), donde no debe omitirse la inclusión de la estimación del periodo tonal (pitch). La función de autocorrelación provee la información necesaria para el desarrollo de este cálculo.
- Los filtros de pre-énfasis proporcionan el soporte necesario a las altas frecuencias, permitiendo representarlas con mayor precisión en el esquema de reconstrucción del residual.
- Es importante seleccionar la representación equivalente a los coeficientes de predicción adecuada, dado que la cuantificación de los mismos no es un procedimiento sencillo. Mediante la distorsión espectral se tiene una medida adecuada para la evaluación de los coeficientes equivalentes al filtro predictor.
- La verificación de la naturaleza de los parámetros extraídos en los modelados determinista y estocástica contribuyó a elegir el tipo de cuantificación requerida para cada uno de ellos.
- Durante el diseño del cuantificador es importante adaptar sus características a las de la función de probabilidad de los datos fuente. Este procedimiento es sencillo de implementar mediante la aplicación del Algoritmo de Lloyds.
- Las señales valoradas durante el diseño del codificador presentan resultados aceptables en comparación con las versiones originales al realizar medidas como el coeficiente de correlación (cercano a 1) y porcentaje de la raíz cuadrada de la diferencia cuadrática media (PRD).

- Las pruebas subjetivas para la evaluación de la calidad del códec deben realizarse en las condiciones adecuadas para los oyentes, es importante verificar su estado anímico durante la prueba.
- El codificador diseñado ha sido valorado dentro de los intervalos de confianza aceptables, al tomar los niveles de confianza en un 95 % . Ha sido superior en la evaluación realizada a otros codificadores de acceso libre existentes en la literatura como OPUS y se ha valorado estadísticamente de manera muy cercana al desempeño de MP3.
- Es muy importante señalar que en los codificadores OPUS y MP3 se emplean algoritmos de cuantificación más sofisticados que los usados en este trabajo y que además se usan técnicas de codificación de fuente. A pesar de esto, el codificador diseñado por esta tesis no ha empleado ninguna de ellas y ha sido valorado de manera muy cercana a estos estándares.

## **7.2. Trabajo futuro de investigación**

En base a los resultados obtenidos en este trabajo de tesis, y con la experiencia adquirida en sus áreas correspondientes, se realizan las siguientes recomendaciones para la realización de trabajos futuros sobre la misma:

- Experimentar las técnicas de Matching Pursuit Molecular en el modelado de eventos cardíacos. Conformar moléculas que sin duda originarán porcentajes mayores de compresión a los obtenidos.
- Implementar en el códec una etapa de segmentación de buen desempeño.
- Dada la naturaleza pasa-bajas de las señales de audio, implementar e investigar técnicas de codificación predictiva lineal deformada (WLPC), dando mejor definición al intervalo de bajas frecuencias.
- Implementar la excitación por voz en los filtros de reconstrucción LPC (VELPC) para ahorrar recursos en el cálculo de la estimación del periodo tonal, además de evitar

las posibles impresiones por dicha estimación. De igual manera explorar el empleo de excitación por código (CELP).

- Realizar un análisis de componentes principales para conocer por medio de la transformada del coseno (DCT) cuántos de sus coeficientes contribuyen para una eficiente reconstrucción del residual mediante VELPC.
- Emplear la cuantificación vectorial sobre algunos de los parámetros extraídos en los modelados MP y LPC, con el objetivo de reducir la tasa de datos final del codificador y lograr mayor compresión.
- Codificar mediante compresión sin pérdidas parámetros con *patrones repetitivos* dados por tener un diccionario base de elementos, tales como las longitudes y las frecuencias de los átomos.
- Conformar un *código de línea* eficiente para representar el codificador del PCG, reducir posteriormente la cantidad de bits generada por medio de algún método compresión sin pérdidas (Huffman por ejemplo).
- Realizar las evaluaciones subjetivas MUSHRA con participantes expertos de la salud.
- Verificar el desempeño del códec en redes de bajas tasas de datos por alguna plataforma de simulación.
- Conformar una base de datos de señales de audio cardíaco propia, empleando el equipo adecuado y tomando un vasto número de muestras con diferentes características.

## Lista de referencias

- Abbas, A. K. y Bassam, R. (2009). Phonocardiography Signal Processing. *Synthesis Lectures on Biomedical Engineering*, **4**(1): 1–194.
- Alajarín, J. M. y Merino, R. R. (2004). Compresión de fonocardiogramas mediante las transformadas wavelet y wavelet packet. En *Actas del XXII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2004)*, páginas 181–184.
- Alajarín, J. M. y Merino, R. R. (2005). Compresión de fonocardiogramas basada en eventos. En *Actas del XXIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2005)*, páginas 423–426. Fundación Rogelio Segovia para el Desarrollo de las Telecomunicaciones.
- Atal, B. y Hanauer, S. L. (1971). Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. *The Journal of the Acoustical Society of America*, **50**(2B): 637–655.
- Blanco-Velasco, M., Cruz-Roldán, F., Godino-Llorente, J. I., Blanco-Velasco, J., Armiens-Aparicio, C., y López-Ferreras, F. (2005). On the use of PRD and CR parameters for ECG compression. *Medical Engineering and Physics*, **27**(9): 798–802.
- Bosi, M. y Goldberg, R. E. (2003). *Introduction to digital audio coding and standards*. Springer.
- Bosi, M., Brandenburg, K., Quackenbush, S., Fielder, L., Akagiri, K., Fuchs, H., y Dietz, M. (1997). ISO/IEC MPEG-2 advanced audio coding. *Journal of the Audio Engineering Society*, **45**(10): 789–814.
- Boutana, D., Benidir, M., y Barkat, B. (2011). Segmentation and identification of some pathological phonocardiogram signals using time-frequency analysis. *IET Signal Processing*, **5**(6): 527–537.
- Candeas, P. (2006). *Desarrollo de técnicas de codificación de audio basadas en modelos de señal paramétricos*. Tesis de doctorado, Universidad de Alcalá.
- Castorena, L. (2012). *Análisis y Adaptación de códecs para la transmisión de audio cardíaco sobre redes inalámbricas con tasas de datos bajas/medias*. Tesis de maestría, Centro de Investigación Científica y Educación Superior de Ensenada.
- Chen, D., Durand, L., y Lee, H. (1997). Time-frequency analysis of the first heart sound. Part 1: Simulation and analysis. *Medical and Biological Engineering and Computing*, **35**(4): 306–310.
- Choi, Y. S., Kang, H. G., y Youn, D. H. (1996). A fast VSELP speech coder based on mutually orthonormal regular pulse vectors. En *Acoustics, Speech, and signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, Vol. 1, páginas 554–557. IEEE. ISBN 0-7803-3192-3.
- Coleman, A., Gleiss, N., Sotscheck, J., Usai, P., y Scheuermann, H. (1989). Subjective performance evaluation of the RPE-LTP codec for the pan-European cellular digital mobile radio system. En *Global Telecommunications Conference and Exhibition Communications Technology for the 1990s and Beyond (GLOBECOM), 1989. IEEE*, páginas 1075–1079. IEEE.

- Dammann, C. L., McDaniel, L., y Maddox, C. L. (1972). D2 channel bank: Multiplexing and coding. *Bell System Technical Journal*, **51**(8): 1675–1699.
- Daumer, W. (1982). Subjective Evaluation of Several Efficient Speech Coders. *IEEE Transactions on Communications*, **30**(4): 665–662.
- Debbal, S. M. y Bereksi-Reguig, F. (2006). Cardiac Murmur Analysis Using the Short-Time Fourier Transform. *Journal of Mechanics in Medicine and Biology*, **6**(03): 273–284.
- Debbal, S. M. y Bereksi-Reguig, F. (2007). Time-frequency analysis of the first and the second heartbeat sounds. *Applied Mathematics and Computation*, **184**(2): 1041–1052.
- Debbal, S. M. y Bereksi-Reguig, F. (2008). Computerized heart sounds analysis. *Computers in Biology and Medicine*, **38**(2): 263–280.
- Djebbari, A. y Reguig, F. B. (2000). Short-time fourier transform analysis of the phonocardiogram signal. En *Electronics, Circuits and Systems, 2000. IECS 2000. The 7th IEEE International Conference on*, Vol. 2, páginas 844–847. IEEE. ISBN 0780365429.
- Doi, K. (2008). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, **31**(4-5): 198–211.
- Elder, B. y Purnhagen, H. (2000). Parametric audio coding. En *Signal Processing Proceedings, 2000. WCC 2000-ICSP 2000. 2000. 5th International Conference on*, Vol. 1, páginas 21–24. ISBN 0-7803-5747-7.
- Elad, M. (2010). *Sparse and Redundant Representations: from theory to applications in signal and image processing*. Springer. ISBN 978-1-4419-7010-7.
- Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, **93**(26): 429–441.
- Gersho, A. y Gray, R. M. (1992). *Vector Quantization and Signal Compression*, Vol. 159. Springer. ISBN 0792391810.
- Gerson, I. y Jasiuk, M. (1990). Vector sum excited linear prediction (VSELP) speech coding at 8 kbps. En *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*, páginas 461–464. IEEE.
- Ghofrani, S., Ayatollahi, A., y Shamsollahi, M. B. (2002). Wigner–Ville Distribution and Gabor transform in Doppler ultrasound signal processing. *Biomedical sciences instrumentation*, **39**: 142–147.
- Gray, R. M. (2006). Toeplitz and Circulant Matrices: A Review. En *Foundations and Trends in Communications and Information Theory*, Vol. 2(3), páginas 155–239. now publishers inc. ISBN 9781933019239.
- Gribonval, R. y Krstulovic, S. (2006). MPTK: Matching Pursuit Made Tractable. En *Acoustics Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, Vol. 3, páginas III–III. IEEE. ISBN 1-4244-0469-X.

- Gribonval, R. y Nielsen, M. (2003). Sparse Representations in Unions of Bases. *Information Theory, IEEE Transactions on*, **49**(12): 3320–3325.
- Gribonval, R. y Nielsen, M. (2007). Highly sparse representations from dictionaries are unique and independent of the sparseness measure. *Applied and Computational Harmonic Analysis*, **22**(3): 335–355.
- Gu, Z. G. Z., Sudhakar, R., y Lee, E. (1995). Energy efficient DSP architectures for VSELP speech coder. En *Southcon '95. Conference Record*, páginas 176–181. IEEE. ISBN 0-7803-2576-1.
- Haar, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, **69**(3): 331–371.
- Härmä, A. y Laine, U. K. (2001). A comparison of warped and conventional linear predictive coding. *Speech and Audio Processing, IEEE Transactions on*, **9**(5): 579–588.
- Hayes, M. H. (2009). *Statistical Digital Signal Processing and Modeling*, Vol. 39. John Wiley & Sons. ISBN 0471594318.
- Hedayioglu, F., Jafari, M. G., Mattos, S. S., Plumbley, M. D., y Coimbra, M. T. (2012). Denoising and segmentation of the second heart sound using matching pursuit. En *Engineering in International Conference of Medicine and Biology Society, San Diego, CA, USA*, páginas 3440–3443. ISBN 9781424441198.
- Huerta, J. M. y Stern, R. M. (2001). Distortion-class modeling for robust speech recognition under GSM RPE-LTP coding. *Speech Communication*, **34**(2): 213–225.
- Hummersone, C. (2011). MUSHRA test GUI. Recuperado el 14 de Agosto del 2014, de <http://iosr.surrey.ac.uk/software>.
- Itakura, F. (1975). Line spectrum representation of linear predictor coefficients of speech signals. *The Journal of the Acoustical Society of America*, **57**(S1): S35–S35.
- Itakura, F. y Saito, S. (1968). Analysis synthesis telephony based on the maximum likelihood method. En *Proceedings of the 6th International Congress in Acoustics*, páginas C17–C20.
- Jasper, J. y Othman, K. R. (2010). Feature extraction for human identification based on envelope signal analysis of cardiac sounds in time-frequency domain. En *Electronics and Information Engineering (ICIEIE), 2010 International Conference On*, Vol. 2, páginas V2–228. IEEE. ISBN 978-1-4244-7679-4.
- Jayant, N. S. (1974). *Digital Coding of Waveforms. Principles and applications to speech and video.*, Vol. 62. Prentice Hall. ISBN 0-13-211913-7 01.
- Kipper, U., Reininger, H., y Wolf, D. (1991). CELP coding with adaptive excitation codebooks. En IEEE, editor, *Signals, Systems & Computers. 1991. Conference Record of the Twenty-Fifth Asilomar Conference on*, páginas 940–943. ISBN 0-8186-2470-1.

- Koishida, K., Tokuda, K., Kobayashi, T., y Imai, S. (1996). CELP coding system based on mel-generalized cepstral analysis. En *Spoken Language, 1996. ICSLP '96 Proceedings., Fourth International Conference on*, Vol. 1, páginas 318–321. IEEE. ISBN 0-7803-3555-4.
- Köymen, H., Altay, B. K., y Ziya, I. Y. (1987). A study of Prosthetic Heart Valve Sounds. *Biomedical Engineering, IEEE Transactions on*, **BME-43**(11): 853–863.
- Leatham, A. (1987). Auscultation and phonocardiography: a personal view of the past 40 years. *British heart journal*, **57**(5): 397–403.
- Litmann (2013). Ruidos cardiacos y pulmonares. Recuperado el 8 de mayo del 2014, de [http://solutions.3m.com.mx/wps/portal/3M/\\_MX/3M-Littmann-LA/home/Education/SoundLibrary/](http://solutions.3m.com.mx/wps/portal/3M/_MX/3M-Littmann-LA/home/Education/SoundLibrary/).
- Mahnke, C. B. (2009). Automated heartsound analysis/computer-aided auscultation: a cardiologist's perspective and suggestions for future development. En *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, páginas 3115–3118. IEEE.
- Makhoul, J. (1973). Spectral Analysis of Speech by Linear Prediction. *Audio and Electroacoustics, IEEE Transactions on*, **21**(3): 140–148.
- Makhoul, J. (1975a). Linear prediction. A tutorial review. *Proceedings of the IEEE*, **63**(4): 561–580.
- Makhoul, J. (1975b). Spectral linear prediction: Properties and applications. *Acoustics, Speech, and Signal Processing, IEEE Transactions on*, **23**(3): 283–296.
- Mallat, S. (1999). *A Wavelet Tour of Signal Processing*. Academic Press. ISBN 9780124666061.
- Mallat, S. G. y Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, **41**(12): 3397–3415.
- Mano, K., Moriya, T., Miki, S., Ohmuro, H., Ikeda, K., y Ikeda, I. (1995). Design of a pitch synchronous innovation CELP coder for mobile communications. *Selected Areas in Communications, IEEE Journal on*, **13**(1): 31–41.
- Martinez-Alajarin, J., Lopez-Candel, J., y Ruiz-Merino, R. (2006). ASEPTIC: Aided system for event-based phonocardiographic tediagnosis with integrated compression. En *Computers in Cardiology, 2006*, páginas 537–540. IEEE. ISBN 978-1-4244-2532-7.
- McAulay, R. y Quatieri, T. (1986). Speech analysis/Synthesis based on a sinusoidal representation. *Acoustics, Speech, and Signal Processing, IEEE Transactions on*, **34**(4): 744–754.
- McLoughlin, I. y Zhong-Qiang, D. Z.-Q. D. (2000). Mandarin speech coding using a modified RPE-LTP technique. En *Circuits and Systems, 2000. IEEE APCCAS 2000. The 2000 IEEE Asia-Pacific Conference on*, páginas 748–751. IEEE. ISBN 0-7803-6253-5.

- Mermelstein, P. (1988). G. 722: a new CCITT coding standard for digital transmission of wideband audio signals. *Communications Magazine, IEEE*, **26**(1): 8–15.
- Messer, S. R., Agzarian, J., y Abbott, D. (2001). Optimal wavelet denoising for phonocardiograms. *Microelectronics Journal*, **32**(12): 931–941.
- Miki, S., Mano, K., Moriya, Y., Oguchi, K., y Ohmuro, H. (1994). A pitch synchronous innovation CELP (PSI-CELP) coder for 2-4 kbit/s. En *Acoustics, Speech and Signal Processing 1994. ICASSP-94., 1994 IEEE International Conference on*, Vol. ii, páginas II–113. IEEE. ISBN 0-7803-1775-0.
- Musmann, H. G. (2006). Genesis of the MP3 audio coding standard. *IEEE Transactions on Consumer Electronics*, **52**(3): 1043–1049.
- Nieblas, C. I. (2014). *Compresión del audio cardíaco mediante diccionarios redundantes y codificación escasa: aplicación a la transmisión en redes de sensores*. Tesis de maestría, Centro de Investigación Científica y Educación Superior de Ensenada.
- Nieblas, C. I., Alonso, M. A., Conte, R., y Villarreal, S. (2013). High performance heart sound segmentation algorithm based on Matching Pursuit. En IEEE, editor, *Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE), 2013 IEEE*, páginas 96–100. ISBN 978-1-4799-1616-0.
- Noah, M. (1984). Optimal Lloyd-Max quantization of LPC speech parameters. En *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, Vol. 9, páginas 29–32. IEEE.
- Noll, P. (1997). MPEG digital audio coding. *Signal Processing Magazine, IEEE*, **14**(5): 59–81.
- O'Reilly, J. (1984). Pulse Code Modulation. En *Telecommunication Principles*, páginas 86–96. Springer.
- Painter, T. y Spanias, A. (2000). Perceptual coding of digital audio. *Proceedings of the IEEE*, **88**(4): 451–515.
- Papoulis, A. (1984). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, segundo edición.
- Pelech, A. N. (2004). The physiology of cardiac auscultation. *Pediatric Clinics of North America*, **51**(6): 1515–1535.
- Plumbley, M. D., Blumensath, T., Daudet, L., Gribonval, R., y Davies, M. E. (2010). Sparse representations in audio and music: From coding to source separation. *Proceedings of the IEEE*, **98**(6): 995–1005.
- Quatieri, T. y McAulay, R. (1986). Speech transformations based on a sinusoidal representation. *Acoustics, Speech, and Signal Processing, IEEE Transactions on*, **34**(6): 1449–1464.
- Rabiner, L., Schafer, R., y Ronald, W. (1978). *Digital processing of speech signals*. Prentice Hall.

- Rabiner, L. L. y Juang, B.-H. B. (1993). *Fundamentals of Speech Recognition*, Vol. 103. Prentice Hall. ISBN 0130151572.
- Roguin, A. (2006). Rene Theophile Hyacinthe Laennec (1781-1826): The man behind the stethoscope. *Clinical Medicine and Research*, **4**(3): 230–235.
- Ruiz Reyes, N. y Candeas, P. (2010). Adaptive Signal Modeling Based on Sparse Approximations for Scalable Parametric Audio Coding. *IEEE Transactions on Audio, Speech, and Language Processing*, **18**(3): 447–460.
- Sava, H., Pibarot, P., y Durand, L. G. (1998). Application of the matching pursuit method for structural decomposition and averaging of phonocardiographic signals. *Medical and Biological Engineering and Computing*, **36**(3): 302–308.
- Schroeder, M. R., Atal, B. S., y Hall, J. (1979). Optimizing digital speech coders by exploiting masking properties of the human ear. *The Journal of the Acoustical Society of America*, **66**(6): 1647–1652.
- Senhadji, L., Carrault, G., Bellanger, J. J., y Passariello, G. (1995). Comparing wavelet transforms for recognizing cardiac patterns. *IEEE Engineering in Medicine and Biology Magazine*, **14**(2): 167–173.
- Series, B. (2014). Method for subjective assessment of intermediate quality level of audio systems.
- Smith, J. O. y Abel, J. S. (1999). Bark and ERB Bilinear Transforms. *Speech and Audio Processing, IEEE Transactions on*, **7**(6): 697–708.
- Soong, F. y Juang, B. (1984). Line spectrum pair (LSP) and speech data compression. En *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'84.*, Vol. 9, páginas 37–40. IEEE.
- Sturbe, H. (1980). Linear prediction on a warped frequency scale. *The Journal of the Acoustical Society of America*, **68**(4): 1071–1076.
- Sunwoo, M. H. y Park, S. (1991). Real-time implementation of the VSELP on a 16-bit DSP chip. *Consumer Electronics, IEEE Transactions on*, **37**(4): 772–782.
- Un, C. U. C. y Magill, D. (1975). The Residual-Excited Linear Prediction Vocoder with Transmission Rate Below 9.6 kbits/s. *Communications, IEEE Transactions on*, **23**(12): 1466–1474.
- Vaidyanathan, P. P. (2008). *The theory of linear prediction*. Morgan & Claypool.
- Valin, J. M., Terriberry, T. B., Montgomery, C., y Maxwell, G. (2010). A high-quality speech and audio codec with less than 10-ms delay. *Audio, Speech and Language Processing, IEEE Transactions on*, **18**(1): 58–67.
- Viswanathan, R. y Makhoul, J. (1975). Quantization properties of transmission parameters in linear predictive systems. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, **23**(3): 309–321.

- Wang, W. y Pan, J. (2004). Decomposition and analysis of the second heart sound based on the matching pursuit method. *Signal Processing, 2004. Proceedings. ICSP'04. 2004 7th International Conference on*, **3**: 2229–2232.
- Watrous, R. L., Thompson, W. R., y Ackerman, S. J. (2008). The impact of computer-assisted auscultation on physician referrals of asymptomatic patients with heart murmurs. *Clinical cardiology*, **31**(2): 79–83.
- WHO (2012). World Health Statistics. Reporte técnico, World Health Organization.
- Wolfe, P., Godsill, S., y Dorfler, M. (2001). Multi-Gabor dictionaries for audio time-frequency analysis. En *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, páginas 43–46. IEEE. ISBN 0-7803-7126-7.
- Wood, J. y Barry, D. (1995). Time-frequency analysis of the first heart sound. *Engineering in Medicine and Biology Magazine, IEEE*, **14**(2): 144–151.
- Wood, J. C., Buda, A. J., y Barry, D. T. (1992). Time-frequency transforms: A new approach to first heart sound frequency dynamics. *Biomedical Engineering, IEEE Transactions on*, **39**(7): 730–740.
- Zhang, X., Durand, L. G., Senhadji, L., Lee, H. C., y Coatrieux, J. L. (1998a). Time-frequency scaling transformation of the phonocardiogram based of the matching pursuit method. *IEEE transactions on bio-medical engineering*, **45**(8): 972–9.
- Zhang, X., Durand, L. G., Senhadji, L., Lee, H. C., y Coatrieux, J. L. (1998b). Analysis-synthesis of the phonocardiogram based on the matching pursuit method. *IEEE transactions on bio-medical engineering*, **45**(8): 962–71.
- Zölzer, U. (2011). *DAFX: Digital Audio Effects: Second Edition*. John Wiley & Sons, Inc. ISBN 9780470665992.