

Tesis defendida por
Carlos Iván Nieblas Hernández

y aprobada por el siguiente Comité

Dr. Miguel Ángel Alonso Arevalo
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. Benito Orozco Serna
Miembro del Comité

Dr. César Cruz Hernández
Coordinador del posgrado
en Electrónica y Telecomunicaciones

Dr. Jesús Favela Vara
Director de la
Dirección de Estudios de Posgrado

Enero de 2014

CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE EDUCACIÓN
SUPERIOR DE ENSENADA



Programa de posgrado en ciencias
en Electrónica y Telecomunicaciones

Compresión del audio cardiaco mediante diccionarios redundates y codificación escasa:
aplicación a la transmisión en redes de sensores.

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de

Maestro en Ciencias

Presenta:

Carlos Iván Nieblas Hernández

Ensenada, Baja California, México,

2014

Resumen de la tesis de Carlos Iván Nieblas Hernández, presentada como requisito parcial para la obtención del grado de Maestro en Ciencias en Electrónica y Telecomunicaciones con orientación en TELECOMUNICACIONES.

Compresión del audio cardiaco mediante diccionarios redundantes y codificación escasa: aplicación a la transmisión en redes de sensores.

Resumen aprobado por:

Dr. Miguel Ángel Alonso Arevalo

Director de Tesis

El objetivo de este trabajo de tesis es el desarrollo de herramientas de procesamiento de señales que modelan adecuadamente la estructura de la señal de audio cardiaco con la finalidad de ser comprimida. Se presenta un estudio teórico y experimental de la representación escasa a través del algoritmo *matching pursuit* (MP) usando diferentes diccionarios tiempo-frecuencia para la compresión del audio cardiaco. De especial interés es la posibilidad de generar el diccionario que mejor se adapte a las características de las señales de fonocardiograma que permita la compresión y segmentación de sonidos cardiacos (ataques) basado en la detección del inicio y finalización de cada sonido. Una vez comprimidas, la calidad de las señales de audio cardiaco es evaluada mediante pruebas subjetivas con la finalidad de cuantificar la distorsión introducida por la descomposición escasa. Los participantes en las pruebas de evaluación subjetiva detectaron en las señales comprimidas una calidad ligeramente superior a los algoritmos del estado-del-arte como MPEG capa 3.

Palabras Clave: **Algoritmo *matching pursuit*, átomos de Gabor, representación escasa, compresión del audio cardiaco.**

Abstract of the thesis presented by Carlos Iván Nieblas Hernández, in partial fulfillment of the requirements of the degree of Master in Sciences in Electronic and Telecommunications with orientation in TELECOMUNICACIONES.

Compression of heart sounds through redundant dictionaries and sparse coding:
application to transmission in sensor networks.

Abstract approved by:

Dr. Miguel Ángel Alonso Arevalo

Director de Tesis

The goal of this thesis is the development of signal processing tools to adequately model the structure of the heart sound signal in order to be compressed. A theoretical and experimental study of the low representation is presented through the algorithm matching pursuit (MP) using different time-frequency dictionaries for the compression of heart sound signal. Of special interest is the possibility of generating the dictionary that best models the characteristics of cardiac sound signals. Such dictionary allows us to compress and segment heart sounds, that is, to detect the beginning and end of each sound. Once compressed, the quality of audio signals heart is assessed via subjective listening tests in order to quantify the distortion introduced by sparse signal representation. During the subjective listening tests, preliminary results suggest that participants have a slightly better preference for the proposed compression scheme over state-of-the-art algorithms such as MPEG Layer 3.

Keywords: Matching pursuit algorithm, Gabor atoms, sparse representation, compression of the heart sounds.

A mi familia

Agradecimientos

A mi padre por sus consejos y a mi madre por su cariño. Agradezco a mi director de tesis, el Dr. Miguel Ángel Alonso Arévalo por su disponibilidad para compartir sus conocimientos. Agradezco al Consejo Nacional de Ciencia y Tecnología por su apoyo económico en este proceso de aprendizaje.

Contenido

	Página
Resumen en español	II
Resumen en inglés	III
Dedicatoria	IV
Agradecimientos	V
Lista de Figuras	VIII
Lista de Tablas	XI
Capítulo 1.. Introducción	1
1.1. Objetivo General	3
1.2. Objetivos Particulares	3
1.3. Justificación	3
1.4. Metodología de la investigación	4
1.5. Organización de la tesis	5
Capítulo 2.. Antecedentes	7
2.1. Introducción	7
2.2. Origen fisiológico de los sonidos cardiacos	7
2.3. Señal de fonocardiograma	8
2.4. Soplos cardiacos	10
2.5. Contenido frecuencial de la señal de fonocardiograma.	10
2.6. Umbral absoluto de audición	12
2.7. Estado del arte	13
Capítulo 3.. Representación escasa	17
3.1. Introducción	17
3.2. Ventajas de la representación escasa	17
3.3. Diccionarios para representación escasa	17
3.4. Diccionarios de transformada	19
3.5. Diccionarios Mixtos	20
3.6. Descomposiciones atómicas	21
3.7. Métodos de cálculo	23
3.8. Métodos iterativos	23
3.9. Métodos paralelos	24
3.9.1. Método de tramas	24
3.10. Basis pursuit	24

	Página
Capítulo 4.. Algoritmo Matching Pursuit	26
4.1. Introducción	26
4.2. Matching Pursuit	26
4.3. Diccionarios tiempo-frecuencia	30
4.4. Diccionario de Gabor	31
4.5. Diccionario MDCT	33
4.6. Principio de incertidumbre de Heisenberg	33
4.7. MPTK (Matching Pursuit ToolKit)	36
Capítulo 5.. Algoritmo de segmentación de sonidos cardiacos	37
5.1. Introducción	37
5.2. Parámetros del algoritmo	44
5.3. Evaluación	46
5.4. Número de átomos en el diccionario	49
5.5. Arquitectura General	50
Capítulo 6.. Compresión del fonocardiograma	59
6.1. Introducción	59
6.2. MUSHRA	59
6.3. Descripción de las señales de prueba	61
6.4. Criterio para la selección de sujetos	63
6.5. Procedimiento del método MUSHRA	65
6.6. Condiciones de audición	65
6.7. Fase de adiestramiento	66
6.8. Preselección	67
6.9. Postselección	67
6.10. Análisis estadístico	68
6.11. Desviación media de cada sujeto	68
6.12. Dispersión individual	69
Capítulo 7.. Conclusiones	72
7.1. Sobre el objetivo de la tesis	74
7.2. Trabajo futuro como continuación a esta investigación.	74
Referencias bibliográficas	76
Apéndice	84

Lista de Figuras

Figura	Página
1. Sección vertical mostrando la estructura interna del corazón. Abbas, A. & Bassam, R. (2009). Phonocardiography Signal Processing	8
2. Rango en frecuencia y nivel de energía de los sonidos y murmullos cardiacos audibles. (La imagen fue tomada de Abbas y Bassam (2009)) .	12
3. Umbral de audición y frecuencia usando la Ecuación (1).	13
4. Primera y segunda iteración para la descomposición de la señal x considerando un diccionario con 3 átomos.	28
5. Forma de onda de Gabor.	32
6. Cajas tiempo-frecuencia ("Heisenberg rectangles") representación de energía de 2 átomos de Gabor.	36
7. Sección de señal de corazón normal mostrando onset y offset para cada ataque.	39
8. Descomposición atómica en el plano tiempo-frecuencia para la señal de la Figura 7.	40
9. Distancia entre átomos sucesivos en el plano tiempo-frecuencia	41
10. Selección de distancias entre átomos óptimos que cumplen el criterio de selección.	42
11. Identificación de la posición u del primer átomo en el plano tiempo-frecuencia.	42
12. Identificación del átomo entre onsets sucesivos.	43
13. Detección de onsets para cada ataque	43
14. Detección de offset en cada ataque de la señal de sonido cardiaco normal (S1,S2).	44
15. Identificación del último átomo en el plano tiempo-frecuencia.	46
16. Umbral mínimo de segmentación en soplo diastólico.	47
17. Plano tiempo-frecuencia para la señal de sonido cardiaco con soplo diastólico.	47

Lista de Figuras (continuación)

Figura	Página
18. Comparación del NRMSE entre diccionarios de Gabor	51
19. Comparación del NRMSE entre el diccionario de Gabor y MDCT	52
20. a) sección de un sonido cardiaco normal a descomponer usando MP, b) átomo óptimo seleccionado en la primera iteración del algoritmo MP, c) Reconstrucción de la señal con el átomo previamente seleccionado, d) Señal residual después de la primera iteración.	55
21. a) Señal a descomponer en la segunda iteración (señal residual)., b) Segundo átomo óptimo seleccionado usando MP, c) Reconstrucción de la señal con los dos átomos previamente seleccionados., d) Señal residual después de la segunda iteración.	56
22. a) Señal a descomponer en la segunda iteración (señal residual), b) 50 átomos seleccionado usando MP., c) Reconstrucción de la señal con los 50 átomos previamente seleccionados., d) Señal residual después de 50 iteraciones.	57
23. Átomo de Gabor plano tiempo-frecuencia.	58
24. Cinco intervalos de escala continua (CQS) usados en el método <i>MUSHRA</i>	60
25. Visualización de los estímulos en la interface de pruebas.	61
26. Intervalos de confianza para cada estímulo.	68
27. Calificación asignada por cada participante a la señal de referencia oculta Split S1 usando diferentes diccionarios de Gabor.	85
28. Calificación asignada por cada participante a la señal referencia oculta Normal Split S2 usando diferentes diccionarios de Gabor.	85
29. Calificación asignada por cada participante a la señal referencia oculta S3 usando diferentes diccionarios de Gabor.	86
30. Calificación asignada por cada participante a la señal referencia oculta S4 usando diferentes diccionarios de Gabor.	86
31. Calificación asignada por cada participante a la señal referencia oculta click de eyección usando diferentes diccionarios de Gabor.	87
32. Calificación asignada por cada participante a la señal referencia oculta soplo sistólico temprano usando diferentes diccionarios de Gabor.	87

Lista de Figuras (continuación)

Figura		Página
33.	Calificación asignada por cada participante a la señal referencia oculta soplo sistólico tardio usando diferentes diccionarios de Gabor.	88
34.	Calificación asignada por cada participante a la señal referencia oculta soplo diastólico retumbante usando diferentes diccionarios de Gabor. . .	88
35.	Calificación asignada por cada participante a la señal referencia oculta estenosis mitral usando diferentes diccionarios de Gabor.	89
36.	Calificación asignada por cada participante a la señal referencia patrón Normal Split S1 usando diferentes diccionarios de Gabor.	89
37.	Calificación asignada por cada participante a la señal referencia patrón a Normal Split S2 usando diferentes diccionarios de Gabor.	90
38.	Calificación asignada por cada participante a la señal referencia patrón S3 usando diferentes diccionarios de Gabor.	90
39.	Calificación asignada por cada participante a la señal referencia patrón S4 usando diferentes diccionarios de Gabor.	91
40.	Calificación asignada por cada participante a la señal referencia patrón click de eyección usando diferentes diccionarios de Gabor.	91
41.	Calificación asignada por cada participante a la señal referencia patrón soplo sistólico temprano usando diferentes diccionarios de Gabor. . . .	92
42.	Calificación asignada por cada participante a la señal referencia patrón soplo sistólico tardio diferentes diccionarios de Gabor.	92

Lista de Tablas

Tabla	Página
I. Caracterización de los ruidos cardiacos	10
II. Comparación de diferentes diccionarios para la segmentación de sonidos cardiacos de la señal normal S1.	38
III. Base de datos de sonidos cardiacos.	48
IV. Características de los diccionarios de Gabor usados en la descomposición MP.	52
V. Base de datos de sonidos cardiacos.	62
VI. Diccionarios de Gabor propuestos.	63
VII. Porcentajes de reconstrucción de las señales originales con diferentes diccionarios de Gabor.	64
VIII. Promedio de intervalos de confianza y desviación para cada uno de los estímulos.	71
IX. Desviación media y dispersión individual por cada sujeto	93
X. Desviación media y dispersión individual por cada sujeto	94

Capítulo 1

Introducción

Una de las prioridades de cualquier sociedad es contar con servicios de salud que les permitan garantizar una buena calidad de vida a sus ciudadanos. De acuerdo con la OMS (Organización Mundial de la Salud) las enfermedades cardiovasculares son la primera causa de muerte a nivel mundial [Organization (Geneva, Switzerland 2011b)]. Actualmente los avances en TIC (Tecnologías de Información y comunicación) y los avances propiamente en la medicina han dotado a la sociedad de los medios y herramientas necesarias para mejorar la calidad de vida. Así mismo hoy en día existe un gran interés en proponer sistemas de telemedicina a la sociedad con el objetivo de facilitar el acceso a diferentes herramientas de diagnóstico a través de una interacción remota entre pacientes y médicos de cualquier especialidad.

Tanto el desarrollo de tecnologías de información (TI) como los nuevos sistemas de monitoreo en la medicina han permitido brindar nuevos servicios para mejorar la calidad de vida. Sin embargo en la mayoría de las veces no es posible proporcionar a toda la sociedad de dichas herramientas y métodos debido a las diversas limitantes existentes, geográficas, de infraestructura, económicas, conocimiento, entre otras. Limitantes como ubicación y accesibilidad ha provocado el surgimiento de áreas tales como telemedicina y telesalud las cuales buscan proveer a la población de servicios básicos de salud. Conjuntamente con el amplio desarrollo de las tecnologías de información los conceptos de telemedicina y telesalud se han fortalecido. Según la UIT (Unión Internacional de Telecomunicaciones) y la OMS (Organización Mundial de la Salud): “La Telemedicina es el suministro de servicios de atención sanitaria, en cuanto la distancia constituye un

factor crítico, por profesionales que apelan a las tecnologías de la información y de la comunicación con el objeto de intercambiar información para hacer diagnósticos”.

La telemedicina puede ofrecer un medio alternativo que permite potencialmente brindar atención adecuada sin traslados y costos innecesarios . Sin embargo aún existen retos que resolver y proponer desde el desarrollo de dispositivos, diseño de sistemas de transmisión y nuevas redes de transmisión.

Es importante describir que la mayor parte de los recursos de salud se encuentran concentrados en las grandes ciudades. Por tal motivo, es prioridad acercar a las diversas comunidades rurales los medios y herramientas necesarias para garantizar cierta calidad mínima de vida en servicios de salud. La incorporación de las tecnologías de información y comunicaciones en los servicios de salud implican un cambio de concepción, organización y gestión de servicios. Telemedicina, telesalud, telemonitorización o teleasistencia son algunos de los términos más utilizados para describir la incorporación de las Tecnologías de la información y la comunicación (TIC) en los servicios de salud.

Actualmente el sector salud ha mejorado y es posible contar con herramientas y técnicas básicas de la medicina general en las regiones rurales.

En el caso de las enfermedades cardiovasculares existe gran cantidad de dispositivos que proveen al médico de información suficiente para el diagnóstico, sin embargo la mayoría de los hospitales y centros de salud no tienen acceso a estos avances, de tal manera que la auscultación con estetoscopio sigue siendo la técnica de valoración más usada por la medicina general. La auscultación es un procedimiento de exploración física que consiste en escuchar de manera directa a través de un estetoscopio el área torácica o del abdomen para valorar los sonidos cardiacos normales o patológicos [Saha y Kumar (2004)].

1.1. Objetivo General

El objetivo General consiste en evaluar el desempeño de la representación escasa a través del algoritmo *Matching Pursuit* (MP) [Mallat y Zhang (1993)] para proponer nuevos mecanismos de análisis tiempo-frecuencia adecuados para la compresión transparente, segmentación y reducción del ruido del audio cardiaco.

1.2. Objetivos Particulares

Investigar la aplicación de la representación escasa en señales de audio cardiaco.

- Generar diccionarios redundantes de forma de onda elementales y seleccionar el diccionario que mejor represente las señales de audio cardiaco.
- Proponer un algoritmo de segmentación de sonidos cardiacos usando el algoritmo MP.
- Evaluar el desempeño del algoritmo MP y diccionarios seleccionados a través de pruebas subjetivas.

1.3. Justificación

En la actualidad existen estetoscopios digitales que permiten la transmisión de señales de audio cardiaco sobre redes de datos, el ancho de banda requerido es del orden de varias decenas de Kbps, lo cual hace que su utilización sea prohibitiva cuando se consideran la capacidades ofrecidas por las redes de datos disponibles en localidades remotas de nuestro país. Por tal motivo se requiere de un mecanismo que permita la transmisión de señales de audio cardiaco en redes inalámbricas con tasa de datos

bajas/medias. Se decidió enfocarse en este tipo de señales debido a que las enfermedades cardiovasculares son la primera causa de mortalidad a nivel mundial [Organization (Geneva, Switzerland 2011a)]. Lo descrito anteriormente permite inferir que una alternativa para la transmisión de datos, sin duda, es la compresión de señales de audio cardiaco mediante la implementación de un esquema de codificación. La codificación escasa ha demostrado ser una herramienta poderosa en el análisis y procesamiento de señales de audio [Plumbley *et al.* (2010)]. Por tal motivo, utilizaremos este algoritmo en el desarrollo de herramienta de análisis de señales de audio cardiaco para su aplicación en la transmisión de redes de sensores.

El creciente interés en las técnicas de representación escasa de señales ha logrado desencadenar variadas mejoras en el análisis tiempo-frecuencia de señales conservando únicamente las características sobresalientes de las señales. Así mismo la representación escasa de señales ha sido ampliamente usada en aplicaciones tales como reconocimiento, y reducción del ruido (*denoising*), separación de fuentes, compresión y segmentación [Cho y Kuo (2011)]. Sin duda esta técnica a través de sus diferentes aplicaciones comparte un aspecto el cual consiste en la simplicidad.

1.4. Metodología de la investigación

En esta sección se presenta una descripción de los pasos realizados para alcanzar los objetivos propuestos de la tesis. La primera parte de la metodología consiste en la formación de fundamentos teóricos sobre la representación escasa de señales, centrandose principalmente en el algoritmo iterativo MP. En base a este concepto se proyecta la investigación de diccionarios tiempo-frecuencia formado por diferentes formas de onda. Posteriormente se evalúan diferentes diccionarios tiempo-frecuencia buscando aquellos

que mejor aproximen las señales de audio cardiaco. Se propone el desarrollo de un algoritmo para la segmentación eficiente de sonidos cardiacos usando el algoritmo *Matching Pursuit* (MP) y diccionarios previamente seleccionados. Finalmente se evalúa el uso del algoritmo MP como herramienta de compresión para señales de audio cardiaco así como una evaluación subjetiva del mismo usando las pruebas perceptuales.

1.5. Organización de la tesis

El presente documento está estructurado de la siguiente forma. En el capítulo 2 se describe brevemente el funcionamiento del corazón haciendo énfasis en el origen de los sonidos cardiacos. Para ello se describen las señales de fonocardiograma con sonidos cardiacos normales y anormales así como sus principales características en tiempo y frecuencia.

El capítulo 3 está dedicado al estudio de la representación escasa de señales. En este capítulo se describen ampliamente los diferentes tipos de diccionarios tiempo-frecuencia usados. Se describen también los diferentes métodos de cálculo para la descomposición atómica de señales.

En el capítulo 4 se presentan los conceptos más relevantes con respecto al algoritmo *Mathching Pursuit* como el soporte fundamental sobre el cual se desarrolla la investigación. Para ello se describe el algoritmo MP como un algoritmo iterativo. Así mismo se incluyen investigaciones sobresalientes sobre el algoritmo MP en el procesamiento de señales biomédicas y se presentan los diccionarios de Gabor como diccionarios óptimos para la descomposición de señales de audio cardiaco.

En el capítulo 5 se presenta una novedosa técnica de segmentación de sonidos de audio cardiaco basada en el algoritmo MP y usando diccionarios de Gabor.

En el capítulo 6 se describe la compresión del audio cardiaco a través de las pruebas subjetivas donde participantes escuchan y evalúan la señal de audio cardiaco.

Finalmente en el capítulo 7 se desarrollan las conclusiones de la tesis y se muestra el trabajo futuro. Se realiza una revisión de las aportaciones durante todo el proceso de investigación de las cuales pueden derivarse futuras tesis en el campo de procesamiento de señales de audio.

Capítulo 2

Antecedentes

2.1. Introducción

En este capítulo se presenta una breve descripción sobre el funcionamiento del corazón haciendo énfasis en el origen de los sonidos cardiacos y en las señales de fonocardiograma. Se estudian los conceptos más relevantes con respecto al umbral de audición. Se describe también la intersección entre la frecuencia y la distribución de energía de los sonidos cardiacos.

2.2. Origen fisiológico de los sonidos cardiacos

En la dinámica cardiaca, los eventos que ocurren desde el inicio de un latido hasta el inicio del próximo se conoce como ciclo cardiaco [Abbas y Bassam (2009)]. El ciclo cardiaco consta de un período de contracción llamado sístole, seguido de un período de relajación durante el cual el corazón se llena de sangre, llamado diástole. El corazón cuenta con cuatro válvulas; las valvulas auriculoventriculares (AV) y las sigmoideas. Las válvulas AV (Mitral y Tricúspide) impiden que de los ventrículos se regrese la sangre a las aurículas durante la sístole, y las válvulas sigmoideas (aórtica y pulmonar) impiden que regrese la sangre de las arterias aorta y pulmonar a los ventriculos durante la diástole. En condiciones normales, la sangre fluye de forma continua de las grandes venas a las aurículas. La Figura 1 muestra el flujo de la sangre a través de las cuatro valvulas del corazón y el proceso de oxigenación de la sangre.

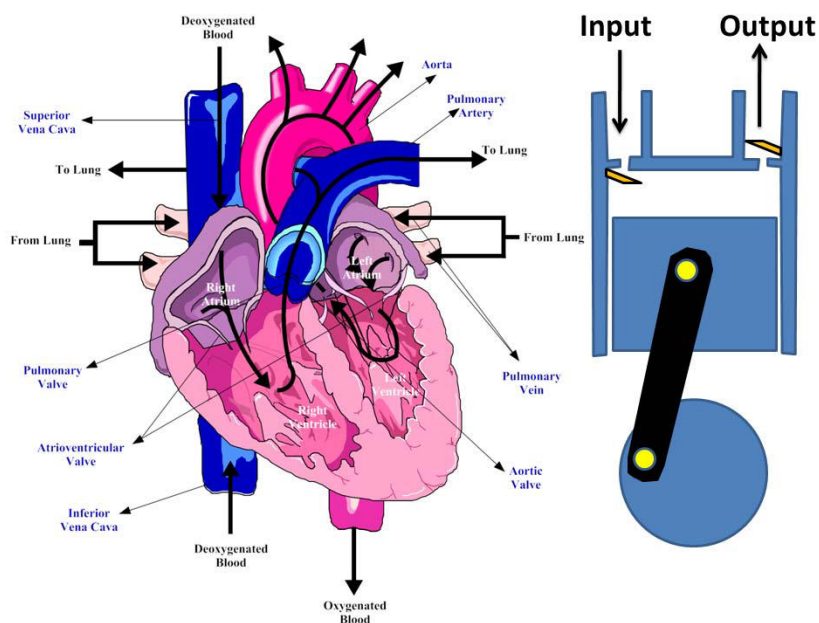


Figura 1: Sección vertical mostrando la estructura interna del corazón. Abbas, A. & Bassam, R. (2009). Phonocardiography Signal Processing

2.3. Señal de fonocardiograma

La señal del fonocardiograma (FCG) se conforma por ciclos cardiacos sucesivos y generalmente puede dividirse en 4 partes: S1, sístole, S2 y diástole. Ocasionalmente, también pueden ser escuchados dos sonidos adicionales, conocidos como S3 y S4.

El primer sonido cardiaco S1 se atribuye al sonido generado por el cierre de las válvulas mitral (M) y tricúspide (T) en el inicio de la sístole ventricular. El segundo sonido cardiaco (S2) ocurre en el fin de la sístole ventricular y al comienzo de la relajación ventricular, formado por dos componentes de alta frecuencia: cierre de las válvulas aórtica A y pulmonar P, y apertura de las válvulas mitral (M) y tricúspide (T). El sonido S2 usualmente presenta componentes de más alta frecuencia que S1, y generalmente su duración es menor que la de S1.

Ocasionalmente se escucha un tercer sonido cardiaco S3, que corresponde a la fase

de llenado ventricular. Es una vibración de baja frecuencia y se puede presentar entre 0.12 a 0.16 s. después de S2. El cuarto sonido S4 es producido al final de la diástole, y se genera por la contracción de las aurículas desplazando el flujo sanguíneo hacia los ventrículos.

La señal de fonocardiograma (FCG) corresponde a la adquisición de las vibraciones mecánicas provenientes del corazón y transmitidas hacia la pared torácica. La explicación valvular afirma que los sonidos del corazón son vibraciones variantes en el tiempo, resultantes de la tensión abrupta que se produce sobre las hojillas de las válvulas cardíacas al final de su cierre y apertura [Abbas y Bassam (2009)]. Por otro lado la explicación hemodinámica, la cual tiene mayor aceptación, atribuye el origen de los sonidos cardíacos a las vibraciones de toda la estructura del corazón, como consecuencia de la aceleración y desaceleración de los fluidos de sangre intracardiacos seguidos por el cierre y la apertura de las válvulas.

La fonocardiografía fue desarrollada para mejorar los resultados obtenidos con el estetoscopio convencional. El estado actual de la tecnología en electrónica ha abierto nuevas posibilidades en el campo de la instrumentación para la auscultación de alta calidad, que incluye el desarrollo de sistemas digitales para la adquisición, registro, almacenamiento, y análisis de los sonidos cardíacos.

En general, la frecuencia de las vibraciones acústicas de los fenómenos cardíacos está determinada por el mecanismo productor, encontrándose variaciones en el rango de 16 a 1000 Hz. La frecuencia de vibración de cada fenómeno determina a su vez la posibilidad de ser detectada por el oído humano. Los sonidos generados por la contracción muscular son de baja frecuencia (1-30 Hz), mientras que los producidos por la acción de las válvulas y el flujo de sangre se consideran de alta frecuencia (30-1000 Hz).

La duración y rango de frecuencias de los diferentes sonidos cardíacos se muestra en

Tabla I: Caracterización de los ruidos cardiacos

Sonido cardiaco	Duración (s)	Rango de frecuencia (Hz)
S1	0.1-.012	20-150
S2	0.08-0.14	50-60
S3	0.04-0.05	20-50
S4	0.04-0.05	< 25

la tabla I [Castorena (2012)].

2.4. Soplos cardiacos

Normalmente, en el silencio de sístole y de diástole no deben existir sonidos adicionales, sin embargo en ciertas condiciones de salud es posible escuchar en estos intervalos ruidos de alta frecuencia, denominados soplos cardiacos, en forma de sonidos ásperos o silbantes que nacen, la mayoría de las veces, por vibraciones anormales del flujo sanguíneo debidas a daños o defectos valvulares, aunque también pueden producirse por defectos en las paredes del corazón.

2.5. Contenido frecuencial de la señal de fonocardiograma.

El espectro de vibración puede ser dividido en varias bandas según [Abbas y Bassam (2009)] .

- **De 0-5 Hz.** Esta banda de vibración corresponde a los movimientos visibles y palpables de la pared torácica. En esta banda se incluyen la palpitación precordial

y otros movimientos de varios espacios intercostales.

- **De 5-25 Hz.** Esta banda incluye vibraciones las cuales son ahora llamada vibraciones de baja frecuencia. Esta banda es parcialmente infrasónica (5-15 Hz) y parcialmente subliminal (15-25 Hz).
- **De 120-500 Hz.** Esta banda corresponde a la mejor área de grabación de la mayoría de los dispositivos y se encuentra en el rango auditivo. De 240-500 Hz corresponde a una buena área de grabación de muchos dispositivos la cual es aproximadamente representada por el método (logarítmico) de Rappaport y Sprague.
- **De 500-1000 Hz.** Esta larga banda corresponde al área del espectro donde los sonidos originados en el corazón y grabados de la pared torácica son extremadamente reducidos en magnitud. Por lo tanto la audibilidad podría ser limitada o incluso nula a causa de una pobre magnitud de los sonidos cardiacos.

En la Figura 2 se muestran 3 aspectos fundamentales a considerar al trabajar con señales de audio cardiaco: Las diferentes bandas de frecuencia, la distribución de energía de los sonidos y murmullos cardiacos y el umbral de audición. De la figura descrita anteriormente se puede observar que la mayor distribución de energía para los sonidos y murmullos cardiacos se encuentra por debajo del umbral de audición. En la Figura 2 se puede apreciar una intersección entre la distribución de energía y el umbral de audición entre aproximadamente 50 y 1000 Hz.

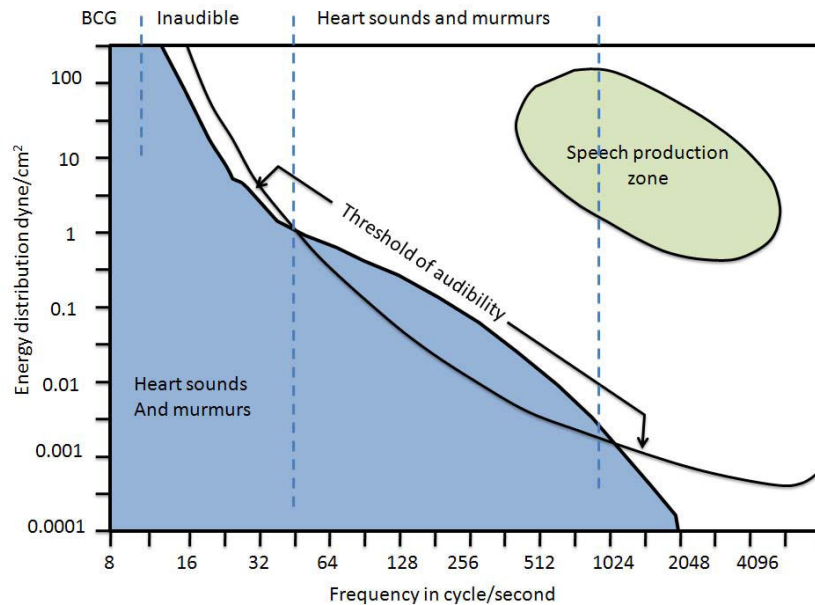


Figura 2: Rango en frecuencia y nivel de energía de los sonidos y murmullos cardiacos audibles. (La imagen fue tomada de Abbas y Bassam (2009))

2.6. Umbral absoluto de audición

El umbral absoluto de audición es la intensidad mínima que debe tener un tono puro para ser detectado por un oyente en un ambiente sin ruido [Prasad *et al.* (2008)]. Generalmente el umbral de audición se expresa en decibeles (dB) de Nivel de presión sonora (Sound Pressure Level, SPL). Este umbral está en función de la frecuencia del tono aplicado. Basado en estudios experimentales sobre un gran número de sujetos, la curva ideal del umbral absoluto de audición (ATH por sus siglas en inglés) puede ser aproximada en función de la frecuencia por la Ecuación (1) [Prasad *et al.* (2008)].

$$T_q = 3,64 \frac{f^{-0,8}}{100} - 6,5e^{-0,6(\frac{f}{1000}-3,3)^2} + 10^{-3} \left(\frac{f}{1000}\right)^4 \text{ dB SPL.} \quad (1)$$

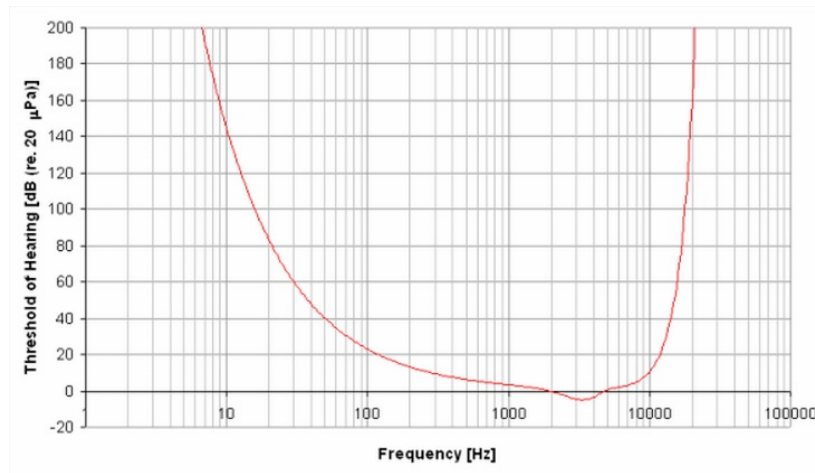


Figura 3: Umbral de audición y frecuencia usando la Ecuación (1).

Esta curva es de gran utilidad en codificación de audio porque las componentes frecuenciales bajo este umbral no pueden ser escuchadas y, por lo tanto, no necesitan ser transmitidas.

La Figura 3 muestra este umbral dependiendo de la frecuencia. Se puede observar cómo el oído es más sensible en el rango de frecuencias de 1 a 5 kHz ., principalmente debido a la acción del oído externo. El umbral crece rápidamente tanto en baja como en alta frecuencia.

2.7. Estado del arte

Las tecnologías de información han tenido grandes avances en el almacenamiento y transmisión de datos de audio. Sin embargo a pesar de los diferentes avances en las tecnologías de información dieron origen a los algoritmos de compresión para reducir el flujo de información sin perder información relevante para los diferentes usuarios de

manera significativa la señal. El análisis de los sonidos cardiacos ha sido un tema central en la investigación por varios años.

La fonocardiografía es un método sencillo, eficiente y no invasivo que permite evaluar la condición del corazón. Esta técnica es ampliamente usada en la evaluación de la condición cardiaca de niños y adultos por su facilidad de operación y bajo costo [Tang *et al.* (2012)]. Debido a la sencillez y eficiencia, este método ha sido ampliamente apoyado por diferentes hipótesis que buscan agrupar los sonidos cardiacos a través de diferentes parámetros atómicos [Tang *et al.* (2012)]. El proceso de segmentación de sonidos cardiacos como un paso previo a procesos de transmisión y compresión ha crecido considerablemente. Sin duda una de las herramientas que ha hecho posible tales avances, es el procesamiento de señales a través de diferentes técnicas y algoritmos de extracción de envolvente [Jiang (2006)]. La auscultación es una de las técnicas más efectivas para el diagnóstico e identificación de los sonidos cardiacos [Hedayioglu *et al.* (2012)]. Sin embargo la auscultación es una habilidad que requiere entrenamiento y experiencia la cual se apoya de la interpretación subjetiva [Mahnke (2009)]. Técnicas de descomposición como como el algoritmo matching pursuit tiene como objetivo la segmentación y denoising para facilitar la identificación de diferentes componentes en los sonidos cardiacos [Hedayioglu *et al.* (2012)]. El problema de segmentación de los sonidos cardiacos en presencia de murmullos ha sido abordado en los últimos años a través de diferentes enfoques de detección de componentes cardiacas [Kumar *et al.* (2011)]. Los resultados de los diferentes algoritmos para la segmentación de sonidos cardiacos se han extendido de tal manera de proponer nuevas técnicas para la detección y análisis de murmullos de sonidos cardiacos [Iwata *et al.* (1980)].

Los sonidos cardiacos son señales clave que permiten modelar la condición mecánica funcional del corazón. De tal manera que el uso de los diferentes técnicas de proce-

samiento de señales para conocer el contenido espectral se ha enfocado en conocer el origen de patologías cardiacas [Kumar *et al.* (2007)].

Las enfermedades del corazón son la mayor causa de muerte en países desarrollados [Mahnke (2009)]. La auscultación usando un estetoscopio es la herramienta más usada por médicos de la medicina general. Además del estetoscopio existe una gran variedad de herramientas como son la cateterización cardiaca, imagen de resonancia magnética, ultrasonido cardiaco, entre otros los cuales proporcionan gran cantidad de información para evaluar las diferentes patologías cardiacas. Sin embargo la disponibilidad de dichas herramientas de diagnóstico están limitadas por su costo y disponibilidad. El avance en las tecnologías de información ha logrado dotar a la sociedad de nuevos dispositivos de ultrasonido cardiaco a bajo costo, sin embargo se requiere de entrenamiento y experiencia para una adecuada interpretación de resultados. Diferentes investigaciones han demostrado una pobre interpretación de la auscultación por parte de los médicos de la medicina general [Mahnke (2009)].

El oído humano es ineficiente para detección de multi-componentes en las señales de fonocardiograma, así mismo la estructura compleja de las señales (PCG) agrega más dificultades en la auscultación [Durand *et al.* (1996)].

El uso de dispositivos sofisticados en el diagnóstico de enfermedades cardiovasculares ha disminuido las habilidades de diagnóstico. Por tal motivo se han propuesto nuevas técnicas para la simulación de sonidos cardiacos como las técnicas de predicción lineal [Agostinho y Souza (1997)] las cuales buscan la relación de funciones y algunos eventos cardiacos.

Las PCG han sido ampliamente estudiadas para investigar el origen de los sonidos cardiacos y detectar enfermedades cardiovasculares. Esto es debido a que las señales PCG presentan características tales como magnitud y frecuencia que permiten identi-

ficar comportamientos de eventos relacionados a enfermedades [Sava y Durand (1997)]

En la auscultación se intenta escuchar y analizar los componentes de sonido cardiaco por separado y entonces sintetizar sus características. El propósito de ciertos estudios es desarrollar algoritmos de segmentación automática los cuales pueden ser menos sensitivos a ruidos ambientales ejemplo de estos son los algoritmos que usan descomposición en ondeletas que aproximen las señales de (PCG) [Huiying *et al.* (1997)].

El interés de modelar los sonidos cardiacos ha aumentado considerablemente en los últimos años. De la misma manera que se busca modelar las señales de PCG, las señales de ECG ha tenido gran impacto en la detección del primer sonido cardiaco (S1) y el segundo sonido cardiaco (s2) a través de diferentes métodos y algoritmos que se centran en conocer la energía de las señales mencionadas [Huiying *et al.* (1997)].

El diagnóstico de enfermedades cardiovasculares con estetoscopio y electrocardiograma son dos métodos fundamentales por su eficiencia, simplicidad y por su propiedad no invasiva [Malarvili *et al.* (2003)]. Encontrar una representación escasa de señales se ha convertido en un área de mayor investigación en los últimos años. La representación escasa son útiles para la compresión de señales y en aplicaciones tales como denosing y separación de fuentes [Plumbley *et al.* (2010)] [Daudet (2006)]. La descomposición atómica escasa ha sido ampliamente usada en aplicaciones tales como el análisis de señales biomédicas [Sturm *et al.* (2008)]

Capítulo 3

Representación escasa

3.1. Introducción

La representación escasa ha demostrado ser una herramienta eficiente en el análisis y procesamiento de señales de audio. A través de los años ha habido un creciente interés en encontrar formas para transformar señales en una representación escasa. La representación escasa es útil en la compresión de señales, así como, en aplicaciones de separación de fuentes, *denoising* y nuevos métodos de codificación, los cuales han demostrado tener un mejor desempeño que los métodos no escasos [Cho y Kuo (2011)].

3.2. Ventajas de la representación escasa

Encontrar una representación escasa para una señal, tiene muchas ventajas para aplicaciones tales como codificación o separación de fuentes [Plumbley *et al.* (2010)]. En codificación, una representación escasa tiene pocos valores diferentes de cero, únicamente estos valores (y su localización) necesitan ser codificados para transmitir o almacenar la señal. Así mismo, el ruido u otros disturbios de la señal en cuestión, no son representados por los mismos coeficientes que la señal *sparse*. Por lo tanto dichos coeficientes no deseados pueden ser removidos permitiendo una mejor reconstrucción de la señal deseada.

3.3. Diccionarios para representación escasa

Representar una señal en forma escasa involucra la selección de un diccionario, el cual es un conjunto de señales elementales o átomos usados para descomponer la señal. Cuando el diccionario forma una base, toda señal es representada únicamente como la combinación lineal de átomos del diccionario. En el caso más simple el diccionario es ortogonal y la representación de coeficientes puede ser calculada como productos internos entre la señal y los átomos; en el caso no-ortogonal, los coeficientes son el producto interno de la señal y el diccionario inverso también llamado diccionario bi-ortogonal [Rubinstein *et al.* (2010)].

Tales diccionarios introducen una interesante representación de una señal. Si se considera el diccionario $D = [d_1, d_2, \dots, d_L] \in \mathbb{R}^{L \times N}$ donde las columnas forman los átomos del diccionario y $L \geq N$. Una señal $x \in \mathbb{R}^N$ usando este diccionario puede ser representada a través del producto interno con los átomos, esto es:

$$\gamma_\alpha = D^T x \quad (2)$$

donde la señal x es una combinación lineal de átomos tal que:

$$x = D\gamma_s \quad (3)$$

Las ecuaciones (2) y (3) coinciden en el caso completo ($L = N$), donde los diccionarios de análisis y síntesis son bi-ortogonales.

Por años los diccionarios ortogonales y bi-ortogonales fueron dominantes debido a su simplicidad matemática. Sin embargo, las debilidades de estos diccionarios eventualmente superan su simplicidad. Esto permitió el desarrollo de diccionarios sobrecompletos que involucran más átomos que las dimensiones de la señal, permitiendo así una

mejor representación de las características de la señal [Rubinstein *et al.* (2010)]. El cambio a diccionarios sobrecompletos fue hecho cautelosamente, en un intento por minimizar la pérdida de propiedades favorables ofrecidas por transformaciones ortogonales [Rubinstein *et al.* (2010)].

3.4. Diccionarios de transformada

Con respecto a las transformadas las bases de Fourier describen una señal en términos de su contenido en frecuencia como una combinación de formas de onda ortogonales. La transformada de Fourier es la técnica de uso más común para el análisis de las señales en el dominio de la frecuencia. Esto es principalmente porque la transformada emplea funciones sinusoidales como funciones base. Así la señal $x(t)$ es proyectada sobre la función base exponencial compleja de frecuencia $w = 2\pi f$, obteniendo la representación de Fourier $X(\omega)$ como:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad (4)$$

Sin embargo para lograr escases, la transformación requiere mejor localización. Los átomos con un soporte concentrado permiten una representación más flexible sobre las características locales de la señal y limita los efectos de irregularidades las cuales son la causa principal de gran cantidad de coeficientes. En este contexto uno de los primeros métodos al ser usado fue la Transformada de Fourier de Tiempo Reducido (Short-time Fourier transform, STFT) [Allen y Rabiner (1977)], la cual emerge como una extensión natural de la transformada de Fourier.

En la STFT la transformada de Fourier es aplicada localmente (posible traslape) a secciones de la señal describiendo la señal en tiempo-frecuencia (o espacio-frecuencia).

Así, el análisis de tiempo-frecuencia es obtenido ya que la localización de la ventana desplazada adiciona la dimensión del tiempo. La expresión para este análisis de tiempo-frecuencia está dado por la ecuación (5) :

$$S(w, t) = \int_{-\infty}^{\infty} x(t)w(\tau - t)e^{-j\omega\tau} d\tau, \quad (5)$$

donde w es una función ventana, comúnmente una ventana de Hann o campana Gaussina. Durante los años 1980 y 1990 , la STFT fue extensamente investigada y generalizada, convirtiendose más conocidamente como la Transformada de Gabor nombre en homenaje a Dennis Gabor [Hill (2007)].

La transformada o banco de filtros a utilizar depende en gran parte de la aplicación a desarrollar o de la señal de entrada. Por lo tanto cuando se busca un modelado sinusoidal mediante una descomposición atómica parece lógico utilizar la Transformada Discreta de Fourier como herramienta de cálculo de las correlaciones, lo que da lugar a un diccionario de exponenciales complejas. De la misma manera es posible utilizar otras transformadas, como Transformada de Coseno Discreta (Discrete Cosine Transform, DCT), paquetes de cosenos o paquetes de senos.

Cuando se desea implementar un modelado de transitorios, las transformadas anteriores no son las más idoneas. En su lugar, es mejor utilizar transformadas como la Transformada Ondeleta o paquetes de ondeletas, que obtiene mejores resultados cuando se busca representar componentes transitorios en la señales de audio.

3.5. Diccionarios Mixtos

Esta unión de diccionarios se ha utilizado con el objetivo de definir un diccionario generalista que incluya el máximo número de comportamientos tiempo-frecuencia para adaptarse a la señal de entrada [Gribonval (2001)]. Así mismo siempre que se propone el uso de un diccionario mixto es necesario reducir al máximo los requisitos de memoria y complejidad de la descomposición atómica. El uso de un diccionario de este tipo es muy restringido, debido a la alta complejidad asociada al cálculo de la descomposición.

3.6. Descomposiciones atómicas

La representación de señales en átomos tiempo-frecuencia es un tema de interés desde la introducción por Gabor en los años 40 [Hill (2007)]. La premisa fundamental de los modelos atómicos es que la señal se puede descomponer en funciones elementales localizadas en tiempo-frecuencia. La descomposición de una señal en funciones presenta un gran número de soluciones lo cual, en la mayoría de los casos, resulta complejo seleccionar la mejor representación.

Si una señal discreta $x[n]$ se descompone en un conjunto de funciones $g_m[n]$ la señal $x[n]$ queda de la forma,

$$x[n] = \sum_{i=1}^I \alpha_i \cdot g_i[n] \quad (6)$$

donde α_i son los coeficientes de ponderación asociados a cada función $g_i[n]$ y el número de funciones es I . Para el procesamiento digital de la señal, la longitud de la señal $x[n]$ debe ser finita, de valor N . Para la mayoría de las aplicaciones las descomposiciones se basan en transformadas, como las de Fourier u ondeleta. En este caso las funciones forman una base y el número de funciones para descomponer la señal es $I = N$. Una

forma de representar el método de cálculo de estas transformadas es como se muestra a continuación:

$$\vec{x} = D\alpha \quad (7)$$

donde x es un vector columna ($N \times 1$), que representa la señal α un vector columna ($N \times 1$) de coeficientes y $D = [g_1 g_2 \dots g_m \dots g_N]$ es una matriz ($N \times N$), cuyas columnas son los vectores columna \vec{g}_m de funciones. El cálculo de los coeficientes, cuando D es invertible, está definido por la ecuación (8):

$$\vec{\alpha} = D^{-1}\vec{x} \quad (8)$$

Cuando las funciones en D forman una base ortonormal se cumple $D^{-1} = D^H$, donde el superíndice H denota el hermitiano, es decir el conjugado transpuesto. En este caso los coeficientes individuales se calculan como,

$$\alpha_m = g_m^H x = \langle x, g_m \rangle \quad (9)$$

Debido a esta fácil manera de calcular la descomposición, las transformadas han sido ampliamente usadas en el procesamiento de señales. Para aplicaciones de análisis y codificación es importante aproximar la señal a partir de un número muy reducido de funciones que representen las características de la señal a procesar. Por tal motivo es necesario trabajar con transformadas para determinar aquella base que mejor represente la señal.

Cuando el número de funciones en el diccionario D es $M > N$, esto es, una matriz ($N \times M$). En este caso ya no es posible el uso de una matriz inversa del diccionario D^{-1} aún cuando el diccionario debe estar formado por funciones linealmente independientes

entre sí.

El principal problema de las descomposiciones atómicas es la alta complejidad asociada a su cálculo. Un aspecto importante a considerar en el cálculo de las descomposición atómica de señales es la necesaria limitación del tamaño del diccionario. Así conociendo la señal a procesar, se utiliza un diccionario u otro, con el propósito de extraer las características de la señal en un tiempo de computación razonable.

A continuación se describen los métodos más usados en la descomposición atómica, ventajas e inconvenientes.

3.7. Métodos de cálculo

Los diferentes métodos para la selección de funciones elementales o átomos que mejor aproximan la señal y calcular los coeficientes de éstas, con el fin de obtener una descomposición atómica, se puede dividir en dos clases principales.

Métodos iterativos. Para la aproximación de una señal, esta puede ser siguiendo un proceso iterativo para la selección de las formas de onda elementales o átomos óptimos, de tal manera que en cada iteración se seleccione el átomo y su coeficiente o factor de ponderación. Se pueden encontrar en la literatura varios ejemplos de esta clase de métodos, donde la mayoría son derivados del algoritmo conocido como *Matching Pursuit* (MP) [Mallat y Zhang (1993)]. El uso de este algoritmo ha aumentado en diversas aplicaciones debido a que permite el cálculo de una descomposición atómica con una complejidad razonable y resultados satisfactorios.

Métodos paralelos. En este caso, no se restringe la búsqueda de los átomos a seleccionar un átomo en cada iteración esto es, se optimiza M el conjunto de los átomos seleccionados de forma simultánea, conduciendo, por tanto a soluciones óptimas.

3.8. Métodos iterativos

Los métodos iterativos se caracterizan por obtener soluciones subóptimas en descomposiciones atómicas de señales y por calcularse mediante métodos que permiten una complejidad reducida. La principal característica de los métodos iterativos es que deciden la elección de uno o varios átomos (y sus coeficientes asociados) en cada iteración, manteniendo estos valores fijos para las siguientes iteraciones. Esta premisa permite reducir la complejidad, porque en el cálculo sólo se optimiza el átomo (o los átomos) a elegir en la iteración actual. El primer método iterativo para la obtención de descomposiciones atómicas encontrando en la bibliografía es *matching pursuit* [Mallat y Zhang (1993)].

3.9. Métodos paralelos

El nombre de métodos paralelos es debido a que en el cálculo de la solución se determina también el valor de los coeficientes.

3.9.1. Método de tramas

El método de tramas *Method of Frames*, (MOF) [Daubechies (1988)] escoge entre todas las soluciones, aquella en la que los coeficientes en su conjunto tienen norma l^2 mínima, es decir, en la que los coeficientes tienen energía mínima. La notación matricial se expresa como:

$$\min \|\alpha\|^2 \tag{10}$$

para la ecuación (7). En esta expresión α es única y es la solución de longitud mínima. Por lo tanto desde el punto de vista geométrico, de todas las soluciones posibles, que pertenecen al espacio multidimensional que representan los átomos del diccionario, por

lo tanto la solución adoptada por el método MOF es aquella que esta más cerca del origen.

3.10. Basis pursuit

La definición del método Basis Pursuit (BP) [Chen y Wigger (1995)] está basada en el MOF, en el sentido de que la solución se basa también en minimizar la norma de los coeficientes, donde la norma utilizada es el valor absoluto de la amplitud o norma l^1 , esto es

$$\min \|\alpha\| \quad (11)$$

para la ecuación (7). Este método resuelve el principal problema del MOF, al cambiar la norma. Ahora los coeficientes con valores pequeños son descartados en la solución, quedando como solución para el caso de la señal formada por un sólo átomo, el mismo átomo. El principal inconveniente del BP radica en que los algoritmos a utilizar para calcular la descomposición son excesivamente complejos, sobre todo para aplicaciones de compresión. Del método anterior en muchas aplicaciones es necesario encontrar una solución que describa la señal por completo, a partir de los átomos del diccionario, sin embargo en muchas ocasiones es suficiente con que se aproxime la señal de entrada como se muestra en la siguiente ecuación:

$$x[n] \approx \sum_{i=1}^I \alpha_i \cdot g_i[n] \quad (12)$$

o en notación matricial

$$\vec{x} \approx D\vec{\alpha}. \quad (13)$$

Capítulo 4

Algoritmo Matching Pursuit

4.1. Introducción

En este capítulo se presenta el algoritmo *Matching Pursuit* (MP) como un algoritmo iterativo el cual descompone una señal en una expansión lineal de formas de onda que pertenecen a un diccionario redundante de funciones. También se describen los diferentes tipos de diccionarios.

4.2. Matching Pursuit

El algoritmo Matching Pursuit [Mallat y Zhang (1993)] es un algoritmo iterativo codicioso o *greedy* el cual descompone una señal $x(t)$ en una combinación lineal de formas de onda y un término residual. Las formas de onda utilizadas en la descomposición pertenecen a un diccionario redundante de funciones y se les conoce comunmente con el nombre de átomos [Mallat y Zhang (1993); Grinbonval y Bacry (2003)]. Matching Pursuit es parte de una clase de algoritmos de análisis de señales conocido como Descomposición Atómica [Krstulovic y Gribonval (2006)] y tiene como objetivo encontrar la descomposición escasa de una señal, es decir, una representación que concentre la mayoría o toda la información de la señal en una combinación lineal de un número reducido de formas de onda elementales (llamadas átomos). La descomposición es realizada por proyección (producto interno) de la señal $x(t)$ sobre un diccionario redundante de funciones $D = g_\gamma(t), \gamma \in \Gamma$ donde $g_\gamma \in \mathcal{L}^2(\mathfrak{R}), \Gamma = \mathbb{R}^+ \times \mathbb{R}^2$ y por la selección de átomos que mejor representen la estructura de la señal. La señal $x(t)$ puede ser reconstruida

a partir de la suma de M átomos y el término residual $R_M(t)$ como se muestra en la siguiente ecuación:

$$x(t) = \sum_{m=1}^M \alpha_m \cdot g_{\gamma_m}(t) + R_M(t) \quad (14)$$

donde $g_{\gamma_m}(t)$ y α_m son el m -ésimo átomo óptimo y el factor de ponderación respectivamente. Debido a que la descomposición MP es un algoritmo iterativo, dos criterios son usados para detener el proceso iterativo. Un criterio consiste en especificar un número de iteraciones M y el otro criterio consiste en determinar el nivel de energía residual. MP es un algoritmo iterativo descendente el cual selecciona el átomo óptimo en cada iteración [E. Ravelli y Daudet (2008)]. Ver Algoritmo 1.

Algorithm 1 Matching Pursuit

input: $x(t)$; $\mathcal{D} = \{g_\gamma(t), \gamma \in \Gamma\}$

output: $\alpha_m, g_{\gamma_m}(t)$

$R = x(t)$

$\alpha_m = 0$

repeat

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

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

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

until Signal to Noise Ratio (SNR) o M atoms have been analyzed.

Suponiendo que H es el espacio de Hilbert, se define un diccionario como una familia $D = (g_\gamma, \gamma \in \Gamma)$ de vectores en H , tales que $\|g_\gamma\| = 1$. Suponga que V es el espacio lineal cerrado del diccionario de vectores. Se dice que un diccionario es completo si y solo si $V = H$. Supongamos que la señal $x \in H$. Por lo tanto es posible calcular la expansión de vectores en D para la mejor aproximación de la estructura interna de la

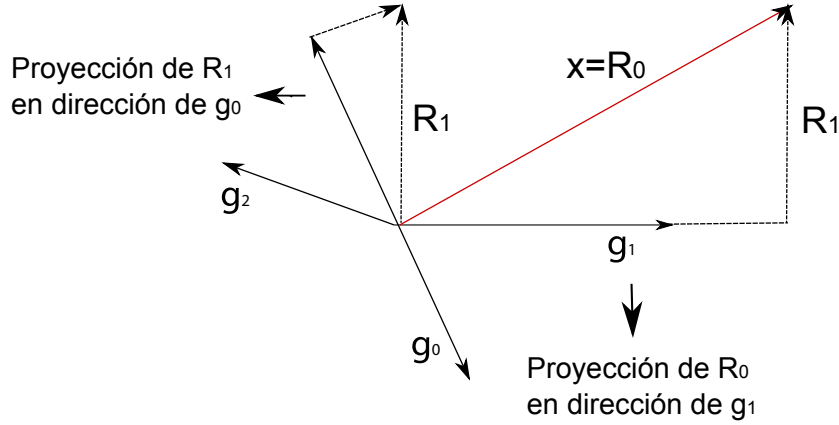


Figura 4: Primera y segunda iteración para la descomposición de la señal x considerando un diccionario con 3 átomos.

señal \vec{x} . Esto es hecho por aproximaciones sucesivas de \vec{x} con proyecciones ortogonales sobre elemento de D . La Figura 4 ilustra la proyecciones de \vec{x} en \vec{g}_0 y en \vec{g}_1

Sea $g_{\gamma_0} \in D$. El vector \vec{x} puede ser descompuesto como se muestra a continuación :

$$\vec{x} = \langle x, g_{\gamma_0} \rangle g_{\gamma_0} + \vec{R}x \quad (15)$$

donde $\vec{R}x$ es el residual después de aproximar \vec{x} en la dirección de g_{γ_0} . Claramente g_{γ_0} es ortogonal a $\vec{R}x$ por lo tanto se cumple la siguiente expresión:

$$\|\vec{x}\|^2 = |\langle x, g_{\gamma_0} \rangle|^2 + \|\vec{R}x\|^2 \quad (16)$$

para minimizar $\|\vec{R}x\|$, debemos seleccionar $g_{\gamma_0} \in D$ tal que $|\langle x, g_{\gamma_0} \rangle|$ es máximo. En algunos casos es posible encontrar un vector g_{γ_0} que es mejor tal que:

$$|\langle x, g_{\gamma_0} \rangle| \geq \alpha \sup_{\gamma \in \Gamma} |\langle x, g_{\gamma} \rangle| \quad (17)$$

donde α es un factor de ponderación que satisface $0 < \alpha \leq 1$. Es importante enfatizar que la selección de un vector g_{γ_0} no es aleatoria. Explicando por inducción como el

algoritmo MP es realizado se tiene lo siguiente. Sea $R^0x = \vec{x}$, suponemos que hemos calculado el residual n-ésimo $R^n x$, para $n \geq 0$. Se selecciona un elemento $g_{\gamma_n} \in D$ el cual se aproxima al residual $R^n x$, esto es:

$$|\langle R^n x, g_{\gamma_n} \rangle| \geq \alpha \sup_{\gamma \in \Gamma} |\langle R^n x, g_\gamma \rangle|. \quad (18)$$

El residual $R^n x$ es descompuesto como se muestra en la ecuación (19):

$$R^n x = |\langle R^n x, g_{\gamma_n} \rangle| g_{\gamma_n}^{\vec{}} + R^{n+1} x, \quad (19)$$

el cual define el residual de orden $n + 1$ por lo tanto $R^{n+1} x$ es ortogonal a $g_{\gamma_n}^{\vec{}}$.

$$\|R^n x\|^2 = |\langle R^n x, g_{\gamma_n} \rangle|^2 + \|R^{n+1} x\|^2. \quad (20)$$

Si llevamos esta descomposición al orden m , es posible descomponer x en una sumatoria

$$\vec{x} = \sum_{n=0}^{m-1} (R^n x - R^{n+1} x) + R^m x. \quad (21)$$

Si sustituimos la ecuación (19) en (21) esto produce

$$\vec{x} = \sum_{n=0}^{m-1} |\langle R^n x, g_{\gamma_n} \rangle| g_{\gamma_n}^{\vec{}} + R^m x. \quad (22)$$

Similarmente $\|\vec{x}\|^2$ es descompuesto en una sumatoria como

$$\|x\|^2 = \sum_{n=0}^{m-1} (\|R^n x\|^2 - \|R^{n+1} x\|^2) + \|R^m x\|^2 \quad (23)$$

Si sustituimos la ecuación (20) en (23) tenemos

$$\|\vec{x}\|^2 = \sum_{n=0}^{m-1} |\langle R^n x, g_{\gamma_n} \rangle|^2 + \|R^m x\|^2. \quad (24)$$

El vector original \vec{x} es descompuesto en una suma de elementos del diccionario que son seleccionados como mejores aproximaciones de sus residuales.

4.3. Diccionarios tiempo-frecuencia

Para el desarrollo de un modelo para la descomposición de señales, los átomos óptimos son seleccionados de forma que correspondan con las características básicas de la señal. Este enfoque es recomendado especialmente para aplicaciones de análisis y codificación de señales debido a que cada átomo puede describir un comportamiento concreto de la señal de entrada. Por lo tanto si el diseño del diccionario es paramétrico, esto es, con parámetros con significado físico, tales como localización temporal, frecuencia de modulación o escala, la relación entre estos parámetros y las características propias de la señal puede ser muy directa.

Las representaciones tiempo-frecuencia son imprescindibles en el procesamiento de señales de audio. En tales diccionarios los átomos tiempo-frecuencia juegan un papel importante en el análisis, síntesis y modificación de señales en algoritmos tales como Matching Pursuit así como en codificación y compresión de señales. La elección de un diccionario en particular depende de la aplicación en cuestión. Por ejemplo, en el diseño de diccionarios, dos factores merecen cierta consideración para el procesamiento de señales. Esto es, cualquier diccionario debe ser completo para la clase de señales a analizar. Sin embargo, es impráctico tener una alta redundancia total. Así que, es deseable conocer con certeza la estructura o comportamiento de la señal para contar con un tamaño de diccionario que nos garantice el mínimo costo computacional.

Cabe distinguir dos enfoques en el diseño del diccionario. Por un lado, los primeros trabajos en descomposiciones atómicas se inclinaron por utilizar un gran diccionario ge-

neralista que incluyera una gran variedad de comportamientos [Mallat y Zhang (1993)]. Sin embargo esta aproximación no es óptima desde el punto de vista de la complejidad por el gran tamaño del diccionario. Como consecuencia se suele diseñar el diccionario en función de la aplicación para la que se utilice la descomposición.

La descomposición de señales en una familia de funciones que están bien localizadas en tiempo y frecuencia tiene una gran variedad de aplicaciones en el análisis y procesamiento de señales. Dependiendo de la elección de átomos tiempo-frecuencia la descomposición podría presentar diferentes propiedades. La transformada de Fourier y la transformada de ondeleta son ejemplos de descomposición de señales tiempo-frecuencia que han sido ampliamente utilizadas. Para extraer información de señales complejas es necesario adaptar la descomposición tiempo-frecuencia a la estructura particular de la señal.

4.4. Diccionario de Gabor

La selección de un diccionario adecuado juega un papel fundamental en el desempeño del algoritmo MP. Varios tipos de diccionarios tiempo-frecuencia han sido propuestos en la literatura, tales como paquetes de ondeletas, paquetes de coseno, transformada de coseno discreta modificada (MDCT) y diccionarios de Gabor [Wolfe *et al.* (2001)]. Se realizaron pruebas preliminares con los diccionarios anteriormente mencionados y se encontró que las funciones de Gabor pueden modelar muy bien los sonidos cardiacos, por tal motivo consideramos que los átomos de Gabor son adecuados para este propósito. Además, la capacidad de las funciones de Gabor para modelar los sonidos cardiacos ya han sido previamente reportada en la literatura [X. Zhang y Coatrieux (1998)].

Los átomos de Gabor se obtienen por dilatación, traslación y modulación de una

ventana madre $w(t)$ como se muestra en la Figura 5 , la cual toma generalmente valores reales, positivos y tiene norma unitaria y está dada por

$$g_\gamma(t) = \frac{1}{\sqrt{s}} w\left(\frac{t-u}{s}\right) e^{i2\pi\xi(t-u)}, \quad (25)$$

donde w es la ventana gaussiana $w(t) = \sqrt[4]{2}e^{-\pi t^2}$, s es usada para controlar la longitud de la envolvente de la forma de onda, el desplazamiento en el tiempo u es usado para especificar la localización temporal del átomo y ξ es la frecuencia de modulación. Definimos entonces $\gamma_m = (s_m, u_m, \xi_m)$, donde el índice γ_m es un elemento del conjunto Γ .

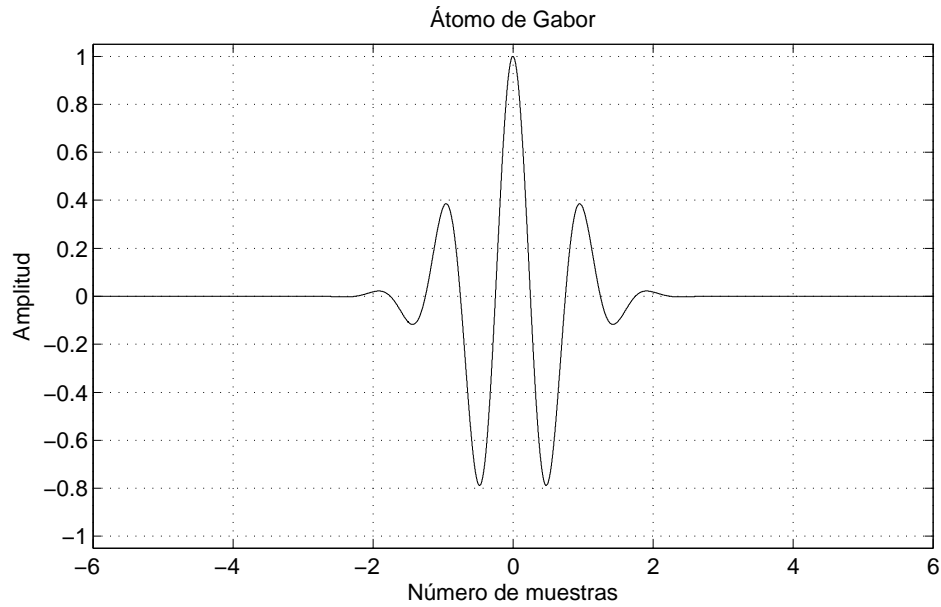


Figura 5: Forma de onda de Gabor.

4.5. Diccionario MDCT

La Transformada Discreta del Coseno Modificada, también conocida por sus siglas en inglés como MDCT (Modified Discrete Cosine Transform) propuesta en [b] y más desarrollada en [b]. Esta transformada es utilizada en el códec AAC [b].

Esta transformada es del tipo cancelación de Aliasing en el Dominio del Tiempo (TDAC por sus siglas en inglés), la cual utiliza bloques con traslape para poder reconstruir la señal [Zhang *et al.* (2013)]. En la ecuación (26) muestra como calcular la MDCT del i -ésimo bloque de datos x de tamaño N

$$X_i[k] = \sum_{n=0}^{N-1} w_a^i[n] x_i[n] \cos\left[\frac{\pi}{2N} \left(2N + 1 + \frac{N}{2}\right) (2k + 1)\right] \quad (26)$$

para $k = 0, \dots, \frac{N}{2} - 1$. w_a^i es la ventana de análisis para el i -ésimo bloque.

Como es denotar, el número de coeficientes de la MDCT es la mitad del tamaño del bloque. La ecuación para calcular la transformada inversa (IMDCT) es la siguiente:

$$x[n] = w_s^i \frac{4}{N} \sum_{k=0}^{\frac{N}{2}-1} X_i[k] \cos\left[\frac{\pi}{2N} \left(2n + 1 + \frac{N}{2}\right) (2k + 1)\right] \quad (27)$$

para $n = 0, \dots, N - 1$. Donde w_s^i en la ecuación (27) es la ventana de síntesis para el i -ésimo bloque.

Para poder reconstruir la señal original es necesario que los bloques este trasladados en un 50% y posteriormente sumar las señales generadas por la IMDCT. Esta técnica es conocida como cancelación de aliasing en el dominio del tiempo.

4.6. Principio de incertidumbre de Heisenberg

Los átomos de Gabor son contruidos por traslación en el tiempo y frecuencia de una ventana g :

$$g_{u,\xi}(t) = g(t - u)e^{i\xi t}. \quad (28)$$

La energía de $g_{u,\xi}$ está concentrada en en la vecindad de u sobre un intervalo de σ_t medida de la desviación estándar de $|g|^2$. La transformada de Fourier \hat{g} es trasladada por ξ

$$g_{u,\xi}(\omega) = g(\omega - \xi)e^{-iu(\omega - \xi)}. \quad (29)$$

La energía de $\hat{g}_{u,\xi}$ está por lo tanto localizad acerca de la frecuencia ξ sobre un intervalo de tamaño σ_ω . En el plano tiempo-frecuencia (t, ω) la energía del átomo $g_{u,\xi}$, es simbólicamente representada por el rectángulo de Heisenberg. Este rectángulo está centrado en (u, ξ) y tiene una duración de tiempo σ_t y un ancho en frecuencia σ_ω . El principio de incertidumbre prueba que su área satisface la desigualdad (30):

$$\sigma_t \sigma_\omega \geq \frac{1}{2}. \quad (30)$$

Esta área es mínima cuando g es Gaussiana, en cuyo caso los átomos $g_{u,\xi}$ son llamados funciones de Gabor [Mallat (2009)].

¿ Es posible contar con una señal $x(t)$ cuya energía está bien localizada en el tiempo y que además la transformada de Fourier $X(\omega)$ concentre la energía en cierta region de la frecuencia? Para reducir el tiempo de duración de una señal es posible escalarla en el tiempo con $s < 1$ de la siguiente manera

$$x_s(t) = \frac{1}{\sqrt{s}} f\left(\frac{t}{s}\right) \quad (31)$$

manteniendo constante la energía de la señal, esto es:

$$\|x_s\|^2 = \|x\|^2 \quad (32)$$

La transformada de Fourier está dada por $X_s(\omega) = \sqrt{s}X(s\omega)$, lo cual provoca una dilatación o expansión $\frac{1}{s}$, i.e, ganamos localización en tiempo pero perdemos localización en frecuencia. La concentración de energía está restringida por el principio de Heisenberg.

La transformada de Fourier definida por la correlación de una señal f con cada átomo $g_{u,\xi}$:

$$Sf(u, \xi) = \int_{-\infty}^{\infty} f(t)g_{u,\xi}^*(t)dt = \int_{-\infty}^{\infty} f(t)g(t-u)e^{-i\xi t}dt \quad (33)$$

es una integral de Fourier que está localizada en la vecindad de u por la ventana $g(t-u)$. Este tipo de integral puede ser escrita como una integral en frecuencia aplicando la expresión de Parseval quedando entonces:

$$Sf(u, \xi) = \int_{-\infty}^{\infty} \hat{f}(\omega)\hat{g}_{u,\xi}^*(\omega)d\omega \quad (34)$$

La transformada $Sf(u, \xi)$ en consecuencia depende únicamente de los valores $f(t)$ y $\hat{f}(\omega)$ en la vecindad del tiempo y frecuencia donde la energía de $g_{u,\xi}$ y $\hat{g}_{u,\xi}$ está concentrada. Gabor interpreta esto como un (quantum of information) en el rectángulo tiempo-frecuencia.

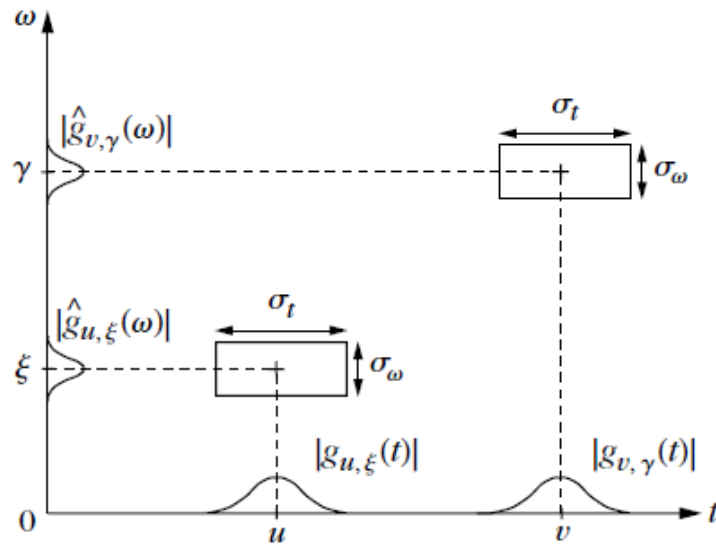


Figura 6: Cajas tiempo-frecuencia ("Heisenberg rectangles") representación de energía de 2 átomos de Gabor.

4.7. MPTK (Matching Pursuit ToolKit)

MPTK (Matching Pursuit ToolKit) [Krstulovic y Gribonval (2006)] permite una implementación rápida del algoritmo Matching Pursuit para la descomposición escasa de señales. MPTK está constituido por una librería en C++ y algunos scripts para ejecutar y graficar en matlab.

Se realizaron pruebas usando MPTK para la creación de diccionarios de Gabor y MDCT de lo cual se encontró que los diccionarios de Gabor presentan una mejor representación de los sonidos cardiacos.

Capítulo 5

Algoritmo de segmentación de sonidos cardiacos

5.1. Introducción

El análisis tiempo-frecuencia aplicado en la identificación y segmentación de sonidos cardiacos ha demostrado ser una herramienta que dota de una gran cantidad de información sobre el comportamiento de señales de fonocardiograma. El proceso de segmentación e identificación de señales de fonocardiograma se ha incrementado considerablemente como un procedimiento previo a la compresión y transmisión de señales de fonocardiograma. El análisis del espectro es una herramienta que proporciona información útil de la localización y características de las señales de fonocardiograma [Boutana *et al.* (2011)]. Otras técnicas como el análisis mediante ondeletas ha sido ampliamente utilizado para la segmentación y discriminación de sonidos cardiacos normales de sonidos cardiacos anormales [Kumar *et al.* (2006)].

En esta investigación se propone un algoritmo de segmentación de sonidos cardiacos normales y anormales, el cual permite la identificación de onsets y offsets de cada ataque.

El algoritmo de segmentación MP-IOI consiste en reconstruir una señal de audio cardiaco con el algoritmo matching pursuit. El diccionario de Gabor utilizado está formado por 3 bloques [128 64 128], [256 128 256], [512 256 512]. En los sonidos cardiacos S1 y S2 (corazón sano), la duración promedio de dichos sonidos es de aproximadamente 0.1 y .08 segundos respectivamente. Esto es, para una frecuencia de muestreo de 11025 Hz entonces tenemos que la duración de los sonidos cardiacos (en número de muestras)

Tabla II: Comparación de diferentes diccionarios para la segmentación de sonidos cardiacos de la señal normal S1.

onset/s.o	offset/s.o	onset/MP	offset/MP	iteración	Longitud de átomo
12	12	7	9	350	256,512,1024
12	12	9	9	350	128, 256,512,1024
12	12	12	0	350	256
12	12	12	12	350	128,256, 512
12	12	12	12	350	64,128,256,512

es de aproximadamente de 1103 y 882 muestras respectivamente, es decir, la duración tanto del sonido cardiaco S1 y S2 está por arriba de la máxima longitud de átomo, permitiendo así una reconstrucción más precisa de cada ataque (únicamente de cada ataque, no de la señal en sí), con el mínimo número de átomos totales en del diccionario.

La tabla II , muestra el número de onset y offset detectados, con diferentes bloques, empleando diccionarios de Gabor.

De la tabla II, es importante observar, que usando diccionarios de Gabor con longitud de átomo (128, 256, 512) y (64, 128, 256, 512) se obtiene, el 100 % de onset y offset para un número de iteraciones fijo. Sí se excede, la longitud de los sonidos S1 y S2 (1103 y 882) no se obtiene el 100 % de onset y offset respectivamente (ver segunda fila y tercera fila). De los 350 átomos seleccionados, considerando el diccionario de Gabor con 4 longitudes de átomo, donde hay 100 % de detección en Onset y offset, únicamente 21 átomos con longitud de 64 fueron seleccionados. Si eliminamos dicho bloque ([64 32

64]) entonces tenemos un diccionario de Gabor con longitudes de átomo (128, 256, 512) (ver fila 4), por lo tanto, el número de átomos totales del diccionario se reduce de 200 000 a 150 000 átomos aproximadamente. Manteniendo así, el número de iteraciones, el 100% de detección y disminuyendo el número de átomos totales del diccionario.

La Figura 7 muestra 2 ciclos cardiacos para una señal de corazón sano (Normal Split S1). En la Figura 7 también se muestra los onsets (inicio de ataque) y offset (parte final del ataque) los cuales fueron identificados escuchando el inicio y fin de cada sonido.

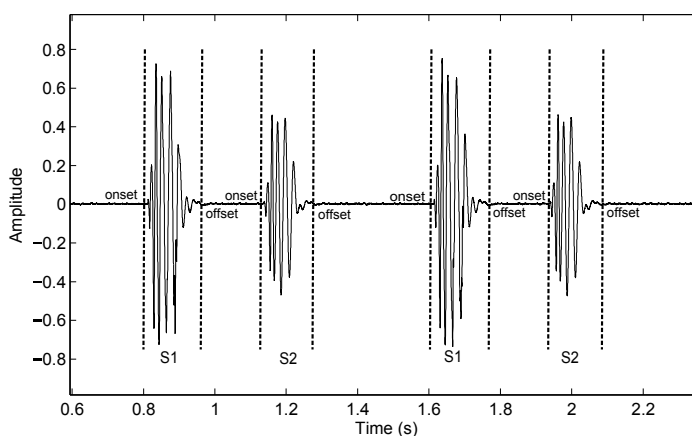


Figura 7: Sección de señal de corazón normal mostrando onset y offset para cada ataque.

La Figura 8 por su parte muestra la descomposición atómica de la fig. 7 usando el algoritmo MP y un diccionario de Gabor con longitudes de átomo (128, 256, 512).

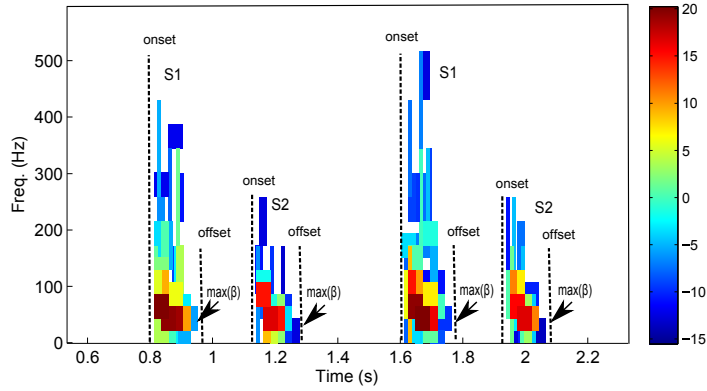


Figura 8: Descomposición atómica en el plano tiempo-frecuencia para la señal de la Figura 7.

En la Figura 8 se puede apreciar la distribución de átomos óptimos en el plano tiempo-frecuencia los cuales se concentran en cada uno de los ataques de la Figura 7.

Una vez que se conoce el diccionario adecuado para la descomposición de sonidos cardiaco se procede con la selección de una señal para la descomposición usando el algoritmo MP.

La descomposición MP produce un conjunto de átomos de Gabor que son óptimos para la representación de la señal previamente analizada. Los átomos óptimos son sumados para reconstruir la señal. El primer paso del algoritmo de segmentación consiste en seleccionar la posición u_m para todos los átomos (tiempo de retraso del m -ésimo átomo con respecto al inicio de la señal) posteriormente las posiciones previamente seleccionadas son ordenadas en forma ascendente. Donde $\{s_n, u_n, \xi_n, \alpha_n, 1 \leq n \leq M\}$ son los parámetros de los átomos óptimos después de ser previamente ordenados con respecto al tiempo de desplazamiento.

El siguiente paso consiste en calcular la distancia entre átomos sucesivos, esto es $\delta_n = u_{n+1} - u_n$. Ver Figura. 9. Por lo tanto se busca únicamente los δ_n que satisfagan

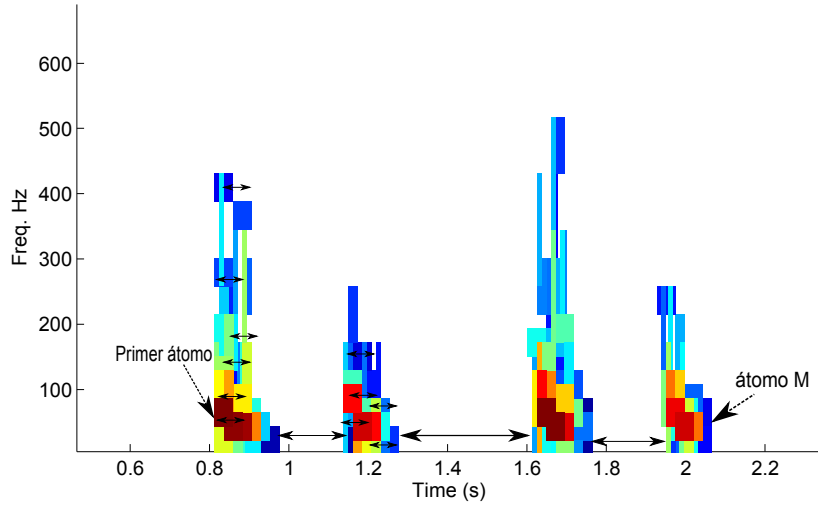


Figura 9: Distancia entre átomos sucesivos en el plano tiempo-frecuencia

la condición $\delta_n \geq \epsilon$ donde el umbral ϵ es el mínimo intervalo permitido entre dos eventos en un ciclo cardiaco (normal o anormal). Ver Fig.10.

Sí $\delta_n \geq \epsilon$ entonces la posición del átomo (u_{n+1}) es considerado como la posición de un onset. Para detectar el primer onset de la señal analizada se considera la posición u_n donde $n = 1$ esto es, la posición del primer átomo óptimo. Ver Fig. 15.

Para detectar los offsets el procedimiento es más elaborado. Una vez que se tienen localizados el conjunto de los onsets detectados en la señal de sonido cardiaco usando el método descrito anteriormente, se procede a calcular las posiciones de los offsets de cada ataque. Se define un nuevo parámetro β_n para cada átomo g_{γ_m} el cual es la suma de la posición en el tiempo y la duración de cada átomo óptimo, i.e. $\beta_n = u_n + longitud(g_{\gamma_m})$.

El siguiente paso consiste en seleccionar únicamente aquellos átomos entre dos onsets sucesivos, esto es $\hat{\beta}_n = \{\beta_n \mid o_i < \beta_n < o_{i+1}\}$ y el resto de los átomos son descartados. Ver Fig. 12.

Finalmente, únicamente $\theta_i = \max(\hat{\beta}_n)$ entre dos onsets sucesivos es seleccionado como un offset valido, ver Fig. 13 y Fig. 14.

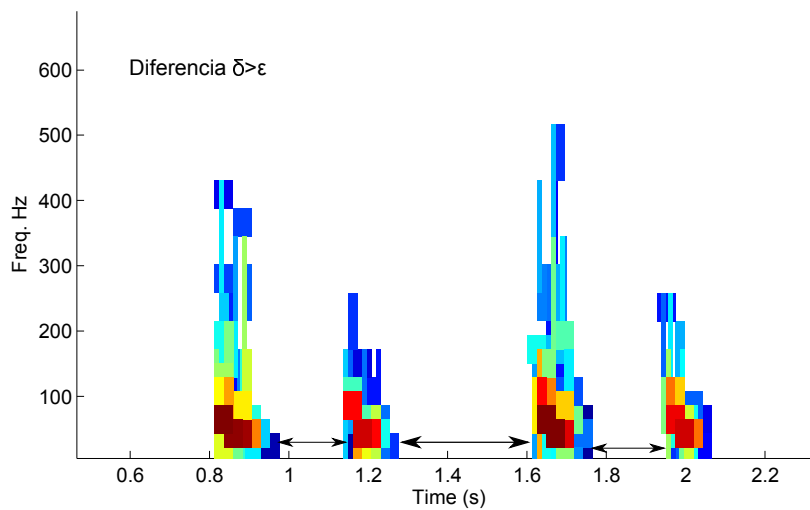


Figura 10: Selección de distancias entre átomos óptimos que cumplen el criterio de selección.

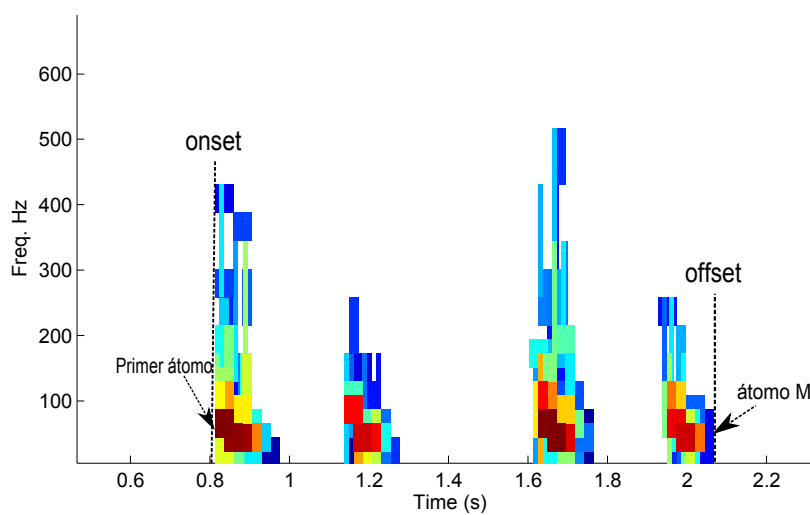


Figura 11: Identificación de la posición u del primer átomo en el plano tiempo-frecuencia.

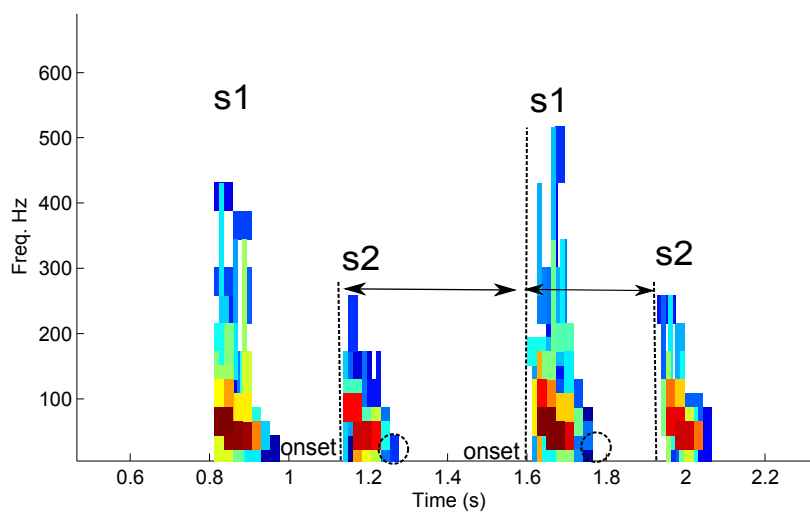


Figura 12: Identificación del átomo entre onsets sucesivos.

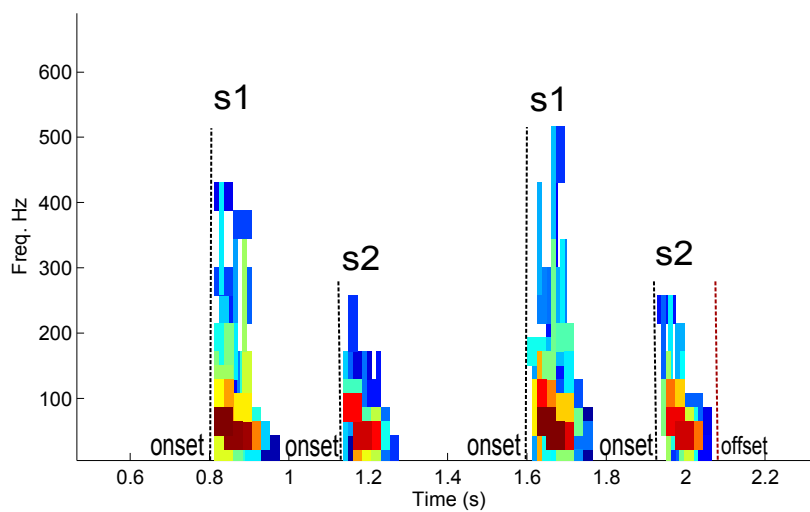


Figura 13: Detección de onsets para cada ataque

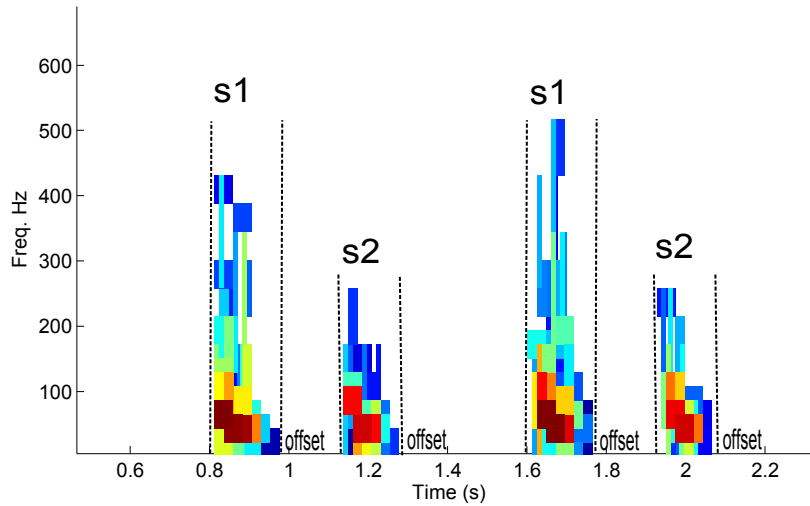


Figura 14: Detección de offset en cada ataque de la señal de sonido cardiaco normal (S1,S2).

El primer onset y el último offset de la señal analizada corresponden a la posición $o_1 = u_{n|n=1}$ del primer átomo y $\theta_I = \max(\beta_n)$ de todos los átomos respectivamente. Ver Algoritmo 2.

5.2. Parámetros del algoritmo

Se aplicó el algoritmo MP para descomponer las señales de sonido cardiaco en una serie de átomos tiempo-frecuencia usando un diccionario de Gabor. Los parámetros del diccionario consisten de 3 bloques diferentes de Gabor donde cada bloque está formado por [longitud de la ventana, desplazamiento de la ventana, tamaño de la [FFT] (en número de muestras)]. El diccionario implementado lo forman bloques con las siguientes características: [128, 32, 128], [256, 32, 256] y [512, 32, 512]. Los parámetros mencionados anteriormente fueron seleccionados de acuerdo al siguiente razonamiento: Por conveniencia computacional únicamente se usaron átomos con longitudes de potencia de 2. Se encontró que usando átomos de longitud mayor a 512 se excede la duración

Algorithm 2 Segmentation

input: $book = \{g_{\gamma_m}, \gamma \in \Gamma\}$
output: $o_i = 0, \theta_i = 0$

$$o_1 = u_{n|n=1}$$

$$\beta_n = u_n + \text{length}(g_{\gamma_n})$$

repeat

$$\delta_n = u_{n+1} - u_n$$

if $\delta_n \geq \epsilon$ **then**

$$o_i = u_{n+1}$$

end if
until $M - 1$ atoms have been analyzed.

repeat
if $o_i < \beta_n < o_{i+1}$ **then**

$$\hat{\beta}_n = \beta_n$$

$$\theta_i = \max(\hat{\beta}_n)$$

end if
until $I - 1$ onsets have been analyzed.

$$\theta_I = \max(\beta)$$

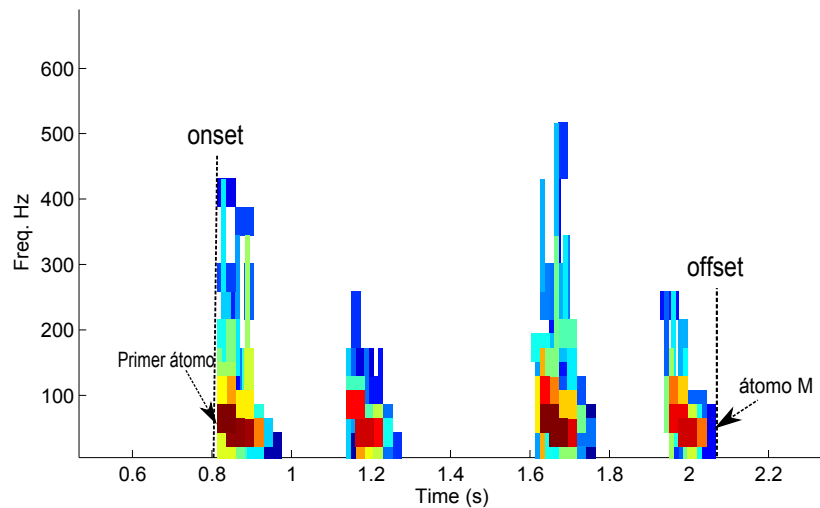


Figura 15: Identificación del último átomo en el plano tiempo-frecuencia.

del sonido cardiaco más corto permitiendo la reconstrucción de parte de la señal no deseada, lo cual contribuye a una imprecisa detección de onsets y offsets. Por otra parte se observó que usando átomos con longitudes por debajo de 128 el desempeño del algoritmo no mejora. La selección del umbral ϵ se realizó en base a los sonidos cardiacos disponibles considerando el mínimo intervalo de tiempo entre eventos consecutivos de sonidos cardiacos.

5.3. Evaluación

El umbral ϵ fue fijado a .0454 seg. La figura 16 muestra un ejemplo del caso del umbral mínimo requerido para la segmentacion de sonidos cardiacos usados. La figura 17 por su parte muestra la descomposición de la fig. 16 usando el algoritmo 2.

Se realizaron un número de experimentos numéricos para evaluar el desempeño del algoritmo propuesto para la segmentación de sonidos cardiacos.

El criterio de tolerancia entre el inicio del ataque y el onset detectado así como la parte final del ataque y el offset detectado fue fijado a .02 seg.

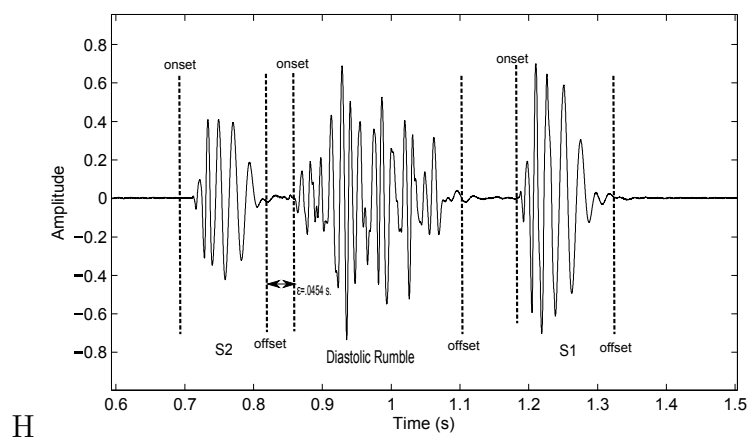


Figura 16: Umbral mínimo de segmentación en soplo diastólico.

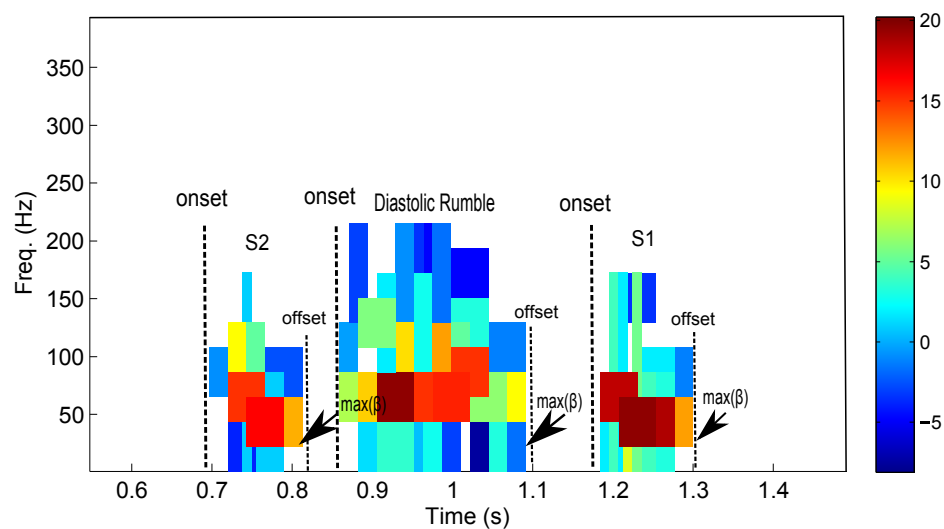


Figura 17: Plano tiempo-frecuencia para la señal de sonido cardiaco con soplo diastólico.

Tabla III: Base de datos de sonidos cardiacos.

Sonidos cardiacos	onset detectado (%)	offset detectado (%)
Normal Split S1	100	100
Normal Split S2	100	91
s3	100	94.4
s4	100	100
Early systolic murmur	100	100
Late systolic murmur	84.6	92.3
Early systolic murmur	100	100
Late systolic murmur	84.6	76.3
Ejection click	100	100
Opening snap	100	100
Diastolic rumble	100	100
total (average)	97.5	96

El desempeño del algoritmo fue probado con diferentes señales de sonido cardiaco. La tabla III muestra el porcentaje de detección (onset y offset) para cada señal empleando el algoritmo 2.

Durante las simulaciones se encontró que para descomponer las señales de sonido cardiaco con una longitud promedio de 6s. es necesario contar con aproximadamente 60 átomos por ciclo para obtener el desempeño mencionado. Se observó de las descomposiciones que para obtener un SNR de al menos 20 *dB* es suficiente para alcanzar una

detección del 96 % o más.

5.4. Número de átomos en el diccionario

Como se describió en el capítulo anterior la selección de un diccionario adecuado es fundamental para obtener una representación escasa eficiente de la señal que cuente con las características de la señal analizada. De la misma manera además de contar con un diccionario que modele adecuadamente las características de la señal en cuestión es importante considerar el número de átomos en el diccionario, buscando siempre encontrar una descomposición adecuada de la señal con el mínimo número de átomos totales en el diccionario propuesto. Esto es, si tenemos un diccionario con las siguientes características [32 16 32] ([window shift ,window length, fft size]) el número de átomos en el diccionario está definido por la ecuación (35) :

$$\text{atomoTotal} = a - \zeta + (a + 1) \cdot \zeta \quad (35)$$

donde a está definido por:

$$a = \lfloor \frac{L}{\zeta} \rfloor - 1 \quad (36)$$

y L y ζ son la longitud de la señal en muestras y *window shift* respectivamente.

Si ahora se considera un conjunto de bloques es decir [32 16 32] [64 32 64] el número de átomos en el diccionario estará dado por la expresión (37):

$$\text{atomoTotal} = \text{atomoTotal}_1 + \text{atomoTotal}_2 \quad (37)$$

donde atomoTotal_1 y atomoTotal_2 hace referencia al número de átomos totales de cada bloque. Por lo tanto el número de átomos en un diccionario formado por n bloques

estará dado por:

$$\text{atomoTotal} = \sum_{i=1}^n \text{atomoTotal}_n \quad (38)$$

De la ecuación (38) se puede inferir que el número de átomos en el diccionario aumentara conforme se agregen bloques al diccionario.

De lo descrito anteriormente el numero de bloques usados en este análisis es fundamental, i.e entre mayor sea el número de bloques generados, mayor será el número de átomos en el diccionario; sin embargo al aumentar el número de átomos el procesamiento de dicho diccionario será mayor.

En la descomposición MP en los sonidos de audio cardiacos es importante buscar la mejor reconstrucción con el mínimo número de átomos para un diccionario con la mínima cantidad de átomos posibles.

5.5. Arquitectura General

El diccionario es implementado como un arreglo de objetos bloque. Cada objeto bloque sabe como modificar los productos internos a través de toda la señal o a través de un soporte de tiempo particular, en una escala particular o para una transformada en particular. Este corresponde por ejemplo a la aplicación de la transformada de Fourier de termino corto (Short Time Fourier Transform, STFT) con una longitud de ventana dada en el caso de usar un diccionario de átomos de Gabor.

Un bloque tiene la capacidad de localizar el máximo producto interno y asociar átomos (como una representación paramétrica o como una forma de onda).

En la figura 18 se muestra la comparación entre un diccionario de Gabor y un diccionario MDCT (Transformada de Coseno Discreto Modificado) con las mismas longitudes

de átomo (en muestras) en ambos diccionarios, esto es, 128 y 16384. Después de realizar la descomposición MP con 350 átomos en una señal de audio cardiaco normal (S1,S2) se observó que el error medio cuadrático normalizado(Normalized Root Mean Square Error, NRMSE) en el diccionario de Gabor tiempo-frecuencia es menor en comparación con el diccionario de MDCT, donde el NRMSE está dado por:

$$NRMSE = \sqrt{\frac{\sum_{n=0}^{N-1} e(n)^2}{\sum_{n=0}^{N-1} x(n)^2}} \quad (39)$$

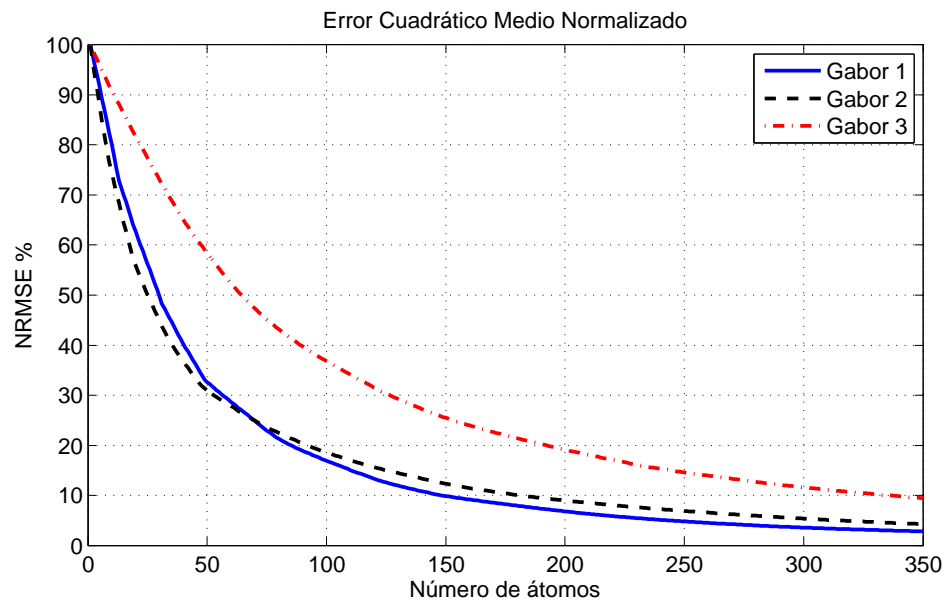


Figura 18: Comparación del NRMSE entre diccionarios de Gabor

Tabla IV: Características de los diccionarios de Gabor usados en la descomposición MP.

Diccionario	Número de átomos	Número de iteraciones	Energía Residual
Gabor 1	742,310	350	1.0664
Gabor 2	197,859	350	2.4447
Gabor 3	1,165,255	350	11.9367

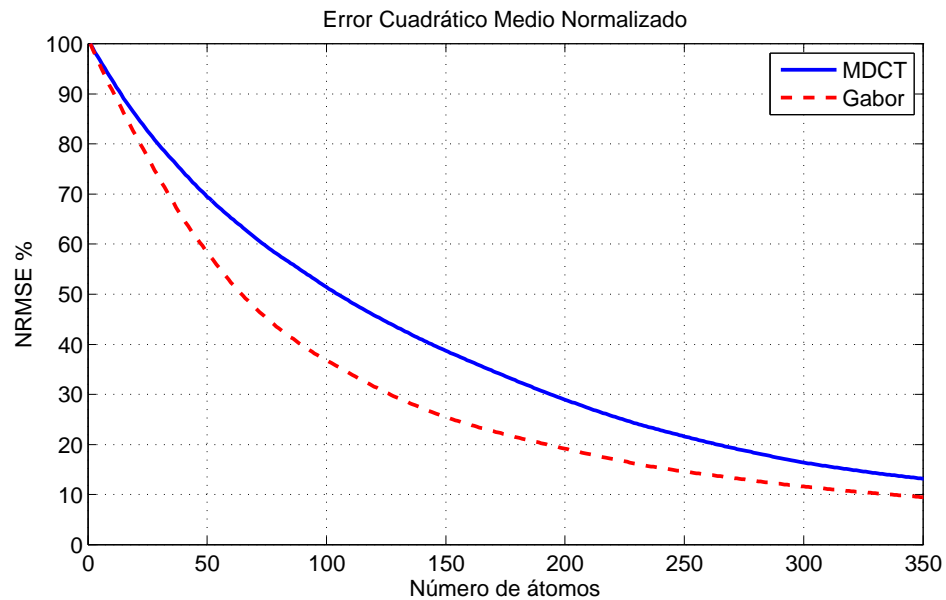


Figura 19: Comparación del NRMSE entre el diccionario de Gabor y MDCT

En la tabla IV se muestra la comparación de 3 diccionarios de Gabor propuesto para la reconstrucción de sonidos de audio cardiaco. Se observa que el diccionario Gabor 2 está formado por 197, 859 átomos y que al aproximar la señal de audio cardiaco normal la energía residual después de la descomposición MP con 350 átomos es de 2.4447. Al comparar este diccionario con el diccionario de Gabor 1, el cual está formado por 742,

310 átomos es posible reducir la cantidad de energía residual con el mismo número de átomos con un diccionario mas grande. Para el diccionario de Gabor 3 se puede observar que el tamaño de diccionario es mucho mayor que ambos diccionarios (Gabor 2 y Gabor1) esto es, 1, 165, 255 átomos, sin embargo el residual es de 11.9367.

Lo anterior permite inferir que es importante conocer el comportamiento de la señal a descomponer para evitar contar con diccionario sobrecompletos donde el número de átomos del diccionario no contribuyan a la reconstrucción de la señal y que además aumenten la cantidad de procesamiento y almacenamiento.

Para obtener un mejor desempeño del algoritmo MP usando diccionarios de Gabor se modificaron los diferentes bloques del diccionario, considerando el NRMSE y el total de átomos en el diccionario. Sin duda es importante describir que los cambios realizados en el tamaño de la FFT, la longitud del átomo y tamaño del traslape (modificación de bloque) aumentan o disminuyen de manera directa el número de átomos en el diccionario.

La descomposición usando el algoritmo MP y diccionarios de Gabor tiempo-frecuencia se describe en la figura 20. Para mostrar esquemáticamente el funcionamiento iterativo del algoritmo MP. Se presenta la señal de audio cardiaco a descomponer en átomos de Gabor, los átomos óptimos, la señal de reconstrucción y el nuevo residual en cada iteración. A continuación se muestra el proceso de descomposición de una señal de sonido de audio cardiaco normal (Normal Split S1) con 3 ciclos $\vec{R}^0 = \vec{x}$, donde \vec{x} es la señal de audio cardiaco a descomponer. El primer paso consiste en seleccionar en la primer iteración $m = 1$ el átomo óptimo, esto es:

$$g_{\gamma_1}^{\vec{}} = \arg \max_{\gamma} |\langle R^0, g_{\gamma} \rangle| \quad (40)$$

Una vez que se conoce el átomo óptimo se calcula el factor de ponderación:

$$\alpha_1 = \langle R^0, g_{\gamma_1} \rangle. \quad (41)$$

Mediante el factor de ponderación y el átomo óptimo $m = 1$ previamente seleccionado es posible reconstruir la señal de la fig. 20 c) a través de $\alpha_1 \cdot g_{\gamma_1}$. Finalmente el nuevo residual fig. 20 d) es calculado usando:

$$\vec{R}^1 = \vec{R}^0 - \alpha_1 \cdot g_{\gamma_1}, \quad (42)$$

En la segunda iteración del algoritmo MP $m = 2$ se observa que la señal a descomponer, fig. 21 a), corresponde al nuevo residual R^1 esto es, fig. 21 d). Nuevamente se busca el átomo tal que $g_{\gamma_2} = \arg \max_{\gamma} = |\langle R^1, g_{\gamma} \rangle|$. El átomo óptimo de la iteración 1 y el nuevo átomo óptimo de iteración 2 se muestran en la fig. 21 y b). Para obtener el factor de ponderación se procede de la misma manera, esto es, $\alpha_2 = \langle R^1, g_{\gamma_2} \rangle$. La señal reconstruida hasta la iteración $m = 2$ se calcula como:

$$x(t) = \sum_{m=1}^{M=2} \alpha_m \cdot g_{\gamma_m}(t) + R_M(t), \quad (43)$$

donde $R_{M=2} = \vec{R}_2 = \vec{R}_1 - \alpha_2 \cdot g_{\gamma_2}$

Para el caso de aproximar la señal de audio cardiaco con 50 átomos ($M = 50$) se observa cierta concentración de átomos óptimos en los ataques de la señal de audio cardiaco (Normal Split S1) y que la señal residual inicia a decrementarse en amplitud. ver Fig. 22.

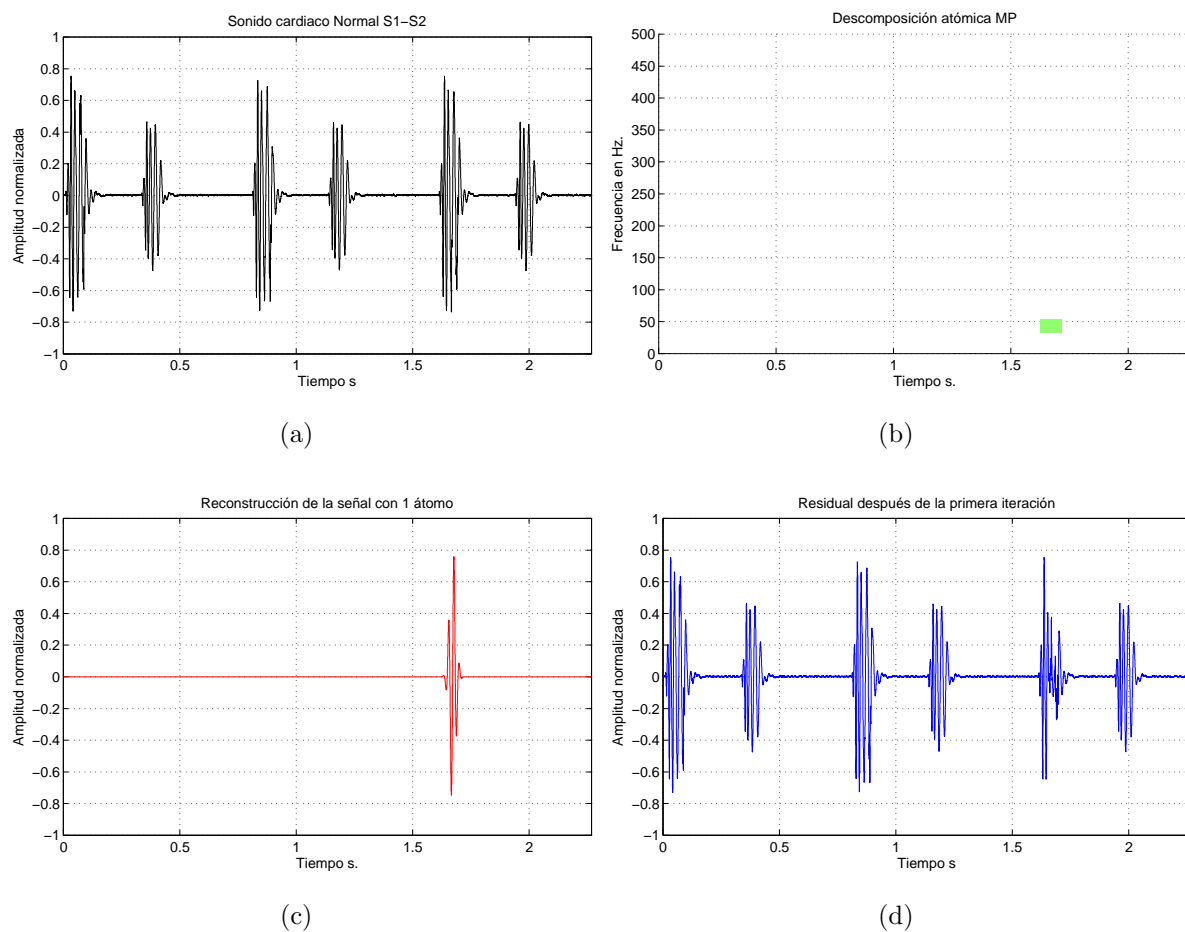


Figura 20: a) sección de un sonido cardiaco normal a descomponer usando MP, b) átomo óptimo seleccionado en la primera iteración del algoritmo MP, c) Reconstrucción de la señal con el átomo previamente seleccionado, d) Señal residual después de la primera iteración.

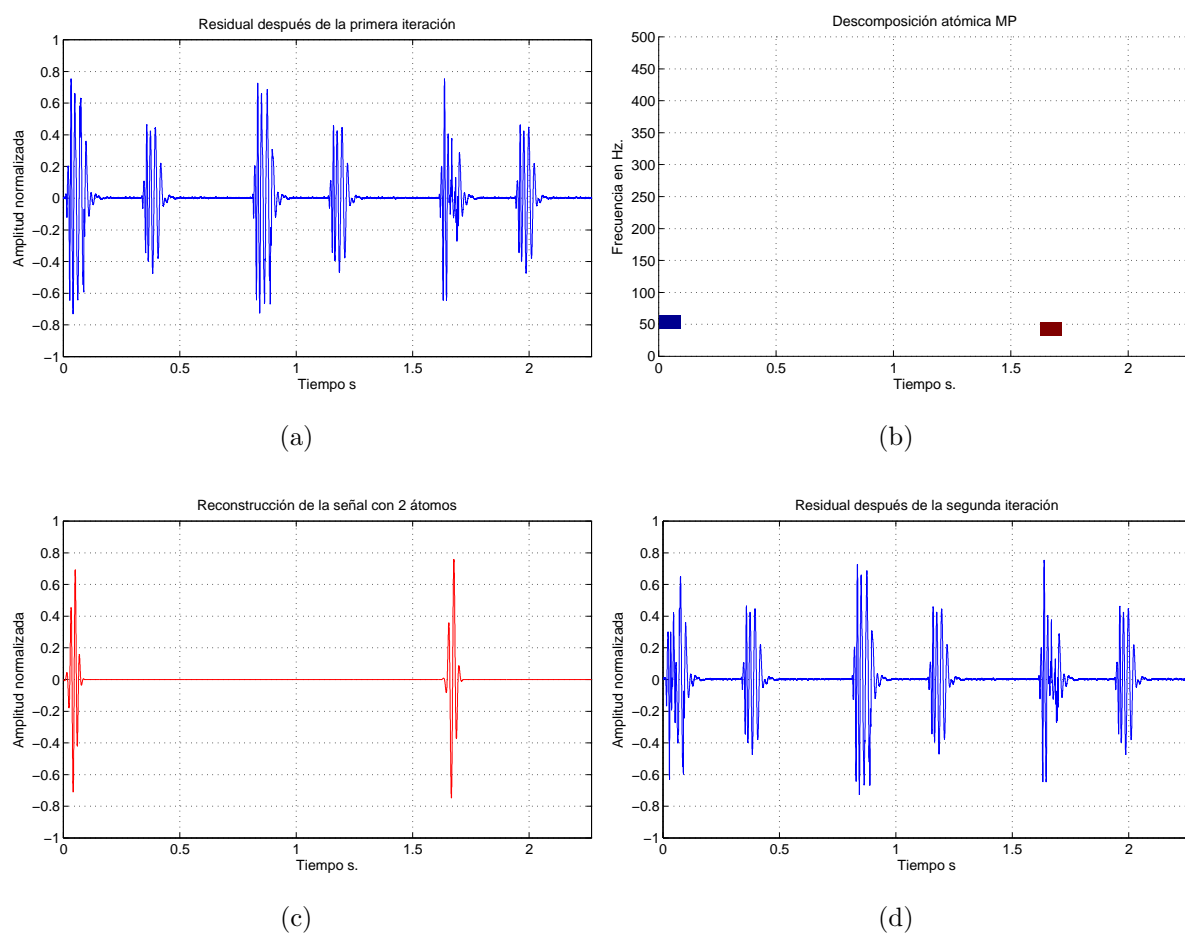


Figura 21: a) Señal a descomponer en la segunda iteración (señal residual)., b) Segundo átomo óptimo seleccionado usando MP, c) Reconstrucción de la señal con los dos átomos previamente seleccionados., d) Señal residual después de la segunda iteración.

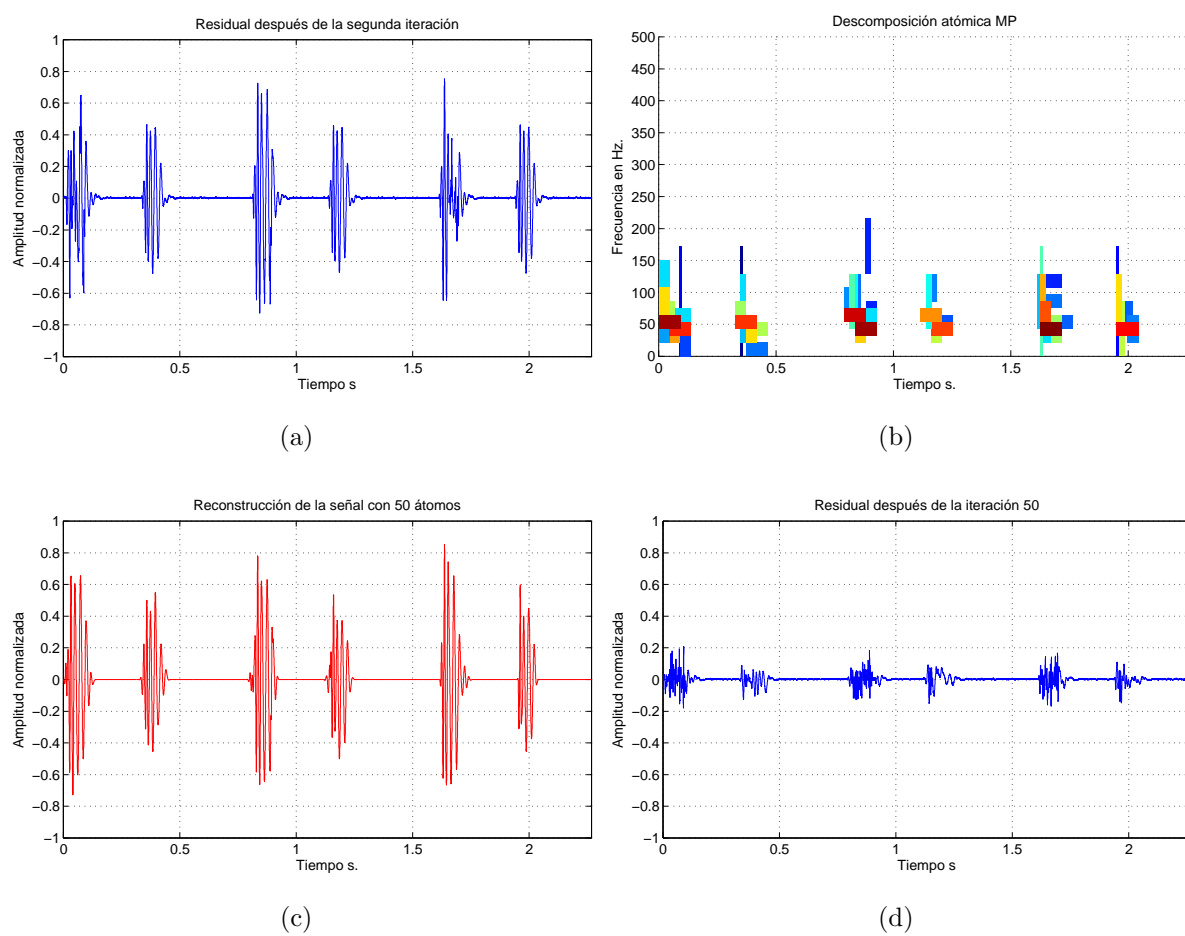


Figura 22: a) Señal a descomponer en la segunda iteración (señal residual), b) 50 átomos seleccionado usando MP., c) Reconstrucción de la señal con los 50 átomos previamente seleccionados., d) Señal residual después de 50 iteraciones.

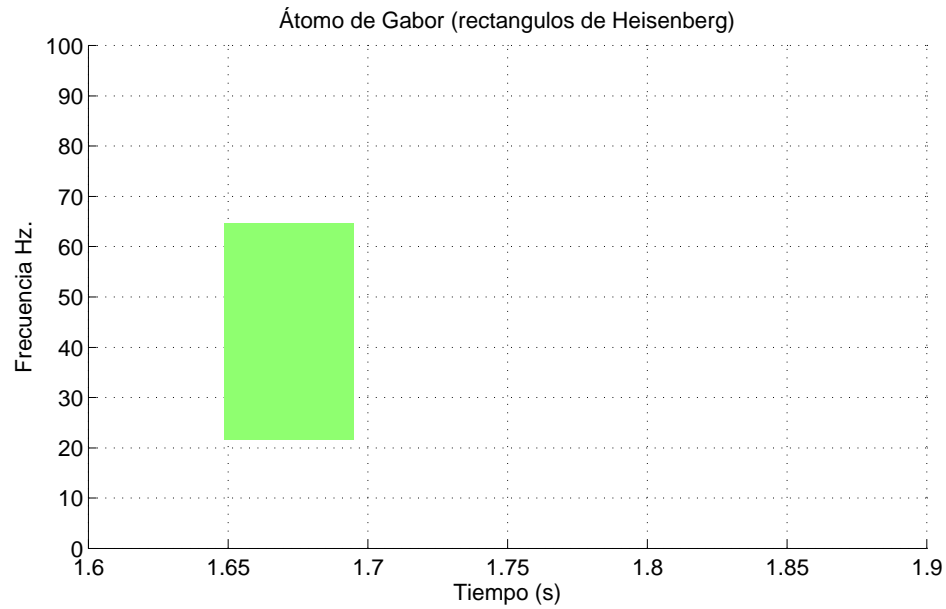


Figura 23: Átomo de Gabor plano tiempo-frecuencia.

En la figura 23 se muestra un átomo de Gabor en el plano tiempo-frecuencia el cual es de longitud en el tiempo $s = .0464s$ esto es, 512 muestras con frecuencia de muestreo de 11025 Hz y con una frecuencia central $\xi = 43\text{Hz}$.

El átomo de Gabor $g_{s,u,\xi}(t)$ está localizado alrededor del tiempo u con una duración del orden de s y su transformada de Fourier $\hat{g}_{s,u,\xi}(\omega)$ la cual está centrada en la frecuencia ξ con una dispersión en frecuencia del orden $\frac{1}{s}$.

Capítulo 6

Compresión del fonocardiograma

6.1. Introducción

El desarrollo reciente en sensores, herramientas de procesamiento de señales y las tecnologías de comunicaciones inalámbricas permiten transmitir y almacenar información de señales biomedicas. Los sistemas de telemedicina requieren de un continuo monitoreo y transmisiones de información relevante de pacientes. Las señales de fonocardiograma son usadas para el diagnóstico de enfermedades cardiovasculares. Estas señales presentan alta tasa de muestreo lo cual produce una cantidad considerable de datos que dificulta su transmisión. Consecuentemente los sistemas de telemedicina requieren algoritmos de compresión sin pérdida de información que permita su transmisión en redes de sensores con limitaciones en el ancho de banda [Manikandan y Dandapat (2007)].

6.2. MUSHRA

El método comúnmente utilizado para medir la calidad de codificadores paramétricos es el método MUSHRA (*Multi Stimulus test with Hidden Reference and Anchor*) [ITU (2001-2003)]. El método MUSHRA está basado en un test doblemente ciego y multi-estimulo, con referencia oculta y señales de prueba (*anchor*), también ocultas, especialmente diseñado para medir señales de audio con medias o grandes degradaciones de codificación. El test de audición se realiza en una o más sesiones. En cada intento se presenta la misma señal de audio procesada de forma diferente (o estímulos). El método MUSHRA se ha ensayado satisfactoriamente. Las pruebas demostraron que el método

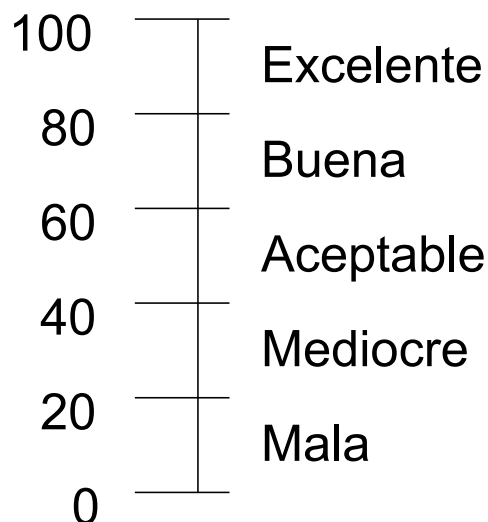


Figura 24: Cinco intervalos de escala continua (CQS) usados en el método *MUSHRA*.

MUSHRA sirve para la evaluación de la calidad de audio intermedia y arroja resultados precisos y fiables.

Durante el proceso de evaluación se pide a un conjunto de participantes que asignen notas a los estímulos (señales de prueba) según la escala de calidad continua (CQS). La CQS consiste en una escala gráfica que se divide en 5 intervalos iguales con adjetivos definidos, ver Figura 24.

El participante registra su evaluación de la calidad en un formulario deslizando el cursor entre los diferentes adjetivos limitándose a asignar una puntuación del fragmento escuchado. El método *MUSHRA* tiene la ventaja de visualizar todos los estímulos al mismo tiempo, como se ilustra en la Figura. 25, de tal manera que el sujeto pueda verificar de manera directa la relación entre los diferentes estímulos. Para realizar estas pruebas se programó una propia versión del estándar, utilizando matlab. Los resultados son más congruentes, lo que da lugar a intervalos de confianza más pequeños.

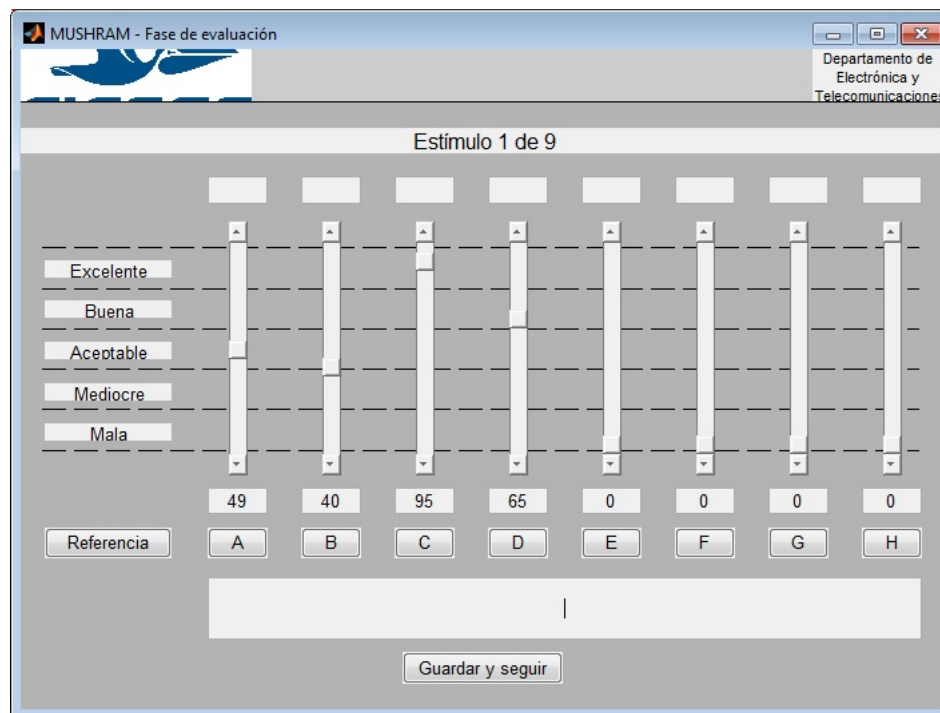


Figura 25: Visualización de los estímulos en la interface de pruebas.

6.3. Descripción de las señales de prueba

Con respecto a las señales de prueba, se recomienda que la longitud de las señales de prueba no debe rebasar 20 s. para evitar fatiga de los oyentes y reducir la duración total de la prueba de audición. El grupo de las señales procesadas consta de todas las señales de prueba y de al menos una señal adicional (patrón) que es una versión filtrada en paso bajo de la señal no procesada. Se presenta al participante una secuencia de ensayos. En cada uno de ellos se le presenta la versión de referencia, así como todas las versiones de señales de prueba procesadas por los sistemas en prueba. Se recomienda no incluir en cada tentativa más de 15 señales.

Las señales de audio cardiaco empleadas en esta prueba se describen en la tabla V. Para la realización de la prueba se reconstruyeron los sonidos cardiacos de la tabla

Tabla V: Base de datos de sonidos cardiacos.

	Señal audio cardiaco	Duración (seg)
Estímulo 1	Normal Split S1	4
Estímulo 2	Normal Split S2	4
Estímulo 3	S3	4
Estímulo 4	S4	5
Estímulo 5	click de eyección	4
Estímulo 6	soplo sistólico temprano	5
Estímulo 7	soplo sistólico tardío	5
Estímulo 8	soplo diastólico retumbante	4
Estímulo 9	Estenosis mitral	5

Tabla VI: Diccionarios de Gabor propuestos.

Nombre	Longitudes de los átomos	Número de átomos en el diccionario
Gabor 1	128, 256, 512	500
Gabor 2	32, 64, 128, 256, 512, 1024	500
Gabor 3	64, 128, 256, 512, 1024	800
Gabor 4	128, 256, 512	800
Gabor 5	32, 64, 128, 256, 512, 1024	400

V usando el algoritmo MP y los diccionarios de Gabor que se muestran en la tabla VI:

Los porcentajes de construcción para cada una de las señales de audio cardiaco usando los diccionarios de Gabor anteriormente mencionados se presentan en la Tabla VII. Se puede observar de la tabla anterior que todos las señales fueron reconstruidas por arriba del 99,6% con los diferentes diccionarios de Gabor

6.4. Criterio para la selección de sujetos

La experiencia ha demostrado que los datos procedentes de alrededor de 20 sujetos suelen ser suficientes para extraer conclusiones adecuadas de la prueba. Sin embargo aumentando el tamaño del grupo de oyentes experimentados puede considerarse en ocasiones más conveniente. En ocasiones hay motivos para introducir una técnica de rechazo ya sea anterior (preselección) o posterior (postselección) a la prueba real. Todo tipo de técnica de rechazo, analizado y aplicado minuciosamente puede conducir a un resultado sesgado. Por tal motivo es importante que se realice una eliminación de datos. Los métodos de postselección pueden básicamente dividirse al menos en dos clases:

Tabla VII: Porcentajes de reconstrucción de las señales originales con diferentes diccionarios de Gabor.

	Gabor 1	Gabor 2	Gabor3	Gabor 4	Gabor 5
Estímulo 1	99.90	99.97	99.93	99.98	99.96
Estímulo 2	99.90	99.98	99.94	99.98	99.96
Estímulo 3	99.80	99.96	99.87	99.96	99.92
Estímulo 4	99.87	99.96	99.91	99.97	99.94
Estímulo 5	99.87	99.96	99.90	99.97	99.94
Estímulo 6	99.75	99.93	99.82	99.94	99.87
Estímulo 7	99.62	99.89	99.73	99.92	99.80
Estímulo 8	99.59	99.92	99.70	99.93	99.84
Estímulo 9	99.84	99.96	99.88	99.96	99.92

- Una se basa en la capacidad de los sujetos para efectuar valoraciones congruentes repetidas.
- La otra se basa en las incongruencias de una apreciación individual comparándola con el resultado medio de todos los sujetos para un elemento determinado.

En las pruebas MUSHRA se recomienda examinar la dispersión individual y la desviación respecto a la valoración media de todos los sujetos. Dado que lo que se prueba es la calidad intermedia, un participante debe ser capaz en principio de identificar muy fácilmente la versión codificada y por lo tanto, dar una valoración que se encuentre dentro de la gama de la manifestada por la mayoría de los participantes [ITU (2001-2003)].

6.5. Procedimiento del método MUSHRA

El procedimiento del método consiste en presentar al participante el conjunto de señales de prueba en la interface previamente descrita. Se solicita al participante que evalúe las diferentes señales de prueba con respecto a la señal de referencia. El conjunto de señales de prueba está formado por señales reconstruidas usando el algoritmo MP y diccionarios de Gabor, así como una señal de referencia oculta, una señal patrón (señal degradada) y una versión de señal comprimida en formato mp3. El procedimiento previamente descrito se repite para cada uno de los estímulos de la Tabla V.

6.6. Condiciones de audición

Las especificaciones de los audífonos utilizados en la prueba son las siguientes:

- HD669 Superlux. Professional Studio Standard Monitoring Headphones. El HD-669 garantiza que un ingeniero de sonido profesional puede tener plena confianza

al grabar o supervisar.

- Sensibilidad: 98 dB SPL (1mW)
 - Respuesta en frecuencia 10-30000 Hz.
 - Impedancia nominal: 56 ohms.
- Sony MDR-ZX100/BLK
 - Sensibilidad: 100 dB SPL (1mW)
 - Respuesta en frecuencia 12-22000 Hz.
 - Impedancia nominal: 24 ohms a 1 KHz.
- Panasonic RP-DJS200-W
 - Sensibilidad: 105 dB SPL (1mW)
 - Respuesta en frecuencia 12-25000 Hz.
 - Impedancia nominal: 24 ohms a 1 KHz.

6.7. Fase de adiestramiento

Antes de iniciar con la apreciación formal es importante realizar la fase de adiestramiento la cual consiste en familiarizar a los participantes con las herramientas empleadas, el entorno de prueba, el procedimiento de la prueba y las señales de audio empleadas en la prueba. Al concluir dicho procedimiento los participantes deben haber adquirido el conocimiento suficiente sobre el dicho procedimiento para una mejor

apreciación formal. Esta fase es fundamental ya que, sí se lleva a cabo de manera correcta es posible que los participantes con una habilidad acústica baja pueden mejorar su condición auditiva y así aportar información relevante en la apreciación formal.

Antes de la fase de adiestramiento o familiarización los participantes fueron informados sobre los objetivos y alcances de esta investigación para aumentar el interés en la realización de las pruebas auditivas.

6.8. Preselección

Se seleccionaron 30 participantes para la realización de la prueba.

6.9. Postselección

La diferencia entre las notas otorgadas a la referencia oculta (señal original) y la señal patrón (señal con SNR 30 dB) es el punto de partida para efectuar el análisis estadístico. Los resultados obtenidos en el proceso de evaluación muestran los diferentes valores asignados a la señal de referencia oculta por cada uno de los participantes. Las figuras en el Apéndice A se muestran los resultados de las pruebas MUSHRA. En esta fase se observan los resultados con los cuales es posible concluir que los sujetos 7,10,21,27 no lograron identificar la referencia.

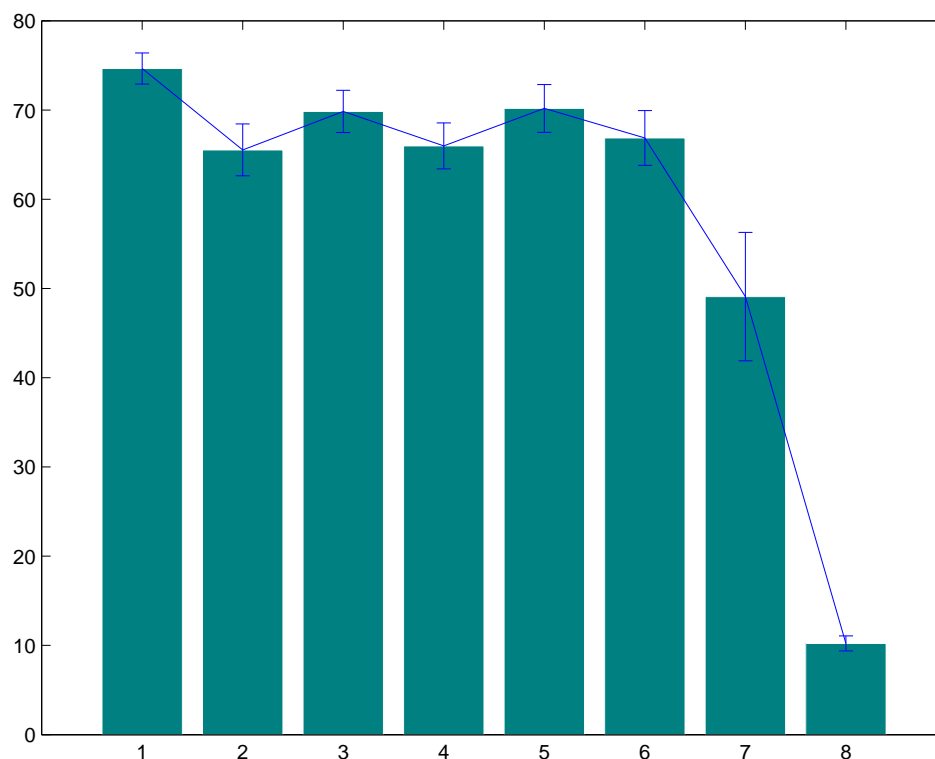


Figura 26: Intervalos de confianza para cada estímulo.

6.10. Análisis estadístico

El objetivo del análisis estadístico es identificar la calidad de funcionamiento medio de cada uno de los estímulos propuestos en esta investigación, así como la fiabilidad de cualquier diferencia entre los valores obtenidos.

6.11. Desviación media de cada sujeto

Es probable que los oyentes con valoraciones en el extremo superior de la escala sean menos críticos y probablemente los sujetos que sólo den valoraciones del extremo inferior de ella sean demasiados críticos. Rechazando los participantes de los extremos, cabe esperar una evaluación más realista de la calidad.

La desviación media de cada sujeto nos sirve para conocer que tanto las calificaciones del oyente se alejan de la calificación media y se calcula:

Para una matriz

$$\begin{bmatrix} & \textit{Sujeto,1} & \textit{Sujeto,2} & \cdots & \textit{Sujeto.n} \\ \textit{Audio,1} & x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ \textit{Audio,2} & x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \textit{Audio.m} & x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{bmatrix} \quad (44)$$

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij} \quad (45)$$

$$Dm_j = \frac{1}{m} \sum_{i=1}^m |x_{ij} - \bar{x}_i| \quad (46)$$

donde

n =número de sujetos

m =número de señales de audios

Así obtenemos la desviación media de cada sujeto donde j representa el número de sujeto al que pertenece.

6.12. Dispersión individual

Los métodos se utilizan principalmente para eliminar participantes que no pueden efectuar discriminaciones adecuadas. La aplicación de un método de postselección

puede esclarecer las tendencias en el resultado de una prueba. No obstante, teniendo presente la variabilidad de las sensibilidades de los participantes a los distintos efectos perturbadores, debe actuarse con cautela.

La dispersión individual nos muestra las incongruencias de una apreciación individual comparándola con el resultado medio de todos los sujetos para un elemento determinado.

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij} \quad (47)$$

$$\sigma_j = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_{ij} - \bar{x}_i)^2} \quad (48)$$

donde

n =número de sujetos

m =número de señales de audios

Así obtuvimos la dispersión individual donde j respresenta el número de sujeto. Los resultados que se obtuvieron para la desviación media y para la dispersión individual se encuentran en la Tabla VIII.

Tabla VIII: Promedio de intervalos de confianza y desviación para cada uno de los estímulos.

	Promedio intervalo de confianza	Desviación estandar
Referencia	74.6570 \pm 1.7424	2.6670
Gabor01	65.5266 \pm 2.9055	4.4473
Gabor02	69.8406 \pm 2.3651	3.6201
Gabor03	65.9758 \pm 2.5789	3.9473
Gabor04	70.1739 \pm 2.6815	4.1043
Gabor05	66.8647 \pm 3.0711	4.7007
mp3	49.0870 \pm 7.1990	11.0189
patron	-30dB 10.2222 \pm 0.8475	1.2972

Capítulo 7

Conclusiones

En este apartado se describen las conclusiones a las que se han llegado después de la investigación realizada. Para ello las conclusiones se han dividido en tres secciones, la primera de ella corresponde a las conclusiones obtenidas sobre los diferentes tipos de diccionarios evaluados para la descomposición de las señales de audio cardiaco; posteriormente se incluyen las conclusiones correspondientes a la implementación de los algoritmos MP para la descomposición escasa de señales de audio cardiaco, para luego concluir con los resultados obtenidos del algoritmo de segmentación propuesto. Dentro de este mismo capítulo se incorpora también las ideas del trabajo futuro.

Se analizaron los algoritmos MP y los diferentes diccionarios tiempo-frecuencia para la compresión y segmentación de señales de fonocardiograma.

De la misma manera se evaluaron y compararon diccionarios MDCT, donde la energía residual para un número de átomos establecido es mayor que los diccionarios de Gabor.

La aproximación de señales de fonocardiograma han sido ampliamente estudiadas en la literatura como una herramienta poderosa para la compresión de diferentes señales [Manikandan y Dandapat (2007)], sin embargo al aproximar las señales de fonocardiograma el diccionario de Gabor para un número de átomos dado presentó mejores resultados. Los diccionarios de Gabor, MDCT, wavelet fueron construidos por un conjunto de átomos los cuales están formados por diferentes parámetros (posición, longitud, frecuencia, Amplitud). El uso del algoritmo MP permitió conocer los parámetros de los átomos óptimos que mejor aproximan la señal.

Se modificaron dichos parámetros de los átomos buscando aquellos átomos que mejor aproximen la señales disminuyendo así los átomos totales del diccionario. Una vez que se encontró que los átomos de Gabor modelan mejor el comportamiento de las señales de fonocardiograma (i. e. menor cantidad de energía con un número fijo de iteraciones). Se construyeron átomos considerando las diversas longitudes de los sonidos cardiacos en las señales de fonocardiograma.

Se propuso un algoritmo de segmentación basado en el algoritmo Matching Pursuit y diccionarios de Gabor para la segmentación de sonidos cardiacos [Nieblas *et al.* (2013)], el cual fue presentado en Napa, CA, USA y publicado en Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE), 2013 IEEE. El objetivo de este algoritmo es segmentar los sonidos cardiacos normales y anormales para las señales de fonocardiograma. En este algoritmo de segmentación se utilizó un diccionario de Gabor y el algoritmo Matching Pursuit [Mallat y Zhang (1993)]. El diccionario de Gabor implementado está formado por 3 bloques [longitud de ventana, traslape de ventana, tamaño FFT] caracterizados en muestras [128 64 128], [256 128 256] [512 256 512]. El algoritmo reportado muestra un alto desempeño logrando detección del 97.5 % y 96 % en onset y offset respectivamente.

Los resultados de las pruebas MUSHRA nos ayudaron a identificar los diccionarios de Gabor que mejor modelan la señales de audio cardiaco. En estos resultados se encontró que el diccionario de Gabor con longitudes de átomo de 32, 64, 128, 256, 512 y 1024 aproximan las señales de audio cardiaco con un 99.90 %. Este porcentaje de reconstrucción fué alcanzado usando aproximadamente 83 átomos por ciclo cardiaco.

7.1. Sobre el objetivo de la tesis

Lo reportado en los apartados anteriores permite afirmar que se ha cumplido de manera satisfactoria el objetivo de la tesis, el cual consiste en evaluar el desempeño de la representación escasa a través del algoritmo Matching Pursuit para proponer nuevos mecanismos de análisis tiempo-frecuencia adecuados para la compresión transparente, segmentación y reducción del ruido del audio cardiaco. Un punto importante sobre la tesis consistió en evaluar los diferentes tipos de diccionarios tiempo-frecuencia buscando aquel diccionario que mejor modelara el comportamiento de los sonidos de audio cardiaco de tal manera que con el menor número de átomos se alcanzará una reconstrucción del 99.9% de la señal en cuestión. Lo descrito anteriormente se logró usando diccionarios de Gabor tiempo-frecuencia.

7.2. Trabajo futuro como continuación a esta investigación.

Tomando en consideración los alcances logrados en esta investigación, el proceso de investigación se puede ampliar considerando los siguientes puntos como trabajo futuro.

- Proponer nuevos algoritmos de segmentación de audio cardiaco basados en el algoritmo MP y comparar su desempeño con el algoritmo de segmentación propuesto en esta tesis.
- Incluir dentro del escenario de procesamiento de señales de audio cardiaco aplicaciones para la reducción de ruido, así como validar los algoritmos MP como una herramienta para la separación de fuentes.
- Investigar e incluir otros esquemas de representación escasa de señales de audio cardiaco para su comparación con los algoritmos MP.

- Investigar y proponer nuevas herramientas de procesamiento de señales para modelar el residual de las señales de audio cardiaco.
- Investigación y desarrollo de un códec con pérdidas para señales de audio cardiaco basado en los algoritmo MP.

Bibliografía

- Abbas, A. y Bassam, R. (2009). *Phonocardiography Signal Processing*. Synthesis Lectures on Biomedical Engineering Series. Morgan & Claypool. ISBN 9781598299755.
- Agostihno, M. y Souza, M. (1997). A new heart sound simulation technique. *IEEE 19th International Conference*, páginas 323–326.
- Allen, J. y Rabiner, L. (1977). A unified approach to short-time fourier analysis and synthesis. *Proceedings of the IEEE*, **65**(11): 1558–1564.
- Amit, G., Gavriely, N., y Intrator, N. (2009). Cluster analysis and classification of heart sounds. *Biomedical Signal Processing and Control. Elsevier Science Direct.*, (4): 26–36.
- Baranek, H., Lee, H., Cloutier, G., y Durand, L. (1989). Automatic detection of sounds and murmurs in patients with Ionescu-shiley aortic bioprostheses. *Medical and Biological Engineering and Computing*, **27**(5): 449–455.
- 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.
- Castorena, L. (2012). *Análisis y Adaptación de códecs para la transmisión de audio cardiaco sobre redes inalámbricas con tasas de datos bajas/medianas.* Tesis de maestría, Centro de Investigación Científica y Educación Superior de Ensenada.
- Chebil, J. y Al-Nabulsi, J. (2007). Classification of heart sound signals using discrete wavelet analysis. *International Journal of Soft Computing*, **2**(1): 37–41.

- Chen, S. y Wigger, J. (1995). Fast orthogonal least squares algorithm for efficient subset model selection. *Signal Processing, IEEE Transactions on*, **43**(7): 1713–1715.
- Cho, N. y Kuo, C.-C. (2011). Sparse music representation with source-specific dictionaries and its application to signal separation. *Audio, Speech, and Language Processing, IEEE Transactions on*, **19**(2): 326–337.
- Daubechies, I. (1988). Time-frequency localization operators: a geometric phase space approach. *Information Theory, IEEE Transactions on*, **34**(4): 605–612.
- Daudet, L. (2006). Sparse and structured decompositions of signals with the molecular matching pursuit. *IEEE Transactions on Acoustics, Speech and Signal Processing on Audio, Speech, and Language Processing.*, **14**: 1808–1816.
- Debbal, S. y Bereksi-Reguig, F. (2007). Time-frequency analysis of the first and the second heartbeat sounds. *Applied Mathematics and Computation*, **184**(2): 1041–1052.
- Durand, L.-G., Zhang, X., Senhadji, L., Lee, H., y Coatrieux, J.-L. (1996). Application of the matching pursuit method for the time-scale transformation of the phonocardiogram. En *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE*, Vol. 3, páginas 1033–1034 vol.3.
- E. Ravelli, G. R. y Daudet, L. (2008). Union of mdct bases for audio coding. *IEEE Transactions on audio, speech, and language processing*, **16**: 1364–1372.
- Gerbarg, D. S., Holcomb Jr., F., Hofler, J., Bading, C., Schultz, G., y Sears, R. E. (1962). Analysis of phonocardiogram by a digital computer. *Circulation Research*, **11**: 569.

- Gribonval, R. (2001). Fast matching pursuit with a multiscale dictionary of gaussian chirps. *Signal Processing, IEEE Transactions on*, **49**(5): 994–1001.
- Gribonval, R., Bacry, E., Mallat, S., Depalle, P., y Rodet, X. (1996). Analysis of sound signals with high resolution matching pursuit. En *Time-Frequency and Time-Scale Analysis, 1996., Proceedings of the IEEE-SP International Symposium on*, páginas 125–128.
- Gribonval, R. y Bacry, E. (2003). Harmonic decomposition of audio signals with matching pursuit. *IEEE Transactions on processing*, **51**: 101–111.
- Groch, M., Domnanovich, J., y Erwin, W. (1992). A new heart-sounds gating device for medical imaging. *IEEE Trans. on Biomedical Engineering*, **39**(3): 307–310.
- Hedayioglu, F., Jafari, M., Mattos, S., Plumbley, M., y Coimbra, M. (2012). Denoising and segmentation of the second heart sound using matching pursuit. En *Int. Conf. of the Eng. in Med. and Bio. Society (EMBS) IEEE*, páginas 3440–3443. IEEE.
- Hill, P. (2007). Dennis gabor - contributions to communication theory; signal processing. En *EUROCON, 2007. The International Conference on num. 34; Computer as a Tool;*, páginas 2632–2637.
- Huiying, L., Sakari, L., y Iiro, H. (1997). A heart sound segmentation algorithm using wavelet decomposition and reconstruction. En *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, Vol. 4, páginas 1630–1633 vol.4.
- ITU (2001-2003). Method for the subjective assessment of intermediate quality level of coding systems.

- Iwata, A., Ishii, N., Suzumura, N., y Ikegaya, K. (1980). Algorithm for detecting the first and the second heart sounds by spectral tracking. *Medical and Biological Engineering and Computing*, **18**(1): 19–26.
- Jabbari, S. y Ghassemian, H. (2009). Sparse modeling of heart sounds and murmurs based on orthogonal matching pursuit. En *Computer Conference, 2009. CSICC 2009. 14th International CSI*, páginas 355–360.
- Jiang, S. C. . Z. (2006). Comparison of envelope extraction algorithms for cardiac sound signal segmentation. páginas 1056–1069.
- Krstulovic, S. y Gribonval, R. (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.
- Kumar, D., Carvalho, P., Antunes, M., Henriques, J., Maldonado, M., Schmidt, R., y Habetha, J. (2006). Wavelet transform and simplicity based heart murmur segmentation. En *Computers in Cardiology, 2006*, páginas 173–176.
- Kumar, D., Carvalho, P., Antunes, M., Henriques, J., Sa e Melo, A., Schmidt, R., y Habetha, J. (2007). Third heart sound detection using wavelet transform-simplicity filter. En *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, páginas 1277–1281.
- Kumar, D., Carvalho, P., Antunes, M., Henriques, J., Sa e Melo, A., y Habetha, J. (2008). Heart murmur recognition and segmentation by complexity signatures. En *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, páginas 2128–2132.

- Kumar, D., Carvalho, P., Antunes, M., Paiva, R., y Henriques, J. (2011). An adaptive approach to abnormal heart sound segmentation. En *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, páginas 661–664.
- Liang, H., Lukkarinen, S., y Hartimo, I. (1997). Heart sound segmentation algorithm based on heart sound envelopogram. En *Computers in Cardiology 1997*, páginas 105–108.
- Mahnke, C. (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.
- Malarvili, M., Kamarulafizam, I., Hussain, S., y Helmi, D. (2003). Heart sound segmentation algorithm based on instantaneous energy of electrocardiogram. En *Computers in Cardiology, 2003*, páginas 327–330. IEEE.
- Mallat, S. (2009). *A wavelet Tour of Signal Processing*.
- Mallat, S. y Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, **41**(12): 3397–3415.
- Manikandan, M. y Dandapat, S. (2007). Wavelet-based ecg and pcg signals compression technique for mobile telemedicine. En *Advanced Computing and Communications, 2007. ADCOM 2007. International Conference on*, páginas 164–169.
- Martinez-Alajarin, J. y Ruiz-Merino, R. (2005). Efficient method for events detection in phonocardiographic signals. En *Microtechnologies for the New Millennium 2005*, páginas 398–409. International Society for Optics and Photonics.

- Nieblas, C. I., Alonso, M. A., Conte, R., y Villarreal, S. (2013). High performance heart sound segmentation algorithm based on matching pursuit. En *Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE), 2013 IEEE*, páginas 96–100.
- Organization, W. H. (Geneva, Switzerland 2011a). Global status report on noncommunicable diseases 2010.
- Organization, W. H. (Geneva, Switzerland 2011b). Global status report on noncommunicable diseases 2010.
- Plumbley, M., Blumensath, T., Daudet, L., Gribonval, R., y Davies, M. (2010). Sparse representations in audio and music: From coding to source separation. *Proceedings of the IEEE*, **98**(6): 995–1005.
- Prasad, R., Koike, T., y Matsuno, F. (2008). Measurement of threshold of hearing using tone trains. En *SICE Annual Conference, 2008*, páginas 2727–2731.
- Ravelli, E., Richard, G., y Daudet, L. (2008). Union of mdct bases for audio coding. *Audio, Speech, and Language Processing, IEEE Transactions on*, **16**(8): 1361–1372.
- Reed, T., Reed, N., y Fritzon, P. (2003). Heart sound analysis for symptom detection and computer-aided diagnosis. *Simulation Modelling, Practice and Theory*, **12**: 129–146.
- Rubinstein, R., Bruckstein, A., y Elad, M. (2010). Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, **98**(6): 1045–1057.
- Saha, G. y Kumar, P. (2004). An efficient heart sound segmentation algorithm for car-

- diac diseases. En *India Annual Conference, 2004. Proceedings of the IEEE INDICON 2004. First*, páginas 344–348.
- Sava, H. y Durand, L.-G. (1997). Automatic detection of cardiac cycle based on an adaptive time-frequency analysis of the phonocardiogram. En *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, Vol. 3, páginas 1316–1319 vol.3.
- Sturm, B., Shynk, J., y Gauglitz, S. (2008). Agglomerative clustering in sparse atomic decompositions of audio signals. En *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, páginas 97–100.
- Tang, H., Li, T., Park, Y., y Qiu, T. (2010). Separation of heart sound signal from noise in joint cycle frequency-time-frequency domain based on fuzzy detection. *IEEE Transactions on Biomedical Engineering*, **57**(10): 2438–2447.
- Tang, H., Li, T., Qiu, T., y Park, Y. (2012). Segmentation of heart sounds based on dynamic clustering. *Biomedical Signal Processing and Control*, **7**(5): 509 – 516.
- Wang, P., Kim, Y., Ling, L. H., y Soh, C. B. (2005). First heart sound detection for phonocardiogram segmentation. En *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, páginas 5519–5522.
- Wang, W., Pan, J., y Lian, H. (2004). Decomposition and analysis of the second heart sound based on the matching pursuit method. En *Signal Processing, 2004. Proceedings. ICSP '04. 2004 7th International Conference on*, Vol. 3, páginas 2229–2232 vol.3.

- 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.
- X. Zhang, L.G. Durand L. Senhadji, H. L. y Coatrieux, J. (1998). Analysis-synthesis of the phonocardiogram based on the matching pursuit method. *IEEE Trans. Biomed. Eng*, **45**: 962–972.
- Yan, Z., Jiang, Z., Miyamoto, A., y Wei, Y. (2010). The moment segmentation analysis of heart sound pattern. *Computer Methods and Programs in Biomedicine*, **98**(2): 140 – 150.
- Yue, Z., Liang, J., y Shen, Y. (2012). Time-frequency analysis of heart sounds in telemedicine consulting system for auscultation. En *Information and Automation (ICIA), 2012 International Conference on*, páginas 652–657.
- Zhang, S., Dou, W., y Yang, H. (2013). Mdct sinusoidal analysis for audio signals analysis and processing. *Audio, Speech, and Language Processing, IEEE Transactions on*, **21**(7): 1403–1414.

Apéndice

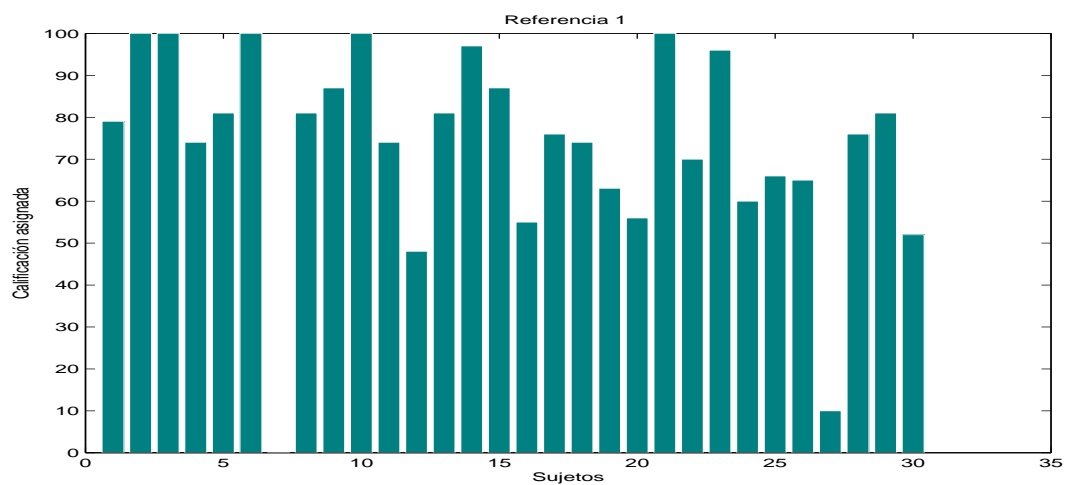


Figura 27: Calificación asignada por cada participante a la señal de referencia oculta Split S1 usando diferentes diccionarios de Gabor.

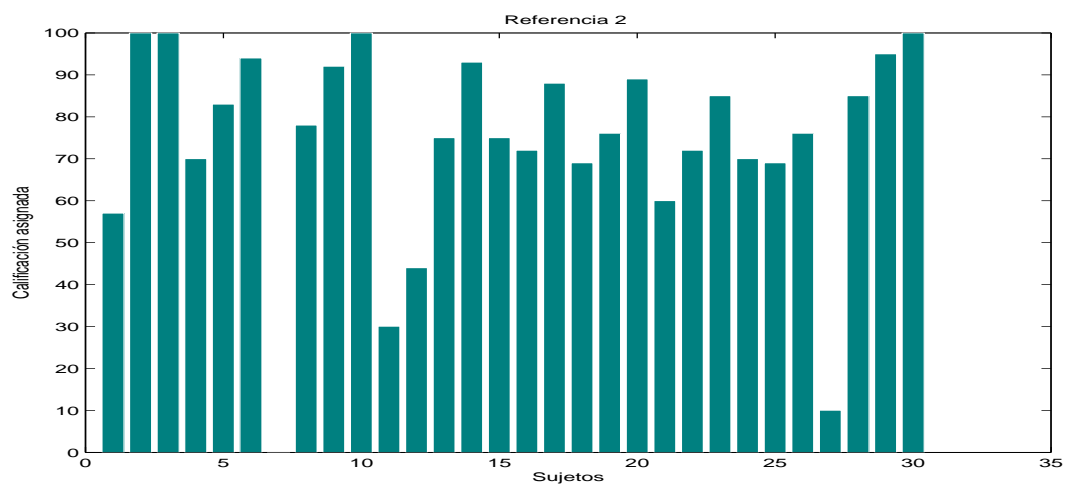


Figura 28: Calificación asignada por cada participante a la señal referencia oculta Normal Split S2 usando diferentes diccionarios de Gabor.

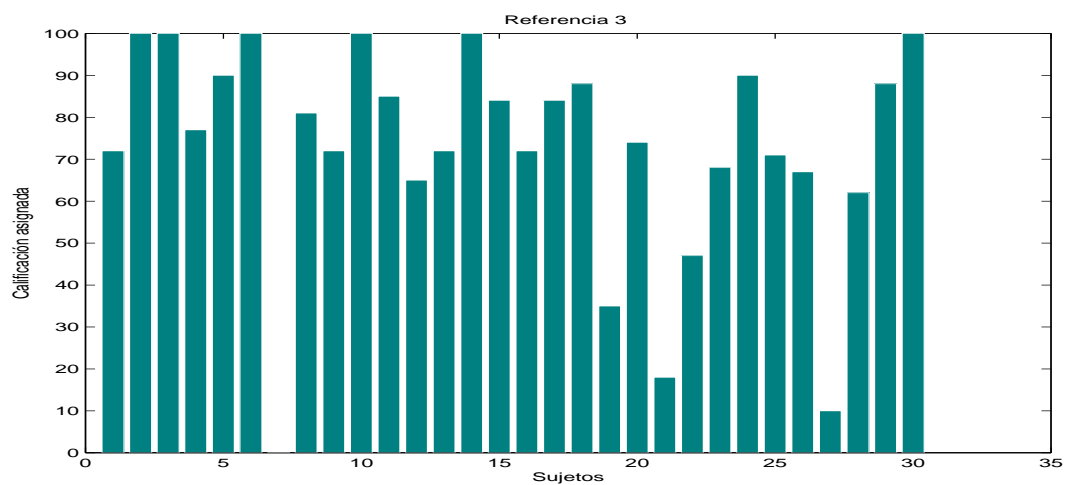


Figura 29: Calificación asignada por cada participante a la señal referencia oculta S3 usando diferentes diccionarios de Gabor.

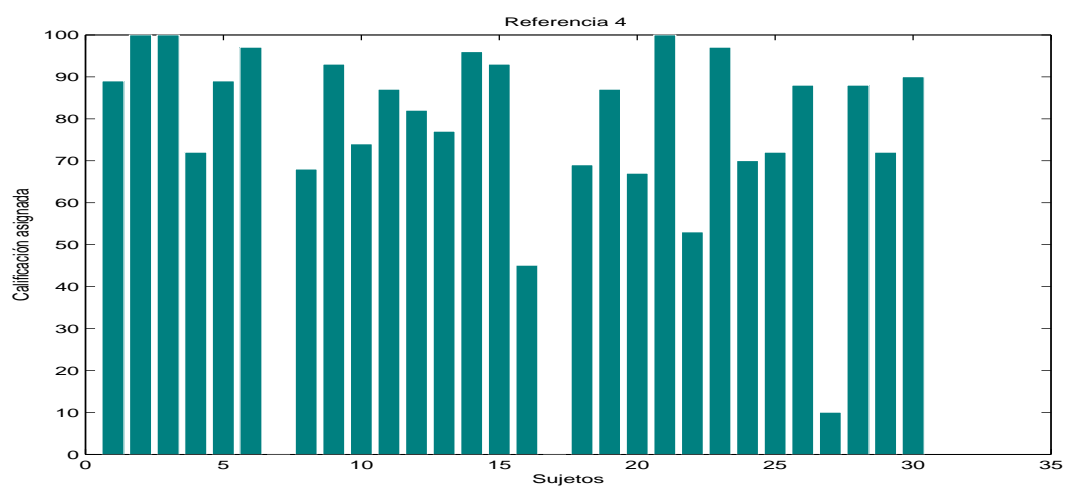


Figura 30: Calificación asignada por cada participante a la señal referencia oculta S4 usando diferentes diccionarios de Gabor.

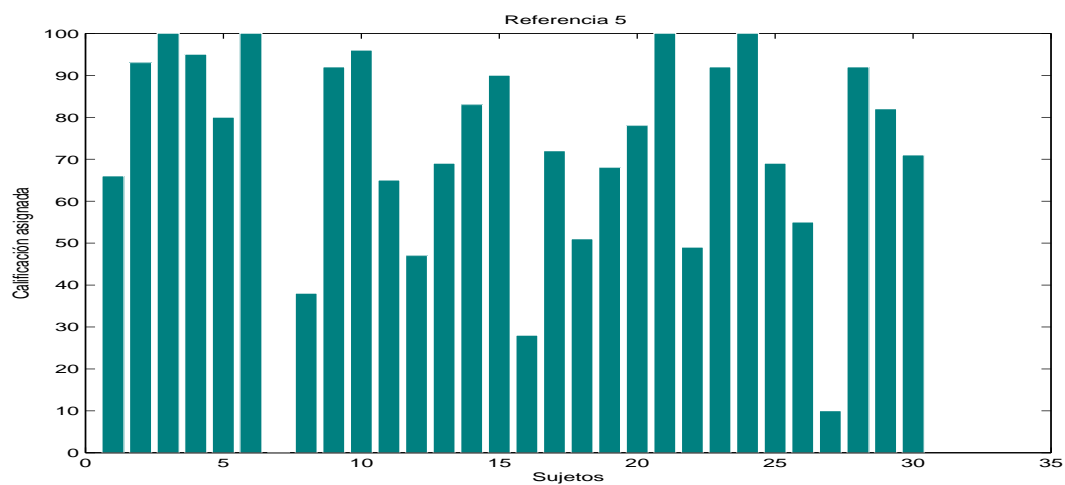


Figura 31: Calificación asignada por cada participante a la señal referencia oculta click de eyección usando diferentes diccionarios de Gabor.

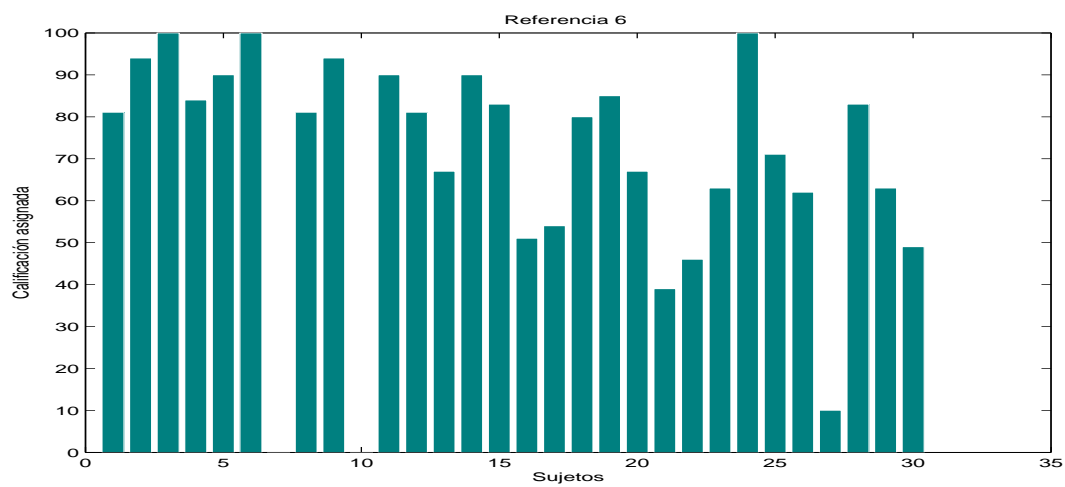


Figura 32: Calificación asignada por cada participante a la señal referencia oculta soplo sistólico temprano usando diferentes diccionarios de Gabor.

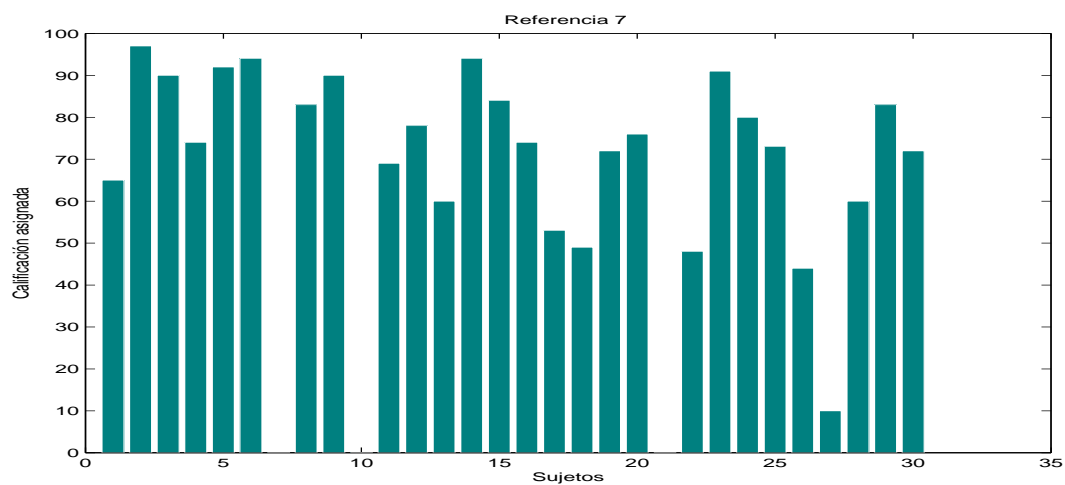


Figura 33: Calificación asignada por cada participante a la señal referencia oculta soplo sistólico tardío usando diferentes diccionarios de Gabor.

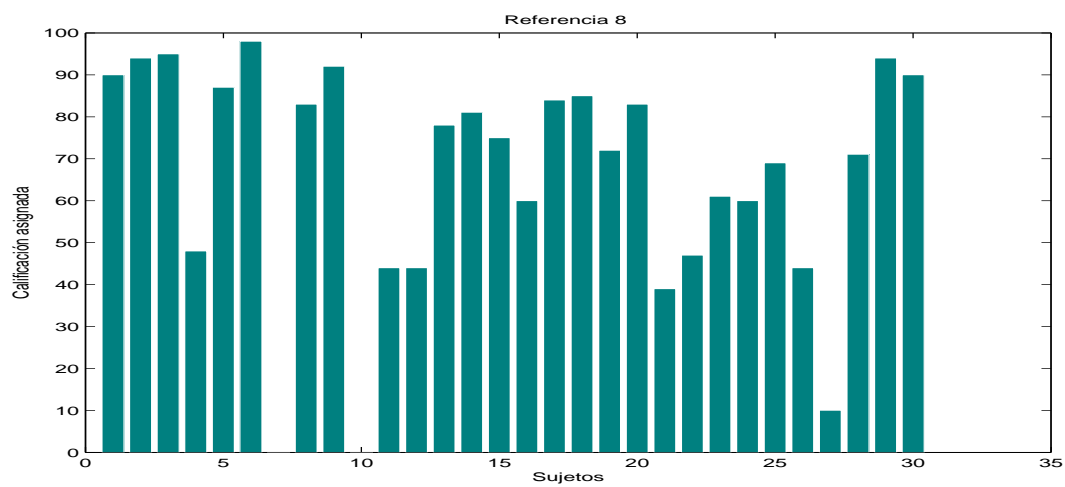


Figura 34: Calificación asignada por cada participante a la señal referencia oculta soplo diastólico retumbante usando diferentes diccionarios de Gabor.

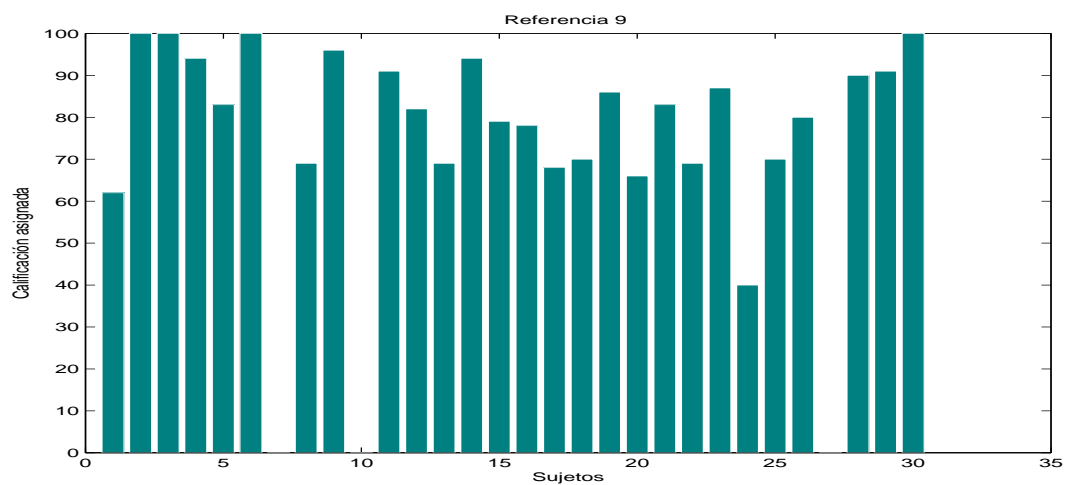


Figura 35: Calificación asignada por cada participante a la señal referencia oculta este-nosis mitral usando diferentes diccionarios de Gabor.

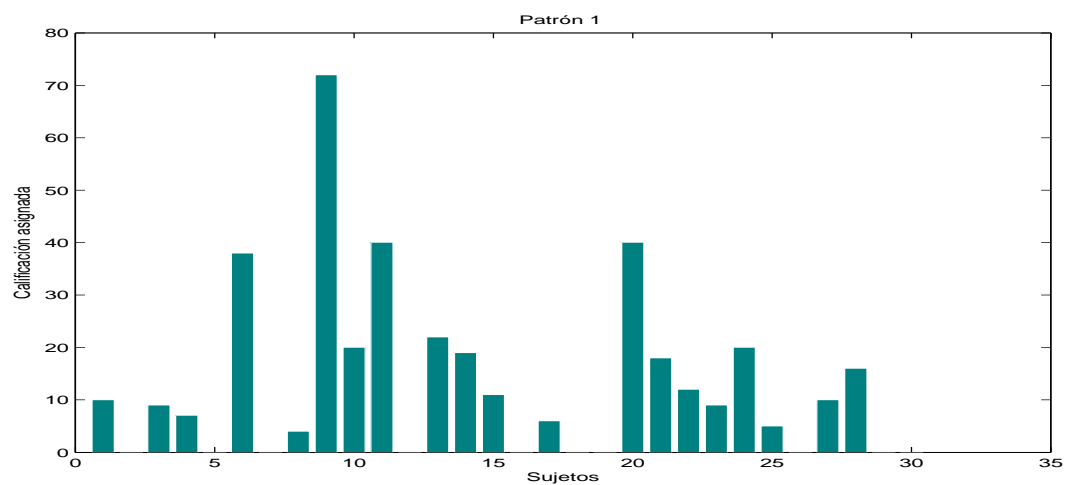


Figura 36: Calificación asignada por cada participante a la señal referencia patrón Normal Split S1 usando diferentes diccionarios de Gabor.

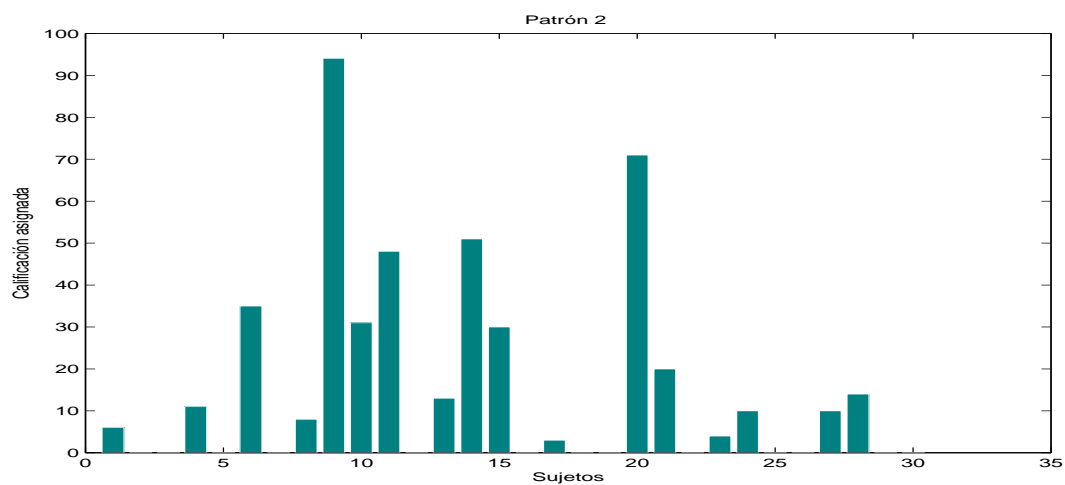


Figura 37: Calificación asignada por cada participante a la señal referencia patrón a Normal Split S2 usando diferentes diccionarios de Gabor.

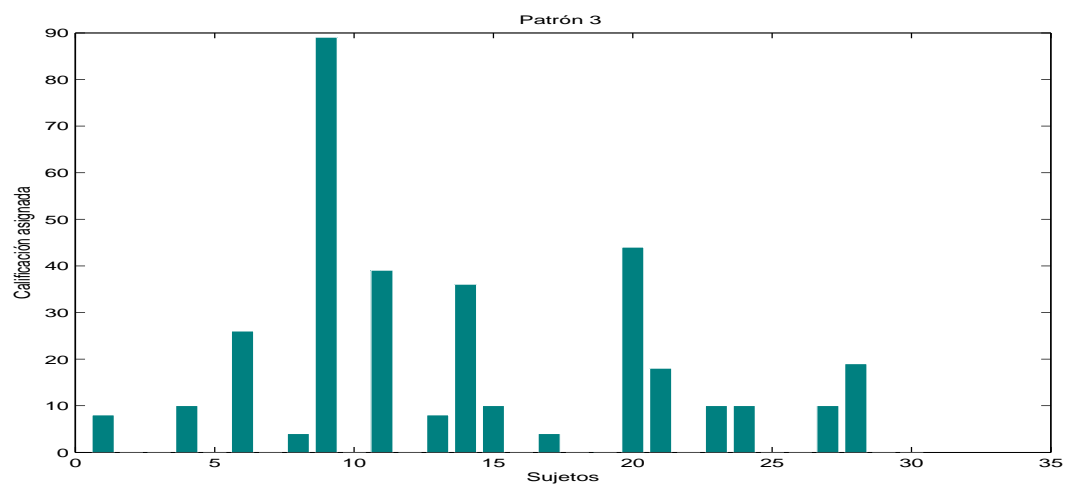


Figura 38: Calificación asignada por cada participante a la señal referencia patrón S3 usando diferentes diccionarios de Gabor.

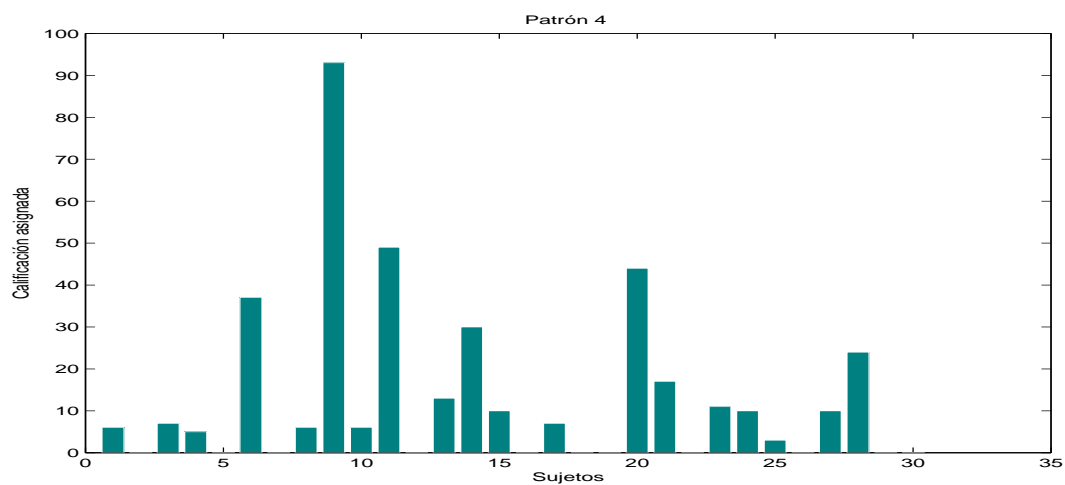


Figura 39: Calificación asignada por cada participante a la señal referencia patrón S4 usando diferentes diccionarios de Gabor.

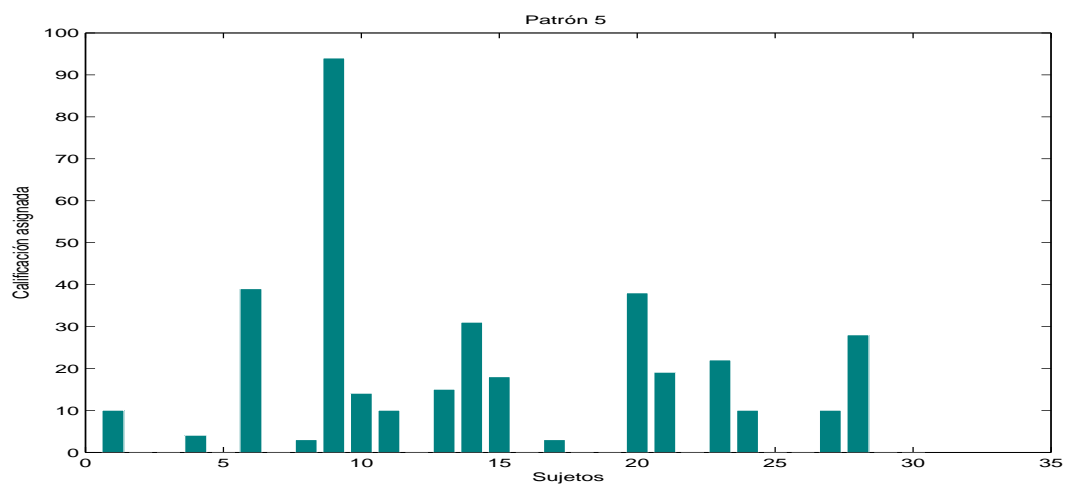


Figura 40: Calificación asignada por cada participante a la señal referencia patrón click de eyección usando diferentes diccionarios de Gabor.

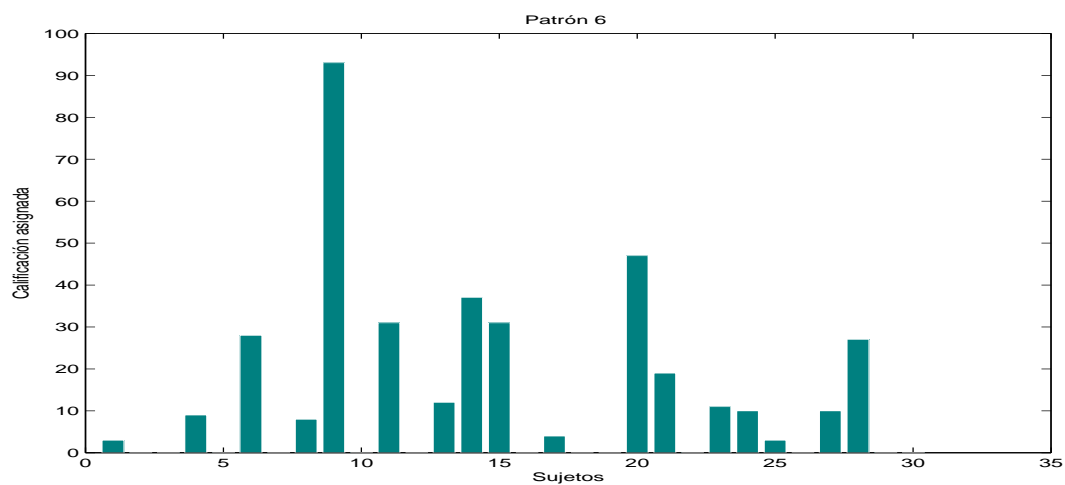


Figura 41: Calificación asignada por cada participante a la señal referencia patrón soplo sistólico temprano usando diferentes diccionarios de Gabor.

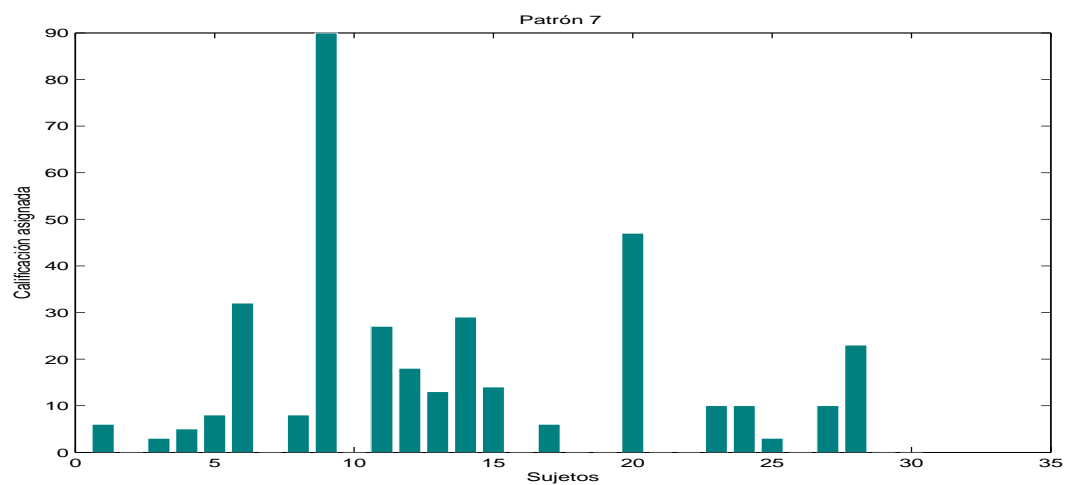


Figura 42: Calificación asignada por cada participante a la señal referencia patrón soplo sistólico tardío diferentes diccionarios de Gabor.

Tabla IX: Desviación media y dispersión individual por cada sujeto

	Desviación media	Dispersión individual
Sujeto 1	-2.8481	13.9335
Sujeto 2	16.1102	29.5065
Sujeto 3	16.1935	25.0665
Sujeto 4	-3.8481	13.8496
Sujeto 5	-16.3759	25.4428
Sujeto 6	27.2074	29.6202
Sujeto 7	-57.0148	59.9834
Sujeto 8	6.1519	17.3715
Sujeto 9	33.4019	38.2486
Sujeto 10	-19.9870	42.2642
Sujeto 11	-2.4454	21.4299
Sujeto 12	-4.2509	20.8944
Sujeto 13	-9.7926	14.7175
Sujeto 14	18.1102	21.3080
Sujeto 15	12.9713	18.6710

Tabla X: Desviación media y dispersión individual por cada sujeto

	Desviación media	Dispersión individual
Sujeto 16	-5.1259	16.1358
Sujeto 17	1.8185	20.3652
Sujeto 18	-3.5426	22.6334
Sujeto 19	0.7769	16.4646
Sujeto 20	16.0685	22.9601
Sujeto 21	1.6102	27.9967
Sujeto 22	-14.7509	20.0245
Sujeto 23	11.6519	22.0933
Sujeto 24	4.5963	15.2108
Sujeto 25	0.1657	10.7701
Sujeto 26	-3.7648	17.0995
Sujeto 27	-47.9870	51.6193
Sujeto 28	5.5546	10.5213
Sujeto 29	12.9852	20.8973
Sujeto 30	6.3602	19.4260