

Instituto Tecnológico de Costa Rica

Escuela de Ingeniería en Electrónica



**Reconocimiento digital en línea de patrones acústicos para la protección del ambiente
por medio de HMM**

**Informe de Proyecto de Graduación para optar por el título de Ingeniero en Electrónica
con el grado académico de Licenciatura**

Luis Esteban Smith Coto

Cartago, Enero de 2008

INSTITUTO TECNOLOGICO DE COSTA RICA

ESCUELA DE INGENIERIA ELECTRONICA

PROYECTO DE GRADUACIÓN

TRIBUNAL EVALUADOR

Proyecto de Graduación defendido ante el presente Tribunal Evaluador como requisito para optar por el título de Ingeniero en Electrónica con el grado académico de Licenciatura, del Instituto Tecnológico de Costa Rica.

Miembros del Tribunal


Ing. Néstor Hernández Hostaller

Profesor lector


Ing. Marvin Hernández Cisneros

Profesor lector


Ing. Pablo Alvarado Moya

Profesor asesor

Los miembros de este Tribunal dan fe de que el presente trabajo de graduación ha sido aprobado y cumple con las normas establecidas por la Escuela de Ingeniería Electrónica

Cartago, 21 de enero de 2008

Declaro que el presente Proyecto de Graduación ha sido realizado enteramente por mi persona, utilizando y aplicando literatura referente al tema e introduciendo conocimientos propios.

En los casos en que he utilizado bibliografía, he procedido a indicar las fuentes mediante las respectivas citas bibliográficas.

En consecuencia, asumo la responsabilidad total por el trabajo de graduación realizado y por el contenido del correspondiente informe final.

Cartago, 21 de enero de 2008



Luis Esteban Smith Coto

Céd: 3-0386-0204

Resumen

La tala y la caza ilegales dentro de zonas protegidas representan pérdidas para Costa Rica en materia de conservación y explotación racional de los recursos naturales debido al impacto de estas a nivel ecológico y económico.

Este informe presenta un prototipo de reconocimiento de patrones acústicos basado en Modelos Ocultos de Markov (HMM) para sonidos de disparos y motosierras dentro del ambiente del bosque, con el propósito de utilizarlo en sistemas de detección de actividades de caza y tala ilegales que se presenten dentro de zonas protegidas.

Modificaciones a las variables de los HMM como el número de estados, así como el largo de las cadenas de observación y la tasa de muestreo para el clasificador, permiten obtener porcentajes de detección superiores al 90%.

Técnicas de filtrado digital de respuesta infinita al impulso (IIR), discriminantes lineales y el método de aglomeración por K-medias se utilizan para otorgar a los modelos de Markov las secuencias para el reconocimiento.

Palabras Claves: HMM, reconocimiento de patrones, discriminación lineal, K-medias, filtros IIR.

Abstract

Illegal hunting and deforestation activities inside protected areas represent for Costa Rica losses in terms of conservation and rational utilization of natural resources due to ecological and economical effects.

This report presents an acoustic pattern recognition prototype based in Hidden Markov Models (HMM) for shots and chainsaws sounds inside the forest environment to use it in surveillance systems of protected areas to detect illegal activities of hunting and tree cutting.

Modifications to HMM variables like the state number, the large of the observation chains and the sample rate for the classifying module, allow the prototype to get detection ratios greater than 90%.

Infinite Impulse Respond (IIR) filters, linear discriminant analysis and K-means clustering are used to give sequences to the Markov Models for the recognition.

Key Words: HMM, pattern recognition, linear discriminant, K-means, IIR filters.

A mis padres

Agradecimiento

Quiero agradecer al cuerpo de profesores del Instituto Tecnológico de Costa Rica, en especial al profesor Ing. Néstor Hernández por darme la oportunidad de trabajar en un proyecto tan visionario. Al Dr.-Ing. Pablo Alvarado por su ayuda durante toda la realización del proyecto. También agradecer a mis compañeros que de una u otra manera contribuyeron en mi formación y me ayudaron a alcanzar esta meta. A mi familia por su incondicional apoyo y por siempre creer en mí. A todos:

Muchas Gracias

Índice General

Capítulo 1: Introducción	1
1.1 Tala ilegal, caza furtiva e incendios forestales en Costa Rica	1
1.1.1 Detección de actividades de tala y caza en zonas protegidas	2
1.1.2 Trabajos previos	3
1.2 Reconocimiento de patrones acústicos	5
1.3 Objetivo y estructura del documento	7
Capítulo 2: Marco Teórico	8
2.1 Extracción de características de la señal acústica	8
2.1.1 Representación de señales acústicas por su contenido de frecuencias	8
2.1.2 Selección de bandas de frecuencias y reducción de dimensiones	15
2.1.3 Aglomeración y método de K-medias	21
2.2 Clasificación de la señal acústica	23
2.2.1 Modelos Ocultos de Markov	23
2.3 Antecedentes Bibliográficos.	30
Capítulo 3: Diseño y evaluación de la etapa de preproceso	31
3.1 Control de Ganancia	31
3.1.1 Estructura del promediador	32
3.1.2 Ajuste de la constante de atenuación	35
3.1.3 Ajuste de la constante de normalización	37
Capítulo 4: Diseño y evaluación de la etapa de extracción de características	39
4.1 Banco de Filtros	39
4.1.1 Comparación entre diseños	39
4.1.2 Diseño de filtros QMF	40
4.1.3 Estimación de energía	42
4.1.4 Selección de bandas de frecuencias	44
4.2 Codificación	46
Capítulo 5: Diseño y evaluación de la etapa de clasificación	48
5.1 Análisis	48
5.1.1 Entrenamiento de los modelos	48
5.1.2 Base de datos para el entrenamiento	49
5.2 Toma de decisiones	49

5.3 Determinación de los parámetros de los modelos	50
5.3.1 S: El conjunto de estados	50
5.3.2 P: La tasa de muestreo para la etapa de clasificación	51
5.3.3 O: El tamaño de la cadena de observaciones	53
Capítulo 6: Conclusiones y recomendaciones	57
6.1 Conclusiones	57
6.2 Recomendaciones	60
Bibliografía	61
Apéndice A.1 Glosario	66

Índice de Figuras

Figura 1.1. Diagrama general del sistema de vigilancia [1].....	3
Figura 1.2 Diagrama de segundo nivel para el módulo de detección [1].....	3
Figura 1.3. Diagrama general del prototipo de reconocimiento.....	6
Figura 1.4. Estructuración del prototipo de reconocimiento.....	6
Figura 2.1. Estructura general de un banco de filtros [12].....	10
Figura 2.2. Implementación de un banco de filtros uniforme.....	11
Figura 2.3. Distribución del espectro para bancos de filtros uniformes.....	12
Figura 2.4. Estructura de un banco de filtros en codificación subbanda [12].....	13
Figura 2.5. División del espectro de frecuencia en un banco de codificación subbanda [12].....	13
Figura 2.6. Respuesta de los filtros espejo en cuadratura.....	14
Figura 2.7. Proyección de datos de dos clases en una línea en dirección de w [25].....	17
Figura 2.8. Secuencia de operaciones para el cálculo de la variable de adelanto $\alpha_{t+1}(j)$. [29].....	27
Figura 2.9. Secuencia de operaciones para el cálculo de $\xi_i(i,j)$ [29].....	28
Figura 3.1. Estructura interna del Control Automático de Ganancia.....	31
Figura 3.2. Configuración analógica del promediador [4, 5].....	32
Figura 3.3. Respuesta del detector ante una entrada sinusoidal.....	32
Figura 3.4. Algoritmo computacional para el promediador.....	34
Figura 3.5. Respuesta del promediador ante una señal sinusoidal.....	34
Figura 3.6. Respuesta del promediador con una constante de atenuación de 0.01.....	36
Figura 3.7. Respuesta del promediador con una constante de atenuación de 0,00058.....	36
Figura 3.8. Respuesta del control de ganancia con una constante de normalización de 0.1 y una constante de atenuación de 0.00058.....	37
Figura 3.9. Respuesta del control de ganancia con una constante de normalización de 0.00001 y una constante de atenuación de 0.00058.....	38

Figura 4.1. Respuesta en frecuencia y respuesta en cuadratura en decibeles para los filtros QMF elípticos de tercer orden.....	42
Figura 4.2. Estructura del banco de filtros.	43
Figura 4.3. Respuesta de un banco de filtros de cuatro niveles.	43
Figura 4.4. Distribución de pesos en las bandas de frecuencia.	45
Figura 4.5. Distribución de una señal símbolos de observación..	47
Figura 5.1. Estructura de la etapa de clasificación.....	49
Figura 5.2. Porcentaje de reconocimiento para modelos con distinto número de estados.....	51
Figura 5.3. Porcentajes de reconocimiento para modelos de tres estados con tasas de muestreo de 44100Hz, 10Hz, 20Hz y 100Hz.....	53
Figura 5.4. Variación del porcentaje de reconocimiento de los modelos en función del tamaño de la cadena de observaciones.....	55

Índice de Tablas

Tabla 3.1. Variables presentes en la etapa de preproceso.....	35
Tabla 4.1. Coeficientes para los filtros elípticos QMF de tercer orden.....	41
Tabla 4.2. Características de frecuencia de la base de datos para el proyecto.....	44
Tabla 4.3. Resultados del análisis de discriminante lineal de Fisher.....	44
Tabla 4.4. Bandas de frecuencia utilizadas en el banco de filtros.....	46
Tabla 5.1. Porcentaje de reconocimiento para modelos con distinto número de estados.....	50
Tabla 5.2. Porcentajes de reconocimiento para los modelos con tres estados y tasas de muestreo de 44100Hz y 10Hz.....	52
Tabla 5.3. Porcentajes de reconocimiento para los modelos con tres estados y tasas de muestreo de 20Hz y 100Hz.....	52
Tabla 5.4. Porcentajes de reconocimiento para cadenas de observación de una y dos muestras.....	54
Tabla 5.5. Porcentajes de reconocimiento para cadenas de observación de tres y cuatro muestras.....	54
Tabla 5.6. Porcentajes de reconocimiento para cadenas de observación de cinco muestras.....	54
Tabla 6.1. Variables para la etapa de extracción de características.....	57
Tabla 6.2. Variables presentes en la etapa de clasificación.....	58
Tabla 6.3. Comparación entre los prototipos de detección anteriores y el proyecto actual.....	59

Capítulo 1: Introducción

1.1 Tala ilegal, caza furtiva e incendios forestales en Costa Rica

Costa Rica es pionera en prácticas de conservación aprovechando el 24% de su territorio (de 51.100km²), el cual pertenece a zonas protegidas de algún tipo (parques nacionales, refugios de vida silvestre, etc.) [6, 7] beneficiándose económicamente de sus condiciones ecológicas mediante prácticas de Desarrollo Sostenible, como la comercialización de Certificados Transferibles de Emisiones de Carbono (CTO's), los pagos por servicios ambientales (PSA's), y el ecoturismo [6, 9, 10]. Se dice que los ecosistemas (bosques, plantaciones forestales, manglares, humedales, arrecifes u otros) brindan bienes y servicios en el ámbito local, nacional y mundial.

Las prácticas humanas, legales e ilegales (como los incendios provocados, la caza y la tala), y los fenómenos climáticos globales están dañando los ecosistemas, en donde uno de los indicadores del deterioro ambiental es la desaparición de especies. Solo en Costa Rica un total de 240 especies se encuentran en peligro de extinción al 2006, mientras que otras han desaparecido por completo, en donde la destrucción y la degradación de los hábitat, la agricultura no sostenible, la sobre caza, la contaminación y las enfermedades están dentro de las principales causas [11]. Por otro lado el 95% de los incendios forestales están relacionados con actividades agropecuarias y en menor porcentaje, con prácticas de cazadores, venganzas y negligencias [8].

Actividades lucrativas como la caza y la tala ilegales unidas a circunstancias como el poco personal de vigilancia y las grandes extensiones de territorio que deben cubrir, provocan que el comercio ilegal de maderas, la extinción de especies animales y la reducción de las zonas protegidas (debido a la deforestación y los incendios) produzcan un desequilibrio de los ecosistemas y por ende un impacto en las actividades económicas y ambientales del país.

1.1.1 Detección de actividades de tala y caza en zonas protegidas

Surge entonces la necesidad de detectar acciones de tala y caza que se presentan en las zonas protegidas de Costa Rica para frenar el deterioro que estas ocasionan a la naturaleza y proporcionar a los encargados de vigilancia una herramienta para la conservación de los bosques.

Dentro de las opciones que se manejan para la detección se presentan los métodos ópticos (mediante una red de cámaras que cubran la zona protegida), los métodos acústicos (distinguiendo sonidos que caractericen la presencia de estas actividades dentro del bosque) o mediante métodos sensoriales (vigilando variaciones de temperatura, presión, etc. que permitan detectar actividades de tala y caza).

Los métodos ópticos permiten observar lo que sucede dentro del bosque, sin necesidad de encontrarse en el lugar. Además, se puede obtener información que permita castigar a los infractores en caso de no llegar a tiempo al sitio. Sin embargo inconvenientes como el radio de cobertura y la necesidad de la línea vista lo hacen poco factible para colocarlo en zonas boscosas.

Actualmente no se conoce qué variables como temperatura, presión, etc. pueden ser analizadas para detectar tala y caza en el ambiente del bosque, por lo que en este momento, no se podrían utilizar métodos sensoriales en un sistema de vigilancia de zonas protegidas.

Un sistema de detección por métodos acústicos tiene la ventaja de tener un mayor radio de cobertura que los métodos ópticos. Se puede realizar un sistema omnidireccional, y no es necesario encontrarse en línea vista con el objetivo. Además, se necesita un menor número de componentes para realizar un transductor de sonido (un micrófono) que para realizar un transductor óptico (cámaras, sistemas de iluminación, filtros de luz, etc.). Dentro de las desventajas que presenta un sistema acústico están: el no poder observar las actividades en el espacio vigilado y el no obtener información de los infractores, argumentos que sí se podrían conseguir por métodos ópticos.

Dado que se puede lograr una cobertura mayor mediante un sistema omnidireccional con una cantidad menor de componentes; se decide realizar un sistema capaz de detectar por medio de sonidos de motosierras y disparos las actividades de tala y caza dentro de zonas protegidas.

1.1.2 Trabajos previos

Este proyecto forma parte de un sistema de vigilancia de zonas protegidas [2, 3] que consta de tres bloques generales (alimentación, detección y comunicación) que se muestran en la Figura 1.1.

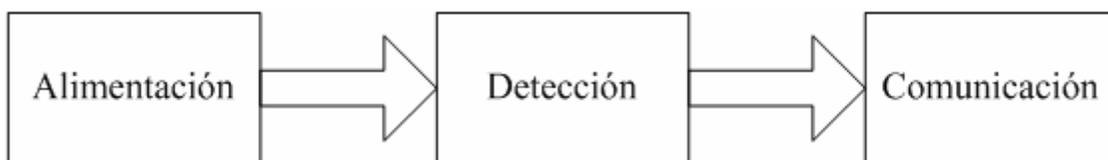


Figura 1.1. Diagrama general del sistema de vigilancia [1].

El bloque de detección consta de tres procesos: la detección de motosierras, la detección de disparos y la detección de fuego (ver Figura 1.2).

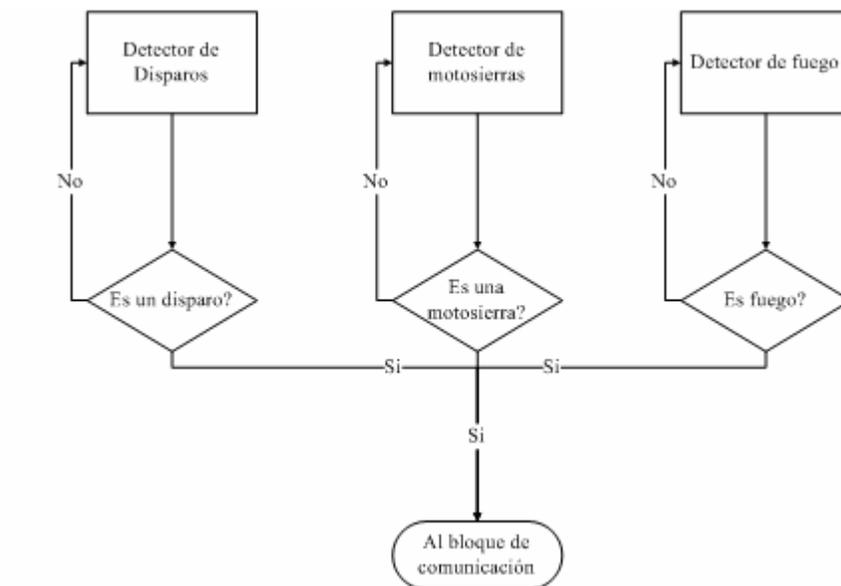


Figura 1.2 Diagrama de segundo nivel para el módulo de detección [1].

Para el sistema de vigilancia se han realizado tres prototipos de reconocimiento de disparos y motosierras (dos analógicos y uno digital). Los prototipos analógicos para la detección de motosierras y la detección de disparos de armas de fuego, se basaron en la teoría de las redes neuronales artificiales para la detección de señales acústicas, las cuales implementaron discriminadores Fisher. Los prototipos constaron de una etapa de acondicionamiento de la señal, un banco de ocho filtros activos y una neurona artificial ajustada por medio de potenciómetros. La simulación y el “entrenamiento” para determinar los parámetros de los diseños se realizaron mediante software en MatLab. Entre las características de los prototipos está la cobertura de 2862m^2 (radio de 30m), un porcentaje de error de un 12% en las detecciones de motosierras y un 10% en las de disparos y un consumo típico de potencia de 1.8W en ambos [4, 5].

La primera implementación algorítmica digital para la etapa de reconocimiento del proyecto, se basó en la teoría “wavelet” para la fase de extracción de características y para la fase de clasificación se utilizaron Modelos Ocultos de Markov (HMM). Como características generales del proyecto se obtuvo una implementación realizada fuera de línea utilizando paquetes de MatLab para el diseño de “wavelets” y los modelos de Markov. En [1] se explica su utilización pero no se presentan resultados con respecto a los diseños utilizados. Para este prototipo se realizaron pruebas y entrenamientos con señales acústicas de disparos de armas de fuego y motosierras a diferentes distancias del sensor grabadas directamente en el bosque.

En síntesis, el problema está en diseñar un sistema capaz de detectar sonidos de motosierras y disparos para indicar actividades ilegales de tala y caza dentro de las zonas protegidas del país. Este debe mejorar los prototipos hasta ahora diseñados, en aspectos como:

- Rango de detección.
- Desempeño en cuanto al porcentaje de detección.
- Costos de implementación.

1.2 Reconocimiento de patrones acústicos

Los sistemas de reconocimiento de patrones han probado ser eficientes en la solución de problemas de detección y clasificación [23, 24].

La meta de estos es asignar señales a distintas clases dependiendo de las características presentes en dichas señales, basándose en un conocimiento anterior de las mismas (entrenamiento) o mediante información estadística extraída de éstas (estimación) [23, 24, 39].

Estos sistemas han sido utilizados en campos como la seguridad y el reconocimiento del habla, en donde por ejemplo, [26] presenta una unidad acústica de vigilancia, para localizar vehículos que emiten sonidos entre 10Hz y 300Hz, así en [27] se utiliza en reconocimiento automático de lenguaje (ASR), o en [28] en donde el reconocimiento de patrones se utiliza para determinar la presencia de un instrumento musical y realizar divisiones por segmentos.

En su mayoría consisten de un transductor que recoge la señal por clasificar, un sistema de extracción de características que transforma la información obtenida en valores simbólicos, y un sistema de clasificación. Dependiendo de las aplicaciones, la señal del transductor es preprocesada para transformarla en un nuevo conjunto de variables y facilitar en cierta forma la solución del problema de reconocimiento [24, 39].

Considerando estas características y el origen de las señales por clasificar se divide el problema en tres etapas generales:

1. Etapa de preproceso de las señales acústicas.
2. Etapa de extracción de características de las señales acústicas.
3. Etapa de clasificación para determinar si la señal acústica pertenece a un disparo, una motosierra o al ambiente normal del bosque.

La estructura general del prototipo se presenta en la Figura 1.3, en donde se muestra la conexión de las etapas y el flujo de las señales entre éstas.

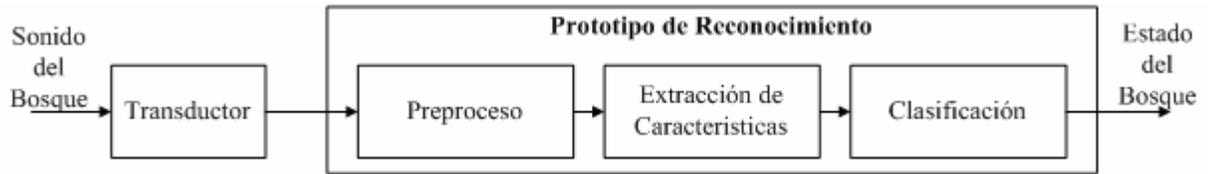


Figura 1.3. Diagrama general del prototipo de reconocimiento.

Cada una de las etapas se divide en módulos que se encargaron de realizar tareas específicas del proceso. La Figura 1.4 muestra la división de las etapas en cinco módulos generales.

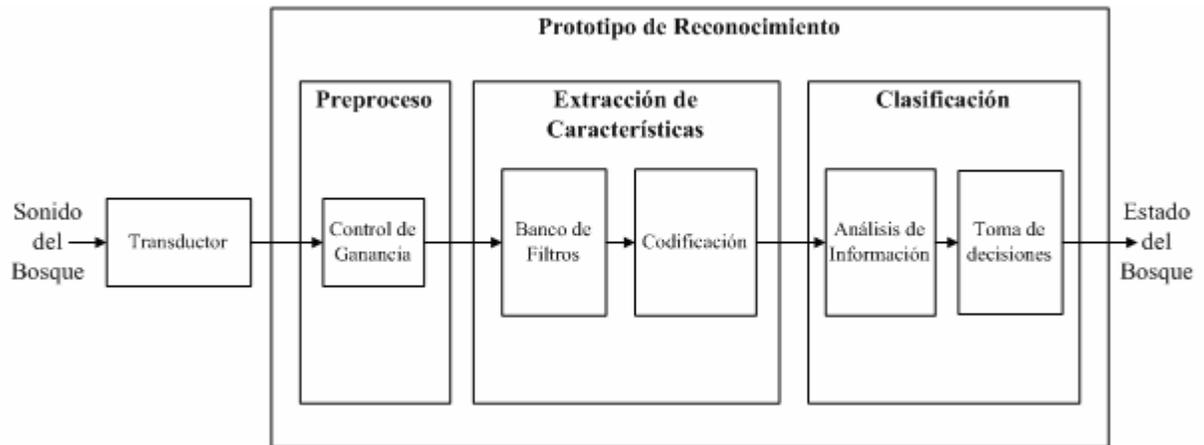


Figura 1.4. Estructuración del prototipo de reconocimiento.

El control de ganancia, se encarga de normalizar la amplitud de la señal de audio, el banco de filtros toma la señal normalizada y la divide en bandas de frecuencia, el módulo de codificación toma estas bandas de frecuencia y convierte estos datos en un grupo de símbolos que luego son interpretados por el módulo de análisis y finalmente el módulo de toma de decisiones indica el estado del bosque.

1.3 Objetivo y estructura del documento

El objetivo de este proyecto es diseñar un prototipo algorítmico para la detección de disparos y motosierras dentro de zonas protegidas, que considere las restricciones de:

- Analizar digitalmente señales acústicas, en tiempo real y
- Desarrollar algoritmos de poca exigencia computacional.

En el capítulo 2 se presentan principios de filtrado digital, discriminación lineal, aglomeración y la teoría matemática de los modelos ocultos de Markov.

Los capítulos 3, 4 y 5 explican el diseño y la evaluación de cada una de las etapas del prototipo de reconocimiento.

Por último, en el capítulo 6 se enumeran las conclusiones obtenidas y algunas recomendaciones para el mejoramiento del prototipo.

Capítulo 2: Marco Teórico

En este capítulo se resumen los principios físicos y matemáticos en los que se basa el proyecto, se presenta teoría de filtrado digital, discriminación lineal, aglomeración y los modelos ocultos de Markov.

2.1 Extracción de características de la señal acústica

Con la etapa de extracción se busca una representación apta de las señales acústicas de disparos, motosierras y el ambiente del bosque, para ser clasificadas. Como se muestra en la Figura 1.4 esta consta de dos módulos, un banco de filtros y un codificador, en esta sección se presentan implementaciones de bancos de filtros digitales, teoría de discriminación lineal y principios de algoritmos de aglomeración como codificadores.

2.1.1 Representación de señales acústicas por su contenido de frecuencias

Los sonidos están constituidos por tres características: la intensidad, la frecuencia fundamental y el número de armónicos. La intensidad está determinada por la amplitud del movimiento oscilatorio, la frecuencia fundamental indica el tono, es decir la manera en la que se percibe un sonido como grave o agudo, mientras que el número de armónicos permite distinguir entre dos sonidos de la misma intensidad y frecuencia fundamental [42].

Aprovechando las características de frecuencia fundamental y número de armónicos, es posible analizar las frecuencias que componen los sonidos para clasificarlos. Entre los métodos de análisis de frecuencias en señales se encuentran: La transformada de Fourier, la transformada de pequeña onda y los bancos de filtros.

La Transformada de Fourier, de la cual se derivan sus aplicaciones discretas (DFT y la STFT), es utilizada para representar señales por medio de su contenido de frecuencias. Este tipo de análisis convierte la información en el dominio del tiempo en información en el dominio de la frecuencia. Esta transformación hace que se pierda información que indica eventos que caracterizan a las señales de origen aleatorio en el tiempo, es decir no se conoce en qué momento se presenta cierta frecuencia, por lo que este método no es apto para sistemas de reconocimiento, ya que estos detalles son necesarios para identificar entre señales. [12, 41].

Los sistemas de análisis mediante transformación de pequeña onda (“*wavelets*”), se basan en dilataciones y traslaciones de una función madre (llamada “*wavelet madre*”), cuyos coeficientes se pueden calcular mediante filtros. Este método permite determinar que frecuencias se presentan durante un intervalo de tiempo, por lo que se maneja información de tiempo y frecuencia [41, 43].

Aunque esta transformación contiene información en el dominio del tiempo, presenta un problema en implementaciones de tiempo real, debido a la utilización de ventanas de muestreo, en donde su tamaño se encuentra enlazado al número de niveles de transformación. Dada esta condición, y los requisitos de diseño resumidas en las secciones 1.1.2 y 1.3, la utilización de ventanas provoca que, en términos de costo de implementación y procesamiento en tiempo real, este método no sea utilizable.

Lo que se busca entonces, es un sistema en el que se pueda manejar información del tiempo y la frecuencia, en donde su implementación pueda realizarse en tiempo real y no requiera de almacenar un número grande de muestras. Dado esto, los bancos de filtros resultan una solución factible en términos de representación tiempo-frecuencia, pueden realizarse en tiempo real sin necesidad de crear ventanas de muestreo y además por medio de ecuaciones de diferencias resultan fáciles de implementar [12].

Bancos de filtros digitales

Cuando un banco de filtros se encarga de dividir una señal en bandas de frecuencias se conoce como *banco de filtros de análisis* y se usan comúnmente en sistemas de análisis espectral [12]. Estos consisten en un conjunto de N filtros caracterizados por su respuesta en frecuencia $H_k(z)$ ($k = 0, 1, \dots, N-1$), tal como se muestra en la Figura 2.1.

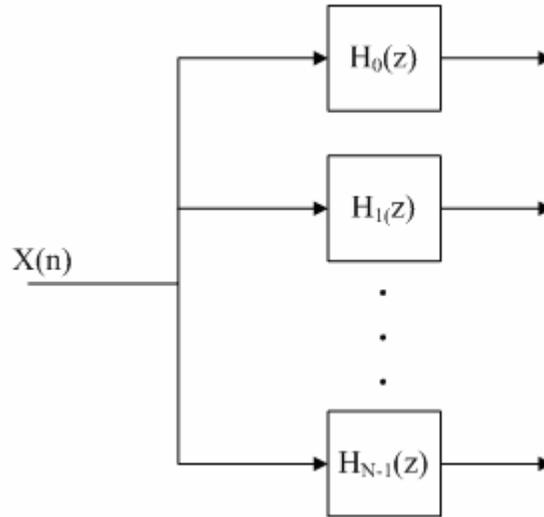


Figura 2.1. Estructura general de un banco de filtros [12].

La implementación de los bancos de filtros de análisis se basa en el diseño de cada filtro $H_k(z)$. A continuación se presentan dos tipos: *los bancos de filtros uniformes* y *la codificación subbanda*.

Bancos de filtros uniformes

Un banco de filtros se considera uniforme cuando los $H_k(z)$ filtros se derivan a partir de un filtro prototipo $H_0(z)$, donde [12]

$$H_k(\omega) = H_0\left(\omega - \frac{2\pi k}{N}\right) \quad k = 0, 1, \dots, N-1 \quad (2.1)$$

Cada una de las $H_k(z)$ respuestas en frecuencia se obtienen desplazando uniformemente la respuesta del filtro prototipo cada $2\pi k/N$. En el dominio del tiempo las respuestas de cada filtro se pueden expresar como [12]:

$$h_k(n) = h_o(n) e^{j2\pi nk/N} \quad k = 0, 1, \dots, N-1 \quad (2.2)$$

Una implementación de este tipo de banco es la que se muestra en la Figura 2.2

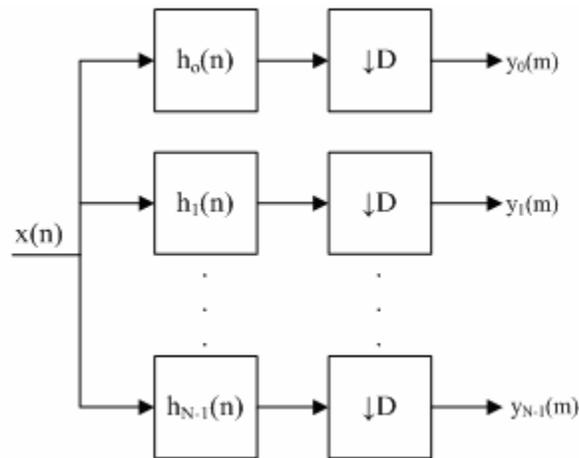


Figura 2.2. Implementación de un banco de filtros uniforme.

Este toma la secuencia $\{x(n)\}$ y la pasa a través de cada filtro $h_k(n)$ para luego realizar una etapa de *diezmado*, esto para mantener la misma cantidad de muestras a la entrada y a la salida del sistema. La señal de salida resultante diezmada se expresa como [40]:

$$y_k(m) = \sum_n h_k(n) x(mN - n) \quad (2.3)$$

Donde $\{y_k(m)\}$ es la salida del k -ésimo filtro paso banda diezmada por un factor de N .

Estos bancos de filtros dividen el espectro de la señal en bandas simétricas, tal como se muestra en la Figura 2.3.

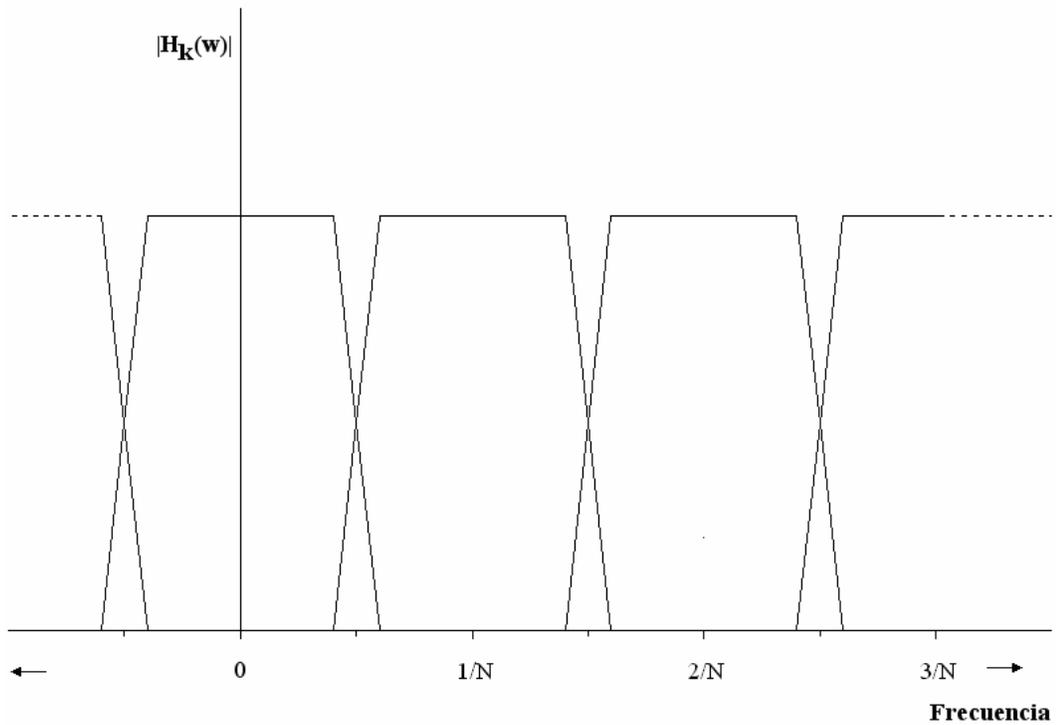


Figura 2.3. Distribución del espectro para bancos de filtros uniformes.

Al momento de diseñar el filtro prototipo es necesario escoger una frecuencia de corte tal que $\omega_0 \leq \pi/N$ y características de corte tales que minimicen la aparición de *alias* en las bandas adyacentes.

Si bien el diezmado es utilizado para trabajar con un mismo número de muestras y no aumentarlo N veces a la salida del banco de filtros, se puede realizar una implementación sin esta etapa y no afectar su comportamiento, por lo que la salida de cada filtro queda como la convolución de $x(n)$ y $h(n)$

$$y_k(n) = h_k(n) * x(n) \quad (2.4)$$

En [12] se explica la utilización de este tipo de banco de filtros para producir una transformación discreta de Fourier (DFT) de una secuencia finita de datos.

Codificación subbanda

Su nombre se debe a que son generalmente utilizados en sistemas de codificación de sonidos, aprovechando la característica que estos poseen de contener la mayoría de la energía en las frecuencias más bajas, los codificadores subbanda utilizan una estructura como la de la Figura 2.4 [12].

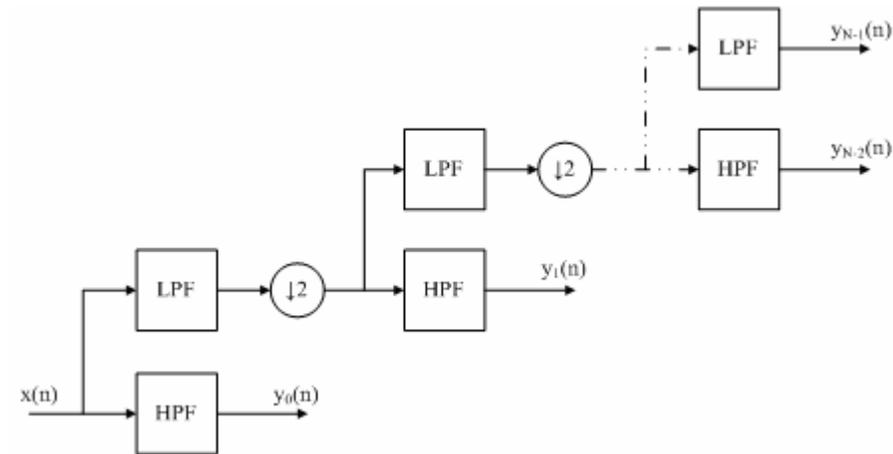


Figura 2.4. Estructura de un banco de filtros en codificación subbanda [12]

La división del espectro de frecuencias se realiza en octavas, en la Figura 2.5 se muestra la salida de un banco de filtros de 4 bandas.

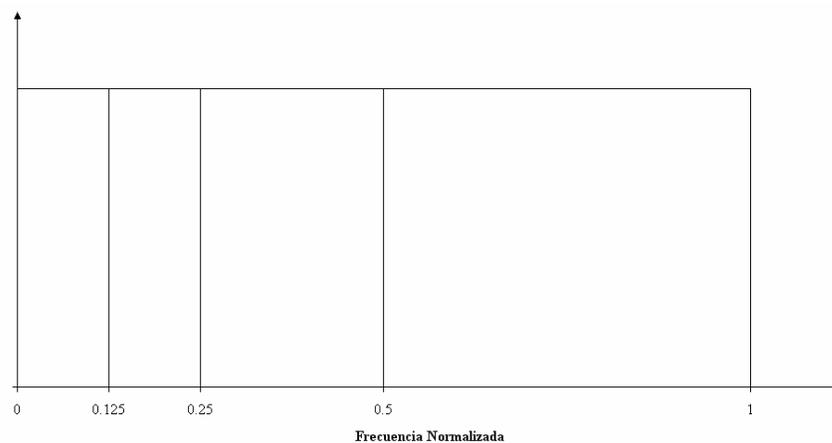


Figura 2.5. División del espectro de frecuencia en un banco de codificación subbanda [12].

Estos toman la señal dada por la secuencia $\{x(n)\}$ y realizan una primera división del espectro en dos, una parte con las componentes altas de frecuencia que van a la salida del banco y una parte con las componentes bajas que es enviada a una etapa de diezmado por 2, que reduce la tasa de muestreo para luego entrar a una nueva etapa de filtrado que separa estas componentes en dos nuevos grupos. El proceso se repite hasta que se alcance el total de bandas del banco de filtros [12].

La realización de los filtros en una codificación subbanda debe hacerse de manera tal que los *alias* producidos por la etapa de diezmado sean despreciables, una solución a este problema es la utilización de filtros espejo en cuadratura [12].

Filtros espejo en cuadratura

La respuesta en frecuencia de los filtros espejo en cuadratura (QMF) se muestra en la Figura 2.6, estos son complementarios, es decir, las frecuencias rechazadas por el filtro paso bajo son tomadas por el filtro paso alto y viceversa.

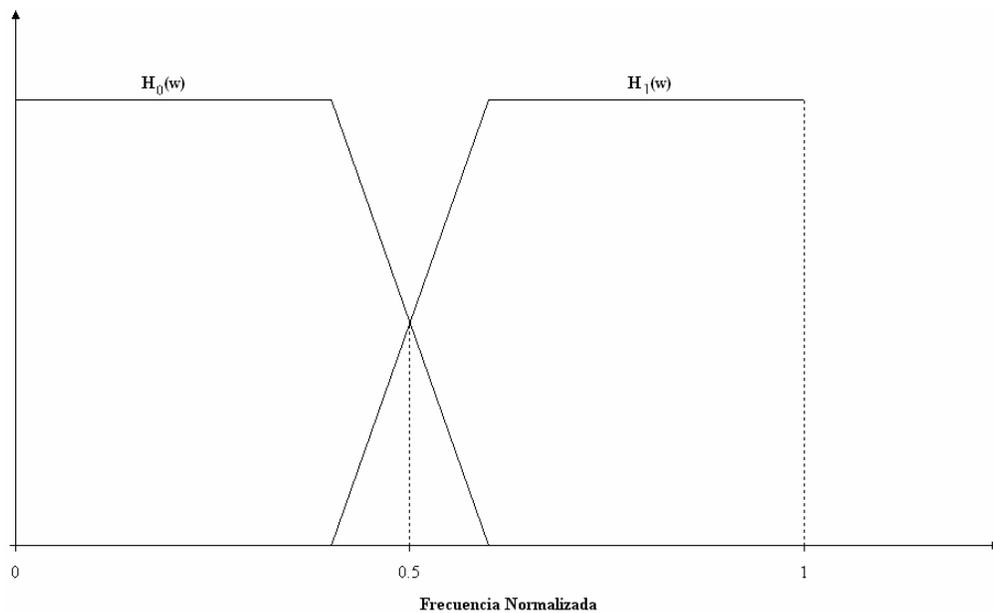


Figura 2.6. Respuesta de los filtros espejo en cuadratura.

La característica general de estos filtros, es que su respuesta total en magnitud cumple con

$$A(\omega) = |H_0(\omega)|^2 + |H_1(\omega)|^2 \approx 1 \quad \text{para todo } \omega \quad (2.5)$$

Estos filtros pueden ser tanto de respuesta finita al impulso (FIR) como de respuesta infinita al impulso (IIR), con sus debidas ventajas y desventajas. En [12] se aborda el diseño de QMF con filtros FIR, mientras que en [37,38] se presenta el diseño de QMF con filtros IIR.

En [12, 15-22] se presentan más detalles acerca del filtrado digital, tipos de filtros así como técnicas de diseño.

La teoría de sistemas de tasa múltiple, diezmado e interpolación de señales digitales se presenta en [12, 40] así como ejemplos e implementaciones.

2.1.2 Selección de bandas de frecuencias y reducción de dimensiones

Para seleccionar el número de bandas de frecuencias que conforman el banco de filtros, es preciso aplicar un método que indique cuáles frecuencias son necesarias para realizar un sistema de clasificación.

Dentro de estos algoritmos se encuentra el método de discriminante lineal de Fisher. Este método indica qué frecuencias son útiles para realizar una clasificación en un número determinado de clases, realiza un mapeo de los datos en un hiperplano y mediante un vector de pesos indica qué dimensiones son necesarias para realizar la mayor separación de los datos en ese hiperplano. A continuación se presenta en manera resumida la teoría del discriminante lineal de Fisher.

Discriminantes lineales de Fisher

La meta de los algoritmos de clasificación es asignar a un vector de entrada x una de C_k clases, donde $k = 1, 2, \dots, K$ y K es el número total de clases. En el mejor de los casos a este vector x le será asignada una y solo una de las clases C_k . [24].

Estos algoritmos dividen el espacio de entrada en regiones de decisión cuyos límites son llamados límites de decisión o superficies de decisión. Cuando los datos pueden ser separados por superficies de decisión lineales se dice entonces que el conjunto es linealmente separable y a los algoritmos que los dividen se les conoce como discriminantes lineales [24].

Los discriminantes lineales de Fisher atacan el problema de clasificación realizando modelos que reducen la dimensionalidad del espacio, y creando regiones de decisión mediante la proyección de los datos en un hiperplano que mejor separe las C_k clases.

Este método determina cuáles de las dimensiones poseen un mayor “peso” al momento de realizar una separación de los datos y, aunque por sí solo no constituye un sistema de reconocimiento completo, es generalmente utilizado para disminuir la complejidad computacional de los algoritmos de clasificación ya que ésta aumenta con la cantidad de dimensiones que se procesen [23, 24].

En [23-25] se introduce teoría de los discriminantes lineales mientras que [4, 5] presentan ejemplos de utilización de los discriminantes lineales de Fisher en problemas de reconocimiento. En esta sección se presenta la información necesaria para entender el papel del discriminante lineal de Fisher en este proyecto, iniciando con su implementación para dos clases y luego su generalización para un mayor número de clases.

Discriminante lineal de Fisher para dos clases

Supóngase un conjunto \mathbf{x} de muestras d -dimensionales, conformado por n_1 muestras de un subconjunto D_1 y n_2 muestras de un subconjunto D_2 . Si se proyectan estos datos en una dimensión utilizando una combinación lineal de los componentes de \mathbf{x} dada por [23, 24]

$$\mathbf{y} = \mathbf{w}^T \mathbf{x} \quad (2.6)$$

Se obtiene un conjunto de n muestras unidimensionales $\mathbf{y} = y_1, y_2, \dots, y_n$ igualmente dividido en dos subconjuntos Y_1 y Y_2 .

Considerando que la proyección a una dimensión genera una pérdida importante de información dado que las clases D_1 y D_2 se encontraban separadas en el espacio original de d -dimensiones y que la proyección a una dimensión hace que se traslapen, se puede encontrar una proyección tal que ajustando los valores de \mathbf{w} se maximice la separación entre clases [23, 24]. En la Figura 2.7 se observa como modificando la dirección de \mathbf{w} se puede mejorar la separación entre los datos de dos clases.

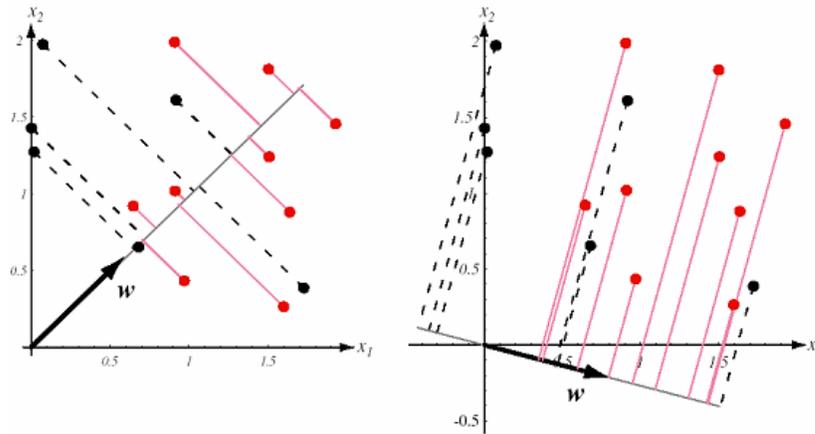


Figura 2.7. Proyección de datos de dos clases en una línea en dirección de \mathbf{w} [25].

Una medida de la separación de los datos es la diferencia entre las medias m_1 y m_2 de los conjuntos de datos D_1 y D_2 dadas por

$$m_1 = \frac{1}{n_1} \sum_{x \in D_1} x \quad m_2 = \frac{1}{n_2} \sum_{x \in D_2} x \quad (2.7)$$

Si se proyectan estas en el vector \mathbf{w} se obtiene una medida de las distancias de las medias dada por [23, 24]

$$|\tilde{m}_1 - \tilde{m}_2| = |\mathbf{w}^T (m_1 - m_2)| \quad (2.8)$$

Esto implica que escalando \mathbf{w} se puede aumentar esta separación. Sin embargo, lo que realmente dará una buena separación de los datos es qué tan lejos esté la media de una medida de desviación estándar para cada clase. En [23] se define la dispersión de las muestras proyectadas como:

$$\tilde{s}_i^2 = \sum_{y \in Y_i} (y - \tilde{m}_i)^2 \quad (2.9)$$

En donde $\frac{1}{n}(\tilde{s}_1^2 + \tilde{s}_2^2)$ es un estimado de la varianza y $\tilde{s}_1^2 + \tilde{s}_2^2$ es conocida como la dispersión total dentro de clases de las muestras proyectadas.

Función de criterio para el discriminante lineal de Fisher

La función de criterio de Fisher está dada por la razón de la dispersión entre clases dada por (2.8) y la dispersión dentro de clases [23, 24]

$$J(\mathbf{w}) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2} \quad (2.10)$$

En donde el vector \mathbf{w} que maximice $J(\cdot)$ es el que mejor separa las proyecciones Y_1 y Y_2 .

Explícitamente, se puede escribir la función de criterio en función de \mathbf{w} como

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \quad (2.11)$$

En donde \mathbf{S}_B es la matriz de dispersión entre clases y esta dada por

$$\mathbf{S}_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^T \quad (2.12)$$

Y \mathbf{S}_W es la matriz de dispersión dentro de clases y está dada por

$$S_W = \sum_{x \in D_1} (x - m_1)(x - m_1)^T + \sum_{x \in D_2} (x - m_2)(x - m_2)^T \quad (2.13)$$

Y el vector w que maximiza $J(\cdot)$ debe satisfacer:

$$w = S_W^{-1}(m_1 - m_2) \quad (2.14)$$

Discriminante lineal de Fisher para múltiples clases

Para un problema de C-clases la generalización natural de los discriminantes lineales de Fisher envuelve C-1 funciones discriminantes. La generalización para la matriz de dispersión dentro de clases es [23, 24]

$$S_W = \sum_{i=1}^C S_i \quad (2.15)$$

Donde

$$S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^T \quad (2.16)$$

Para S_B hay que suponer un vector de media total m y una matriz de dispersión total S_T como

$$m = \frac{1}{n} \sum_x x = \frac{1}{n} \sum_{i=1}^C n_i m_i \quad (2.17)$$

Y

$$S_T = \sum_x (x - m)(x - m)^T = S_W + \sum_{i=1}^C n_i (m_i - m)(m_i - m)^T \quad (2.18)$$

En donde el segundo término es llamado la matriz general de dispersión entre clases S_B , por lo que

$$S_T = S_W + S_B \quad (2.19)$$

La proyección de un espacio vectorial de d-dimensiones a uno de (c-1)-dimensiones se realiza con las c-1 funciones discriminantes

$$y_i = \mathbf{w}_i^T \mathbf{x} \quad \text{con } i = 1, \dots, c-1 \quad (2.20)$$

Si los y_i son vistos como componentes de un vector \mathbf{y} y los vectores de peso \mathbf{w}_i son vistos como columnas de una matriz de d por (c-1) \mathbf{W} , entonces la proyección se puede escribir como una operación matricial [23, 24]

$$\mathbf{y} = \mathbf{W}^T \mathbf{x} \quad (2.21)$$

Se puede demostrar además que:

$$\begin{aligned} \tilde{S}_W &= \mathbf{W}^T S_W \mathbf{W} \\ \tilde{S}_B &= \mathbf{W}^T S_B \mathbf{W} \end{aligned} \quad (2.22)$$

Función de criterio

Al igual que para dos clases se tiene una función de criterio para encontrar la matriz \mathbf{W} que mejor separe las C clases. Esta cumple con [23, 24]

$$J(\mathbf{W}) = \frac{|\tilde{S}_B|}{|\tilde{S}_W|} = \frac{|\mathbf{W}^T S_B \mathbf{W}|}{|\mathbf{W}^T S_W \mathbf{W}|} \quad (2.23)$$

La forma de encontrar el \mathbf{W} óptimo se resume en un problema de eigenvalores y eigenvectores, los cuales deben cumplir

$$S_B \mathbf{w}_i = \lambda_i S_W \mathbf{w}_i \quad (2.24)$$

Donde estos eigenvalores λ pueden encontrarse como las raíces del polinomio

$$|S_B - \lambda_i S_W| = 0 \quad (2.25)$$

Resolviendo la ecuación

$$(S_B - \lambda_i S_W)w_i = 0 \quad (2.26)$$

Se obtienen los vectores que constituyen la matriz de pesos \mathbf{W} .

2.1.3 Aglomeración y método de K-medias

Dentro del análisis de patrones se encuentran las técnicas de análisis de conglomerados (“*clustering analysis*”) las cuales se encargan de estudiar cómo se organiza una colección de patrones (generalmente representados como vectores o puntos en un espacio multidimensional) en conjuntos, basados en la similitud entre datos. Intuyendo que los patrones dentro de un grupo son más similares entre ellos que entre otros que pertenezcan a otros grupos [36].

Los algoritmos de aglomeración han demostrado ser útiles en sistemas de análisis de patrones incluyendo segmentación de imágenes y clasificación, particularmente en estudios de interrelación de los datos produciendo un acercamiento a la estructura de los mismos [24, 33 - 36].

En sí, la meta de los algoritmos de aglomeración, es encontrar un número de puntos dentro de un conjunto de datos que puedan representar dicho conjunto. Estos puntos son conocidos como centros, prototipos, centroides o palabras de código (“*codewords*”), en donde, luego de obtenerlos pueden usarse para tareas como la compresión de datos y la clasificación [33-36].

En [23, 24, 33-36] se presentan los principios de los algoritmos de aglomeración. En esta sección se presenta de manera resumida el algoritmo de K-medias.

Algoritmo de K-medias

La estimación inicial de los centroides es la base del algoritmo de K-medias por lo que una pobre inicialización puede generar problemas como los “centros muertos” o la redundancia [34].

Se dice que se genera un “centro muerto” cuando no se asignan puntos a un centroide, esto sucede cuando un centro se ubica entre dos centroides activos o bien alejado del conjunto de datos. Una manera de evitar esto es tomar los centros iniciales del conjunto de muestras de manera aleatoria, sin embargo esto no garantiza que los centros sean igualmente activos [34].

La redundancia se presenta al querer obtener un mejor modelado de los datos por lo que se asignan un número mayor de centroides para representar el conjunto de datos, pero un alto número de centroides incrementa la probabilidad de colocarlos en la misma posición o muy cercanos entre sí [34].

Dado esto se puede resumir que el problema de aglomeración por K-medias es en sí un problema de optimización [33 - 36].

El algoritmo de K-medias se describe como sigue [33, 34, 36]:

1. Seleccionar K centros iniciales para los conjuntos.
2. Asignar cada punto x del total de datos con el centroide más cercano.
3. Aplicar la función de criterio:

$$J = \sum_{j=1}^K \sum_{i=1}^N \|x_i - c_j\|^2 \quad (2.27)$$

4. Si J es mejorada entonces se calculan los nuevos centros con los datos asignados a cada conglomerado.

5. Repetir los pasos de 2 a 4 hasta que se acote el valor de J o hasta alcanzar un máximo de iteraciones.

El método de K-medias es explicado con detalle en [33-36] donde se muestran distintas formas de determinar las distancias de los datos así como su implementación en sistemas de clasificación y compresión.

2.2 Clasificación de la señal acústica

La etapa de clasificación se encarga de asignar a las señales acústicas provenientes del bosque, previamente tratadas por las etapas de preproceso y extracción de características, una de tres clases: disparos, motosierras o ambiente normal.

Dado que los sonidos presentes en un ambiente natural son aleatorios, se necesita de un sistema que pueda distinguir entre la cantidad de acontecimientos que se presenten dentro del bosque, de manera que se reduzca la aparición de falsas alarmas y los eventos no detectados.

Los modelos ocultos de Markov son una herramienta probabilística que permite el estudio de señales observables de procesos del mundo real. Éstas pueden ser puras o corruptas (por ruido, distorsiones o reverberación), de naturaleza discreta o continua y cuya fuente puede ser estacionaria o aleatoria. [29]

En esta sección se explican los conceptos básicos para la aplicación de los modelos de Markov en sistemas de reconocimiento. La teoría completa puede revisarse en [23, 24, 29] además [14, 30-32] presentan ejemplos de utilización de los HMMs en distintas aplicaciones así como bases teóricas.

2.2.1 Modelos Ocultos de Markov

Estos consisten de un sistema de estados los cuales están interconectados entre sí, en donde es posible moverse de cualquier estado a otro con transiciones regidas por densidades probabilísticas.

Elementos de un HMM

Un HMM discreto completo está definido por los siguientes elementos [23, 24, 29]:

1. **S**: el conjunto de estados en el modelo. Aunque estos están ocultos, para muchas aplicaciones prácticas existe algún significado físico enlazado a los estados o a su configuración en el modelo. Se denotan como $\mathbf{S} = \{S_1, S_2, \dots, S_N\}$ y en el tiempo t como q_t .
2. **V**: el conjunto de símbolos por estado y define el tamaño del alfabeto discreto. Los símbolos corresponden a la salida física del sistema por ser modelado, estos se denotan como $\mathbf{V} = \{V_1, V_2, \dots, V_M\}$.

3. $\mathbf{A} = \{a_{ij}\}$: la distribución de probabilidad para la transición entre estados, donde:

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i] \quad ; 1 \leq i, j \leq N \quad (2.28)$$

Para el caso especial donde cualquier estado puede alcanzar cualquier otro en un solo paso se tiene que $a_{ij} > 0$. Para otro tipo de HMM se puede tener un $a_{ij} = 0$ para uno o más pares (i, j) .

4. $\mathbf{B} = \{b_j(k)\}$: la distribución de probabilidad de un símbolo de observación en el estado j , donde:

$$b_j(k) = P[V_k \text{ ent} | q_t = S_j] \quad 1 \leq j \leq N, 1 \leq k \leq M \quad (2.29)$$

5. $\boldsymbol{\pi} = \{\pi_i\}$: la distribución para el estado inicial donde:

$$\pi_i = P[q_1 = S_i] \quad 1 \leq i \leq N \quad (2.30)$$

En resumen para una especificación completa de un HMM se necesitan establecer dos parámetros del modelo N y M , establecer los símbolos de observación \mathbf{V} y establecer las tres medidas de probabilidad \mathbf{A} , \mathbf{B} y $\boldsymbol{\pi}$. [29] utiliza la notación completa $\boldsymbol{\lambda} = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ para indicar los parámetros completos de un modelo dado, [23, 24] utilizan su propia notación pero en este documento se utiliza la notación dada por [29] para referirse a un modelo completo de Markov.

Los tres problemas básicos para los HMM

Existen tres problemas de interés que deben ser resueltos para que el modelo sea útil en el mundo real y estos son [23, 24, 29]:

Problema 1: Dada una secuencia de observaciones \mathbf{O} y un modelo λ , ¿cómo se computariza correctamente $P(\mathbf{O}|\lambda)$, la probabilidad de la secuencia dado λ ?

Problema 2: Dada una secuencia de observaciones \mathbf{O} y el modelo λ , ¿cómo se escoge una secuencia de estados $\mathbf{Q} = q_1, q_2, \dots, q_t$ que sea óptima de manera razonable, es decir, que mejor explique la secuencia \mathbf{O} ?

Problema 3: ¿Cómo se ajustan los parámetros del modelo $\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ para maximizar $P(\mathbf{O}|\lambda)$?

Solución a los tres problemas básicos de los HMM

A continuación se presentan las soluciones al primer y tercer problema, la solución al segundo problema no es necesaria en sistemas de reconocimiento de patrones por lo que no se explica en el presente documento. En [23, 24, 29] se explican las soluciones a los tres problemas básicos de los HMM.

Solución al primer problema: El procedimiento adelante-atrás (*Forward-Backward procedure*)

Aunque para resolver el problema 1 se necesita únicamente el procedimiento hacia adelante (*forward*), para la solución de los dos problemas restantes la implementación del procedimiento hacia atrás es necesaria, ambos procedimientos se explican en [23, 24, 29].

Procedimiento hacia adelante (*forward*)

Para este procedimiento se considera la variable de adelanto $\alpha_t(i)$ que se define como:

$$\alpha_t(i) = P(O_1 O_2 \cdots O_t, q_t = S_i | \lambda) \quad (2.31)$$

Que representa la probabilidad de la secuencia parcial de observaciones hasta el tiempo t y el estado S_i en el tiempo t , dado el modelo λ .

El cálculo de $\alpha_t(i)$ se realiza inductivamente en tres pasos como sigue:

1. Inicialización:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (2.32)$$

Se inicializa la variable de adelanto como la probabilidad de la unión del estado S_i y la observación inicial O_1 .

2. Inducción:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad 1 \leq t \leq T-1 \quad (2.33)$$
$$1 \leq j \leq N$$

Esta es la etapa central del procedimiento, en donde la Figura 2.8 muestra como es alcanzado el estado S_j en el tiempo $t+1$ desde cualquier estado de los N posibles.

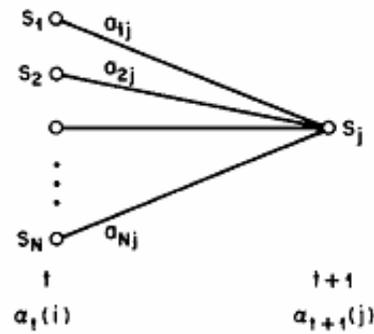


Figura 2.8. Secuencia de operaciones para el cálculo de la variable de adelanto $\alpha_{t+1}(j)$. [29]

3. Terminación

$$P(O | \lambda) = \sum_{i=1}^N \alpha_T(i) \quad (2.34)$$

Se realiza el cálculo de la probabilidad $P(O|\lambda)$ como la suma de las variables α en el instante T.

Procedimiento hacia atrás (*backward*)

De manera similar al procedimiento hacia adelante, se define la variable de atraso $\beta_t(i)$ definida como:

$$\beta_t(i) = P(O_{t+1} O_{t+2} \dots O_T | q_t = S_i, \lambda) \quad (2.35)$$

Y se resuelve inductivamente por dos pasos.

1. Inicialización:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (2.36)$$

2. Inducción:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \dots, 1 \quad (2.37)$$

$$1 \leq i \leq N$$

Solución al tercer problema: El método de Baum-Welch

No existe una manera analítica de establecer un modelo que maximice la probabilidad de una secuencia de observaciones; más aún, con un conjunto finito de datos de entrenamiento no hay manera óptima de estimar los parámetros del modelo.

Sin embargo, se puede obtener un $\lambda=(\mathbf{A},\mathbf{B},\boldsymbol{\pi})$ que haga que $P(\mathbf{O}|\lambda)$ sea máximo para el conjunto de datos de entrenamiento utilizando un procedimiento iterativo como el método de Baum-Welch.

Como primer paso del método se define $\xi_t(i,j)$ como la probabilidad de encontrarse en el estado S_i en el instante t y en estado S_j en el tiempo $t+1$ dado el modelo y la cadena de observaciones.

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \quad (2.38)$$

La Figura 2.9 muestra las operaciones necesarias para el cálculo de $\xi_t(i,j)$ en donde se puede observar la implicación de las variables de atraso y adelanto del procedimiento adelante-atrás.

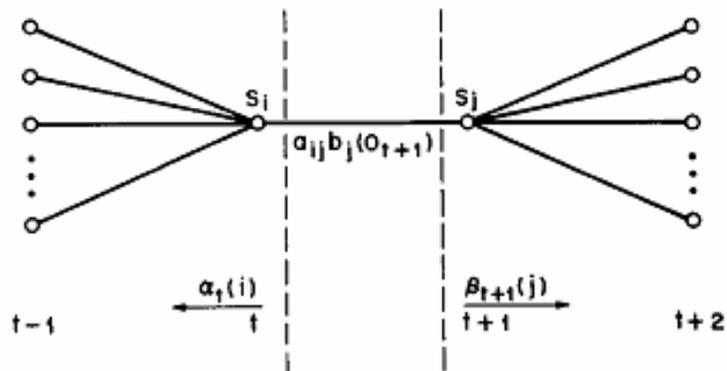


Figura 2.9. Secuencia de operaciones para el cálculo de $\xi_t(i,j)$ [29].

Conociendo esto puede escribirse $\xi_t(i,j)$ como:

$$\begin{aligned}\xi_t(i, j) &= \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O|\lambda)} \\ &= \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}\end{aligned}\quad (2.39)$$

Seguido el cálculo de $\xi_t(i,j)$ se define $\gamma_t(i)$ como la probabilidad de estar en el estado S_i en el tiempo t .

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (2.40)$$

Dadas estas probabilidades se pueden reestimar los valores de \mathbf{A} , \mathbf{B} y $\boldsymbol{\pi}$ como:

$$\bar{\pi}_i = \text{número esperado de veces en el estado } S_i \text{ en el tiempo } (t = 1) = \gamma_1(i) \quad (2.41)$$

$$\begin{aligned}\bar{a}_{ij} &= \frac{\text{número esperado de transiciones desde el estado } S_i \text{ al estado } S_j}{\text{número de transiciones desde el estado } S_i} \\ &= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}\end{aligned}\quad (2.42)$$

$$\begin{aligned}\bar{b}_j(k) &= \frac{\text{número esperado de veces que se observa el simbolo } V_k \text{ en el estado } S_j}{\text{número de veces en el estado } S_j} \\ &= \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^{T-1} \gamma_t(j)}\end{aligned}\quad (2.43)$$

Si se define el modelo actual como λ y se utiliza para el cálculo de las ecuaciones (2.41, 2.42 y 2.43) y se define el modelo re-estimado $\bar{\lambda}$, entonces se puede llegar a dos conclusiones:

1. El modelo inicial define un punto crítico en la función de probabilidad, en tal caso $\lambda = \bar{\lambda}$, o bien,
2. El modelo $\bar{\lambda}$ es más probable que el modelo λ , y se ha encontrado un nuevo modelo para el que es más probable que se produzca la secuencia de observaciones de entrenamiento.

El procedimiento puede continuar hasta que se alcance el punto crítico.

2.3 Antecedentes Bibliográficos.

Las redes neuronales artificiales, los Modelos Ocultos de Markov y los clasificadores estadísticos son ampliamente utilizados en el área de reconocimiento de patrones en donde el procesamiento de la señal depende completamente de la aplicación.

En [13,29-32] se presentan las bases teóricas de los Modelos Ocultos de Markov y su utilización en sistemas de reconocimiento del habla, además en [14] se explica la utilización de las cadenas de Markov en modelos de predicción de mercados.

Se utilizaron, además, bibliotecas computacionales en código MatLab para el desarrollo del prototipo, principalmente

- *Signal Processing Toolbox: IIR Filter Design*: Herramienta de diseño de filtros digitales de respuesta infinita al impulso [16].
- *MatLab Arsenal*: Contiene algoritmos de discriminación lineal [41].
- *HMM Toolbox for MatLab*: Herramientas para el diseño de modelos ocultos de Markov [42].

Capítulo 3: Diseño y evaluación de la etapa de preproceso

La etapa de preproceso consta de un módulo que se encarga de normalizar la señal de audio proveniente del módulo de transducción, tal como muestra la Figura 1.4, Este se conoce como módulo de control de ganancia y se presenta a continuación.

3.1 Control de Ganancia

La implementación del módulo de control de ganancia se basa en la ecuación

$$\text{Salida} = \frac{\text{Entrada}}{\text{Promedio} + \text{Constante}} \quad (3.1)$$

Mientras que su estructura interna se muestra en la Figura 3.1.

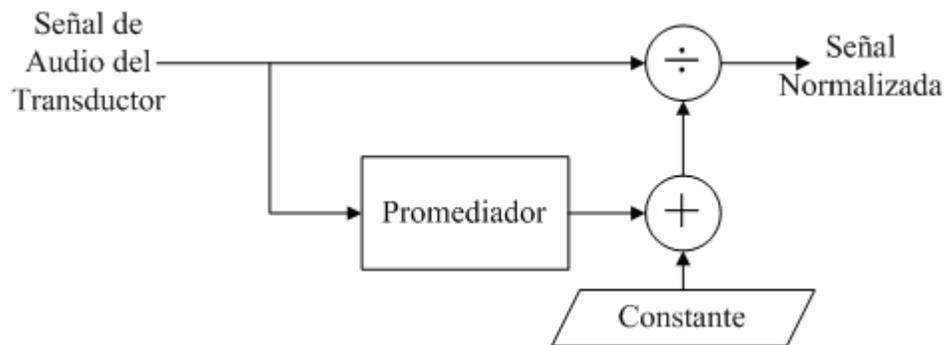


Figura 3.1. Estructura interna del Control Automático de Ganancia.

El diseño del módulo consiste en establecer el promediador así como el valor de la constante, conocida como constante de normalización.

3.1.1 Estructura del promediador

La estructura del promediador se basa en la presentada por [4, 5] donde, analógicamente, el promedio de la señal de entrada se tomó mediante un circuito detector como el de la Figura 3.2

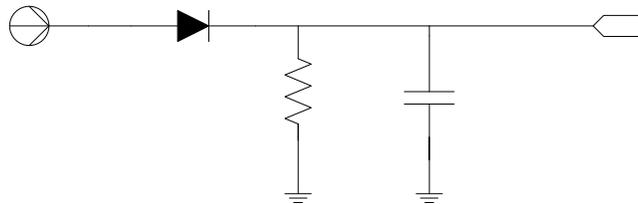


Figura 3.2. Configuración analógica del promediador [4, 5].

El comportamiento de este circuito ante una entrada sinusoidal se muestra en la Figura 3.3.

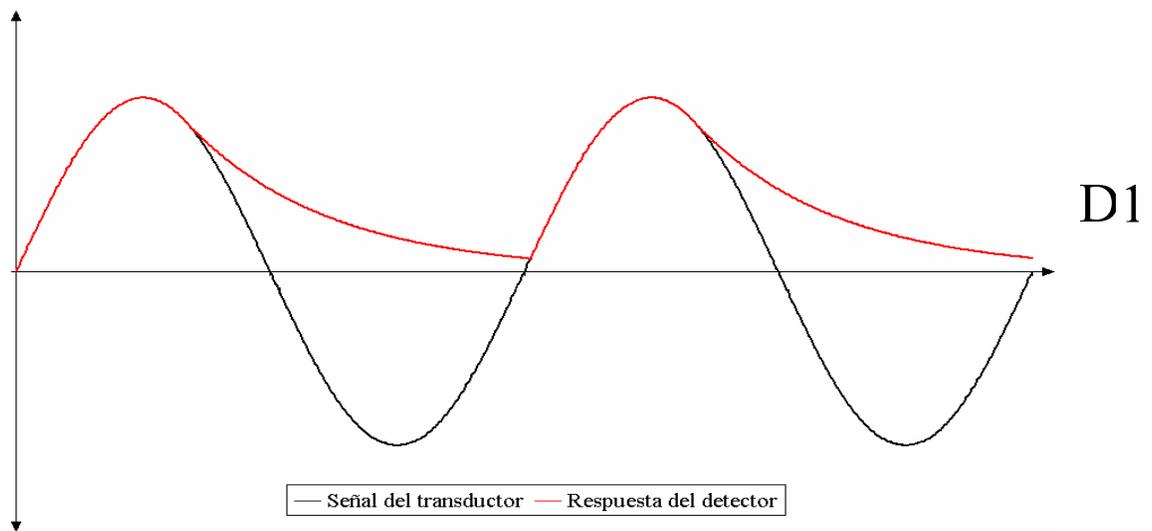


Figura 3.3. Respuesta del detector ante una entrada sinusoidal.

Asumiendo el diodo D1 como ideal y el condensador sin carga, se puede explicar el funcionamiento de este circuito de la siguiente manera.

Al ingresar un valor mayor a cero al módulo, el condensador se carga a este valor de entrada. Mientras que la señal del transductor sea menor que el valor de carga del condensador la resistencia descarga el condensador a una tasa $\tau = RC$ segundos hasta alcanzar un valor de carga del condensador que sea menor al valor de la entrada del módulo.

Este comportamiento se adapta a la aplicación mediante un algoritmo computacional. Este no toma enteramente el funcionamiento del circuito de la Figura 3.2, sino que se realiza una aproximación, esto porque además de las condiciones de funcionamiento se cumple con los siguientes requisitos de diseño:

- El algoritmo debe tener una descarga lineal: Esto para reducir la complejidad computacional del control de ganancia, ya que esto implica que la descarga se realiza mediante una suma y no mediante una multiplicación como lo establece el modelo discreto de un circuito RC.
- Además debe realizar una rectificación completa de la señal de entrada: Con esto se evitan saturaciones a la salida del módulo que se puedan ocasionar debido a variaciones mayores de energía en la parte negativa de la señal de entrada.

Dadas estas condiciones, el funcionamiento del algoritmo computacional se resume en la Figura 3.4. Mientras que la Figura 3.5 muestra la respuesta del promediador ante una señal sinusoidal.

La constante k , conocida como constante de atenuación, determina la tasa de disminución para la salida del promediador.

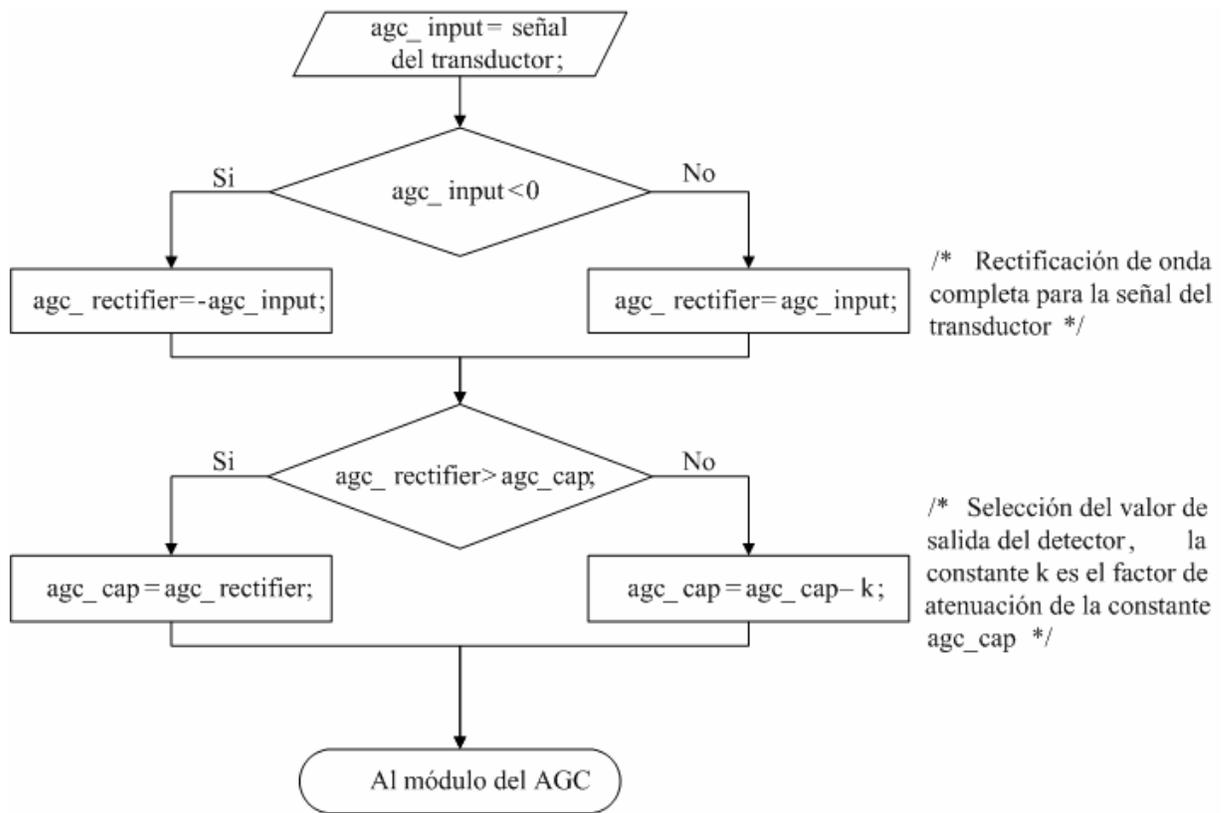


Figura 3.4. Algoritmo computacional para el promediador.

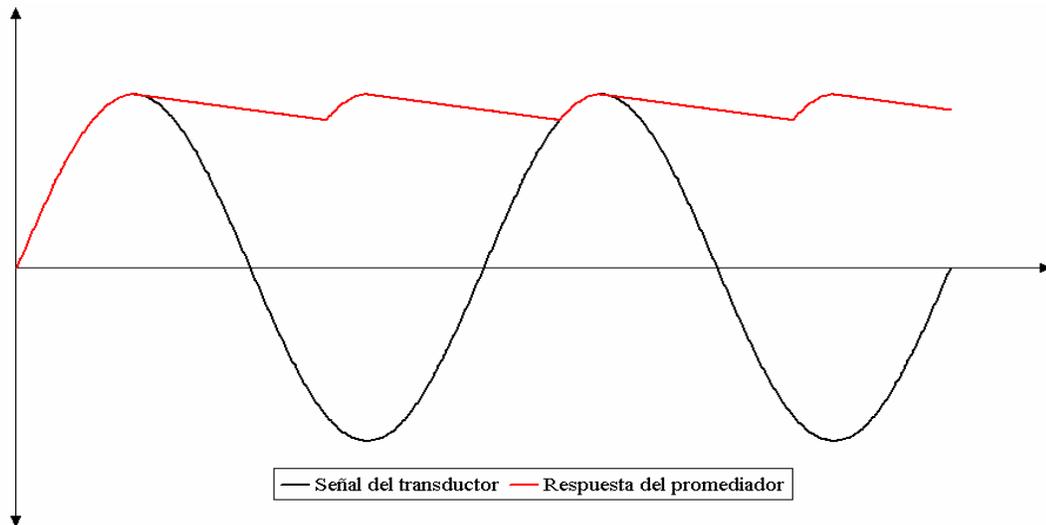


Figura 3.5. Respuesta del promediador ante una señal sinusoidal.

Para la etapa de preproceso se reconocen entonces dos valores que deben ser ajustados para obtener un comportamiento deseado del control de ganancia, la constante de normalización y la constante de atenuación, en la Tabla 3.1 se presentan las variables y su función dentro de la etapa.

Tabla 3.1. Variables presentes en la etapa de preproceso.

Variable	Función
Constante de Normalización	<ul style="list-style-type: none"> • Evitar la saturación de la salida del control de ganancia
Constante de Atenuación	<ul style="list-style-type: none"> • Determinar la tasa de disminución del valor de salida para el promediador

3.1.2 Ajuste de la constante de atenuación

El ajuste de la constante de atenuación sigue los lineamientos que se enumeran a continuación:

- El promediador no debe reaccionar a altas frecuencias: Mientras esto se dé, se tendrá una normalización a la salida del control de ganancia que no dependerá de los pequeños cambios de la señal sino que será gobernada por los picos de la señal de entrada.
- De igual manera la tasa de disminución debe ser tal que cuando se presente un salto brusco en la entrada (ej. un disparo a corta distancia) no se afecte los eventos siguientes otorgándoles una menor ganancia.

En resumen la constante de atenuación debe permitir que la salida del promediador se ajuste al contorno de la señal de entrada

El efecto de modificar el valor de la constante de atenuación sobre la salida del promediador se presenta a continuación.

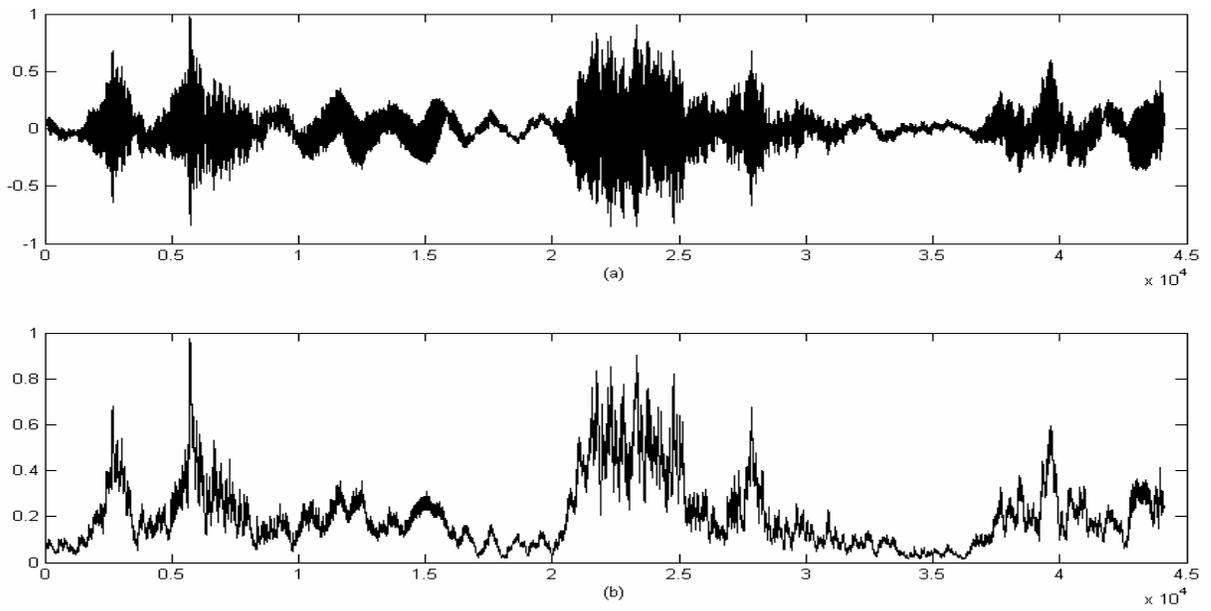


Figura 3.6. Respuesta del promediador con una constante de atenuación de 0.01.

La Figura 3.6.b presenta la salida del promediador con una constante de atenuación de 0.01, en esta se observa como la salida se ajusta a las variaciones de alta frecuencia de la entrada

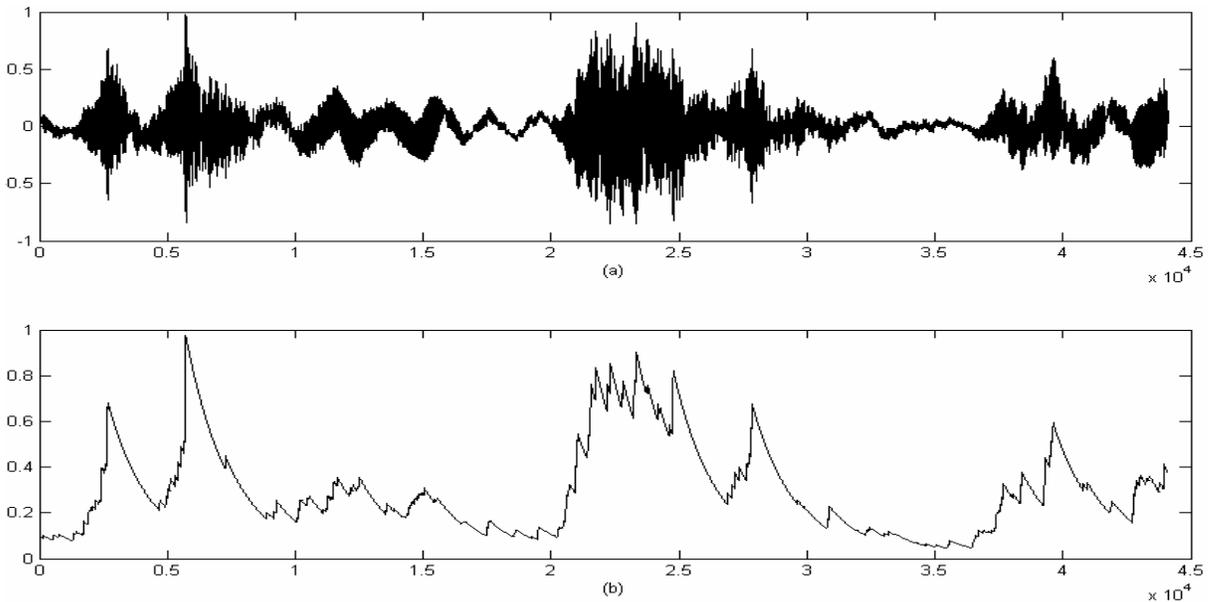


Figura 3.7. Respuesta del promediador con una constante de atenuación de 0,00058

En el caso de la Figura 3.7.b se presenta la salida del promediador con una constante de atenuación de 5.8×10^{-4} , en este caso se observa que el valor de salida del promediador se ajusta al contorno de la señal de entrada provocando que el promediador sea menos susceptible a las variaciones de alta frecuencia.

3.1.3 Ajuste de la constante de normalización

La constante de normalización, se utiliza para prevenir la saturación de la salida del control de ganancia por valores del promediador iguales o cercanos a cero. Esta debe ser tal que no afecte el valor del promedio que determina la ganancia del módulo.

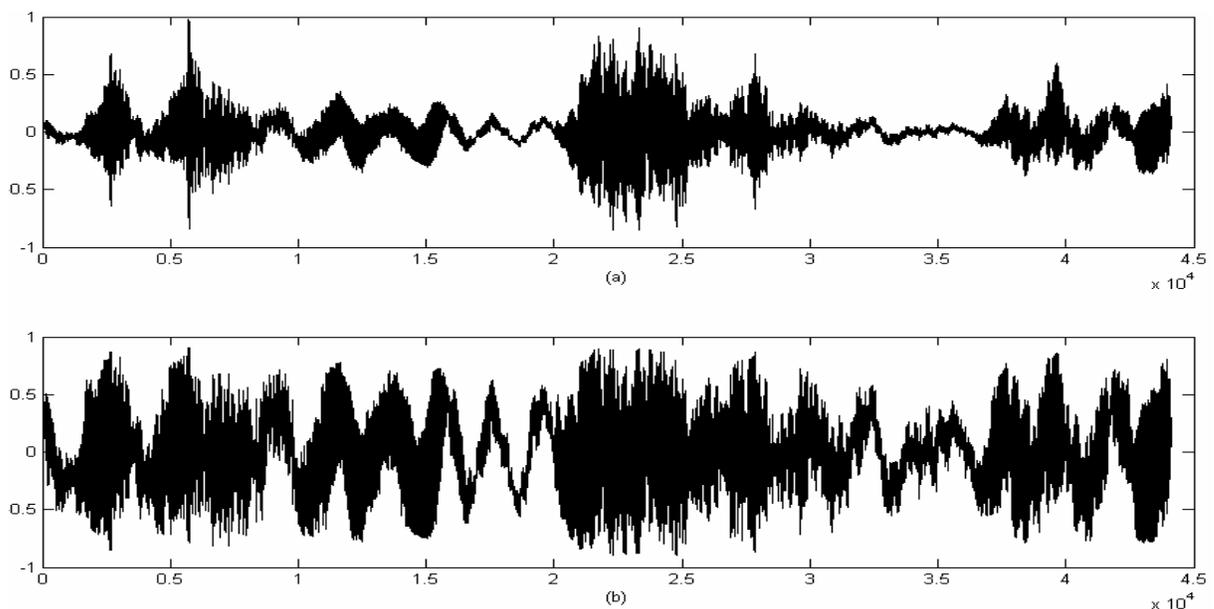


Figura 3.8. Respuesta del control de ganancia con una constante de normalización de 0.1 y una constante de atenuación de 0.00058.

La Figura 3.8 muestra como un valor de 0.1 de la constante de normalización puede afectar la ganancia en zonas donde la amplitud de la señal de entrada es baja en comparación con los picos de la señal, provocando que la normalización no se ejecute de manera uniforme.

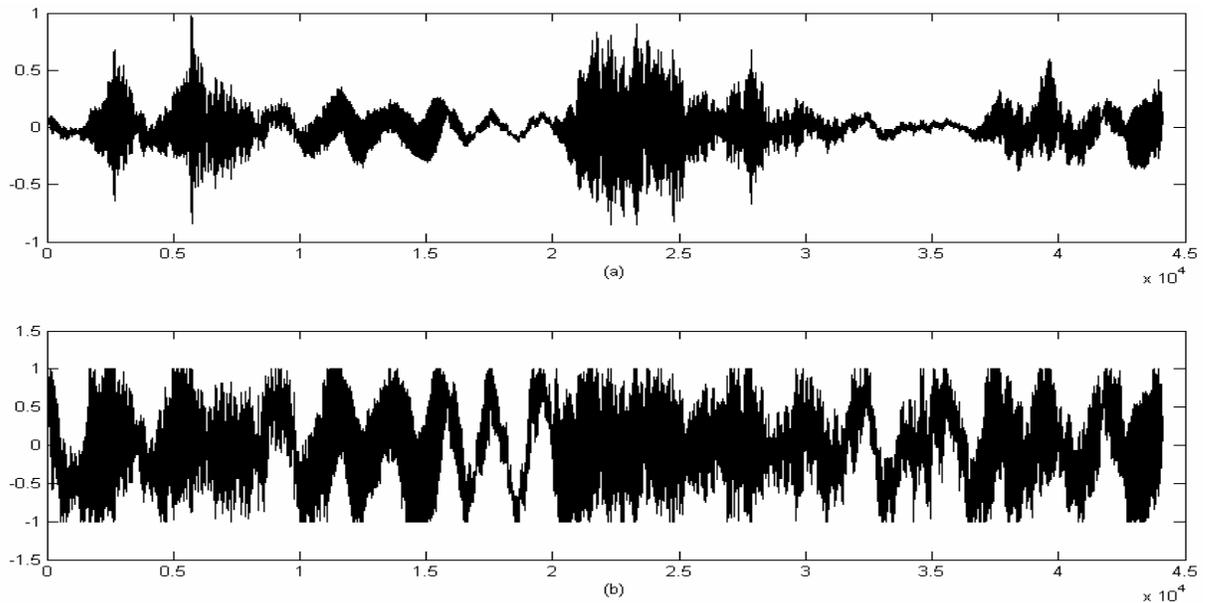


Figura 3.9. Respuesta del control de ganancia con una constante de normalización de 0.00001 y una constante de atenuación de 0.00058.

Un valor de 10^{-5} en la constante de normalización permite que el control de ganancia actúe uniformemente sobre toda la señal, tal como se observa en la Figura 3.9 donde las zonas de menor amplitud son amplificadas de manera proporcional a las zonas de mayor amplitud de la señal de entrada.

Como resultado se tiene a la salida de este módulo, una señal de audio normalizada que entra a la etapa de extracción de características para ser dividida en bandas de frecuencias por el banco de filtros.

Capítulo 4: Diseño y evaluación de la etapa de extracción de características

Este capítulo presenta la estructura de los módulos que constituyen la etapa de extracción de características, presentada en la Figura 1.4. Cada módulo se encarga de dividir y codificar la señal normalizada proveniente de la etapa de preproceso

4.1 Banco de Filtros

Para la elección del banco de filtros a utilizar se valora preliminarmente las ventajas de cada tipo y las condiciones de la aplicación. A continuación se comparan los sistemas de filtros uniformes y de codificación subbanda presentados en la sección 2.1.1 en términos de tamaño y complejidad computacional.

4.1.1 Comparación entre diseños

a) Tamaño

El sistema de codificación subbanda necesita de una conexión en cascada de los niveles de filtrado, es decir la salida de un sistema es la entrada del siguiente, por lo que aparenta ser de un tamaño mayor, pero todos los niveles de filtrado son idénticos por lo que es posible reutilizar los filtros y el sistema de diezmado. Para el sistema de filtros uniformes es necesario el diseño del filtro de la ecuación (2.1), pero para cada nivel de filtrado es necesario desplazar la respuesta del filtro por lo que en realidad se necesitan de N filtros paso-banda, donde N es el número de bandas.

b) Complejidad computacional

Para lograr una división de N bandas de frecuencia es necesario una repetición de $N-1$ niveles de filtrado para un codificador subbanda y N filtros distintos para un sistema de filtros uniformes.

Debido al menor tamaño y complejidad computacional se decide utilizar un codificador subbanda para el banco de filtros, debido a que su diseño e implementación se limita a dos filtros y no a N como un banco de filtros uniformes.

El diseño de un codificador subbanda se basa en determinar un par de filtros espejo en cuadratura (QMF), como se resume en la sección 2.1.1. La evaluación y el diseño de estos filtros se presentan a continuación.

4.1.2 Diseño de filtros QMF

Los filtros QMF pueden ser tanto IIR como FIR [12], para esta aplicación se decide implementar el banco mediante filtros IIR, dado que un filtro IIR de orden menor (entiéndase tercer, cuarto, quinto orden) presenta características de corte similares a un filtro FIR de mayor orden, y aunque la distorsión de fase que estos introducen a la señal filtrada es considerable, en este proyecto se desprecia dado que su salida se utiliza para determinar la energía presente en la banda y no para la reconstrucción de la señal original [12, 15, 20, 21]. Esto permite un tratamiento de la señal digital en línea con una exigencia computacional menor que los filtros FIR.

Para el banco, se demandan filtros que cumplan con los siguientes requisitos con el menor orden posible:

- Rizado de la banda pasante menor a 1dB.
- Rizado de la banda no pasante menor a 30dB
- Banda de transición entre 0.4 y 0.7 de la frecuencia normalizada.

Determinadas estas condiciones, se eligió un filtro elíptico o de Cauer [17] dadas sus características de corte en comparación con los filtros Butterworth [18] y Chebyshev [19].

Vaidyanathan [37, 38] presenta un algoritmo para el diseño de filtros QMF IIR en donde es necesario únicamente el diseño del filtro paso-bajo para obtener un filtro paso-alto de respuesta complementaria. Para determinar el filtro complementario el algoritmo solicita que el numerador del filtro paso-bajo cumpla con las siguientes condiciones:

1. Debe ser simétrico y
2. Si el orden es impar entonces el numerador del filtro paso-alto será antisimétrico, en caso contrario el numerador será simétrico.

La Tabla 4.1 presenta los coeficientes de los filtros QMF complementarios que constituyen el codificador subbanda. Tanto el filtro paso bajo como el filtro paso alto comparten los polos (denominador).

Tabla 4.1. Coeficientes para los filtros elípticos QMF de tercer orden.

Número de coeficiente	Denominador	Numerador Filtro Paso Bajo	Numerador Filtro Paso Alto
1	1	0.211933103	0.210320841
2	0.006150841	0.502613340	-0.497393977
3	0.422261262	0.502613340	0.497393977
4	0.000680784	0.211933103	-0.210320841

Como resultado se obtuvo un par complementario de tercer orden. El numerador del filtro paso bajo es simétrico de orden impar, por lo que su complemento paso alto es antisimétrico.

Las características de corte y respuesta total en magnitud en decibeles para los filtros espejo en cuadratura se presentan en la Figura 4.1.

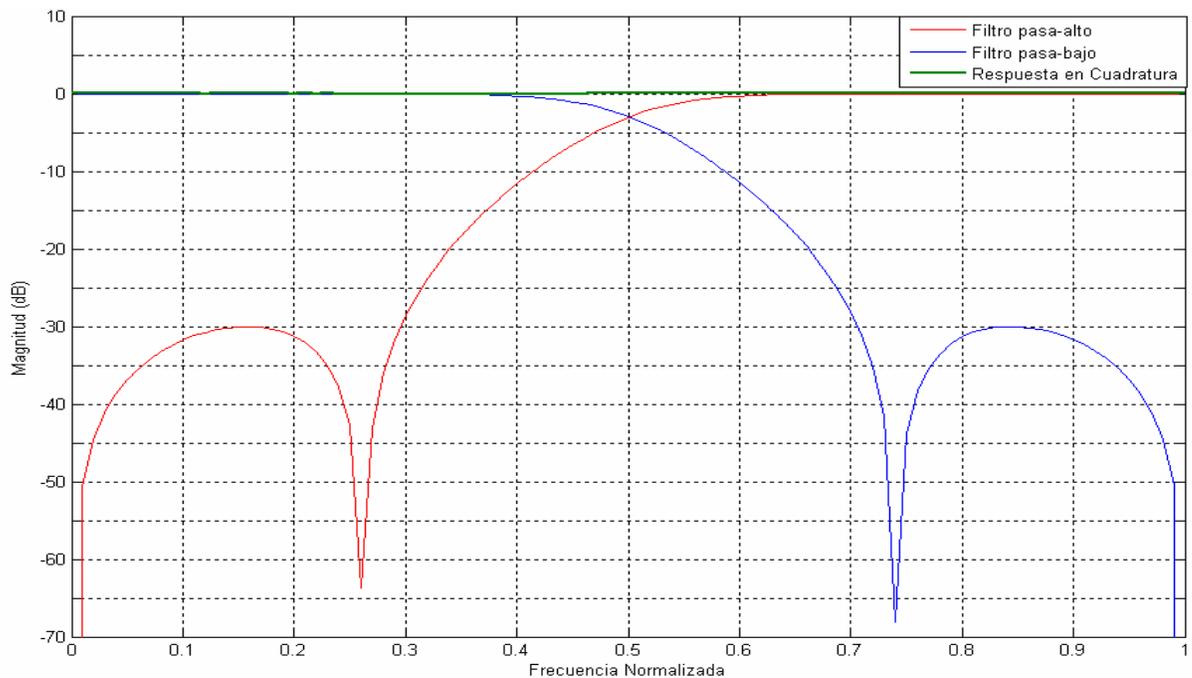


Figura 4.1. Respuesta en frecuencia y respuesta en cuadratura en decibeles para los filtros QMF elípticos de tercer orden

La frecuencia de corte se encuentra en 0.5 de la frecuencia normalizada, corte de media banda. El rizado de la banda pasante es menor a 1dB para ambos filtros, mientras que el rizado de la banda no pasante se encuentra por debajo de los 30dB, y aunque la banda de transición se encuentra fuera del rango establecido, se considera que un 0.01 por fuera de esta banda no perjudica considerablemente los parámetros de diseño. Se observa que la respuesta total en magnitud (línea verde en la Figura 4.1) cumple con lo establecido por la ecuación (2.5) por lo que los filtros diseñados son filtros QMF.

4.1.3 Estimación de energía

La estimación de la energía presente en las bandas de frecuencia se realiza mediante promediadores como el presente en el control de ganancia (Figura 3.2) en cada una de las bandas de salida del banco de filtros. En la Figura 4.2 se presenta la estructura interna del banco de filtros y la conexión de los promediadores.

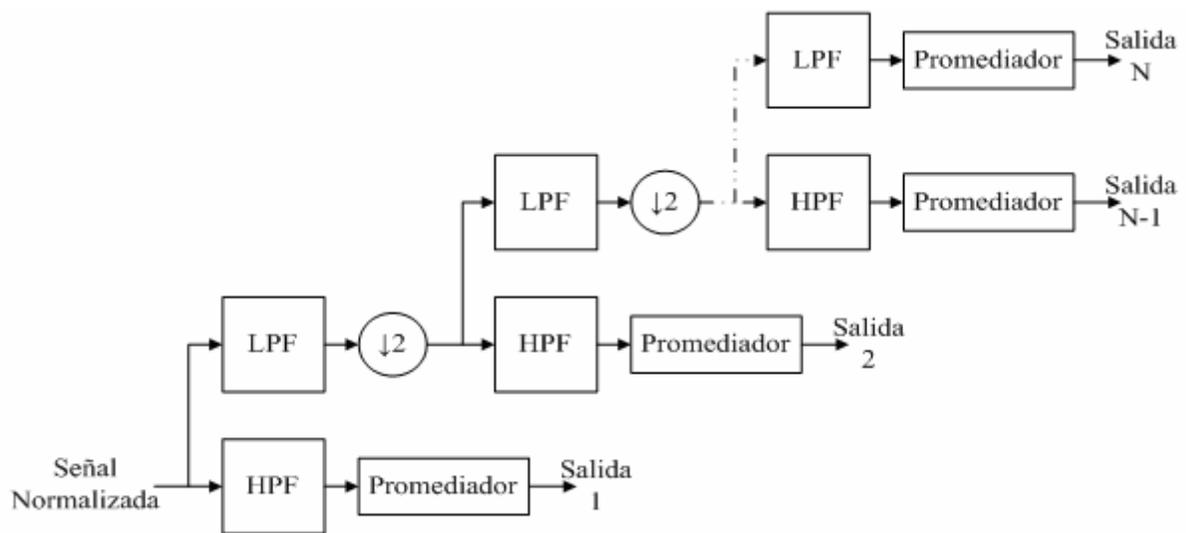


Figura 4.2. Estructura del banco de filtros.

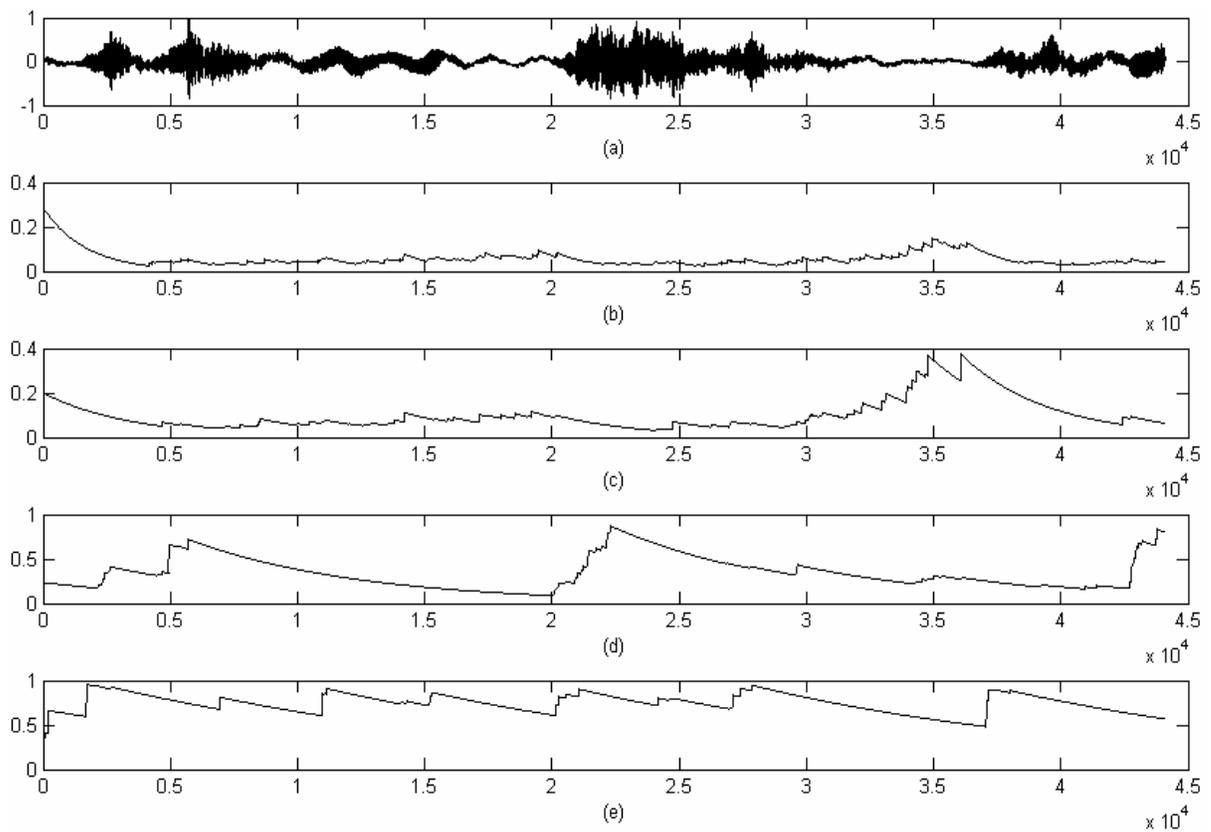


Figura 4.3. Respuesta de un banco de filtros de cuatro niveles.

La Figura 4.3 presenta la respuesta de un banco de filtros de cuatro niveles. En esta se observa como se obtiene la energía de las bandas conforme se presenta la señal acústica y como se extrae información de tiempo y frecuencia de la señal.

4.1.4 Selección de bandas de frecuencias

La selección del número de bandas necesarias para solucionar el problema de reconocimiento se realiza mediante un proceso de discriminación lineal de Fisher. Para conocer las bandas en las que se divide el espectro, es necesario establecer las características de frecuencia de la base de datos con que se cuenta para el proyecto. Estas se resumen en la Tabla 4.2.

Tabla 4.2 Características de frecuencia de la base de datos para el proyecto.

Característica	Valor
Frecuencia de Muestreo (Hz)	44100
Frecuencia de Corte Baja (Hz)	20
Frecuencia de Corte Alta (Hz)	20000

Se establecen entonces, once bandas iniciales de frecuencia con las que se realiza la discriminación lineal de Fisher, las frecuencias de corte para cada banda y el resultado del análisis de discriminación lineal para la proyección a una dimensión se presentan en la Tabla 4.3, la Figura 4.4 presenta de manera gráfica el de peso asignado a cada banda de frecuencia.

Tabla 4.3. Resultados del análisis de discriminante lineal de Fisher.

Banda	Frecuencia Alta	Frecuencia Baja	Pesos LDA
1	20000.00	11025.00	-0.0076801
2	11025.00	5512.50	-0.0001915
3	5512.50	2756.25	0.0154513
4	2756.25	1378.13	-0.0178226
5	1378.13	689.06	0.0063133
6	689.06	344.53	0.1057255
7	344.53	172.27	-0.0130055
8	172.27	86.13	-0.9939808
9	86.13	43.07	0
10	43.07	21.53	0
11	21.53	-	0.0000117

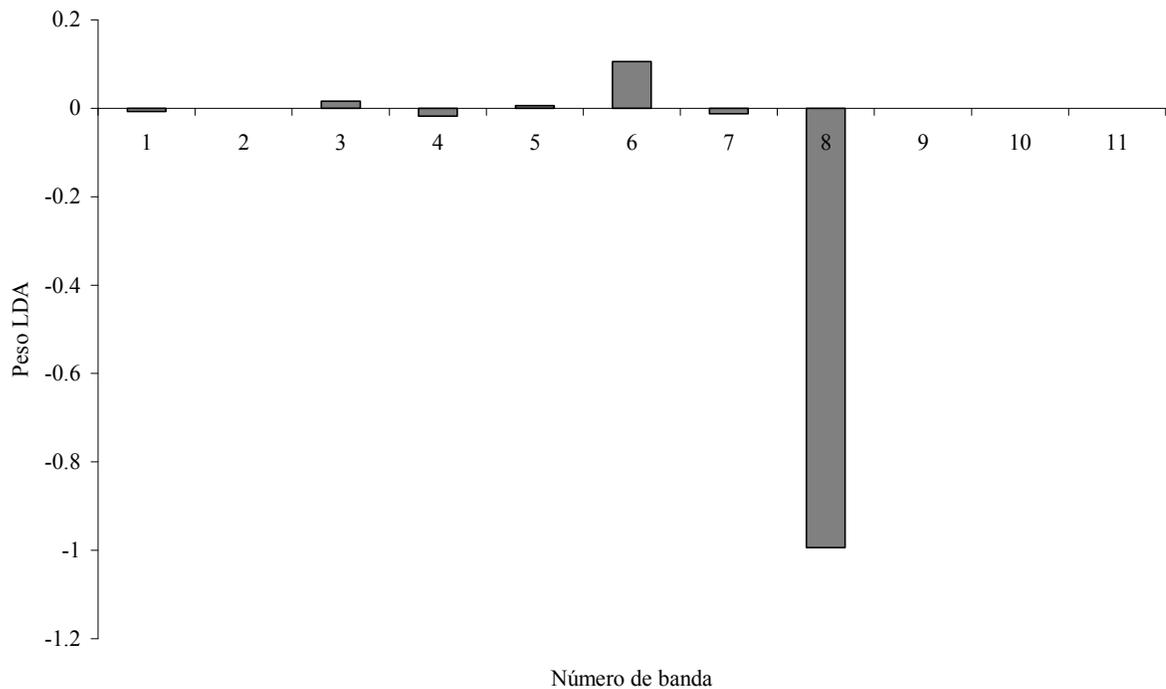


Figura 4.4. Distribución de pesos en las bandas de frecuencia.

Se presentan resultados del análisis de discriminante lineal para la proyección a una dimensión debido que los resultados arrojados por el LDA hasta un total de proyecciones a seis dimensiones son similares a los presentados en la Tabla 4.3 y la Figura 4.4. En estos se observa como la banda correspondiente a las frecuencias entre los 86.13Hz y 172.27Hz presenta el mayor peso de todas las bandas, esto indica que dentro del espacio vectorial de once dimensiones, los datos se dividen mejor en esa dirección.

Además, el LDA determina que las bandas 9 (86.13-43.07Hz) y 10 (43.07-21.53Hz) no entregan información relevante para la clasificación, ya que presentan un cero en el análisis lineal, mientras que la banda 11 (21.53-20Hz) presenta un peso de 1.17×10^{-5} que es bajo en comparación con las otras bandas de frecuencia, por estas razones se descartan estas tres bandas de frecuencia del banco de filtros.

Aunque las bandas 1 (20K-11025Hz), 2 (11025-5012.5Hz) y 5 (1378.13-689.06Hz), presentan pesos bajos en comparación a la banda 8, en comparación con las bandas 3, 4 ,6 y 7 no son despreciables, por lo que se decide mantener estas bandas dentro de la etapa de extracción. En la Tabla 4.4 se presentan las bandas utilizadas en el banco de filtros.

Tabla 4.4. Bandas de frecuencia utilizadas en el banco de filtros.

Banda	Frecuencia Alta	Frecuencia Baja
1	20000.00	11025.00
2	11025.00	5512.50
3	5512.50	2756.25
4	2756.25	1378.13
5	1378.13	689.06
6	689.06	344.53
7	344.53	172.27
8	172.27	86.13

Como salida de este módulo se obtiene la señal de audio normalizada dividida en ocho bandas de frecuencia que son enviadas al módulo de codificación.

4.2 Codificación

La función del módulo de codificación es producir los símbolos del alfabeto discreto para los modelos ocultos de Markov.

Este módulo se encarga de describir el comportamiento de los datos de salida del banco de filtros, que en teoría posee un número infinito de posibles variantes, mediante un número finito de conglomerados, dado que los HMM discretos necesitan que la salida del sistema a modelar tenga un número finito de salidas posibles.

El tamaño del alfabeto discreto V es determinado por el número de centroides K asignados al algoritmo de K-medias. Para estimar los centroides y determinar la distribución de los vectores de salida del banco de filtros se utilizaron distancias euclidianas cuadradas [35, 36].

El prototipo utiliza un total de 32 centroides para la codificación. Rabiner [29] recomienda utilizar un número mayor o igual a 32 centroides para los sistemas de reconocimiento, debido a que la distorsión de los datos se reduce cuanto mayor es el número de conglomerados del sistema de cuantificación [23, 24, 29, 33-36]. En la Figura 4.5 se muestra un ejemplo de la distribución de una señal acústica en símbolos de observación.

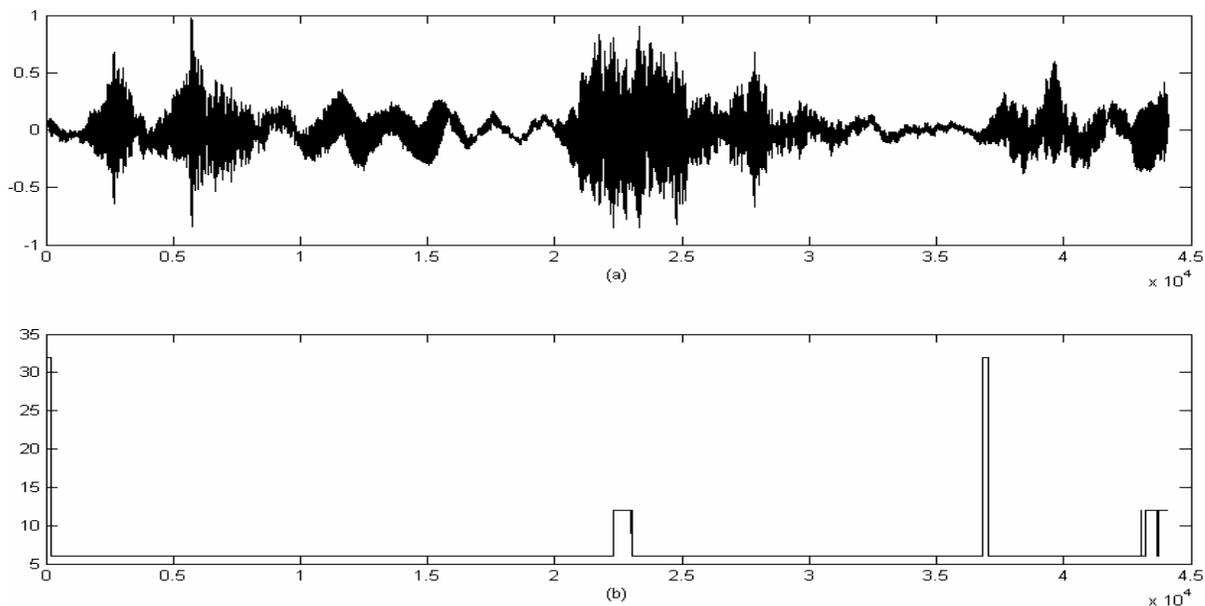


Figura 4.5. Distribución de una señal símbolos de observación..

A la salida de este módulo se tienen los vectores provenientes del banco de filtros codificados en 32 posibles valores que van a la etapa de clasificación.

Capítulo 5: Diseño y evaluación de la etapa de clasificación

La implementación de esta etapa se basa en los modelos ocultos de Markov, descritos en la sección 2.2. Esta consiste de dos módulos, análisis y toma de decisiones, como lo describe la Figura 1.4, el diseño y los resultados obtenidos para estos módulos se presentan en este capítulo.

5.1 Análisis

Este módulo se encarga de tomar los datos provenientes del módulo de codificación y mediante el procedimiento hacia adelante, presentado en la sección 2.2.1, determina la probabilidad de que una cadena de observaciones pertenezca a un modelo dado.

Para el prototipo se generaron tres modelos de Markov: un modelo para sonidos de disparos, uno para sonidos de motosierras y otro para los sonidos del ambiente normal del bosque.

5.1.1 Entrenamiento de los modelos

Como se muestra en [23, 24, 29] para determinar un modelo oculto de Markov es necesario establecer los parámetros \mathbf{S} , \mathbf{V} , \mathbf{A} , \mathbf{B} y $\boldsymbol{\pi}$ presentados en la sección 2.2.1. En el módulo de codificación se fijó el parámetro \mathbf{V} que corresponde al alfabeto discreto de observaciones. Los parámetros \mathbf{A} , \mathbf{B} y $\boldsymbol{\pi}$ son matrices aleatorias que se determinan mediante el algoritmo de Baum-Welch. El parámetro \mathbf{S} , que corresponde al conjunto de estados, es el que describe el comportamiento de los modelos a estimar.

Además de estos parámetros, para utilizar un modelo de Markov en problemas de reconocimiento es necesario definir la variable \mathbf{O} que corresponde a la longitud de la cadena de observaciones.

Para el presente prototipo se define además la variable **P** que corresponde al periodo de tiempo en el que se toma una muestra del módulo de codificación para formar la cadena de observaciones, esto para poder estimar si es posible trabajar a una tasa de muestreo distinta a la utilizada para las etapas de preproceso y extracción de características.

5.1.2 Base de datos para el entrenamiento

Para el entrenamiento de los modelos se contó con sonidos grabados en el Parque Nacional Braulio Carrillo, con disparos y motosierras capturados a distancias de 30m, 90m, 250m y 600m de la fuente.

5.2 Toma de decisiones

El módulo de toma de decisiones se encarga de clasificar la cadena de observaciones en una de las tres clases definidas: disparo, motosierra y ambiente normal del bosque.

La clasificación se realiza mediante la comparación de las probabilidades de salida de cada modelo. La cadena se clasifica dependiendo del modelo que lance la mayor probabilidad.

La estructura interna de la etapa de clasificación se muestra en la Figura 5.1.

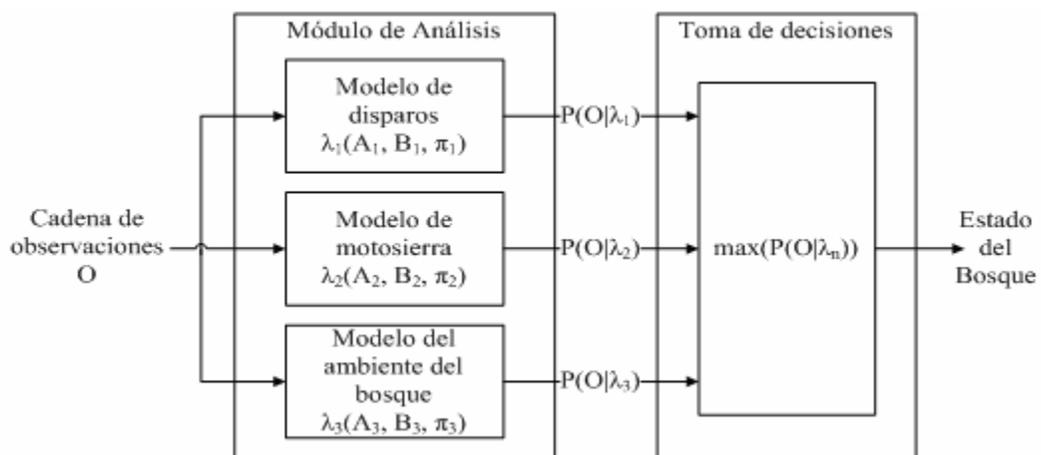


Figura 5.1. Estructura de la etapa de clasificación

5.3 Determinación de los parámetros de los modelos

5.3.1 S: El conjunto de estados

Para determinar el número de estados de los modelos se tomaron en cuenta las siguientes recomendaciones hechas por Rabiner [29] para sistemas de reconocimiento:

1. Asignar el mismo número de estados para todos los modelos, con esto se obtiene un mejor desempeño de los modelos cuando evalúan cadenas del mismo largo.
2. El error de reconocimiento de los modelos no varía de manera significativa con el número de estados, pero es posible encontrar algunas tasas de error mínimas.

En la Tabla 5.1 se presentan porcentajes de reconocimiento para modelos de disparos, motosierras y el ambiente del bosque, con números de estados de dos a cinco. La Figura 5.2 muestra como varían estos porcentajes con respecto al número de estados para los modelos.

Tabla 5.1. Porcentaje de reconocimiento para modelos con distinto número de estados.

Número de estados	Disparo	Motosierra	Bosque
2	0.8571	0.7528	0.8275
3	0.8571	0.7977	0.8275
4	0.9047	0.6292	0.8275
5	0.8571	0.7303	0.8275

Se observa como el porcentaje de reconocimiento no varía de manera significativa cuando se modifica el número de estados, salvo para el de motosierras en donde se tiene una variación de casi 13% al cambiar de tres a cuatro estados. Para los modelos de disparos y ambiente del bosque esto implica que no importa el número de estados de los modelos, no es posible aumentar el porcentaje de reconocimiento alterando únicamente el número de estados.

Aún con esto, en la Figura 5.2 se observa como con tres estados, se obtiene un punto en el cuál los porcentajes de reconocimiento de los tres modelos rondan el 80%. Representando un máximo en el porcentaje de reconocimiento de motosierras y un valor normal de los modelos de disparos y ambiente del bosque. Dado esto se decide que los modelos a utilizar sean de tres estados.

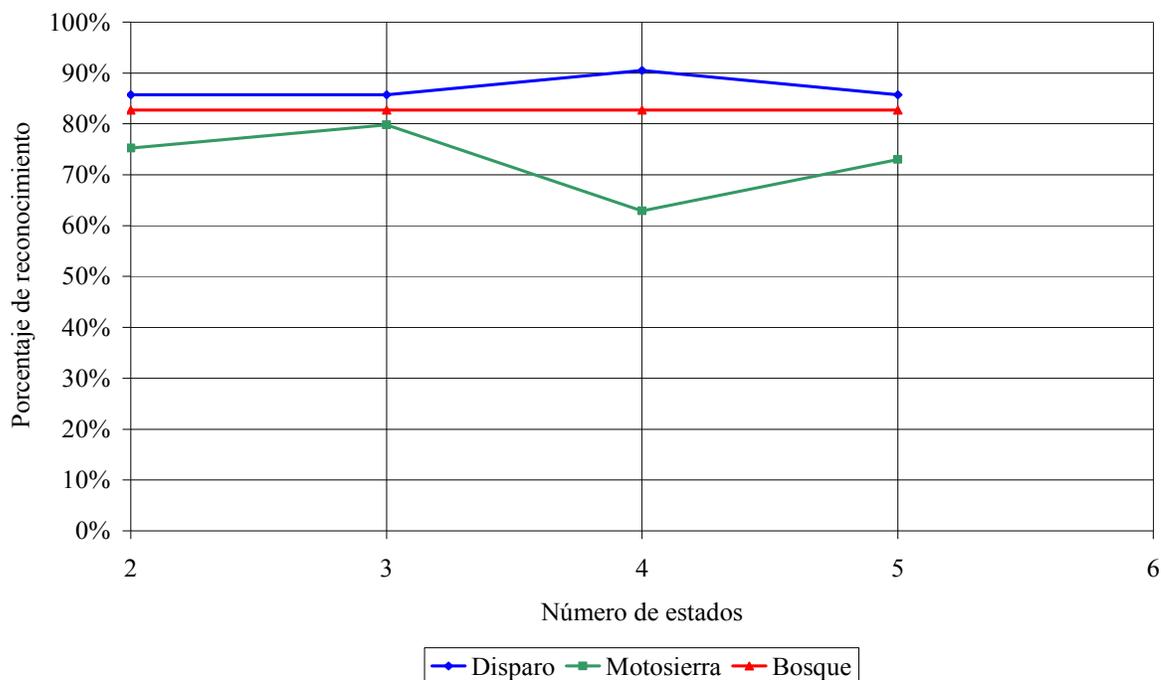


Figura 5.2. Porcentaje de reconocimiento para modelos con distinto número de estados.

5.3.2 P: La tasa de muestreo para la etapa de clasificación

Las Tablas 5.2 y 5.3 presentan porcentajes de reconocimiento para los modelos de tres estados para disparos, motosierras y ambiente del bosque, en estas se muestra como es posible modificar el porcentaje de reconocimiento de los modelos al variar el periodo con que se toman las muestras para crear la cadena de observaciones.

Tabla 5.2. Porcentajes de reconocimiento para los modelos con tres estados y tasas de muestreo de 44100Hz y 10Hz.

	44100Hz			10Hz		
	Reconocido como			Reconocido como		
Sonido	Disparo	Motosierra	Bosque	Disparo	Motosierra	Bosque
Disparo	52	3	3	51	2	5
Motosierra	1	10	2	0	13	0
Bosque	5	3	52	5	6	49
Tasa de Reconocimiento	89.66%	76.92%	86.67%	87.93%	100.00%	81.67%

Tabla 5.3. Porcentajes de reconocimiento para los modelos con tres estados y tasas de muestreo de 20Hz y 100Hz.

	20Hz			100Hz		
	Reconocido como			Reconocido como		
Sonido	Disparo	Motosierra	Bosque	Disparo	Motosierra	Bosque
Disparo	51	1	6	54	1	3
Motosierra	2	10	1	2	11	0
Bosque	7	0	53	9	1	50
Tasa de Reconocimiento	87.93%	76.92%	88.33%	93.10%	84.62%	83.33%

En la Figura 5.3 se observa cómo la variación de la tasa de muestreo afecta el porcentaje de reconocimiento de los modelos. Para el modelo de disparos, se mejora la detección en un 5.17% entre 10Hz y los 100Hz, mientras que para el modelo del bosque se presenta una mejora de un 6.66% entre los 10Hz y 20Hz, el modelo de motosierras presenta la mayor alteración al mejorar en un 23.08% en el porcentaje de reconocimiento al pasar de una tasa de muestreo de 44100Hz a una de 10Hz.

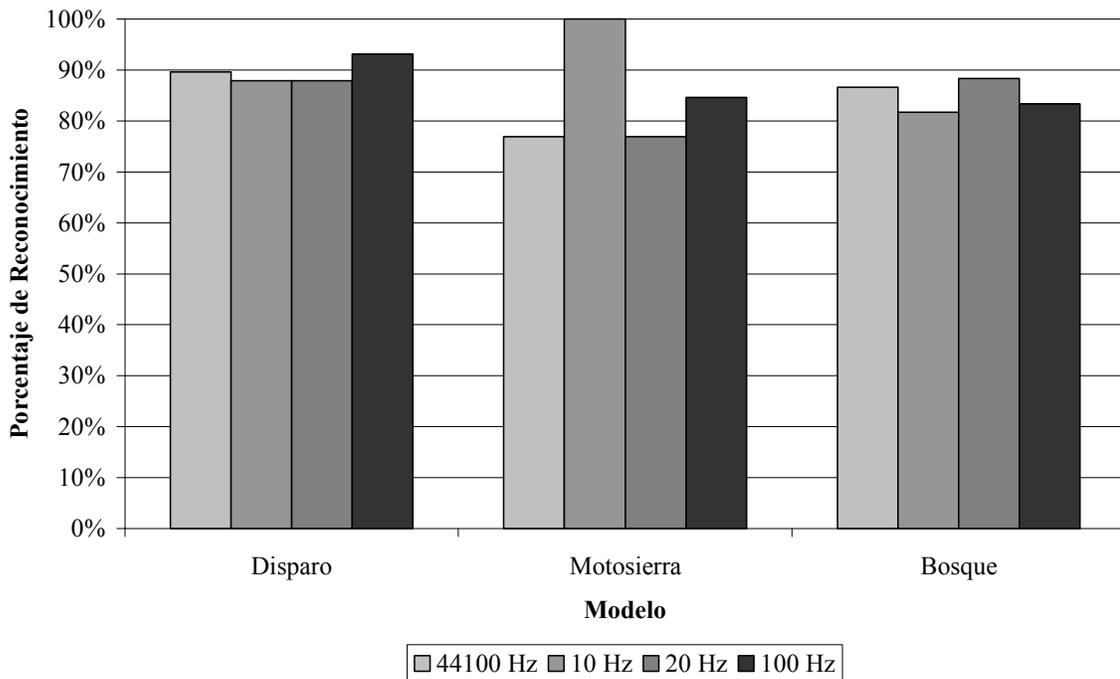


Figura 5.3. Porcentajes de reconocimiento para modelos de tres estados con tasas de muestreo de 44100Hz, 10Hz, 20Hz y 100Hz.

Por lo tanto, es posible utilizar en la etapa de clasificación una tasa de muestreo menor que la utilizada en las etapas de preproceso y de extracción de características. Este prototipo utiliza un periodo de muestreo de 0.1s, dado que para este se tiene una mejora en el porcentaje de reconocimiento del modelo de motosierras, que presenta las tasas de reconocimiento menores de los tres modelos entrenados.

5.3.3 O: El tamaño de la cadena de observaciones

Los porcentajes de reconocimiento para los modelos de tres estados con una tasa de muestreo de 0.1s y cadenas de observación de una a cinco muestras se muestran en las Tablas 5.4, 5.5 y 5.6.

Tabla 5.4. Porcentajes de reconocimiento para cadenas de observación de una y dos muestras

	Una Observacion			Dos Observaciones		
	Reconocido como			Reconocido como		
Sonido	Disparo	Motosierra	Bosque	Disparo	Motosierra	Bosque
Disparo	3	12	2	11	4	2
Motosierra	7	60	22	18	53	18
Bosque	7	97	523	7	97	523
Tasa de Reconocimiento	17.65%	67.42%	83.41%	64.71%	59.55%	83.41%

Tabla 5.5 Porcentajes de reconocimiento para cadenas de observación de tres y cuatro muestras

	Tres Observaciones			Cuatro Observaciones		
	Reconocido como			Reconocido como		
Sonido	Disparo	Motosierra	Bosque	Disparo	Motosierra	Bosque
Disparo	15	0	2	16	0	1
Motosierra	21	60	8	15	63	11
Bosque	7	97	523	7	97	523
Tasa de Reconocimiento	88.24%	67.42%	83.41%	94.12%	70.79%	83.41%

Tabla 5.6. Porcentajes de reconocimiento para cadenas de observación de cinco muestras

	Cinco Observaciones		
	Reconocido como		
Sonido	Disparo	Motosierra	Bosque
Disparo	16	0	1
Motosierra	14	68	7
Bosque	7	97	523
Tasa de Reconocimiento	94.12%	76.40%	83.41%

En términos de porcentaje de reconocimiento, se presenta un aumento en los modelos de disparos y motosierras conforme se aumentan las muestras que componen las cadenas de observación. Para el modelo de disparos se tiene un porcentaje de reconocimiento de un 17.65% para una cadena de una muestra (Tabla 5.4), mientras para una de cuatro observaciones se obtiene un 94.12% (Tabla 5.5). Para el modelo de motosierras se pasa de un 59.55% con una cadena de dos observaciones a un 76.40% con una de cinco muestras.

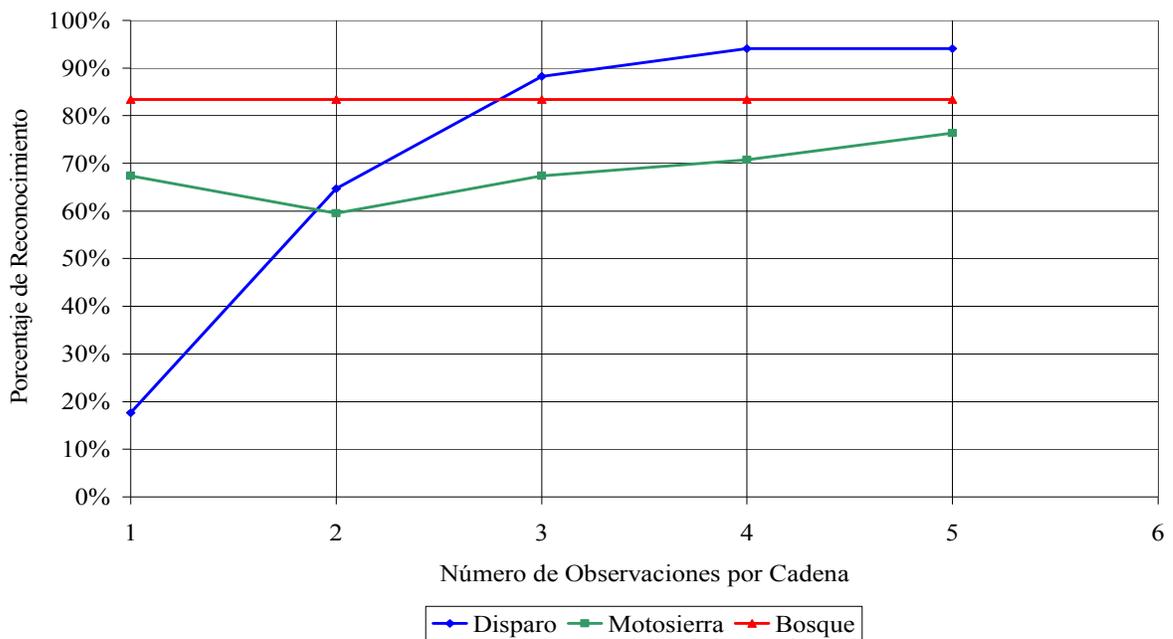


Figura 5.4. Variación del porcentaje de reconocimiento de los modelos en función del tamaño de la cadena de observaciones.

En la Figura 5.4 se observa como después de un determinado número de muestras el porcentaje de reconocimiento llega a un valor máximo. En este caso para el modelo de disparos se alcanza un máximo de 94.12% en la detección con una cadena de observaciones de cuatro muestras, mientras que la tasa de reconocimiento del modelo de motosierras aumenta con el número de observaciones.

Para el entrenamiento del modelo de disparos, la mayor parte de las cadenas utilizadas son menores a 1s y se obtienen porcentajes de reconocimiento de un 94.12% con cuatro muestras, es decir con cadenas de observación de 0.4s de duración. Además se presenta un aumento en el porcentaje de reconocimiento de motosierras conforme se aumenta el tamaño de la cadena de observaciones, lo que hace pensar que con un número de muestras cercano al utilizado para el entrenamiento del modelo se puede llegar a obtener un máximo en el porcentaje de detección.

Como salida de este módulo se tienen los símbolos provenientes del módulo de codificación clasificados en disparos, motosierras o en ambiente normal, con lo que se determina el estado del bosque.

Capítulo 6: Conclusiones y recomendaciones

6.1 Conclusiones

Para la etapa de preproceso se tiene un algoritmo de control de ganancia que se basa en un circuito analógico; esto permite realizar un acondicionamiento en tiempo real de la señal de entrada mediante un algoritmo de bajo nivel computacional realizado por medio de sumas de variables y una división. Se presenta además que mediante la variación de las constantes de atenuación y de normalización se puede modificar el comportamiento de esta etapa.

La bandas de frecuencia necesarias para el reconocimiento de disparos y motosierras en el ambiente del bosque se seleccionan mediante los discriminantes lineales de Fisher; con esto se realiza una reducción de dimensiones que disminuye la complejidad computacional del sistema de reconocimiento, específicamente del modulo de codificación. Como resultado se tiene un total de ocho bandas de frecuencia que se resumen en la Tabla 4.4 y que van de los 86.13Hz a los 20KHz.

Con la utilización de un banco de filtros IIR se tiene un sistema de análisis en tiempo real que permite extraer información de tiempo y frecuencia de las señales acústicas. Además, se presenta un menor tamaño y complejidad computacional con la implementación de un codificador subbanda en comparación con un banco de filtros uniforme. En la tabla se resumen las variables que constituyen la etapa de extracción de características y cuya modificación puede alterar la respuesta de la misma.

Tabla 6.1. Variables para la etapa de extracción de características.

Variable	Valor
Orden de los filtros	3
Constante de atenuación de los promedidores (Figura 4.2)	5.8×10^{-4}
Número de bandas de frecuencia	8
Número de centroides de codificación	32

Para poder utilizar los modelos ocultos de Markov en sistemas de reconocimiento es necesario establecer un número de variables que optimicen la tasa de detección del sistema. Tal como se muestra en el capítulo 5, modificando variables como la tasa de muestreo de la etapa de clasificación así como el tamaño de las cadenas de observación se pueden obtener porcentajes óptimos de detección de los modelos.

La utilización de modelos de pocos estados permite reducir la complejidad computacional del sistema de reconocimiento, ya que aumentar el número de estados aumenta también el tamaño de las matrices \mathbf{A} , \mathbf{B} y $\boldsymbol{\pi}$ que caracterizan a un modelo λ por lo que se incrementa el número de operaciones necesarias para calcular las probabilidades mediante el procedimiento hacia adelante. Además se comprueba que el número de estados que se utilizan para los modelos no afecta en gran medida la tasa de reconocimiento del sistema (Tabla 5.1), pero que existen máximos que se pueden utilizar para mejorar el comportamiento de los modelos.

Al reducir la tasa de muestreo para la etapa de clasificación se cuenta con un mayor tiempo para el cálculo de las probabilidades, esto produce que se tenga una menor rigidez para la implementación del clasificador ya que se cuenta con un mayor tiempo de procesamiento.

En cuanto al número de muestras necesarias para clasificar una señal, se tiene que una cadena de cuatro muestras es suficiente para obtener un porcentaje de reconocimiento superior al 90%, con cadenas de entrenamiento menores a 1s para el modelo de disparos. Dado que cada muestra se toma cada 0.1s, se tiene un análisis del estado del bosque cada 0.4s. En la Tabla 6.2 se resumen las variables presentes en la etapa de clasificación.

Tabla 6.2. Variables presentes en la etapa de clasificación.

Variable	Valor
S: El conjunto de estados	3 estados
P: La tasa de muestreo	10 muestras por segundo
O: El tamaño de la cadena de observaciones	4 muestras por cadena

Para comprobar que el prototipo diseñado mejora las condiciones de rango de detección y desempeño obtenidos con los prototipos analógicos presentados en [4] y [5] se presenta la Tabla 6.3 que contiene los valores de los proyectos anteriores y el actual.

Tabla 6.3. Comparación entre los prototipos de detección anteriores y el proyecto actual.

	Prototipo analógico de detección de motosierras [5]	Prototipo analógico de detección de disparos [4].	Este proyecto
Rango de detección	30m / 2862m ²	30m / 2862m ²	600m / 1,131km ²
Desempeño en cuanto al porcentaje de detección	88%	90%	Se pueden alcanzar porcentajes de detección superiores a un 90%

Debido que se pueden detectar sonidos de motosierras y disparos a 600m de la fuente, se mejora el rango de detección, pasando de un radio de 30m (que corresponde a un área de 2862m²) a uno de 600m cubriendo un total de 1.131km².

En cuanto al porcentaje de detección, éste puede superar el 90% presentado por el prototipo de detección de disparos que se presenta en [4], para el presente proyecto se obtuvo un 94.12% en la tasa de reconocimiento de disparos y una tasa creciente para el modelo de motosierras, esto implica que mediante un proceso de optimización se puede mejorar las tasas de reconocimiento y reducir los porcentajes de error de cada modelo entrenado.

6.2 Recomendaciones

Dentro de las recomendaciones que se hacen para el mejoramiento de los resultados obtenidos con este proyecto se establece la utilización de herramientas como el análisis multivarianza o el factor lambda de Wilk para seleccionar las bandas de frecuencia del banco de filtros y estimar el efecto de la reducción de dimensiones en el reconocimiento de los patrones.

Además, se puede estimar la razón que existe entre el tamaño de las cadenas de entrenamiento y las cadenas de observación en los modelos ocultos de Markov. Por último, se recomienda realizar un procedimiento de optimización para determinar una combinación apta de variables que mejoren la detección de los eventos.

Bibliografía

- [1]. Sáenz, María Gabriela. **“Reconocimiento de patrones acústicos para la protección del ambiente utilizando wavelets y Modelos Ocultos de Markov”**. Proyecto de graduación. Cartago: Instituto Tecnológico de Costa Rica. Noviembre 2006
- [2]. Hernández, Néstor. **“Diseño de una red inalámbrica de telecomunicaciones para la protección ambiental en el bosque”**. Propuesta de proyecto de investigación. Cartago, Costa Rica. Setiembre 2004.
- [3]. Hernández, Néstor. **“Optimización de la red inalámbrica de telecomunicaciones para la protección ambiental en el bosque”**. Propuesta de proyecto de investigación. Cartago, Costa Rica. Mayo 2006.
- [4]. Leiva, Arnoldo. **“Diseño e implementación de la etapa de detección de disparo de armas de la red inalámbrica de telecomunicaciones para la protección de zonas protegidas”**. Proyecto de graduación. Cartago: Instituto Tecnológico de Costa Rica, 2005.
- [5]. Salas, Walter. **“Diseño e implementación de un sensor para la detección de motosierras de la red inalámbrica de telecomunicaciones de la protección ambiental en el bosque”**. Proyecto de graduación. Cartago: Instituto Tecnológico de Costa Rica, 2005.
- [6]. **“Costa Rica inicia venta de oxígeno para preservar el planeta”**. Artículo del diario La Prensa, Honduras. 15 de mayo de 1997. Versión en línea en: <http://www.laprensahn.com/caarc/9705/c15002.htm>. Última revisión: 10 de abril de 2007

- [7]. 1ra. Reunión Mesoamericana de Incendios Forestales. “**Manejo del fuego en Costa Rica**”. Ciudad de Guatemala, julio 2002. Versión en línea en: <http://www.fire.uni-freiburg.de/GlobalNetworks/MesoAmerica/Costa%20Rica%20Overview%202002.PDF>. Última revisión: 10 de abril de 2007
- [8]. Villalobos, Roberto y otros. “**El Niño y los incendios forestales en Costa Rica**”. Gestión de Desarrollo, IMN. Enero 2000. Versión en línea en: <http://www.imn.ac.cr/publicaciones/revista/Topicos-incendios%20forestales1200.pdf>. Última revisión: 10 de abril 07
- [9]. Suárez, Nelson. “**Se arriendan bosques tropicales**”. Colombia 2002. Febrero 2002. Versión en línea en: <http://www.premioreportagem.org.br>. Última revisión: 10 de abril 07
- [10]. Jiménez, Francisco y otros. “**Experiencias de pago por servicios ambientales en cuencas en Costa Rica**”. Departamento de Recursos Naturales y Ambiente. CATIE. Turrialba, Costa Rica. Versión en línea en: www.rlc.fao.org. Última visita: 10 de abril de 07
- [11]. “**El estado de las plantas y animales se deteriora progresivamente, lo revela la Lista Roja de la UICN de Especies Amenazadas de 2006**”. Comunicado de prensa de la Unión Mundial para la Naturaleza. 1º de mayo de 2006. Ginebra, Suiza. Versión en línea en: <http://www.iucn.org>. Última visita: 10 de abril de 2007
- [12]. Proakis, John y Manolakis, Dimitris. “**Tratamiento Digital de Señales**”. Tercera edición. Prentice Hall, Madrid, España, 1998.
- [13]. Ackenhusen, John G. “**Signal Processing Technology and Applications**”. IEEE technology update series. New York, USA. 1995.
- [14]. Dintel, John y otros. “**Administración científica**”. Representaciones y servicios de ingeniería, S.A. México D.F. 1980.

- [15]. “**Filtro Electrónico**”. Wikipedia la enciclopedia libre. Versión en línea en: <http://es.wikipedia.org>. Última revisión: 22 de Agosto de 2007.
- [16]. “**Signal Processing Toolbox: IIR Filter Design**”. MathWorks. Versión en línea en: www.mathworks.com. Última Revisión: 24 de Agosto de 2007.
- [17]. “**Filtro de Cauer**”. Wikipedia la enciclopedia libre. Versión en línea en: <http://es.wikipedia.org>. Última revisión: 12 de Septiembre de 2007.
- [18]. “**Filtro de Butterworth**”. Wikipedia la enciclopedia libre. Versión en línea en: <http://es.wikipedia.org>. Última revisión: 12 de Septiembre de 2007.
- [19]. “**Filtro de Chebyshev**”. Wikipedia la enciclopedia libre. Versión en línea en: <http://es.wikipedia.org>. Última revisión: 12 de Septiembre de 2007.
- [20]. “**Finite Impulse Response**”. Wikipedia la enciclopedia libre. Versión en línea en: <http://en.wikipedia.org>. Última revisión: 12 de Septiembre de 2007.
- [21]. “**Infinite Impulse Reponse**”. Wikipedia la enciclopedia libre. Versión en línea en: <http://es.wikipedia.org>. Última revisión: 12 de Septiembre de 2007.
- [22]. Alvarado, Pablo. “**Capítulo 2: Señales y Sistemas de Variable Discreta**”. Material Didáctico para el curso de Procesamiento Digital de Señales. ITCR. 2007. Versión en línea en: <http://www.ie.itcr.ac.cr/palvarado/PDS>. Última revisión: 12 de Septiembre de 2007.
- [23]. Duda, Richard O y otros. “**Pattern Clasification**”. Segunda edición. John Wiley & Sons. Singapore. 2001.
- [24]. Bishop, Christopher M. “**Pattern recognition and machine learning**”. Primera edición. Springer. Singapore. 2006.
- [25]. Srihari, Sargur N. “**Discriminant Analisis**”. CSE 555 Srihari. Versión en línea en: www.cedar.buffalo.edu. Última revisión: 18 de octubre de 2007.
- [26]. Pirchio, F. “**An acoustic surveillance unit for energy aware sensor network**”. XII Workshop Iberchip: pp 191-194. San Jose, Costa Rica: 2006.

- [27]. B, Jelinek. “**Generalized hierarchical search in the ISIP ASR system: Signals, Systems and Computers**”. Conference Record of the Thirty-Fifth Asilomar. Volumen 2: pp 1553-1556. Estados Unidos: IEEE, 2001.
- [28]. Brown, Judith. “**Feature dependence in the automatic identification of musical woodwind instruments**”. Versión en línea en: <http://scitation.aip.org/>. Última revisión: Febrero 2006].
- [29]. Lawrence, R. “**A tutorial on HMM and selected applications in speech recognition**”. Volumen 77. No2, Estados Unidos: IEEE, febrero 1989.
- [30]. Veeravalli, Anant. “**A tutorial on using hidden markov models for phoneme recognition**”. Volumen, 20-22 pp: 154-157. Estados Unidos: IEEE, Marzo de 2005.
- [31]. Mohamed, M; Gader, P. “**Generalized hidden Markov models. Theoretical frameworks**” Volumen 8 pp: 67-81. Estados Unidos: IEEE, Febrero del 2000.
- [32]. Lim, S; Clements, M. “**Pseudo-continuous hidden Markov modeling for automatic speech recognition**”. Volumen 2 pp: 482-88. Estados Unidos: IEEE, Abril de 1992.
- [33]. “**K-means clustering: Data clustering and pattern reconigton**”. Versión en línea en: <http://neural.cs.nthu.edu.tw>. Última Revisión: 12 de noviembre de 2007.
- [34]. Yusoff Mashor, Mohd. “**Improving the Performance of K-Means Clustering Algorithm to Position Centres of RBF Network**”. International Journal of the Computer, the Internet and Management. Versión en línea en: <http://www.journal.au.edu>. Última revisión: 12 de noviembre de 2007.
- [35]. “**Cluster Análisis**”. StatSoft, Inc. Versión en línea en: <http://www.statsoft.com>. Última revisión: 12 de noviembre de 2007.
- [36]. Jain, A.K y otros. “**Data Clustering: A Review**”. ACM Computing Surveys, Vol. 31 No. 3. Septiembre 1999. Versión en línea en: www.cs.rutgers.edu. Última revisión: 12 de noviembre de 2007.

- [37]. Vaidyanathan, P.P. “**Design of Doubly-Complementary IIR Digital Filters Using a Single Complex Allpass Filter, With Multirate Applications**”. IEEE transactions on circuits and systems, Vol. 34 No. 4, Abril 1987
- [38]. Vaidyanathan, P.P. “**A New Approach to the Realization of Low-Sensitivity IIR Digital Filters**”. IEEE transactions on circuits and systems, Vol. 34 No. 2, Abril 1986.
- [39]. Meyer, Ernesto A.”**Glosario de términos técnicos**”. Grupo de Informática Aplicada al Inglés Técnico, Argentina, 1995. Versión en línea en: <http://es.wikipedia.org>. Última revisión: 15 de noviembre de 2007.
- [40]. Purcell, Jerry E. Ph.D.”**Multirate Filter Design: An Introduction**”. Momentum Data Systems. Versión en línea en: <http://www.mds.com>. Última revisión: 18 de mayo de 2007.
- [41]. Kouro, Samir. Musalem, Rodrigo. “**Tutorial introductorio a la teoría de wavelet**”. Universidad Técnica Federico Santa María. Chile. Julio 2002. Versión en línea en: <http://www.elo.uffs.cl>. Última revisión: 26 de junio de 2007.
- [42]. Monografías. “**Señales acústicas: su propagación y características**”. 1997. Versión en línea en: <http://www.monografias.com/trabajos15/acustica/acustica.shtml> Última revisión: 15 de noviembre de 2007.
- [43]. “**Wavelet**”. Versión en línea en: <http://www.wikipedia.org> .Última revisión: 13 de enero de 2008.

Apéndice A.1 Glosario

ASR: Del inglés, *Automatic Speech Recognition*. Reconocimiento Automático del Habla

Wavelet: ent. Pequeña onda.

DFT: Del inglés, *Discrete Fourier Transformation*. Transformación Discreta de Fourier

FFT: Del inglés, *Fast Fourier Transformation*. Transformación Rápida de Fourier

STFT: Del inglés, *Short Time Fourier Transformation*. Transformación de Fourier de Tiempo Corto.

Diezmado: dic. Sacar de diez uno. Reducir el número de muestras con que se representa una señal discreta [40].

Octava: ing. La mitad de una frecuencia dada.

Alias: Frecuencias fuera de la frecuencia de Nyquist que se presentan en el espectro de una señal discreta.

QMF: Del inglés, *Quadrature Media Filter*. Filtros Medios en Cuadrature

LDA: Del inglés, *Linear Discriminant Analysis*. Análisis de Discriminante Lineal

HMM: Del inglés, *Hidden Markov Models*. Modelos Ocultos de Markov