



HAL
open science

Analizador de espectro psicoacústico

Alejandro Osses

► **To cite this version:**

Alejandro Osses. Analizador de espectro psicoacústico. Acoustics [physics.class-ph]. Universidad Tecnológica de Chile INACAP, 2010. Español. NNT: . hal-03871034

HAL Id: hal-03871034

<https://hal.science/hal-03871034>

Submitted on 29 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**UNIVERSIDAD TECNOLÓGICA DE CHILE, INACAP
INGENIERÍA CIVIL EN SONIDO Y ACÚSTICA**

**PROFESOR GUÍA
VÍCTOR ESPINOZA C.**

**TESIS DE TITULACIÓN
ANALIZADOR DE ESPECTRO PSICOACÚSTICO**

**TESIS DE TITULACIÓN PRESENTADA
COMO PARTE DE LOS REQUISITOS
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN SONIDO Y ACÚSTICA**

ALEJANDRO OSSES VECCHI

SANTIAGO DE CHILE, 2010

AGRADECIMIENTOS

En primer lugar quisiera agradecer a mis padres –Luis y Alicia– y hermanos –Roberto y Carolina– por su eterna e incondicional ayuda tanto en mi proceso de formación como en el apoyo para el desarrollo de esta tesis. No puedo dejar de mencionar a algunas de las personas que me acompañaron en mi proceso universitario como son mis compadres Gabriel *Raios* y Omar *Iban-iz*, la “gran Liz” y Fran, quienes fueron los principales afectados por mis ataques de terquedad. También fui descubriendo otras grandes personas como son Maurito *Satriani*, Karin Andrea, Mônica Costa y Catita Luz, quienes además de haber acompañado mi crecimiento como estudiante, lo han hecho como amigos.

Tampoco puedo dejar de agradecer los conocimientos y consejos recibidos de Víctor Espinoza, Rolito Venegas, Pachi Raveau, Hernán Ángel, Sergio “Eddie” Floody, Paulo Massarani, Swen Müller, Christian Budde, Zemar Soares, Mayra Cardozo, Sergio Matamala y Karin Riedemann quienes fueron los responsables de ir perfilando mi orientación profesional. Finalmente, quisiera agradecer al equipo de trabajo de las empresas Control Acústico y Acustical, con quienes he compartido y de quienes he aprendido durante estos últimos meses de mi etapa universitaria.

Este trabajo, el cual he tardado en desarrollar poco más de 1 año, está hecho con el mayor de los esmeros y confío en que pueda ser un buen aporte al desarrollo del área de la psicoacústica en nuestra casa de estudios.

Índice general

Índice general	ii
0 Resumen	1
1 Introducción	2
1.1. Formulación del problema	2
1.2. Objetivo General	3
1.3. Objetivos Específicos	3
2 Marco Teórico	4
2.1. Banco de Filtros	4
2.1.1. Bandas de octava y fracciones de octava	5
2.1.2. IEC 61260: Bandas de 1/3 de octava	7
2.2. Psicoacústica	9
2.2.1. Aspectos generales: Fisiología del oído humano	10
2.2.2. Bandas Críticas	11
2.2.3. Modelo de Enmascaramiento de Terhardt	13
2.2.4. Ponderación de amplitud: Sonoridad o <i>Loudness</i>	14
2.2.5. Contornos de Igual Sonoridad	15
2.2.6. Curvas de Ponderación en Frecuencia	17
2.2.7. Ponderación de sonoridad (<i>Loudness</i>) usando Redes Neu- ronales Artificiales	19
2.2.8. Suma energética: Presión Sonora y Nivel de Presión Sonora	21
2.2.9. Suma de sonoridad: Sonoridad Específica y Sonoridad Total	22
2.2.10. Distribución de energía de una señal de audio	23

2.2.11. Modelos de Sonoridad	24
2.2.12. Estándar ISO 532:1975-B	25
2.3. Procesamiento digital de señales	29
2.3.1. Filtros	29
2.3.2. Diseño de Filtros	33
2.3.3. La técnica de <i>Warping</i> de frecuencias	41
2.3.4. Redes Neuronales Artificiales	44
3 Desarrollo: Diseño e implementación	53
3.1. Ponderación de Amplitud	54
3.1.1. Implementación de las curvas de ponderación en frecuencia	55
3.1.2. Ponderación de amplitud en Fones y cálculo energético . . .	56
3.2. Aproximación a un banco de filtros en bandas críticas	59
3.2.1. Parámetros de diseño	59
3.2.2. Diseño de los filtros pasa-bajos	61
3.2.3. Diseño de los filtros pasa-altos	64
3.2.4. Construcción de los filtros pasa-banda	64
3.3. Filtrado multibanda de señales de audio	67
3.3.1. Banco de filtros en el dominio lineal de frecuencias	68
3.3.2. Implementación Recursiva del Banco de Filtros	68
3.3.3. Filtrado en el dominio de la Frecuencia	69
3.3.4. Filtrado usando Redes Neuronales Artificiales de Funciones Radial Base	70
3.3.5. Filtrado en tiempo real: Comparación entre metodologías .	71
3.3.6. Precisión del algoritmo de filtrado usando RBFN	73
3.3.7. Ponderación psicoacústica de frecuencias considerando la técnica de <i>Warping</i>	75
3.4. Comparación entre modelos de Sonoridad: ISO 532B y Método Propuesto	76
3.5. Prototipo para el Analizador de Espectro Psicoacústico en Tiempo Real	78

<i>ÍNDICE GENERAL</i>	IV
3.5.1. Implementación en un plug-in VST	80
4 Análisis y Resultados	85
4.1. Análisis de los archivos de audio: consideraciones	86
4.2. Metodología para el análisis de temas musicales	88
4.3. Análisis de los fragmentos musicales	89
5 Conclusiones y discusión	93
Bibliografía	96
A Transformación Bilineal	100
A.1. <i>Warping</i> de frecuencias	101
A.2. Mapeamiento del dominio analógico al dominio digital	102
A.2.1. Conversión de filtros pasa bajos a filtros pasa altos	104
B Diseño de Filtros Pasa banda	105
B.1. Filtro Pasa bajos: Transformación Bilineal de un Sistema Analógico $H(s)$	105
B.2. Filtro Pasa altos: Diseño de un Filtro Butterworth	106
B.3. Filtro Pasa banda	110
C Plug-ins de audio VST	111
C.1. SDK VST	112
C.1.1. Principales funciones para la implementación del plug-in VST	112
C.2. Creando un proyecto en Microsoft Visual Studio 2008	113
C.3. VST <i>Wizard</i>	113
Índice de figuras	116
Índice de tablas	122

Capítulo 0

RESUMEN

En la presente investigación se propone un análisis psicoacústico aplicado a temas musicales, es decir, evalúa parámetros de carácter perceptivo asociados al oído humano y los compara con metodologías convencionales de análisis de señales. En este contexto, se trabaja con un análisis en bandas de frecuencia cuyas amplitudes son expresadas en dB ponderados en las curvas A, B, C y D (dB(A), dB(B), dB(C), dB(D)) y se comparan con la utilización de amplitudes expresadas en Fones, la que corresponde a la unidad de medición de la denominada Sonoridad o *Loudness*. Este descriptor manifiesta cuán intenso es percibido un sonido, de manera que una mayor sonoridad se asocia a un sonido *perceptivamente* más intenso. De aquí la motivación de estudiar este interesante descriptor acústico, el cual entrega información relevante cuando se utiliza como herramienta de monitoreo durante el proceso de producción musical, tal como se plantea en las hipótesis de esta investigación y como se corroboró durante el proceso de análisis y conclusiones de este trabajo.

Para la elaboración de un modelo que permita efectuar este tipo de análisis en tiempo real y considerando un banco de filtros auditivos, se conjugaron técnicas como el diseño de filtros analógicos y digitales, Redes Neuronales Artificiales (RNA) y otras herramientas para el procesamiento digital de señales como son: la técnica de *Warping* de frecuencias y los conceptos de suma energética.

Capítulo 1

INTRODUCCIÓN

1.1. Formulación del problema

Los analizadores de espectro utilizados comúnmente para la producción musical, descomponen una determinada señal de audio en valores de amplitud en función de la frecuencia, generalmente en tiempo real.

La resolución en el eje de frecuencias puede presentarse como muestras equiespaciadas, es decir, a un intervalo constante de Δ Hertz, o bien, puede presentarse como un conjunto de i intervalos de frecuencia asociados a los anchos de cada una de las i bandas pertenecientes a un banco de filtros. Los bancos de filtros más populares están definidos en bandas de octava y en fracciones de octava ($1/3$ de octava, $1/6$ de octava).

El trabajar el dominio de la frecuencia tomando las consideraciones anteriores, presenta la desventaja de no contemplar la agrupación de frecuencias tal y como el oído humano lo efectúa. El comportamiento no lineal del oído humano se manifiesta, además, en la percepción de la intensidad sonora, la cual no presenta una relación lineal con respecto a la sensación de amplitud percibida por nuestro sistema auditivo.

Un concepto clave en el estudio de esta no linealidad es el concepto de sonoridad (o *loudness*) propuesto por Fletcher y Munson [10], que contempla la intensidad con que un sonido es percibido como función de la frecuencia. Los denominados contornos de igual sonoridad o curvas isofónicas, denotan la mayor

sensibilidad de nuestro sistema auditivo a las frecuencias medias que a las frecuencias bajas o altas. Debido a lo anterior es que aparecerían, años más tarde, las curvas de ponderación en frecuencia (curvas A y C), las cuales intentan compensar este efecto. La consideración de variables ligadas a la percepción auditiva humana es estudiada en la rama de la psicoacústica.

En la presente investigación se profundizará en el funcionamiento del sistema auditivo humano, con el fin de elaborar un analizador de espectro, que permita una lectura directa de parámetros que afectan de manera significativa nuestra percepción como auditores y que, por tanto, podría resultar de utilidad para profesionales del audio vinculados al área de *mastering*.

1.2. Objetivo General

Diseñar e implementar una herramienta computacional de análisis espectral en tiempo real basado en parámetros psicoacústicos.

1.3. Objetivos Específicos

1. Diseñar e implementar distintos prototipos de analizador de espectro basados en bandas de tercio de octava y en bandas críticas.
2. Comparar y ajustar el modelo de *warping* de frecuencias al modelo de bandas críticas.
3. Incorporar los filtros de ponderación A, C [17], D [1] y Z (respuesta plana).
4. Buscar patrones de balance espectral en temas musicales mediante la utilización del *Analizador de espectro psicoacústico*.

Capítulo 2

MARCO TEÓRICO

Gran parte de este capítulo está dedicado al tratamiento de conceptos relacionados con **psicoacústica**: ciencia que estudia los fenómenos asociados a la percepción auditiva. No obstante, se comenzará definiendo el concepto de banco de filtros el cual tendrá una gran trascendencia en la presente investigación, considerando que se propondrá un modelo perceptual para un banco de filtros en bandas críticas.

2.1. Banco de Filtros

Un banco de filtros se define como un sistema que permite dividir una señal de audio en tiempo discreto $x[n]$ en una serie de señales $\{x_1[n], x_2[n], \dots\}$ de manera que cada una de ellas corresponda a una región de frecuencias diferente del espectro de la señal original $x[n]$. En la Figura 2.1 se muestra la respuesta en frecuencia de un banco de filtros ideal que permite separar una señal de audio en 3 regiones no solapadas [4].

En la práctica, la amplitud de respuesta de una banda en sus respectivas frecuencias de corte corresponde a 0,707 veces la amplitud máxima del filtro (-3 dB), por lo que las bandas contiguas a ella presentan un cierto grado de entrecruzamiento. Pese a esto, la superposición de todas las bandas del banco de filtros reconstruye aproximadamente la señal de audio originalmente considerada.

Tanto las frecuencias de corte f_i y f_s como la frecuencia central f_c de cada banda en un banco de filtros, así como sus pendientes de atenuación, adquieren

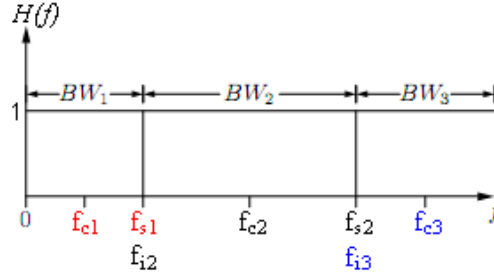


Figura 2.1: Banco de filtros ideal con 3 bandas. Las bandas poseen una pendiente de atenuación infinita en las frecuencias de corte inferior f_i y superior f_s de cada banda, por lo que idealmente las bandas no están solapadas.

diferentes valores de acuerdo al tipo de análisis que se desea llevar a cabo. Algunos de estos bancos de filtros se encuentran rotulados en estándares internacionales, como es el caso de las bandas de octava, que revisaremos a continuación.

2.1.1. Bandas de octava y fracciones de octava

En variadas aplicaciones de ingeniería se hace utilización del análisis en bandas de octava o de fracciones de octava (como es el caso de las bandas de 1/3 de octava) con el fin de analizar la distribución de energías en cada banda, no sólo aplicadas a señales de audio. Debido a lo anterior es que en el estándar internacional IEC 61260 [16], publicado en 1995, se especifican las características de este tipo de banco de filtros.

La división del espectro de frecuencias audibles es efectuada de manera que la frecuencia central f_c de una banda corresponda al doble de la frecuencia central f_{c-1} de la banda anterior. Esto es reflejado en la Ecuación 2.1, considerando una constante $1/b = 1$. A este factor se le denomina *Designador de Ancho de Banda*.

$$\frac{f_c}{f_{c-1}} = 2^{1/b} \quad (2.1)$$

Si, en lugar de utilizar un *designador de ancho de banda* igual a 1 ($1/b = 1$), se considera un *designador* igual a 1/3 ($1/b = 1/3$), entonces la Ecuación 2.1 establece la relación entre 2 frecuencias centrales contiguas correspondientes a un banco de filtros en bandas de 1/3 de octava.

De manera general, las frecuencias centrales de una banda pueden ser referenciadas a una frecuencia f_r , definida arbitrariamente en $f_r = 1.000$ Hz, de manera que el cálculo de una frecuencia central cualquiera se efectúa considerando la Ecuación 2.2.

$$f_c[k] = 2^{k/b} \cdot f_r = 2^{k/b} \cdot 1.000 \quad (2.2)$$

La constante k adopta valores enteros de manera que f_c pertenezca al rango de frecuencias auditivo. La constante b determina si las frecuencias centrales corresponden a un banco de filtros en bandas de octava ($b = 1$) o en fracciones de octava ($b = 3, b = 6$, etc.). Particularizando al caso de un banco de filtros en bandas de un tercio de octava:

$$f_c[k] = 2^{k/3} \cdot 1.000 \quad (2.3)$$

Las frecuencias de corte superior f_s e inferior f_i son definidas por las Ecuaciones 2.4 y 2.5, considerando un $b = 3$ para el caso de un banco de filtros en bandas de un tercio de octava.

$$f_s[k] = 2^{+1/(2b)} \cdot f_c[k] \quad (2.4)$$

$$f_i[k] = 2^{-1/(2b)} \cdot f_c[k] \quad (2.5)$$

También es posible interpretar estas frecuencias de corte como la media geométrica entre las frecuencias centrales de bandas adyacentes:

$$f_s = \sqrt{f_c[k+1] \cdot f_c[k]} \quad (2.6)$$

$$f_i = \sqrt{f_c[k] \cdot f_c[k-1]} \quad (2.7)$$

A partir de las Ecuaciones 2.6 y 2.7, es posible obtener la expresión que entrega el ancho de banda (BW^1), en Hz, de cada filtro pasabanda como función de su frecuencia central f_c :

¹BW, del inglés bandwidth.

$$BW[k] = f_c[k] \frac{2^{1/3} - 1}{2^{1/6}} \quad (2.8)$$

Las Ecuaciones 2.1 a 2.5 son la base del estándar internacional IEC 61260:1995 [16], aplicados a la definición de un banco de filtros en bandas de 1/3 de octava. En la Tabla 2.1 se muestran las frecuencias aproximadas que caracterizan el banco de filtros de 1/3 de octava considerando sólo aquellas bandas que caen dentro del rango auditivo.

2.1.2. IEC 61260: Bandas de 1/3 de octava

Al considerar un *Designador de ancho de banda* $b = 3$, se obtiene el más popular de los bancos de filtros de fracciones de octava: las bandas de un tercio (1/3) de octava. En la Tabla 2.1 se presentan tanto sus frecuencias centrales como las frecuencias de corte asociadas a este banco, de acuerdo al estándar IEC 61260:1995.

Banda Nºk	Frecuencia[Hz]			Banda Nºk	Frecuencia[Hz]		
	f _i	f _s	f _c		f _i	f _s	f _c
-16	22	28	25	-1	713	898	800
-15	28	35	31	0	891	1.123	1.000
-14	36	45	40	1	1.114	1.403	1.250
-13	45	56	50	2	1.425	1.796	1.600
-12	56	71	63	3	1.782	2.245	2.000
-11	71	90	80	4	2.227	2.806	2.500
-10	89	112	100	5	2.806	3.536	3.150
-9	111	140	125	6	3.564	4.490	4.000
-8	143	180	160	7	4.455	5.612	5.000
-7	178	225	200	8	5.613	7.072	6.300
-6	223	281	250	9	7.127	8.980	8.000
-5	281	354	315	10	8.909	11.225	10.000
-4	356	449	400	11	11.136	14.031	12.500
-3	445	561	500	12	14.254	17.959	16.000
-2	561	707	630	13	17.818	22.449	20.000

Tabla 2.1: Bandas de 1/3 de octava. Tanto los números de banda k como las frecuencias centrales nominales aquí mostradas se encuentran estandarizadas en la norma IEC 61260:1995 [16].

Para proceder en el diseño de un banco de filtros de 1/3 de octava sólo falta definir la pendiente de atenuación presentada por cada banda. En la norma IEC 61260 se entrega la Tabla 2.2. Este cuadro establece las atenuaciones mínimas y máximas que debe cumplir un banco de filtros de 1/3 de octava para satisfacer una precisión Clase 0, Clase 1 o Clase 2.

Frecuencia Normalizada, f/f_m (base 2)	Mínimo; Máximo límite de atenuación [dB]		
	Clase del Filtro		
	0	1	2
1,00000	-0,15; +0,15	-0,3; +0,3	-0,5; +0,5
1,02676 ó 0,97394	-0,15; +0,2	-0,3; +0,4	-0,5; +0,6
1,05594 ó 0,94702	-0,15; +0,4	-0,3; +0,6	-0,5; +0,8
1,08776 ó 0,91958	-0,15; +1,1	-0,3; +1,3	-0,5; +1,6
<1,12246 ó >0,89090	-0,15; +4,5	-0,3; +5,0	-0,5; +5,5
>1,12246 ó <0,89090	+2,3; +4,5	+2,0; +5,0	+1,6; +5,5
1,29565 ó 0,77181	+18,0; +∞	+17,5; +∞	+16,5; +∞
1,88695 ó 0,52996	+42,5; +∞	+42; +∞	+41; +∞
3,06955 ó 0,32578	+62,0; +∞	+61; +∞	+55; +∞
≥5,43474 ó ≤0,18400	+75,0; +∞	+70; +∞	+60; +∞

Tabla 2.2: Límites en la atenuación relativa para filtros de 1/3 de octava, según IEC 61260.

En la Figura 2.2 se muestran los puntos de atenuación mínima estandarizada para un filtro de 1/3 de octava. La línea continua representa un filtro que satisface con dichas exigencias. La banda mostrada corresponde a la banda centrada en 5.000 Hz ($k = 7$) y corresponde a un filtro de respuesta impulsiva infinita (IIR) (Ver Sección 2.3.1) del tipo Butterworth, de orden 6.

Mediante la utilización de un algoritmo preprogramado de MATLAB ® fue

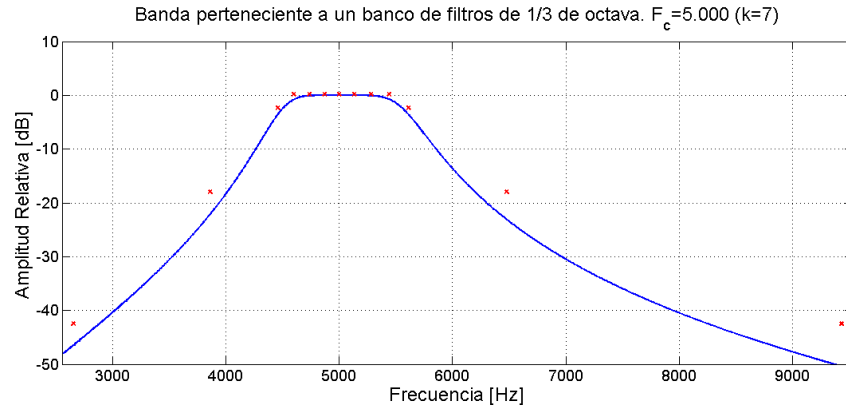


Figura 2.2: La línea continua (azul) corresponde a la respuesta en frecuencia de una banda centrada en 5.000 Hz, la cual cumple con los requerimientos de atenuación mínima exigidos por la IEC 61260 (cruces en rojo).

posible corroborar la posibilidad de obtener un banco de filtros Clase 0, para cada una de las bandas de frecuencia, al diseñar considerando filtros IIR de orden 6 del tipo Butterworth.

2.2. Psicoacústica

El entorno sonoro en que vivimos es extremadamente complejo. Donde quiera que vayamos existe una infinidad de fuentes sonoras provocando una serie de sensaciones, muchas veces condicionando nuestro comportamiento frente a ello.

Si lo miramos desde otro punto de vista, podemos preguntarnos cómo podemos extraer esta información y hacer uso de ella si se encuentra codificada en un arreglo tan complejo y altamente estructurado de ondas acústicas. Pero hay más que eso, nuestra percepción sonora no depende sólo del espacio físico en el que estamos inmersos sino que también depende de las características de nuestro receptor de ondas acústicas: el sistema auditivo.

La ciencia que estudia el comportamiento de nuestro sistema auditivo, es decir, la relación entre estímulos acústicos y su sensación auditiva primaria es la ciencia de la **Psicoacústica**.

Como un primer paso, entonces, se hará una revisión de la fisiología y funcionamiento del sistema auditivo humano.

2.2.1. Aspectos generales: Fisiología del oído humano

El oído es dividido en 3 regiones: oído externo, oído medio y oído interno. Es en este último (Fig. 2.2.1²), más específicamente en la membrana basilar (ubicada en la coclea), donde se produce la descomposición espectral de las ondas acústicas en función del punto en que ésta es estimulada.

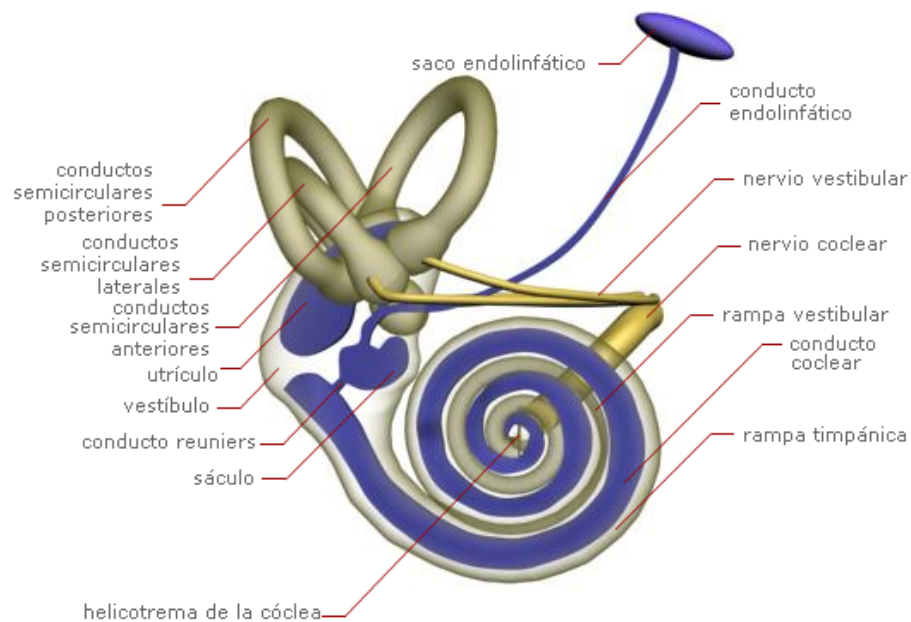


Figura 2.3: Oído interno. Está compuesto principalmente por los conductos semicirculares y la coclea o caracol.

En 1990, Greenwood [11] introdujo una expresión analítica para la determinación del punto x de la membrana basilar que es estimulado en función de la frecuencia f que es percibida:

$$x = 16,67 \log_{10} \left(\frac{f + 165,4}{165,4} \right) \quad (2.9)$$

Donde f representa la frecuencia en Hz y x está expresado en mm.

²Imagen extraída del sitio: http://www.elmundo.es/elmundosalud/documentos/2007/09/oido_anatomia.html

2.2.2. Bandas Críticas

El concepto de banda crítica está asociado al fenómeno psicoacústico de enmascaramiento. Ante la presencia de un ruido, una banda crítica se define como una región de frecuencias de paso (filtro pasabanda) en la que cualquier tono puro cuya amplitud se encuentre por debajo de este filtro será inaudible al oído humano, fenómeno correspondiente al enmascaramiento de señales.

El sistema auditivo humano es modelado usualmente como una serie de 24 bandas de frecuencia solapadas, denominadas bandas críticas [42], donde cada una de ellas es capaz de estimular la misma cantidad de células en la membrana basilar (150 células ciliadas de un total de 3.600), produciendo un desplazamiento proporcional en ella en relación a la frecuencia expresada usando la razón de banda crítica, definida analíticamente por la Ecuación 2.10 [42]. Esta relación de proporcionalidad es esquematizada en la Figura 2.4.

$$\frac{z}{\text{Bark}} = 13 \arctan\left(0,76 \frac{f}{\text{kHz}}\right) + 3,5 \arctan\left(\frac{f}{7,5 \text{ kHz}}\right)^2 \quad (2.10)$$

Donde z se mide en Barks y representa la razón de banda crítica y f representa la frecuencia en kHz.

El ancho de cada banda crítica está definido por la Ecuación 2.11.

$$\text{CB/Hz} = 25 + 75[1 + 1,4(f_c/\text{kHz})^2]^{0,69} \quad (2.11)$$

Donde CB corresponde al denominado ancho de banda crítico, expresado en Hz, y f_c es la frecuencia central en kHz del filtro. Esta expresión posee una tolerancia de un 10% respecto de los anchos de banda experimentales, rotulados en [40]. Al convertir el ancho de banda crítico CB al dominio Bark, todas las bandas críticas quedan caracterizadas por un ancho de banda constante de 1 Bark.

En la Tabla 2.3 son mostradas las frecuencias de corte y frecuencias centrales de cada banda crítica, determinadas experimentalmente y redondeadas al múltiplo de 10 más cercano [42]. Estos experimentos se realizaron sobre la base de

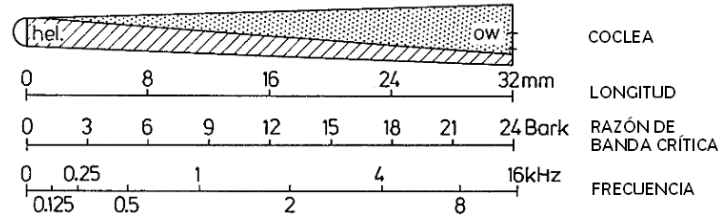


Figura 2.4: Escalas de alturas en relación a la longitud de la coclea, sin daños [9]. Las escalas de frecuencia y posición (longitud) mostradas se relacionan mediante la Ecuación 2.9.

mediciones del umbral de enmascaramiento en relación a tonos de características complejas.

Banda Nº	Frecuencia[Hz]			Banda Nº	Frecuencia[Hz]		
	f_i	f_s	f_c		f_i	f_s	f_c
0	0	100	50	13	2.000	2.320	2.150
1	100	200	150	14	2.320	2.700	2.500
2	200	300	250	15	2.700	3.150	2.900
3	300	400	350	16	3.150	3.700	3.400
4	400	510	450	17	3.700	4.400	4.000
5	510	630	570	18	4.400	5.300	4.800
6	630	770	700	19	5.300	6.400	5.800
7	770	920	840	20	6.400	7.700	7.000
8	920	1.080	1.000	21	7.700	9.500	8.500
9	1.080	1.270	1.170	22	9.500	12.000	10.500
10	1.270	1.480	1.370	23	12.000	15.500	13.500
11	1.480	1.720	1.600	24	15.500	22.050	18.775
12	1.720	2.000	1.850				

Tabla 2.3: Bandas críticas. Son mostradas las especificaciones de 25 filtros pasa-banda. El último de ellos es una extensión de los valores definidos por Zwicker.

En la Tabla 2.3, el número de banda de cada filtro puede ser interpretado como el valor asociado a la frecuencia de corte inferior, en dominio Bark. También se puede diferir el ya mencionado ancho de banda crítica constante de 1 Bark.

En la Figura 2.5 se muestra la relación entre la frecuencia en Hz y su equivalente en Barks³.

³La relación existente entre 2 escalas de frecuencias se denomina Mapeamiento

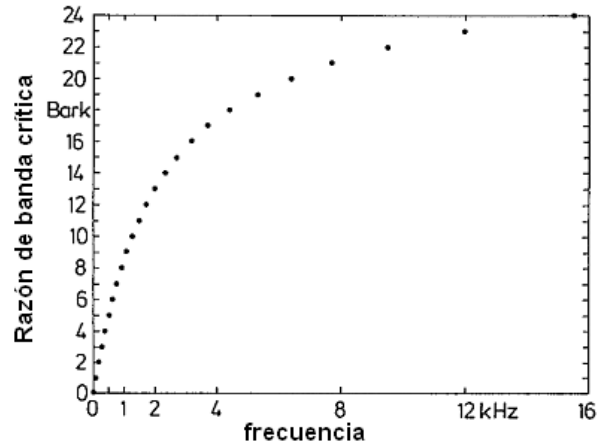


Figura 2.5: Razón de banda crítica Z , medida en Barks, en función de la frecuencia [kHz]. El parámetro Z se asocia a las bandas críticas adyacentes. Sólo se muestran los primeros 24 valores.

2.2.3. Modelo de Enmascaramiento de Terhardt

El modelo de enmascaramiento de Terhardt permite la obtención del umbral de enmascaramiento asociado a una señal enmascaradora del tipo sinusoidal (tono puro). Para ello, entrega las pendientes de atenuación de una banda de frecuencias, que se aproximan a líneas rectas al considerar un eje de frecuencias expresado en Barks, y que decaen desde la frecuencia de sintonía de dicha banda de enmascaramiento.

Si consideramos un tono puro de 1 kHz a 80 dB, se obtiene una banda de enmascaramiento con frecuencias de corte de 920 y 1.080 Hz (banda crítica N° 8, Tabla 2.3), tal como se muestra en la Figura 2.6.

Las diferencias entre la pendiente superior e inferior manifiestan la naturaleza asimétrica de una banda crítica. El modelo de enmascaramiento de Terhardt⁴ considera lo anterior y establece las expresiones analíticas para la determinación de dichas pendientes en función de la frecuencia y del nivel del tono enmascarador (Ecuaciones 2.12 y 2.13)

⁴El modelo fue publicado en el año 1979 [37], sin embargo, también es detallado en [38]. Otra referencia al mismo modelo de enmascaramiento puede ser encontrada en [20]

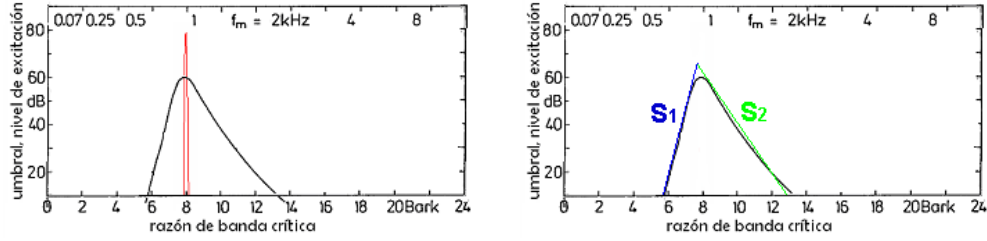


Figura 2.6: (a) Todas aquellas señales cuya amplitud caiga bajo la curva de enmascaramiento (curva negra) no será percibida ante la presencia del tono puro de 1 kHz a 80 dB (curva roja). (b) Aproximación a las pendientes de una banda crítica mediante líneas rectas.

$$S_1 = 27 \quad f < F_M \quad (2.12)$$

$$S_2 = -24 - 230/F_M + 0,2 \cdot L_p \quad f > F_M \quad (2.13)$$

Donde S_1 y S_2 están expresados en dB/Bark (véase la Figura 2.6.b), F_M representa la frecuencia de enmascaramiento en Hz, que coinciden con las frecuencias centrales de las bandas críticas definidas por Zwicker en [40] y L_p es el nivel de presión sonora en decibeles.

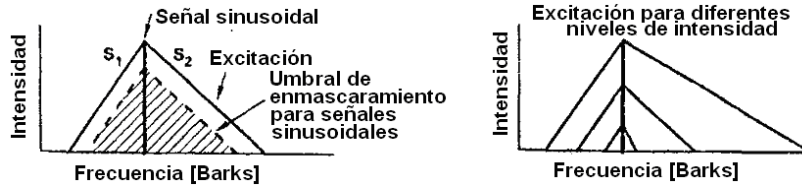


Figura 2.7: Modelo de Terhardt para el umbral de enmascaramiento de señales sinusoidales. Se puede apreciar la variación de las pendientes superiores en función de la intensidad del tono enmascarador [20].

2.2.4. Ponderación de amplitud: Sonoridad o Loudness

La sonoridad o *loudness* es una medida subjetiva de la intensidad de un sonido, de manera que 2 tonos puros con igual sonoridad son percibidos a igual intensidad por el oído humano. Se define el **nivel de sonoridad P** , como el nivel de presión sonora (L_p) a 1 kHz que es percibido a igual intensidad que un L_p a alguna

determinada frecuencia. La unidad empleada para el nivel de sonoridad es el fon (Phon). El análogo subjetivo a la presión sonora es la **sonoridad**, medida en Sones (Sone), y está definida por las Ecuaciones 2.14 y 2.15 [3].

$$S = 2^{\frac{P-40}{10}} \quad \text{Si } P \geq 40 \quad (2.14)$$

$$S = \left(\frac{P}{40}\right)^{2,642} \quad \text{Si } P < 40 \quad (2.15)$$

De estas ecuaciones, es posible deducir la relación inversa entre la Sonoridad S y el Nivel de Sonoridad P , lo que se expresa en las ecuaciones 2.16 y 2.17:

$$P = 10 \cdot \log_2(S) + 40 \quad \text{Si } S \geq 1 \quad (2.16)$$

$$P = 40 \cdot S^{0,3785} \quad \text{Si } S < 1 \quad (2.17)$$

El concepto de sonoridad fue inicialmente definido para señales sinusoidales (tonos puros). En este contexto, resulta más sencillo obtener el Nivel de Sonoridad P asociado a un tono y luego convertirlo a Sonoridad usando las Ecuaciones 2.14 o 2.15, según corresponda. La obtención del Nivel de Sonoridad P se efectúa mediante la lectura de los llamados Contornos de Igual Sonoridad [19].

2.2.5. Contornos de Igual Sonoridad

Los contornos de igual sonoridad o curvas isofónicas representan la respuesta del oído humano ante diferentes niveles y frecuencias. Fueron obtenidas empíricamente en condiciones de campo libre (ambiente sin reflexiones) para personas sin daño auditivo. Además, como señales de estímulo al sistema auditivo humano, se usaron tonos puros provenientes desde una dirección normal⁵ al oyente. Como consecuencia de los numerosos estudios desarrollados desde la década de los 30 para la determinación estas curvas, en 2003 se publicaron los contornos que permiten obtener el Nivel de Sonoridad P de un tono puro sintonizado en una frecuencia de f Hz con un Nivel de Presión Sonora de L_p dB.

⁵Sonido proveniente desde el frente del auditor

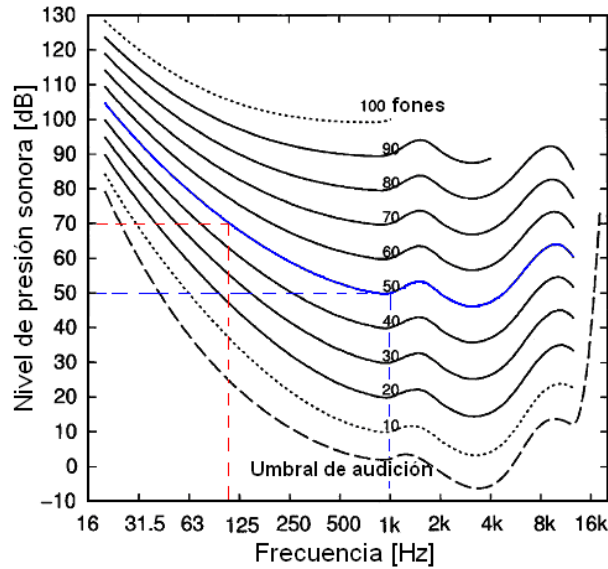


Figura 2.8: Curvas de igual sonoridad, según Suzuki y Takeshima[35]. Se tiene el nivel de presión sonora en función de la frecuencia. Cada curva es caracterizada por un valor en Fones que es igual al nivel de presión sonora a 1 kHz.

En la Figura 2.2.5 se muestran las curvas isofónicas de acuerdo a [19]. También se muestra que el Nivel de Sonoridad P correspondiente a un tono puro de alrededor de 70 dB @ 125 Hz es igual a 50 Fones, por lo que ese tono es percibido con igual intensidad que otro tono puro, sintonizado en 1 kHz, pero de amplitud 50 dB.

En otras palabras, el rótulo de cada curva isofónica coincide con el nivel de presión sonora L_p presentado por la respectiva curva en la frecuencia de 1 kHz.

Los contornos de igual sonoridad denotan la sensibilidad variable que presenta el oído humano en función de la frecuencia de un sonido y, por tanto, plantea la necesidad de otorgar una importancia diferente a cada componente frecuencial cuando se efectúa una suma de energías.

Debido a lo anterior, cuando se analiza la calidad sonora de una señal de audio, parece ser más apropiado efectuar una suma energética basada en el concepto de sonoridad que en el de presión sonora, sin embargo, existe la limitación de que

los contornos de igual sonoridad no están definidos para señales que presentan un espectro de frecuencias complejo (sólo se definen para tonos puros) y no están caracterizados analíticamente por una función matemática.

Como etapa anterior a la consideración de los contornos de igual sonoridad para el análisis de señales y en un afán de desarrollar descriptores acústicos de mayor significación psicoacústica, se propusieron las curvas de ponderación en frecuencia.

2.2.6. Curvas de Ponderación en Frecuencia

Las curvas de ponderación en frecuencia asumen que el nivel de sonoridad de una señal acústica es aproximadamente constante. Por ejemplo, la curva A –que es la más popular de estas curvas de ponderación– se correlaciona con la curva isofónica de 40 fones. La curva B está relacionada con la isofónica 70, mientras que la curva C se relaciona con la isofónica 100.

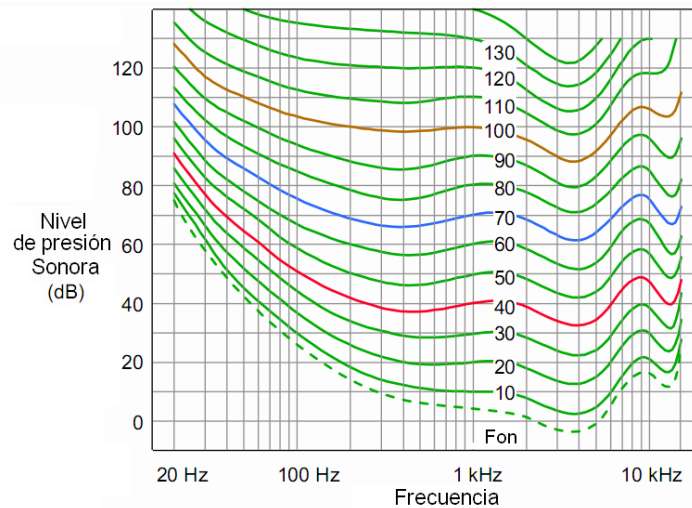


Figura 2.9: Contornos de igual sonoridad según Robinson y Dadson [30]. Son destacadas las curvas de 40, 70 y 100 Fon, a partir de las cuales se infirieron las curvas de ponderación A, B y C: curvas inversas a la desviación en dB de la respectiva isofónica usando como referencia la frecuencia de 1.000 Hz.

Tanto la curva de ponderación A (Figura 2.10.b) como el resto de las curvas

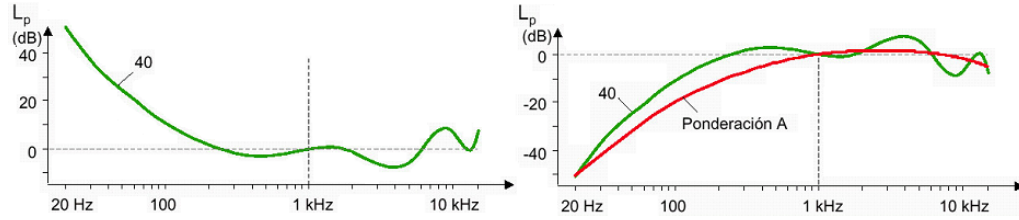


Figura 2.10: (a) Curva isofónica 40, normalizada a 0 dB@ 1 kHz. (b) Curva isofónica 40 complementaria, comparada con la curva de ponderación A.

de ponderación están representadas por una serie de funciones de transferencia, definidas inicialmente en el dominio análogo de frecuencias (Ecuaciones 2.18 a 2.21). Las funciones de transferencia asociadas a las curvas de ponderación en frecuencia A, B y C, son respectivamente:

$$H(s) = \frac{7,3901 \cdot 10^9 s^4}{(s + 129,43)^2 (s + 676,70) (s + 4636,36) (s + 76617,16)^2} \quad (2.18)$$

$$H(s) = \frac{5,9862 \cdot 10^9 s^4}{(s + 129,43)^2 (s + 995,88) (s + 76617,16)^2} \quad (2.19)$$

$$H(s) = \frac{5,9124 \cdot 10^9 s^2}{(s + 129,43)^2 (s + 76617,16)^2} \quad (2.20)$$

Las Funciones de Transferencia $H(s)$ de las Ecuaciones 2.18 a 2.21 son obtenidas al considerar una combinación específica de 5 polos ubicados en las frecuencias mostradas en la Tabla 2.4.

p_i	f_i	$\Omega_i = 2\pi f_i$	Definido en
p_1	20,6	129,43	IEC 61672-1:2002
p_2	107,7	676,70	IEC 61672-1:2002
p_3	737,9	4.636,36	IEC 61672-1:2002
p_4	12.194	76.617,16	IEC 61672-1:2002
p_5	158,5	995,88	IEC 60651:1979

Tabla 2.4: Polos considerados en la definición de las curvas de ponderación en frecuencia A, B y C. Los estándares considerados corresponden a la norma IEC 60651:1979 [15] la cual fue reemplazada por la norma IEC 61672 [17]. En este caso se hace referencia a una norma derogada debido a que en ella se definió la curva de ponderación B.

Otra curva de ponderación, la curva D, fue utilizada como descriptor para el ruido de aviones. Se estandarizó en 1976 [14], siendo posteriormente derogada. No se asocia a ninguna curva isofónica, sin embargo otorga una alta ponderación a las frecuencias entre 2 y 6 kHz. Su función de transferencia, en dominio analógico, está definida por la Ecuación 2.21.

$$H(s) = \frac{91103,487 \cdot s \cdot (s^2 + 6532s + 4,0975 \cdot 10^7)}{(s + 1773,6)(s + 7288,5)(s^2 + 21514s + 3,8836 \cdot 10^8)} \quad (2.21)$$

Para la suma energética de presiones ponderadas, se sigue considerando el mismo principio de suma de presiones. Se expresan en decibeles y debe indicarse entre paréntesis la curva de ponderación aplicada: dB(A), dB(C), dB(B) o dB(D)⁶.

En el presente estudio se decidió incluir las 4 curvas de ponderación, además de la denominada curva Z (respuesta plana). La curva B, pese a estar obsoleta, considera que la señal en análisis posee una sonoridad de 70 fones, es decir, se asocia a sonidos de intensidad sonora media, a diferencia de la curva A que establece como premisa que las componentes de frecuencia de la señal son de bajo nivel. Por otro lado, la curva D se incorporó debido a ser un exigente descriptor acústico al otorgar una importante ponderación a las frecuencias más sensibles del oído humano.

2.2.7. Ponderación de sonoridad (*Loudness*) usando Redes Neuronales Artificiales

Tal como se mencionó anteriormente, los contornos de igual sonoridad no tienen asociada una expresión analítica que los caracterice. Además, las curvas se encuentran definidas en pasos discretos de 10 fones, por lo que se dificulta la deducción del Nivel de Sonoridad P exacto asociado a un tono cualquiera. Esta situación motivó a Espinoza, Venegas y Floody [8] a implementar una Red Neu-

⁶Como se mencionó anteriormente, tanto la curva B, como la curva D no están actualmente vigentes.

ronal Artificial (RNA)⁷ que permitiera la obtención del Nivel de Sonoridad P asociado a un tono puro, tanto en condiciones de coincidencia con los contornos de igual sonoridad rotulados en [19], como en los puntos intermedios a estas curvas (interpolación). La RNA propuesta corresponde a un Perceptrón Multicapa (MLP) en configuración 2-11-9-1. La red devuelve una única salida correspondiente al Nivel de Sonoridad P , determinado a partir la frecuencia f y el nivel de presión sonora L_p . Las capas ocultas poseen 11 y 9 neuronas y funciones de transferencia tangente sigmoideal y logarítmica sigmoideal, respectivamente.

En la Figura 2.11 se muestran las curvas de igual Nivel de Presión Sonora L_p , las cuales fueron determinadas mediante la implementación de la mencionada RNA.

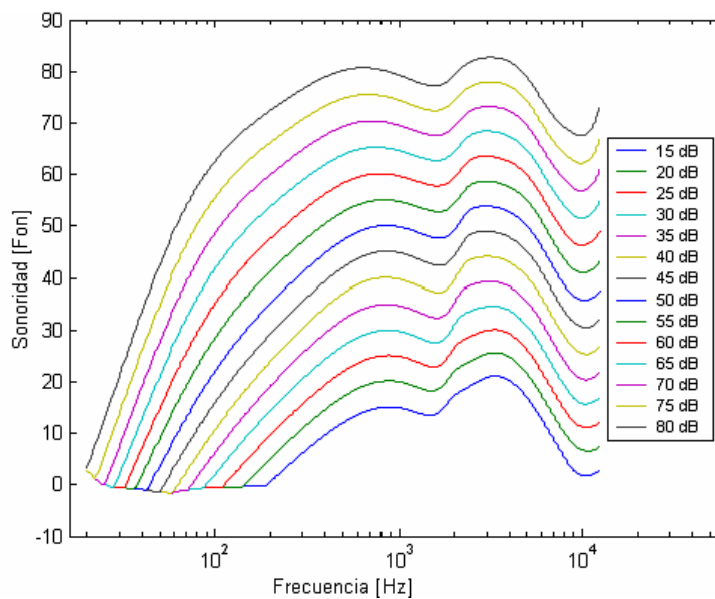


Figura 2.11: Curvas de igual nivel de presión sonora, determinados usando la Red Neuronal Artificial *feed-forward* 2-11-9-1. En el eje de las ordenadas se lee el Nivel de Sonoridad P asociado a cada punto de frecuencia y nivel de presión sonora L_p (rotulado en las leyendas).

A diferencia de las curvas de ponderación en frecuencia (proceso de ponderación estática), al usar la RNA, se efectúa un proceso de ponderación dinámica, ya

⁷Las Redes Neuronales Artificiales serán revisadas en la sección 2.3.4.

que se obtiene un nivel de sonoridad P que tiene una directa relación con cada nivel de presión sonora y cada frecuencia de una señal. Además, la implementación de la RNA MLP es una operación computacionalmente económica.

2.2.8. Suma energética: Presión Sonora y Nivel de Presión Sonora

Lo que realmente percibe el oído humano cuando oye un sonido son las variaciones de presión sonora en un medio. La mínima presión acústica capaz de estimular a nuestro sistema auditivo es del orden de los $2 \cdot 10^{-5}$ Pascales (Pa). El rango de percepción se extiende aproximadamente hasta los 20 Pa, correspondiente a una presión un millón de veces mayor que el umbral auditivo. Se hace necesario utilizar una escala logarítmica de manera de reducir numéricamente este gran rango dinámico de presiones. Esta escala logarítmica, medida en decibeles (dB) se extiende aproximadamente desde 0 dB ($2 \cdot 10^{-5}$ Pa) hasta 120 dB (20 Pa). La relación entre el Nivel de Presión Sonora L_p en dB y la presión sonora p en Pa está dada por:

$$L_p = 10 \cdot \log_{10} \left(\frac{p^2}{p_0^2} \right) = 20 \cdot \log_{10} \left(\frac{p}{p_0} \right) \quad (2.22)$$

Donde p_0 es la denominada presión de referencia correspondiente a $p_0 = 2 \cdot 10^{-5}$ Pa.

Cuando se desea sumar la presión sonora de dos fuentes diferentes, con un nivel de presión sonora L_{p1} y L_{p2} , se debe aplicar la Ecuación 2.23.

$$L_{p_{\text{total}}} = 10 \cdot \log_{10} \left(10^{\frac{L_{p1}}{10}} + 10^{\frac{L_{p2}}{10}} \right) \quad (2.23)$$

Esto corresponde a la suma de las presiones (expresadas en pascales) al cuadrado. Si se poseen n fuentes, la suma energética de la presión sonora resultante puede generalizarse como:

$$L_{p_{\text{total}}} = 10 \cdot \log_{10} \left(\sum_{i=1}^n 10^{\frac{L_{pi}}{10}} \right) \quad (2.24)$$

En el caso de las señales de audio digital, la amplitud $x[n]$ de la señal puede ser trabajada de manera análoga a la cantidad (p/p_0) . De esta manera, el nivel de amplitud aplicado a este caso cumple con la Ecuación 2.25.

$$L_p = 20 \cdot \log_{10} (|x[n]|/N) \quad (2.25)$$

La unidad de medida de L_p , en este caso, es el decibel a escala completa dBFS, donde el nivel máximo se da cuando $x[n]$ tiene una amplitud de 1 o -1, lo que corresponde a 0 dBFS. Para una equivalencia entre la escala dBFS y la escala en dB debe aplicarse una calibración de la señal de audio digital.

2.2.9. Suma de sonoridad: Sonoridad Específica y Sonoridad Total

El proceso de suma de sonoridad es un proceso que se efectúa en el dominio de la frecuencia. Se debe tener el espectro de frecuencias de la señal en análisis convertido a Sones. La curva que se obtiene al graficar la sonoridad en Sones en función de la frecuencia es lo que se denomina la curva de sonoridad específica, donde el eje de frecuencias debe estar expresado en Barks (Recordar la Ecuación 2.10). De esta manera, la Sonoridad Específica S' se mide en Sones/Bark.

La Sonoridad Total S de una señal corresponde al área bajo la curva de sonoridad específica S' , por lo que se determina efectuando la integral de S' a lo largo de todo el eje de frecuencias audibles [9], tal como lo expresa la Ecuación 2.26.

$$S = \int_0^{24 \text{ Barks}} S' dz \quad (2.26)$$

El diferencial dz denota que el eje de frecuencias debe estar expresado en Barks. En una aproximación discreta, la integral de la Ecuación 2.26 puede ser expresada como:

$$S = \sum_{k=0}^{K-1} S'_k \Delta z_k \quad (2.27)$$

Donde K es el número de componentes de la respuesta en frecuencia de $x[n]$ ($X[k]$), S'_k es la sonoridad asociada a la componente k de $X[k]$ y Δz_k es la distancia, en Barks, entre la componente de frecuencia $z(k)$ y $z(k+1)$.

2.2.10. Distribución de energía de una señal de audio

Resulta de gran utilidad en diversas áreas de la ingeniería analizar el espectro de frecuencias de una señal determinada. Una de las herramientas de monitoreo de la distribución energética es la aplicación de un banco de filtros, en donde se divide el rango de frecuencias en diferentes bandas, determinándose la amplitud de energía para cada una de ellas.

En el caso de las señales de audio discretas, un popular indicador de la cantidad de energía es el Valor Cuadrático Medio (RMS), el cual entrega un valor de amplitud promediado para una señal temporal $x[n]$, con N muestras de longitud:

$$X_{\text{RMS}} = \sqrt{\frac{1}{N} \sum_{i=1}^N x^2[n]} \quad (2.28)$$

La cantidad X_{RMS} también puede ser determinado utilizando la información en el dominio de la frecuencia de la misma señal $x[n]$, que denotaremos por $X[k]$:

$$X_{\text{RMS}} = \sqrt{\frac{1}{N} \left(\frac{1}{K} \sum_{k=1}^K X^2[k] \right)} \quad (2.29)$$

La Ecuación 2.29 es una consecuencia del Teorema de Parseval aplicado al caso discreto, el cual está dado por la Ecuación 2.30.

$$E = \sum_{n=1}^N |x[n]|^2 = \frac{1}{K} \sum_{k=1}^K |X[k]|^2 \quad (2.30)$$

Donde $X[k]$ corresponde a la transformada de Fourier de $x[n]$. Sólo se utilizan las $K = N/2$ primeras muestras de la respuesta en frecuencia $X[k]$, debido a que la segunda mitad entrega información redundante.

La descomposición de una señal de audio para analizar la distribución de energía como función de la frecuencia se hace más evidente al filtrar la señal

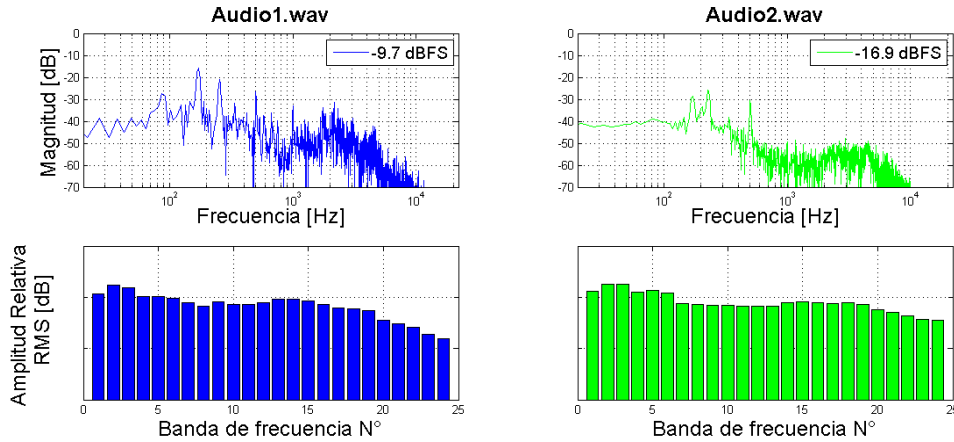


Figura 2.12: Transformada de Fourier para 2 señales de audio digital (arriba) y descomposición en un banco de 24 bandas de frecuencia (abajo).

original considerando algún banco de filtros. En la Figura 2.12 se puede notar que la comparación entre señales de audio se hace menos compleja comparando i valores de amplitud (de las i bandas de frecuencia) que comparando $K = N/2$ amplitudes.

2.2.11. Modelos de Sonoridad

Existen variados modelos para el cálculo de la sonoridad de una señal acústica compleja, sin embargo, sólo 2 de ellos se encuentran estandarizados, ambos en la norma ISO 532:1975 [18]. Estas metodologías fueron bautizadas como Método A y Método B y están basadas en los trabajos de investigación de los científicos S.S. Stevens y Eberhard Zwicker, ambos de gran influencia en el mundo de la psicoacústica. Algunos alcances relativos a ambos métodos se resumen en la Tabla 2.5.

El Método A es un método sencillo recomendado cuando se cuenta con un análisis en bandas de octava de la señal en evaluación⁸, sin embargo, se utilizará como referente el **Método B**, que considera un análisis en bandas críticas estimado a partir de un análisis en bandas de 1/3 de octava. Además, este método

⁸Para una revisión detallada del cálculo de sonoridad usando el Método A, se recomienda la lectura de [31].

Método	Cantidad medida		Ancho de banda	Campo sonoro
	Sonoridad	Nivel de Sonoridad		
A	Sones (OD)	Fones (OD)	octava	difuso
B	Sones (GD)	Fones (GD)	1/3 de octava	difuso
	Sones (GF)	Fones (GF)		libre

Tabla 2.5: Se especifica la nomenclatura que debe utilizarse al usar el Método A o el Método B en el cálculo de sonoridad. El Método A es recomendable en conjunto a un análisis en bandas de 1 octava y sólo está disponible para condiciones de campo difuso. El método B considera bandas de 1/3 de octava y entrega la posibilidad de cálculo en condiciones de campo libre.

permite la evaluación de la sonoridad en condiciones de campo libre, es decir, en un ambiente libre de reflexiones. Si bien una condición de campo difuso es más cercana a los ambientes acústicos de los lugares en que usualmente se escucha música, en esta investigación se trabajará asumiendo condiciones de campo libre, compatibilizando los resultados que se obtendrán durante la evaluación del **Método B** respecto del modelo que se propondrá a partir de los nuevos contornos de sonoridad, los cuales están definidos en campo libre.

2.2.12. Estándar ISO 532:1975-B

Este procedimiento para el cálculo de sonoridad es un procedimiento gráfico que se complementa con la medición física (previa) de los valores por bandas de frecuencia. Es recomendable efectuar un análisis en bandas de 1/3 de octava.

El método considera una aproximación a los contornos de igual sonoridad y las características de transmisión del oído externo y medio hasta el oído interno. Permite obtener un valor de Sonoridad Total S en condiciones de campo libre y de campo difuso.

Este método utiliza como parámetros de entrada los niveles de presión sonora medidos para 28 de las 31 bandas de 1/3 de octava que se encuentran dentro del rango auditivo (bandas -16 a 11, ver Tabla 2.1). A partir de ellas se efectúa la aproximación a 20 de las bandas críticas del oído humano. Para ejemplificar el cálculo de la sonoridad de una señal medida, se considerará el análisis en bandas

de 1/3 de octavas mostrado en la Tabla 2.6.

Banda Nºk	Frecuencia f _c [Hz]	L _p [dB]	Banda Nº	Frecuencia f _c [Hz]	L _p [dB]
-16	25	26,1	-2	630	51,3
-15	31	35,8	-1	800	49,0
-14	40	42,5	0	1.000	45,5
-13	50	55,7	1	1.250	41,7
-12	63	60,7	2	1.600	40,1
-11	80	68,7	3	2.000	48,4
-10	100	62,1	4	2.500	44,5
-9	125	55,4	5	3.150	42,9
-8	160	64,7	6	4.000	41,7
-7	200	55,7	7	5.000	40,9
-6	250	63,9	8	6.300	40,7
-5	315	61,3	9	8.000	39,3
-4	400	53,7	10	10.000	35,9
-3	500	50,8	11	12.500	35,8

Tabla 2.6: Niveles de Presión Sonora L_{pi} en bandas de 1/3 de octava asociados a una fuente de ruido estacionaria. Se determinará el nivel de sonoridad total P usando el procedimiento ISO 532-B.

Paso 1

Las bandas de 1/3 de octava que se encuentran por debajo de los 280 Hz poseen un ancho de banda muy inferior al de las bandas críticas del oído humano en este rango de frecuencias. Los L_{pi} de las 11 bandas que se encuentran en esta región deben agruparse como se describe a continuación para obtener los *niveles en bandas críticas* L_1 , L_2 y L_3 , usando la Ecuación 2.24. L_1 se obtiene agrupando las bandas -16 a -11, L_2 se obtiene agrupando las bandas -10 a -8 y L_3 se obtiene agrupando las bandas -7 y -6.

$$L_1 = 10 \cdot \log_{10} \left(10^{\frac{26,1}{10}} + 10^{\frac{35,8}{10}} + 10^{\frac{42,5}{10}} + 10^{\frac{55,7}{10}} + 10^{\frac{60,7}{10}} + 10^{\frac{68,7}{10}} \right) = 69,5 \text{ dB}$$

$$L_2 = 10 \cdot \log_{10} \left(10^{\frac{62,1}{10}} + 10^{\frac{55,4}{10}} + 10^{\frac{64,7}{10}} \right) = 66,9 \text{ dB}$$

$$L_3 = 10 \cdot \log_{10} \left(10^{\frac{55,7}{10}} + 10^{\frac{63,9}{10}} \right) = 64,5 \text{ dB}$$

Se obtiene una nueva tabla de valores de niveles de presión sonora (Tabla 2.7), incluyendo los niveles L_1 , L_2 y L_3 recientemente calculados.

Banda Nº	Frecuencia f_c [Hz]	L [dB]	Banda Nº	Frecuencia f_c [Hz]	L [dB]
1	63	69,5	11	1.600	40,1
2	125	64,7	12	2.000	48,4
3	224	64,5	13	2.500	44,5
4	315	61,3	14	3.150	42,9
5	400	53,7	15	4.000	41,7
6	500	50,8	16	5.000	40,9
7	630	51,3	17	6.300	40,7
8	800	49,0	18	8.000	39,3
9	1.000	45,5	19	10.000	35,9
10	1.250	41,7	20	12.500	35,8

Tabla 2.7: Niveles de Presión Sonora en Bandas Críticas L_i .

Paso 2

Teniendo en consideración el máximo nivel L_i de la Tabla 2.7 y las condiciones en las que se desea modelar la sonoridad de la señal en análisis (campo libre o campo difuso), se debe escoger uno de los 10 gráficos (5 gráficos para una situación de campo libre y 5 para una situación de campo difuso) que provee el estándar ISO para poder desarrollar esta tarea. Debemos preocuparnos de que la escala del gráfico escogido permita marcar en líneas horizontales todos los valores L_i , como se muestra en la Figura 2.13. Cuando los escalones formados por estas líneas horizontales son adyacentes y crecientes en la medida que aumenta la frecuencia, ellos deben ser unidos con una línea vertical. Si, por el contrario, los escalones son decrecientes, los escalones deben ser unidos trazando una línea oblicua paralela a la línea segmentada de la respectiva banda. Esta línea oblicuas manifiesta el enmascaramiento entre bandas de frecuencia.

En la Figura 2.13, casi todas las bandas presentan escalones decrecientes, salvo en el caso de las bandas centradas en 500 Hz (sobre la de 630 Hz) y la de 2,5 kHz (sobre la de 3,15 kHz). Además debe observarse que la banda crítica

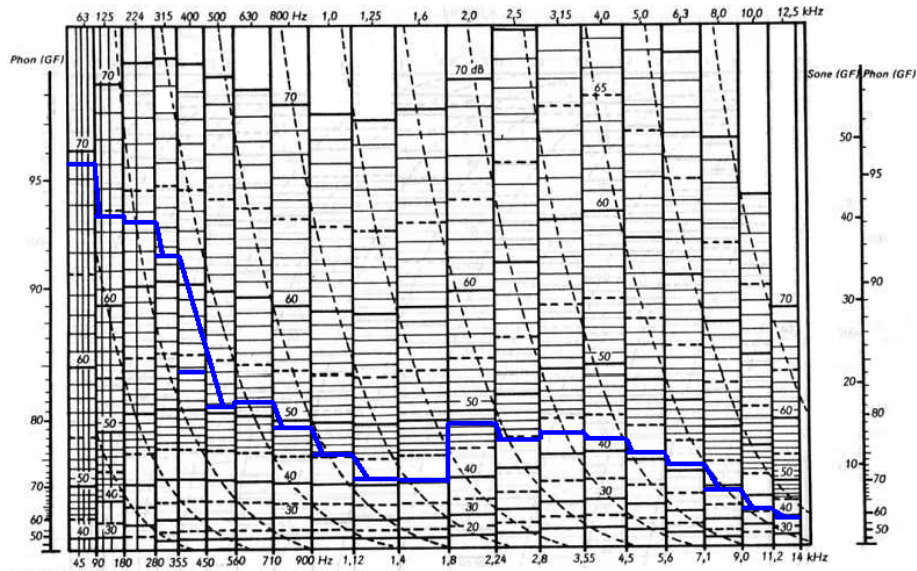


Figura 2.13: Gráfico escogido para condiciones de campo libre, considerando un nivel máximo $L = 69,5 \text{ dB @ } 63 \text{ Hz}$.

centrada en 400 Hz resulta inaudible debido al enmascaramiento producido por la banda precedente.

Paso 3

Debe determinarse manualmente el área bajo la curva de sonoridad específica determinando la altura de un rectángulo de área equivalente cuya base es del ancho de la gráfica escogida. Para esta etapa se debe ignorar la escala en Sones y Fones rotulada en el extremo derecho, ya que ellas no poseen una escala lineal y serán utilizadas sólo en el último momento para la lectura de la sonoridad S y/o el nivel de sonoridad P .

Como se puede observar, la principal desventaja de este método es la naturaleza gráfica del método, dificultando su utilización. Pese a esto, en 1991 [41] se publicó un algoritmo basado en el lenguaje BASIC que permite implementar este estándar en una rutina computacional.

El **Método B**, recién descrito, será comparado con un método que se construirá a partir no sólo de los conceptos psicoacústicos revisados, sino también

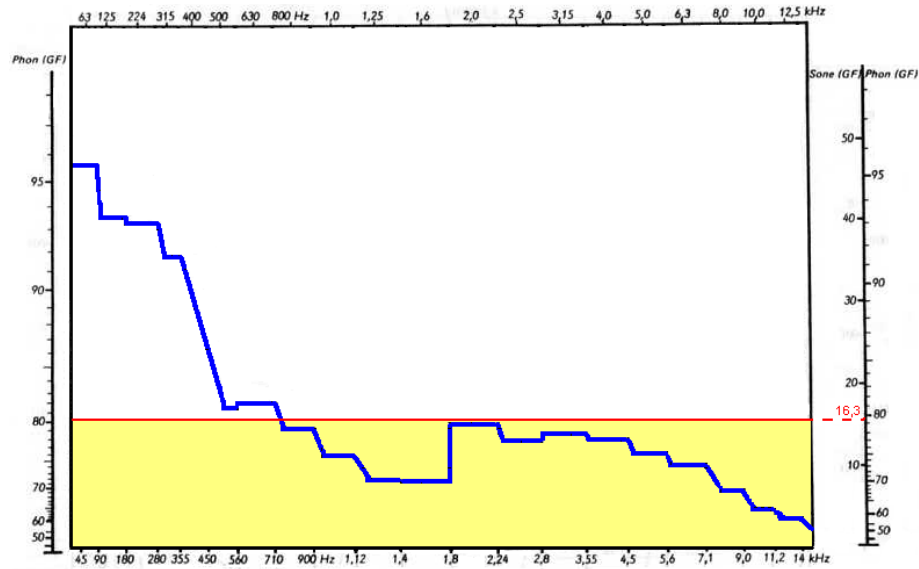


Figura 2.14: Se muestra el rectángulo de área equivalente (región coloreada) al área bajo la curva de sonoridad específica (área bajo curva azul), su altura leída en Sones entrega la sonoridad total S . En este caso, $S = 16,3$ Sones (GF).

de algunas convenientes técnicas de procesamiento digital de señales y diseño de filtros.

2.3. Procesamiento digital de señales

Los principales tópicos que serán revisados en esta sección son: el diseño de filtros digitales, la aplicación de un sistema digital que permite el pre-*warping* de frecuencias y algunos conceptos relacionados con las Redes Neuronales Artificiales (RNA). Todos estos elementos serán utilizados para la construcción de un modelo de sonoridad en bandas críticas que será comparado con el descrito **Método B** (ISO 532-B).

2.3.1. Filtros

Por definición, un filtro tiene la capacidad de alterar la ganancia de una señal de audio actuando sobre una o más partes dentro de su rango de frecuencias. Entre las utilidades de los filtros se encuentra la restauración de señales y la separación de componentes de frecuencia presentes en una misma señal [34]. En

la presente investigación se hace utilización de filtros digitales para concretar la separación de componentes de frecuencia permitiendo el análisis de la distribución energética de una señal de audio digital.

La estructura básica que permite el diseño de un filtro son los denominados filtros pasa-bajos (LPF), los cuales atenúan las componentes de frecuencia que se encuentran sobre una frecuencia de corte preestablecida. A partir de un LPF es posible obtener un filtro pasa-altos (HPF) y, a través de la disposición en cascada de un LPF y un HPF es posible obtener un filtro pasa-banda (BPF), tal como se muestra en la Figura 2.15. Los bancos de filtros se componen de una serie de BPF conectados en paralelo.

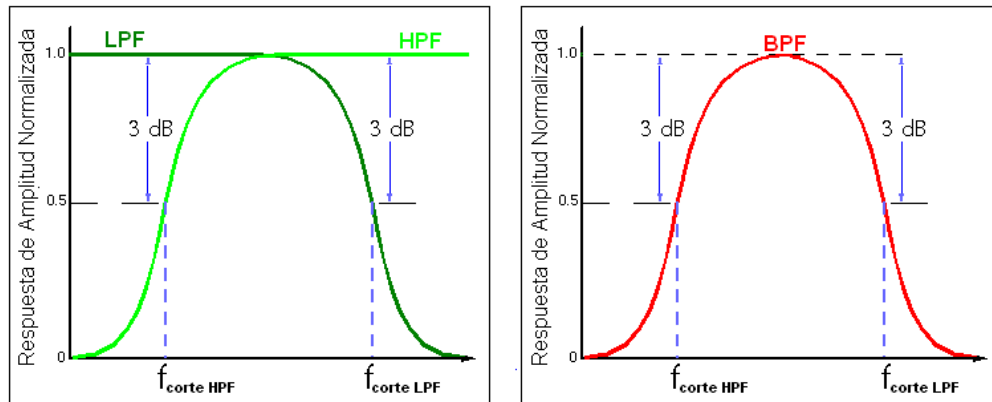


Figura 2.15: (a) Filtros pasa bajos (LPF) y pasa altos (HPF); (b) Filtro pasa banda (BPF) formado de la disposición en cadena de un HPF y un LPF, donde $f_{\text{corte HPF}} < f_{\text{corte LPF}}$.

Filtros Analógicos

Son sistemas que trabajan en un dominio temporal continuo, actuando sobre una señal de entrada $x(t)$ y entregando la señal modificada $y(t)$. Los filtros analógicos son utilizados principalmente en la electrónica. Pese a lo anterior, el procesamiento de señales en tiempo continuo es un campo paralelo al procesamiento digital de señales (DSP) en donde cada una de las operaciones y técnicas consideradas son muy similares, son análogas entre sí [34]. Esto permite hacer

extensible mucha de la teoría utilizada en los filtros analógicos a los filtros digitales.

El proceso introducido por un filtro analógico está caracterizado por su función de transferencia:

$$H(s) = \frac{b_n s^n + b_{n-1} s^{n-1} + \dots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0} \quad (2.31)$$

Para concretar un sistema analógico se utilizan componentes eléctricos y/o electrónicos. Un filtro analógico se presenta en la Figura 2.16. Este sistema es el conocido filtro RC, el cual debe su nombre a la resistencia R y al condensador C que son parte del sistema.

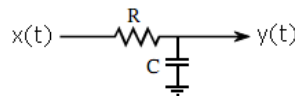


Figura 2.16: Circuito Analógico RC, compuesto por una resistencia de R Ohms y un condensador de C Faradios.

La función de transferencia para el sistema de la Figura 2.16 está dada por:

$$H(s) = \frac{1}{1 + RCs} \quad (2.32)$$

Nuestro interés por los filtros analógicos en el marco de la presente investigación, radica en la posibilidad de obtener un filtro digital equivalente al aplicar una transformación bilineal sobre la Ecuación 2.32 y sobre cualquier otra función de transferencia $H(s)$.

Filtros Digitales

Un filtro digital es un sistema discreto lineal e invariante en el tiempo (LTI) que permite concretar el filtrado de una señal $x[n]$, entregando una señal alterada representada comúnmente por $y[n]$. El proceso introducido por el filtro es representado por la ecuación en diferencias lineales:

$$y[n] + a_1y[n-1] + \dots + a_Ny[n-N] = b_0x[n] + b_1x[n-1] + \dots + b_Nx[n-N] \quad (2.33)$$

Donde $x[n]$ e $y[n]$ representan las muestras actuales de entrada y salida del filtro digital, mientras que los valores $x[n-k]$ e $y[n-k]$ denotan los valores anteriores adquiridos por ambas series y que están k muestras antes de la muestra n . Los coeficientes numéricos a_k y b_k son los encargados de caracterizar la modificación espectral introducida por el filtro digital. La función de transferencia asociada también está caracterizada por los coeficientes a_k y b_k y está dada por la Ecuación 2.34.

$$H(z^{-1}) = \frac{b_0 + b_1z^{-1} + \dots + b_Nz^{-N}}{1 + a_1z^{-1} + \dots + a_Nz^{-N}} \quad (2.34)$$

Si todos los coeficientes a_k son nulos, se dice que el filtro digital posee una respuesta impulsiva finita (FIR) debido a que la salida $y[n]$ dependería sólo de la señal de entrada $x[n]$. En el caso contrario, es decir, si algún valor de a_k es no nulo, se habla de un filtro de respuesta impulsiva infinita (IIR), pues la salida $y[n]$ depende no sólo de las muestras actuales y pasadas de $x[n]$ sino también de las salidas anteriores $y[n-k]$, tal como se muestra en la Figura 2.17. Es posible implementar un algoritmo computacional que realice el cálculo recursivo de la salida $y[n]$.

En el presente trabajo se comenzará trabajando con filtros digitales del tipo IIR y, posteriormente, se evaluará la posibilidad de implementar filtros FIR de similares características. En general, los filtros IIR poseen una implementación más compleja, pero más eficaz si se evalúa la relación costo-beneficio entre el orden del filtro conseguido y el rechazo en dB ofrecido por el mismo. De todas formas, una revisión detallada de filtros digitales FIR e IIR puede ser encontrada en los textos [25] [23].

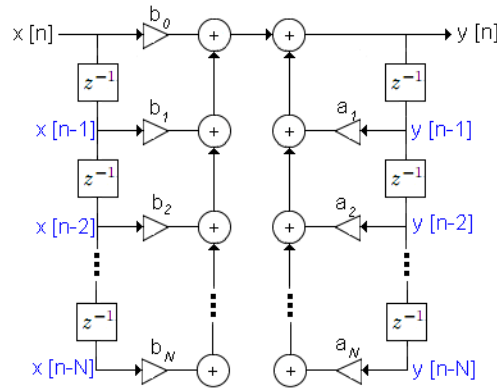


Figura 2.17: Estructura digital que permite la implementación recursiva de los filtros de respuesta impulsiva infinita (IIR). Si los coeficientes a_k fueran nulos, entonces la estructura digital carecería de realimentación y su implementación sería más sencilla. Se hablaría, entonces, de un filtro de respuesta impulsiva finita (FIR).

2.3.2. Diseño de Filtros

El diseño de filtros digitales se efectúa en 3 etapas: (a) especificación de las propiedades deseadas para el sistema, (2) aproximación del sistema usando un sistema digital causal y (c) realización del sistema [25].

Como la estructura base para el diseño de filtros son los filtros pasa-bajos (LPF), a través de ellos se mostrarán las especificaciones propias de un filtro, las que se muestran en la Figura 2.18. Generalmente se parte de un filtro con especificaciones en el dominio analógico de frecuencias, para pasar a un sistema equivalente en el dominio digital.

En primer lugar, en la respuesta en frecuencia de un filtro se pueden distinguir 3 regiones: región de pasa banda, región de transición y región de rechaza banda. En un caso ideal, la región de transición no existe ya que la respuesta de amplitud pasaría de un valor unitario en la región de pasa banda a un rechazo absoluto en la región de rechaza banda. En la práctica, la amplitud en la región de pasa banda tiene una amplitud que oscila en δ_1 en torno a 1 para frecuencias menores

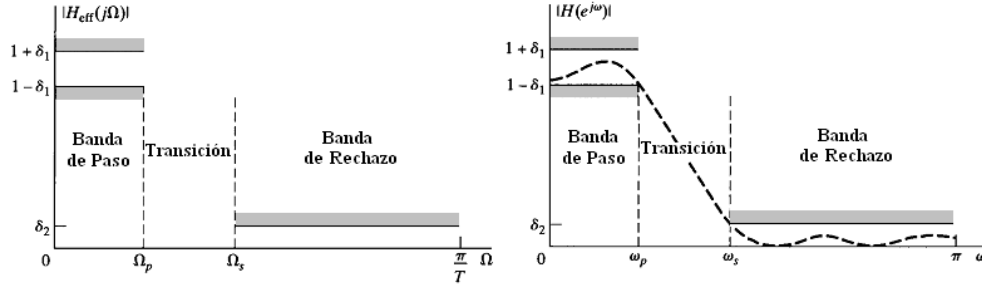


Figura 2.18: (a) Especificaciones para la respuesta en frecuencia analógica de un filtro pasa bajos (LPF) y (b) sus respectivas especificaciones en dominio digital.

a la frecuencia angular analógica Ω_p ⁹ que tiene su equivalente digital ω_p .

$$\Omega = 2 \cdot \pi \cdot f \quad \text{Con } 0 < \Omega < \infty \quad (2.35)$$

$$\omega = 2 \cdot \pi \cdot \frac{f}{f_s} \quad \text{Con } 0 < \omega < \pi \quad (2.36)$$

Donde f es la frecuencia en Hz, f_s corresponde a la frecuencia de muestreo en Hz, Ω se mide en rad/s y ω se mide en rad/muestras.

Para que la frecuencia ω_p corresponda a una frecuencia de corte, debe producirse que la respuesta de amplitud $H(\omega_p)$ en dB debe ser igual a -3 dB. Esto ocurre a un valor específico de δ_1 . Por otro lado, la selectividad del filtro está definida tanto por la frecuencia Ω_s ¹⁰ como por la atenuación mínima en la zona de rechaza banda definida por la amplitud δ_2 .

Filtro pasa-bajos analógicos, del tipo Butterworth

Los filtros pasa bajos Butterworth tienen la propiedad de poseer una respuesta de amplitud que se asemeja a una respuesta plana en la región de pasa banda. La atenuación a partir de una frecuencia angular de corte Ω_c depende del orden N del filtro. El orden N de un filtro Butterworth es un entero positivo que brinda aproximadamente 6 dB de atenuación por octava.

⁹El subíndice p indica que Ω_p es la frecuencia límite de la región de pasa banda

¹⁰El subíndice s indica que Ω_s es la frecuencia en donde comienza la región de rechaza banda (*stopband*).

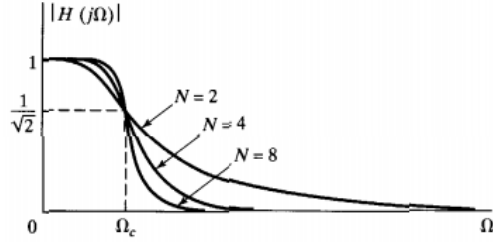


Figura 2.19: Respuesta de amplitud para filtros Butterworth de diferentes órdenes.

El módulo de la respuesta de amplitud para un filtro pasa bajos Butterworth cumple con la Ecuación 2.37.

$$|H(j\Omega)|^2 = \frac{1}{1 + \left(\frac{\Omega}{\Omega_c}\right)^{2N}} \quad (2.37)$$

Para determinar el valor apropiado de N , como función de los parámetros de diseño, es posible deducir la Ecuación 2.38 [25].

$$N = \frac{\log_{10} \left(\frac{\left(\frac{1}{\delta_2}\right)^2 - 1}{\left(\frac{1}{1-\delta_1}\right)^2 - 1} \right)}{2 \cdot \log_{10} (\Omega_s/\Omega_p)} \quad (2.38)$$

Para simplificar la etapa de diseño, se pueden hacer coincidir las frecuencias de pasa banda y de corte ($\Omega_p = \Omega_c$). Esta consideración hace que $(1 - \delta_1) = 1/\sqrt{2}$, por lo que la Ecuación 2.38 se reduce a:

$$N = \frac{\log_{10} \left(\left(\frac{1}{\delta_2}\right)^2 - 1 \right)}{2 \cdot \log_{10} (\Omega_s/\Omega_p)} \quad (2.39)$$

Donde Ω_p y Ω_s corresponden a las frecuencias angulares de corte e inicio de la zona de rechaza banda¹¹, respectivamente y el parámetro δ_2 determina la atenuación mínima del filtro en la zona de rechaza banda (para $\Omega \geq \Omega_s$).

Teniendo la frecuencia angular de corte Ω_c y el orden N del filtro Butterworth se conoce la respuesta de amplitud del filtro pasa bajos analógico. Pese a ello, debemos obtener la función de transferencia $H(s)$ y no sólo su magnitud. La

¹¹Debe cumplirse que $\Omega_s > \Omega_p$.

deducción de $H(s)$ se puede diferir diseñando un filtro pasa bajos con frecuencia angular $\Omega_c = 1$ rad/s [32]. De esta forma, la Ecuación 2.37 queda expresada simplificada como:

$$|H(j\Omega)|^2 = \frac{1}{1 + \Omega^{2N}} \quad (2.40)$$

En esta expresión, se debe sustituir la variable Ω por (s/j) o, lo que es equivalente, $\Omega^2 = -s^2$:

$$|H(j\Omega)|^2 = \frac{1}{1 + \Omega^{2N}} \quad (2.41)$$

$$|H(s)|^2 = \frac{1}{1 + (-1)^N \cdot s^{2N}} \quad (2.42)$$

Igualando a 0 el denominador de la Ecuación 2.42, es posible encontrar los $2N$ polos asociados a la función de transferencia $H(s)$, de los cuales sólo deben considerarse los polos ubicados en la mitad izquierda del plano complejo s [32]. Ellos están definidos por la Ecuación 2.43.

$$p_k = \exp \left[j \left(\frac{2k + N - 1}{2N} \right) \pi \right] \quad k = 1, 2, \dots, N \quad (2.43)$$

La función de transferencia analógica del LPF Butterworth de orden N y frecuencia angular de corte 1 rad/s queda definida por la Ecuación 2.44.

$$H(s) = \frac{1}{\prod_{k=1}^N (s - p_k)} \quad (2.44)$$

Finalmente, se debe efectuar una conversión al filtro pasa bajos obtenido, de manera de que posea la frecuencia de corte deseada. Ello se hace reemplazando la variable s por (s/Ω_c) en la Ecuación 2.44. La estructura de la nueva función de transferencia seguirá siendo la misma, sólo se obtendrán valores numéricos diferentes.

Teniendo $H(s)$, sólo nos queda convertirla a una función de transferencia discreta $H(z^{-1})$, para lo que se utilizará la técnica de la transformación bilineal.

Transformación bilineal: del dominio analógico al dominio digital

La transformación bilineal es una técnica interesante que se utiliza en esta investigación desde diferentes puntos de vista¹². Como se mencionó en el apartado anterior, para el diseño de un filtro Butterworth con frecuencia angular de corte Ω_c [rad/s] es necesario encontrar un orden N apropiado para cumplir con ciertas especificaciones. La idea es iniciar un diseño de filtros considerando parámetros de frecuencia digital para obtener un filtro digital, pasando por una técnica de diseño en dominio analógico. Nuestro problema puede resolverse aplicando una transformada bilineal a la función de transferencia analógica $H(s)$ en donde la variable independiente s debe reemplazarse por la expresión de la Ecuación 2.45 [25].

$$s = \frac{2}{T} \cdot \left(\frac{1 - z^{-1}}{1 + z^{-1}} \right) \quad (2.45)$$

Donde T es una constante que, en general, consideraremos igual a 1.

La función de transferencia $H(z^{-1})$ resultante corresponde a un filtro digital de respuesta impulsiva infinita (IIR), por lo que la transformación bilineal de la Ecuación 2.45 corresponde a la principal herramienta de diseño de este tipo de filtros.

Al aplicar la transformación de la Ecuación 2.45, se obtiene que la relación entre la frecuencia angular analógica Ω y la frecuencia angular digital ω está dada por la Ecuación:

$$\Omega = 2 \cdot \tan(\omega/2) \quad (2.46)$$

Por lo que:

$$\Omega_p = 2 \cdot \tan(\omega_p/2) \quad (2.47)$$

$$\Omega_s = 2 \cdot \tan(\omega_s/2) \quad (2.48)$$

¹²Ver el Anexo A.

Al reemplazar el cociente Ω_s/Ω_p en la Ecuación 2.39, se obtiene la expresión que permite el cálculo del orden N requerido para un LPF Butterworth a partir de las frecuencias digitales ω_p y ω_s :

$$N = \frac{\log_{10} \left(\left(\frac{1}{\delta_2} \right)^2 - 1 \right)}{2 \log_{10} (\tan(\omega_s/2) / \tan(\omega_p/2))} \quad (2.49)$$

El parámetro δ_2 determina la atenuación mínima en la zona de rechaza banda. En general, la atenuación requerida en un filtro es entregada en dB, por lo que resultaría más práctico convertir δ_2 a una amplitud logarítmica.

$$A = 20 \log_{10}(\delta_2) \quad (2.50)$$

Donde A representa la atenuación mínima en la zona de rechaza banda, en dB. Si evaluamos la Expresión 2.50 en la ecuación deducida para N :

$$N = \frac{\log_{10} (10^{(A/10)} - 1)}{2 \log_{10} (\tan(\omega_s/2) / \tan(\omega_p/2))} \quad (2.51)$$

Recordemos que una atenuación de A dB es equivalente a aplicar una ganancia de $-A$ dB, por lo que la Ecuación B.11 recibe como parámetro un valor $A > 0$ dB.

Filtros pasa-altos Butterworth

Entre otros filtros, los pasa altos (HPF) y pasa banda (BPF) pueden ser diseñados a partir de un filtro pasa bajos (LPF) al que se aplica una conveniente transformación. En el caso de los HPF, existen dos técnicas de diseño a partir de un LPF: Inversión espectral y reversión espectral [34].

La inversión espectral se logra implementando el sistema digital de la Figura 2.20.

La técnica de reversión espectral intercambia las muestras de la respuesta en frecuencia de un LPF, mediante la modificación de su respuesta al impulso $h[n]$.

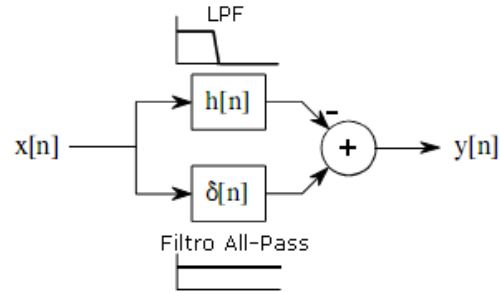


Figura 2.20: Sistema digital temporal que produce un HPF sobre la base de un LPF.

El efecto producido por la técnica de reversión espectral es mostrado en la Figura 2.21.

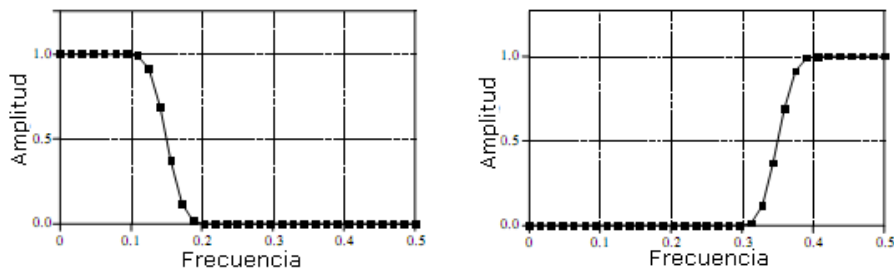


Figura 2.21: Respuesta en frecuencia de un LPF (a) original y (b) con reversión espectral (HPF).

Una manera de concretar una reversión espectral es aplicar la transformación bilineal presentada en la Ecuación 2.52 (y demostrada en el Anexo A) al diseño de un LPF.

$$z^{-1} \rightarrow -z^{-1} \quad (2.52)$$

La ventaja de usar la técnica de reversión espectral es que se puede conocer la atenuación exacta existente entre 2 frecuencias cualquiera pertenecientes al nuevo filtro pasa altos. Basta con saber que las amplitudes asociadas a las frecuencias ω_1 y ω_2 del LPF original son equivalentes a las amplitudes de las frecuencias $(\pi - \omega_1)$ y $(\pi - \omega_2)$ del nuevo HPF.

Con el antecedente anterior, para el diseño de un HPF Butterworth dejaremos de lado la técnica de inversión espectral para usar la de reversión espectral. Al usar la reversión espectral, el diseño de un HPF¹³ Butterworth se limita a diseñar un LPF Butterworth que posea una frecuencia digital de paso igual a $(\pi - \omega_p)$ y una frecuencia digital de rechazo igual a $(\pi - \omega_s)$ y luego aplicar la transformación de la Ecuación 2.52.

Filtros en secciones de 2^o orden

Las funciones de transferencia $H(z^{-1})$ obtenidas en el proceso de diseño usando la transformación bilineal corresponderán a filtros IIR de orden N , es decir, tanto su sección directa (numerador de H) como su sección realimentada (denominador H) estarán caracterizadas por la misma cantidad (N) de unidades de retardo. La Ecuación 2.53 expresa la función $H(z^{-1})$ como el producto de N filtros IIR, con ceros y polos en c_k y d_k , respectivamente.

$$H(z^{-1}) = \frac{b_0 (1 - c_1 z^{-1})(1 - c_2 z^{-1}) \cdots (1 - c_N z^{-1})}{a_0 (1 - d_1 z^{-1})(1 - d_2 z^{-1}) \cdots (1 - d_N z^{-1})} \quad (2.53)$$

La Ecuación 2.53 puede ser expresada como una productoria de polinomios de segundo orden, como se muestra en la Ecuación 2.54.

$$H(z^{-1}) = \prod_{k=1}^{N/2} \frac{(1 - c_{2k-1} z^{-1})(1 - c_{2k} z^{-1})}{(1 - d_{2k-1} z^{-1})(1 - d_{2k} z^{-1})} \quad (2.54)$$

Lo que, al desarrollar, puede ser representado como:

$$H(z^{-1}) = \prod_{k=1}^{N/2} H_k(z^{-1}) \quad (2.55)$$

Donde $H_k(z^{-1})$ es una función de la forma definida por la Ecuación 2.56 y el subíndice k está entre 0 y $N/2$.

$$H_k(z^{-1}) = \frac{b_{0k} + b_{1k} z^{-1} + b_{2k} z^{-2}}{1 + a_{1k} z^{-1} + a_{2k} z^{-2}} \quad (2.56)$$

¹³Esta técnica se aplica para el diseño de cualquier HPF a partir de un LPF. Aquí se particulariza a los filtros Butterworth por ser la variedad de filtros escogida para el presente trabajo.

La respuesta total del sistema $H(z^{-1})$ de orden N expresada como una multiplicación de $K = N/2$ sistemas $H_k(z^{-1})$ de orden 2 equivale a la implementación en cascada de los filtros H_k , tal como se esquematiza en la Figura 2.22.

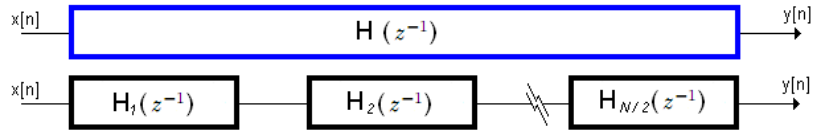


Figura 2.22: Esquema de implementación de un sistema caracterizado por la función de transferencia $H(z^{-1})$ como (arriba) un sistema de orden N y, (abajo) como $N/2$ sistemas de orden 2. La implementación en cascadas de filtros de orden 2 es más robusta ante la aparición de errores de redondeo.

La implementación de un sistema utilizando filtros de 2^o orden es común cuando se tienen sistemas de orden elevado. Con esto se evita la aparición de errores de redondeo que se manifiestan en la señal de salida $y[n]$. El sistema digital resultante continúa estando caracterizado por los mismos ceros c_k y polos d_k de la Ecuación 2.53.

2.3.3. La técnica de *Warping* de frecuencias

La implementación de la técnica de *Warping* de frecuencias, permite efectuar un cambio en la resolución de frecuencias de una señal discreta de audio sin alterar su amplitud (ganancia unitaria).

Teóricamente, se aplica una transformación bilineal¹⁴, dependiente de un parámetro λ . Este sistema digital está definido por la Ecuación 2.57 y su diagrama en bloques es presentado en la Figura 2.23 [13].

$$D(z^{-1}) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}} \quad (2.57)$$

La función de transferencia $D(z^{-1})$ posee una amplitud unitaria¹⁵, sin embargo, debido a las características de fase del sistema, al implementar la técnica

¹⁴Ver Apéndice A

¹⁵El sistema $D(z^{-1})$ no enfatiza ni atenúa ningún rango de frecuencias, por lo que se trata de un sistema *AllPass* (“pasa todo”).

de *warping* se introduce una nueva resolución de frecuencias ν , en rad/muestras, cuya relación con la frecuencia digital original ω está dada por la Ecuación 2.58¹⁶.

$$v(\omega) = \omega + 2 \cdot \arctan \left(\frac{\lambda \sin(\omega)}{1 - \lambda \cos(\omega)} \right) \quad (2.58)$$

La constante arbitraria λ , entonces, es el parámetro encargado de relacionar las frecuencias ω y ν , motivo por el cual también se le conoce como factor de *warping* λ o factor de *AllPass*.

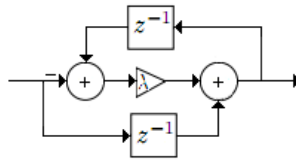


Figura 2.23: Sistema digital con Función de Transferencia $D(z^{-1})$.

En la Figura 2.24 se muestra la relación entre la frecuencia en dominio lineal, ω , y en dominio *warp*, ν . Tanto la frecuencia angular ω como la frecuencia distorsionada ν toman valores entre 0 y π .

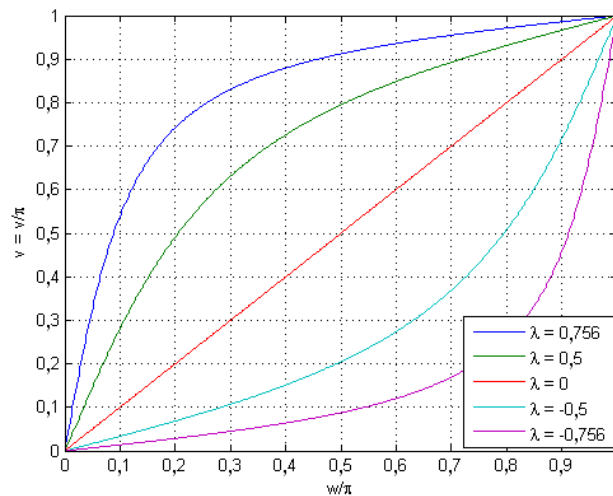


Figura 2.24: Warping de frecuencias para diferentes valores de λ , normalizadas (ω/π , ν/π).

¹⁶Expresión deducida a partir de la función de fase de la estructura *AllPass* de la Ecuación 2.57

La elección del factor de *warping* λ entrega un grado de libertad al mapeamiento entregado por la transformación bilineal, por lo que éste debe ser escogido convenientemente dentro del rango de valores que va entre -1 y 1. En 1999, Smith y Abel [33] desarrollaron una expresión que determina el factor λ , logrando un ajuste muy próximo entre la escala de frecuencia en dominio *warp*, ν , y la escala en bandas críticas, z .

$$\lambda(f_s) = 1,0674 \left[\frac{2}{\pi} \arctan(0,06583 f_s) \right]^{\frac{1}{2}} - 0,1916 \quad (2.59)$$

Si se considera una frecuencia de muestreo $f_s = 44.100$ Hz, el factor de *warping* óptimo λ es igual a 0,7564. En la Figura 2.25 se puede apreciar la gran similitud entre $z = z(\omega)$, según la Ecuación 2.10, y $\nu = \nu(\omega)$.

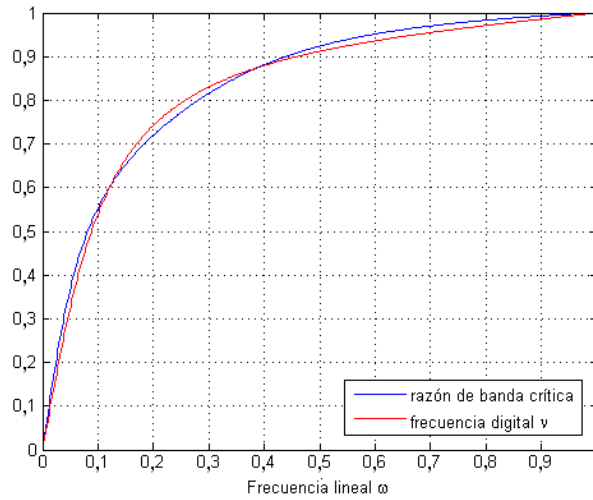


Figura 2.25: Razón de banda crítica z y frecuencia digital ν como función de la frecuencia ω . *Warping* de frecuencias utilizando el factor λ apropiado. Las frecuencias están normalizadas.

Para la implementación en tiempo real de la estructura digital representada por la función de transferencia de la Ecuación 2.57, se debe disponer una cadena de filtros *AllPass* en cascada, tal como se muestra en la Figura 2.26. La lectura de los valores entregados por este sistema producen una distorsión en el eje de frecuencias manifestada al determinar la transformada de Fourier de la salida $w[n]$ del sistema.

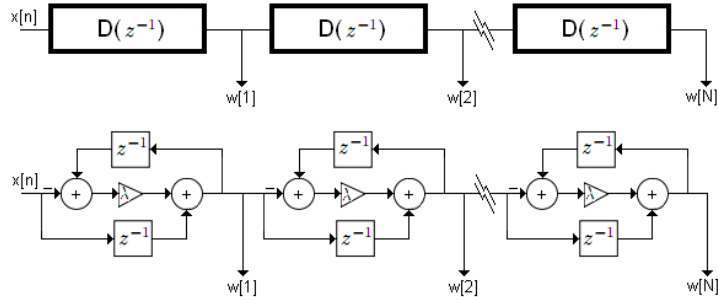


Figura 2.26: Cadena de filtros *AllPass*. Hace posible aplicar la técnica de *warping* a una señal de audio cualquiera.

La cadena de filtros *AllPass* mostrada en la Figura 2.26, sin embargo, no representa una estructura con respuesta en frecuencia plana, debido al efecto producido por el retardo de grupo asociado a $D(z^{-1})$. Se hace necesario, entonces, la inclusión de una etapa de pre-énfasis $A(z^{-1})$ de la señal temporal $x[n]$ para compensar este efecto [24].

$$A(z^{-1}) = \frac{\sqrt{1 - \lambda^2}}{1 - \lambda z^{-1}} \quad (2.60)$$

El pre-énfasis $A(z^{-1})$ también es dependiente del factor de *warping* λ .

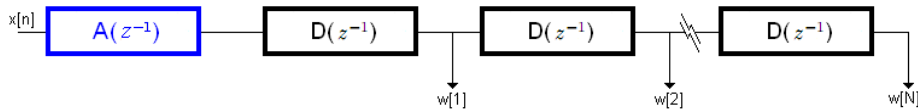


Figura 2.27: Cadena de filtros que implementa el *warping* de frecuencias a una señal $x[n]$. Esta estructura *AllPass* es conocida como filtro de Laguerre. La etapa de pre-énfasis está caracterizada por la función de transferencia $A(z^{-1})$ de la Ecuación 2.60

2.3.4. Redes Neuronales Artificiales

Una Red Neuronal Artificial (RNA) es una estructura interconectada de unidades simples de procesamiento (Neuronas Artificiales) que realiza el aprendizaje, memorización, generalización o abstracción de características relevantes que relacionan los parámetros de entrada con los de salida de la red [39] mediante una

cantidad finita de operaciones matemáticas, principalmente sumas y multiplicaciones.

La capacidad de procesamiento de una RNA está almacenada en valores numéricos denominados **pesos sinápticos**, los que manifiestan la fuerza de conexión entre procesadores elementales (neuronas artificiales) [12]. El ajuste de estos valores se efectúa durante el proceso de aprendizaje de la red a partir de un set de ejemplos de entrenamiento.

Las RNA no poseen un campo de acción definido, sin embargo, en general, son utilizadas para modelar problemas que presentan un alto grado de no linealidad y que incluyen 1 o más factores de incidencia (variables dependientes). Principalmente se utilizan en problemas de carácter físico-matemático.

Para la utilización de una RNA se debe considerar (a) el diseño estructural o topología a utilizar (ver Figura 2.28), (b) el procedimiento de almacenamiento de la información: elección del algoritmo de entrenamiento y, (c) el procedimiento de recuperación de la información: como aproximador funcional o como clasificador.

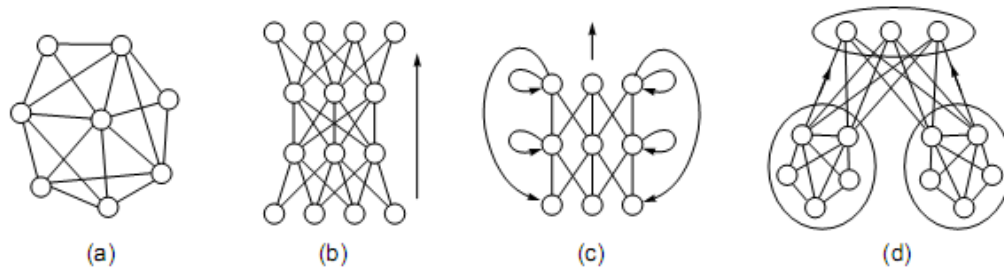


Figura 2.28: Topologías de redes [36]: (a) No estructuradas, (b) en capas, (c) recurrentes, (d) modulares.

Esta metodología ofrece solución a problemas científicos poco estudiados o que presentan un alto grado de no linealidad.

La Neurona como Procesador Elemental

Todo el procesamiento y cálculo numérico en una RNA se realiza en cada una de las neuronas artificiales o procesadores elementales (PE) que la componen.

A la entrada de una neurona, se realiza siempre un proceso de agregación de n señales de entrada, de manera de obtener un nivel de activación numérico net . Este proceso tiene asociado lo que se denomina **función de agregación** (Tabla 2.8), siendo la función de suma ponderada la más utilizada.

Función de Transferencia	$net = f(\omega_{ik}, x_i)$
Sumatoria	$\sum_{i=1}^n \omega_{ki} x_i$
Productoria	$\prod_{i=1}^n \omega_{ki} x_i$
Distancia Euclidiana	$\ \mathbf{x} - \mathbf{c}_k\ $

Tabla 2.8: Funciones de agregación a una neurona artificial k .

Posteriormente, la salida de la neurona se determina a partir del nivel de activación net y una función de transferencia de activación $Y = f(net)$. Las funciones de activación más usadas son: Lineal, Sigmoidal, Tangente Hiperbólica y Radial base (ver Tabla 2.9).

La función de activación más simplificada corresponde a la función escalón o umbral, definida como:

$$Y = \begin{cases} 1 & \text{Si } net \geq \theta \\ 0 & \text{Si } net < \theta \end{cases} \quad (2.61)$$

Esta función sólo discrimina si la entrada net supera o no supera un umbral numérico θ . No obstante la simplicidad de la función umbral, es la que más se aproxima a la capacidad de procesamiento de una neurona biológica, que –al ser excitada– emite o –al ser inhibida– no emite un impulso nervioso.

Función de Transferencia	$Y = f(net)$
Escalón	$\begin{cases} 1 & \text{Si } net \geq \theta \\ 0 & \text{Si } net < \theta \end{cases}$
Signo	$\begin{cases} +1 & \text{Si } net \geq 0 \\ -1 & \text{Si } net < 0 \end{cases}$
Lineal	net
Sigmoidal	$1/(1 + \exp(-net))$
Tangente Hiperbólica	$2/(1 + \exp(-2 \cdot net)) - 1$
Radial Base	$\exp(-net^2)$

Tabla 2.9: Algunas funciones de activación, también llamadas funciones de transferencia.

Conexiones

Las conexiones entre neuronas son generalmente dirigidas y a cada una de ellas se le corresponde un peso sináptico ω . Si la RNA está organizada en capas, se pueden diferenciar las conexiones hacia adelante (*Feed-forward*), las de retroalimentación (*feedback*) y también configuraciones con conexiones intercapas, es decir, con conexiones a neuronas de otras capas.

Teniendo en cuenta estos conceptos, a continuación se pasará a revisar las 2 topologías utilizadas durante esta investigación. Ambas responden al modelo en capas con propagación hacia adelante mostrado en la Figura 2.28(b).

Perceptrón Multicapa, MLP

El perceptrón multicapa (MLP¹⁷) es la topología de redes neuronales artificiales más utilizada. Comúnmente es referida como red de propagación hacia

¹⁷Del inglés: *Multi-Layer Perceptron*

adelante (red *Feed-Forward*), sin embargo, esta definición es genérica y extensible para cualquier RNA que tenga más de 1 capa y cuyo flujo de la información es hacia adelante. El MLP consta de 3 o más capas¹⁸.

La función de agregación corresponde, generalmente, a la sumatoria de las entradas a la neurona ponderada con los pesos sinápticos de las conexiones involucradas (Función Sumatoria, Tabla 2.8). Las funciones de activación más populares son la función Sigmoidal, Tangente Hiperbólica y Lineal (ver Tabla 2.9).

Los perceptrones multi-capa corresponden a la topología de redes neuronales artificiales más abordadas en la bibliografía. Por este motivo, existen innumerables ejemplos de este tipo de RNA que permiten comprender el funcionamiento de los MLP.

Por otro lado, las Redes de Funciones Radial Base (RBFN) representan un pilar fundamental en la presente investigación, motivo por el cual, se abordarán a continuación y en mayor profundidad que los MLP.

Redes de Funciones Radial Base

Una Red de Funciones Radial Base (RBFN) es una configuración específica de las redes de propagación hacia adelante (Redes *Feed-Forward*), que posee 3 capas. La capa de entrada no procesa información (sólo la entrega), la capa oculta posee una función de transferencia del tipo radial y la capa de salida posee una neurona con una función de transferencia lineal. La topología de una RBFN con n parámetros de entrada y 1 salida es mostrada en la Figura 2.29.

Cada RBFN puede interpretarse como un modelo f , que se obtiene de la combinación lineal de un set de m funciones del tipo radial. Por analogía con la representación de un vector \vec{f} como combinación lineal de vectores base \vec{h}_j , se dice que f es una función base. Donde f está dada por la Ecuación 2.62.

¹⁸En esta investigación contabilizamos la capa de entrada dentro del número total de capas de la red, es decir, una red de 3 capas posee 1 capa de entrada, 1 oculta y 1 de salida.

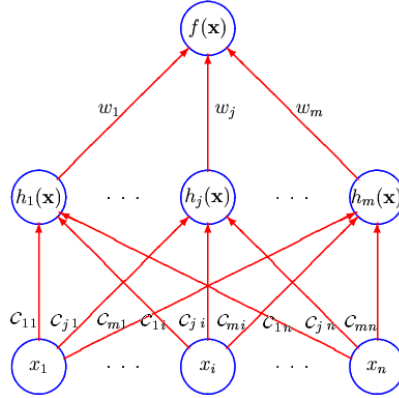


Figura 2.29: Red de funciones Radial Base, RBFN, con n parámetros de entrada, m neuronas en la capa oculta y 1 salida. Imagen extraída de [26]

$$f(\vec{x}) = \sum_{j=1}^m \omega_j \cdot h_j(\vec{x}) \quad (2.62)$$

Donde el valor ω_j representa el peso sináptico asociado a la conexión entre la neurona j de la capa oculta y la única neurona de salida. Además, la función h_j corresponde a una función del tipo radial asociada a la neurona j de la capa oculta.

Las funciones radiales son funciones que crecen o decrecen monótonamente a partir de un punto central. Los parámetros que describen una función radial son su centro, su escala de distancia y su forma exacta [26].

La función de transferencia radial $h_j(\text{net}_j)$ más utilizada es la función Gaussiana y está dada por la Ecuación 2.63.

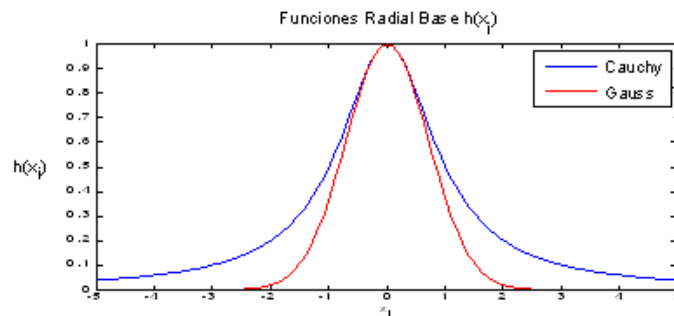
$$h_j(\text{net}_j) = \exp(-\text{net}_j^2) \quad (2.63)$$

Otra función de activación radial es la función de Cauchy, presentada en la Ecuación 2.64. En la Figura 2.30 se muestran las funciones de activación Gaussiana y de Cauchy, suponiendo una entrada unidimensional.

Centro \vec{c}_j	$\vec{c}_j = (c_{j1}, c_{j2}, \dots, c_{jn})$	Define la coordenada espacial del centro de la función radial
Radio r_j	r_j	Concepto similar al de ancho de banda, pero asociado a la función radial. A mayor $1/r_j$, se tendrá una función radial más ancha
Función de Agregación a la Neurona j	$\text{net}_j = \frac{\ \vec{x} - \vec{c}_j\ ^2}{r_j}$	Se calcula la distancia euclidiana en el espacio entre los vectores x y c_j . A menor distancia, mayor nivel de activación
Función de Agregación a la Neurona de salida	$\text{net} = \sum_{j=1}^m h_j(\text{net}_j) \cdot \omega_j$	Producto punto entre las funciones de activación radial h_j y los pesos sinápticos ω_j

Tabla 2.10: Parámetros asociados a una red de funciones radial base (RBFN).

$$h_j(\text{net}_j) = \frac{1}{1 + \text{net}_j^2} \quad (2.64)$$

Figura 2.30: Funciones radial base de Cauchy y Gauss, para el caso unidimensional (1 sola entrada) con centros en $c = 0$ y $r = 1$.

Ejemplo de Implementación de una RBFN

Se implementará una Red Neuronal Artificial de Funciones Radial Base para una entrada unidimensional con el fin de comprender el flujo de la información en la red y su relación con los parámetros de la misma. La red escogida es mostrada

en la Figura 2.31. Se considerará que las funciones radiales $h_j(x)$ ¹⁹ son del tipo Gaussianas (Ecuación 2.63).

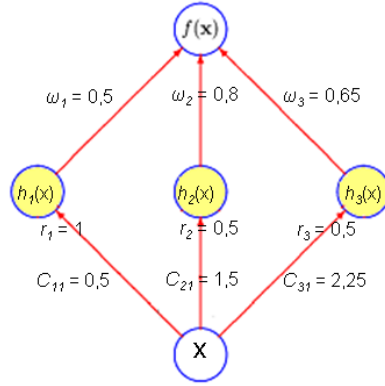


Figura 2.31: Ejemplo de una red de funciones radial base (RBFN), con 1 parámetro de entrada y $m = 3$ neuronas en la capa oculta. Las funciones radiales $h_j(x)$.

Tal como es definido en la Figura 2.31, las funciones radiales h_1 , h_2 y h_3 están centradas en $x = 0,5$, $x = 1,5$ y $x = 2,25$ respectivamente. Si se evalúan los centros c_j , además de los valores escogidos para los radios r_j de cada función radial en la Ecuación 2.63, considerando la agregación net_j según la Tabla 2.10, se obtiene:

$$\begin{aligned}
 h_1(x) &= \exp\left(-\frac{(x-c_{11})^2}{r_1^2}\right) = \exp\left(-\frac{(x-0,50)^2}{1^2}\right) = \exp[-(x-0,50)^2] \\
 h_2(x) &= \exp\left(-\frac{(x-c_{21})^2}{r_2^2}\right) = \exp\left(-\frac{(x-1,50)^2}{0,5^2}\right) = \exp[-4(x-1,50)^2] \\
 h_3(x) &= \exp\left(-\frac{(x-c_{31})^2}{r_3^2}\right) = \exp\left(-\frac{(x-2,25)^2}{0,5^2}\right) = \exp[-4(x-2,25)^2]
 \end{aligned}$$

Supongamos que se desea obtener la salida $f(x)$ de la RBFN para una entrada $x = 0,6$. Al reemplazar este valor en las expresiones anteriores se tiene que:

$$\begin{aligned}
 h_1(0,6) &= 0,990 \\
 h_2(0,6) &= 0,039 \\
 h_3(0,6) &= 1,864 \cdot 10^{-5}
 \end{aligned}$$

¹⁹Al tener un vector \vec{x} unidimensional, éste corresponderá a un escalar, por lo que se omitirá la notación vectorial.

Los valores de activación devueltos por las funciones radiales h_j manifiestan que la entrada $x = 0,6$ está mucho más próxima al centro de la primera función radial $c_1 = 0,5$ que a los centros $c_2 = 1,5$ y $c_3 = 2,25$.

La combinación lineal de funciones radiales que determina la salida $f(x)$ de la red está determinada por los pesos sinápticos ω_j :

$$\begin{aligned} f(x) &= \omega_1 h_1(x) + \omega_2 h_2(x) + \omega_3 h_3(x) \\ f(x) &= 0,5 h_1(x) + 0,8 h_2(x) + 0,65 h_3(x) \\ f(x) &= 0,5 \cdot 0,990 + 0,8 \cdot 0,039 + 0,65 \cdot 1,864 \cdot 10^{-5} \\ f(x) &= 0,526 \end{aligned}$$

La combinación lineal definida por los pesos ω_j entrega una mayor ponderación a la función radial h_2 (0,8), luego a la función h_3 (0,65) y finalmente a la función radial h_1 (0,5). El resultado final para una entrada $x = 0,6$ es $f(0,6) = 0,526$.

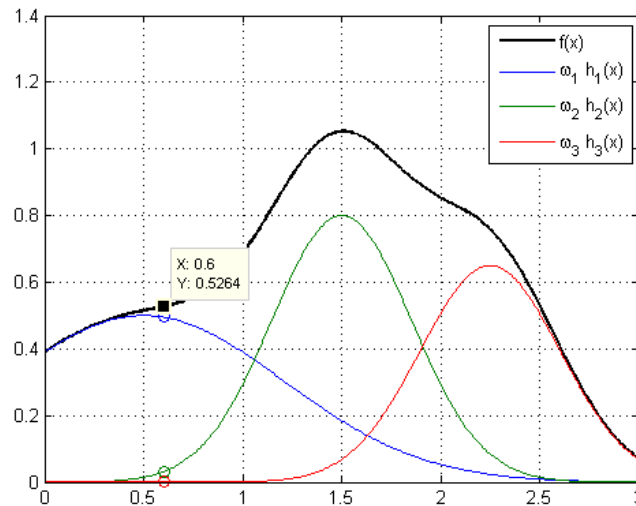


Figura 2.32: Salida $f(x)$ de la RBFN de la Figura 2.31. Se muestran las funciones radiales que generan la salida $f(x)$. La curva en negro representa la salida de la red. Se destaca el valor adquirido cuando $x = 0,6$.

Capítulo 3

DESARROLLO: DISEÑO E IMPLEMENTACIÓN

La elaboración del prototipo que permite realizar un análisis psicoacústico basado en los conceptos de bandas críticas, fue realizada en 3 etapas: construcción del modelo para el análisis espectral, elaboración del algoritmo de procesamiento (en modalidad *off-line*) e implementación en un lenguaje compilado para el procesamiento en tiempo real.

Tanto en la etapa de construcción del modelo para el análisis en bandas críticas como en la etapa primaria de implementación del mismo, se utilizó el *software* de análisis matemático MATLAB®. En estas fases se realizaron una serie de análisis comparativos del modelo en elaboración con respecto a otros modelos de sonoridad: el Método B (ISO 532-B).

En una última etapa, se procedió a adaptar el algoritmo diseñado en MATLAB® para su implementación en tiempo real, como plug-in de audio VST®. Se programó en el lenguaje C++, utilizando el programa Microsoft Visual Studio 2008 sobre la base del SDK 2.4¹ que provee la compañía Steinberg.

A continuación se presenta el diseño del banco de filtros psicoacústico, propuesto a partir de los conceptos definidos en el capítulo anterior.

¹Un SDK es un *Software Development Kit*, lo que corresponde a una serie de rutinas pre-programadas (en este caso por la compañía Steinberg), que permiten el desarrollo de un tipo de aplicación específica (un plug-in VST).

Propuesta de una metodología para el análisis psicoacústico de señales

La elaboración del modelo para el análisis psicoacústico de señales supone 2 problemáticas: la ponderación de amplitud en función de la frecuencia y la ponderación de frecuencias en bandas críticas.

3.1. Ponderación de Amplitud

Se implementarán las curvas de ponderación en frecuencia A, B, C y D, caracterizadas por las funciones de transferencia de las Ecuaciones 2.18 a 2.21. Además, se implementará una ponderación de amplitud en Fones. Para las curvas A, B, C y D, se debe efectuar una transformación bilineal a las funciones de transferencia asociadas a ellas, lo que responde a una serie de filtros de respuesta impulsiva infinita (IIR) con coeficientes conocidos. Estos coeficientes son mostrados en la Tabla 3.1.

Curva	b_0	b_1	b_2	b_3	b_4	b_5	b_6
A	0,2557	-0,5115	-0,2557	1,0230	-0,2557	-0,5115	0,2557
B	0,2173	-0,2173	-0,4345	0,4345	0,2173	-0,2173	-
C	0,2170	$1,9984 \cdot 10^{-15}$	-0,4340	$-3,4694 \cdot 10^{-16}$	0,2170	-	-
D	0,7802	-1,4381	-0,1071	1,4381	-0,6731	-	-
Curva	a_0	a_1	a_2	a_3	a_4	a_5	a_6
A	1,0000	-4,0196	6,1894	-4,4532	1,4208	-0,1418	0,0044
B	1,0000	-3,1123	3,3663	-1,4003	0,1511	-0,0048	-
C	1,0000	-2,1347	1,2793	-0,1496	0,0049	-	-
D	1,0000	-3,2765	4,0919	-2,3215	0,5070	-	-

Tabla 3.1: Coeficientes b_k y a_k de las funciones de transferencia digital $H(z^{-1})$ para las curvas de ponderación en frecuencia A, B, C y D.

La implementación de los filtros IIR con los coeficientes de la Tabla 3.1 se efectuará recursivamente y no constituye un costo computacional considerable debido a que estos filtros poseen un orden bajo (6, 5, 5 y 4 para las curvas A, B, C y D respectivamente).

Para la ponderación de amplitud en Fones, se utilizará la función NeuLoud –propuesta por Espinoza et al. en [8]– la cual entrega el nivel de sonoridad P

en Fones a partir del Nivel de Presión Sonora L_p en dB y la frecuencia f en Hz. Debido a que esta interpolación depende de la frecuencia f , el cálculo de sonoridad debe ser hecho en el dominio de la frecuencia y se efectúa frecuencia por frecuencia.

Como consecuencia de la dependencia de la función NeuLoud respecto del parámetro de entrada L_p (Nivel de Presión Sonora) se requiere que la señal de audio que se está analizando se encuentre calibrada. Para esta calibración se debe aplicar una corrección de Δ dB de manera que el valor RMS de la señal de entrada, medido en dBFS, equivalga a un valor arbitrario de dB (Re $20\mu\text{Pa}$). El criterio utilizado en la presente investigación corresponde a la recomendación AES para el ajuste del nivel de monitoreo en salas de cine [21] en donde, al reproducir un ruido rosa, el nivel de presión sonora debe alcanzar los 83 dB(C) instantáneos. En estas condiciones, el nivel RMS equivalente se encuentra en torno a los 70 dB(Z). Fue, entonces, a este nivel que se ajustó el nivel RMS de cada señal de audio digital. Si la señal de audio presenta un nivel de -22 dBFS, la corrección de nivel Δ dB es de $70 \text{ dB} - (-22 \text{ dBFS}) = +92 \text{ dB}$.

3.1.1. Implementación de las curvas de ponderación en frecuencia

Si bien las curvas de ponderación de amplitud son definidas en el dominio de la frecuencia, su implementación puede realizarse indistintamente en el dominio temporal o en el dominio de la frecuencia. Ambos dominios están relacionados a través del teorema de Parseval que, en el caso de señales de audio digital, consideran su aproximación discreta, la cual fue revisada en la Sección 2.2.10.

La implementación temporal se realiza al implementar la estructura digital mostrada en la Figura 2.17 usando los coeficientes de la Tabla 3.1. La lectura de la salida $y[n]$ no permite la visualización directa de la implementación de una curva de ponderación determinada. Sin embargo, si se determina la Transformada de Fourier a la señal $x[n]$ (sin filtrar, obteniendo $X(\omega)$) y luego se determina la Transformada de Fourier de la señal de salida $y[n]$ (obteniendo $Y(\omega)$) puede

notarse como la amplitud $\|Y(\omega)\|$ tanto en baja como en alta frecuencia es menor que la de $\|X(\omega)\|$, debido a la aplicación del filtro que implementa las curvas de ponderación A, B, C o D.

3.1.2. Ponderación de amplitud en Fones y cálculo energético

La ponderación de amplitud en Fones implica una implementación algo más compleja que la de las curvas A, B, C y D. En esta investigación se utiliza la propuesta de Espinoza et al. [8] para la conversión de cada nivel de presión sonora L_p , en dB, en nivel de sonoridad P , en Fones. Esta conversión se basa en una interpolación de los contornos de igual sonoridad rotulados en la norma ISO 226:2003 [19] usando una Red Neuronal Artificial *Feed-Forward* 2-11-9-1. Para la suma energética, se debe obtener la curva de sonoridad específica, medida en Sones/Bark. Para ello, simplemente se debe obtener el valor de sonoridad en Sones, frecuencia por frecuencia, usando las Ecuaciones 3.1 y 3.2 sobre los obtenidos Niveles de Sonoridad P expresados en Fones.

$$S = 2^{\frac{P-40}{10}} \quad \text{Si } P \geq 40 \quad (3.1)$$

$$S = \left(\frac{P}{40}\right)^{2,642} \quad \text{Si } P < 40 \quad (3.2)$$

Al efectuar la suma de sonoridad, en el dominio de la frecuencia, se debe considerar un eje expresado en unidades compatibles a la razón de banda crítica z , en Barks.

Por otra parte, debe advertirse que los contornos de igual sonoridad considerados para la ponderación de amplitud en Fones se encuentran determinados para una señal de excitación del tipo tono puro y no para señales de espectro complejo como lo son las señales de audio. Sin embargo, si se considera un criterio de enmascaramiento dependiente de la amplitud de cada componente de frecuencia puede extenderse la aplicación de la ponderación propuesta a señales de espectro complejo. El criterio de enmascaramiento que se considerará es el mismo modelo

utilizado en la aproximación de bandas críticas: el modelo de Terhardt (Sección 2.2.3).

Si denominamos $M(f_i)$ a la curva de enmascaramiento asociada a una frecuencia f_i con una amplitud L_{pi} dB y consideramos una señal de audio $x[n]$ de N muestras de longitud a la cual se aplica una FFT de K puntos, entonces se obtendrán K umbrales de enmascaramiento $M(f_i)$ con $i = 1$ a K . Se define un Umbral de Enmascaramiento Global $M_{\text{global}}(f)$, el cual se obtiene al superponer los K umbrales de enmascaramiento $M(f_i)$ y leer los máximos valores [7], tal como se presenta en la Figura 3.1.

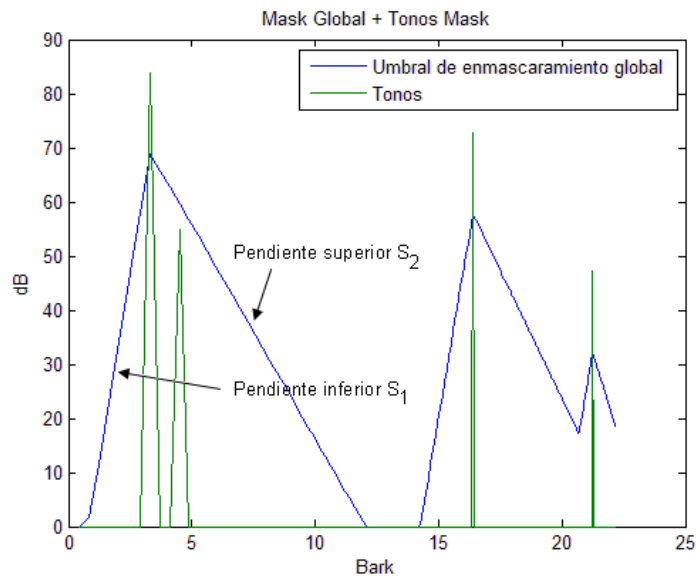


Figura 3.1: Curva de enmascaramiento global usando el modelo de Terhardt. De los 4 tonos puros presentados, el segundo tono puro es enmascarado por el umbral generado por el primer tono, por lo que el tono de amplitud 55 dB no será percibido por el oído humano.

Este procedimiento será aplicado a cada bloque de audio de N muestras, ingresando el máximo entre la amplitud $\|X(\omega)\|$ y el umbral $M_{\text{global}}(f)$ a la RNA del tipo Perceptrón Multicapa 2-11-9-1 para interpolación del nivel de sonoridad P . Esta consideración permite la aplicación de los contornos de igual sonoridad a señales de espectro de banda ancha. Los resultados serán comparados al procedimiento rotulado en la norma ISO 532-B, en condición de campo libre.

Si, además, se añade un análisis en bandas críticas, tal como se propuso en 2009 [27], el diagrama en bloques para la determinación de la sonoridad total de una señal de audio a partir de la ponderación de amplitud usando Redes Neuronales Artificiales y la determinación del umbral de enmascaramiento global $M_{\text{global}}(f)$ queda representado por la Figura 3.2.

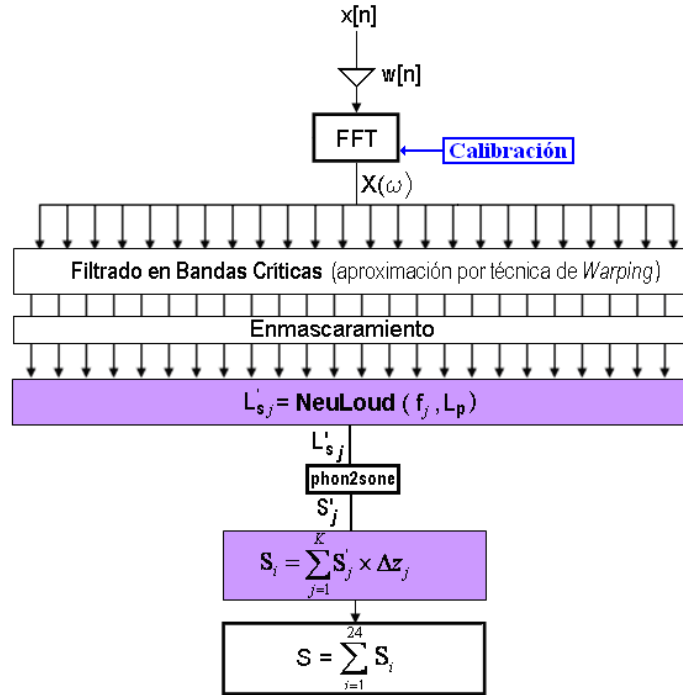


Figura 3.2: Modelo para el cálculo de sonoridad propuesto en [27]. Se debe considerar una etapa de calibración para la señal $x[n]$ que permita la conversión de dBFS a dB (re.20 μPa)

Esta propuesta considera un análisis efectuado mayoritariamente en el dominio de la frecuencia. Previo a la aplicación de una FFT, sólo es aplicada una ventana $w[n]$.

Las etapas del modelo, mostradas en la Figura 3.2, son: (a) Aplicación de una ventana temporal a la señal de audio de entrada; (b) Conversión del dominio temporal al dominio de la frecuencia; (c) Filtrado de la señal usando un banco de filtros en bandas críticas; (d) Cálculo del enmascaramiento global de la señal

de audio por cada banda crítica. Se usa el modelo de Terhardt considerando una separación de 17 dB entre el tono enmascarador con su respectiva banda enmascarada; (e) Ponderación dinámica de la señal usando la RNA de Espinoza et al. [8] (función `NeuLoud` de MATLAB®), entregando el nivel de sonoridad específico P' (Fon/Bark); (f) Conversión de nivel de sonoridad específica P' a sonoridad específica S' (Son/Bark); (g) Cálculo de la sonoridad por cada banda crítica: área bajo la curva de sonoridad específica; (h) cálculo de la sonoridad total: suma aritmética de la sonoridad de cada una de las 24 bandas críticas.

3.2. Aproximación a un banco de filtros en bandas críticas

Para efectuar una aproximación al análisis en bandas críticas del oído humano, según lo propuesto por Zwicker (Tabla 2.3), se debe introducir una distorsión del eje de frecuencias de manera de obtener un eje con resolución similar a la razón de banda crítica z , medida en Barks. Con esta nueva resolución es posible diseñar un banco de filtros, a partir de las frecuencias centrales propuestas por Zwicker, definiendo convenientemente las pendientes de atenuación de cada banda.

Para la distorsión del eje de frecuencias se usará la técnica de *Warping* de frecuencias, mientras que para la aproximación de las pendientes de cada filtro pasabanda, se hará utilización del modelo de enmascaramiento de Terhardt.

3.2.1. Parámetros de diseño

En la sección 2.3.3, Figura 2.25, se muestra gráficamente la similitud entre la razón de banda crítica z [Bark] y la frecuencia en dominio *warp* ν [rad/samples], al considerar un eje normalizado de frecuencias. Si expresamos z y ν como función de la frecuencia f en Hz, se obtiene:

$$z(f) = 13 \arctan(7,6 \cdot 10^{-4} f) + 3,5 \arctan(f/7500)^2 \quad (3.3)$$

$$\nu(f) = 2\pi f/f_s + 2 \cdot \arctan\left(\frac{\lambda \sin(2\pi f/f_s)}{1 - \lambda \cos(2\pi f/f_s)}\right) \quad (3.4)$$

Donde f_s es la frecuencia de muestreo en Hz, que determina la máxima frecuencia representable en un sistema digital, correspondiente a $f = f_s/2$ (Teorema de Shannon).

Usando las Ecuaciones anteriores para una $f_s = 44.100$ Hz, el rango de valores adquiridos por la razón de banda crítica z y la frecuencia ν en dominio warp son:

$$0 < z < 24,7 \quad [\text{Bark}]$$

$$0 < \nu < \pi \quad [\text{rad/samples}]$$

En la Figura 3.3 se muestra la relación aproximadamente lineal entre las variables z y ν .

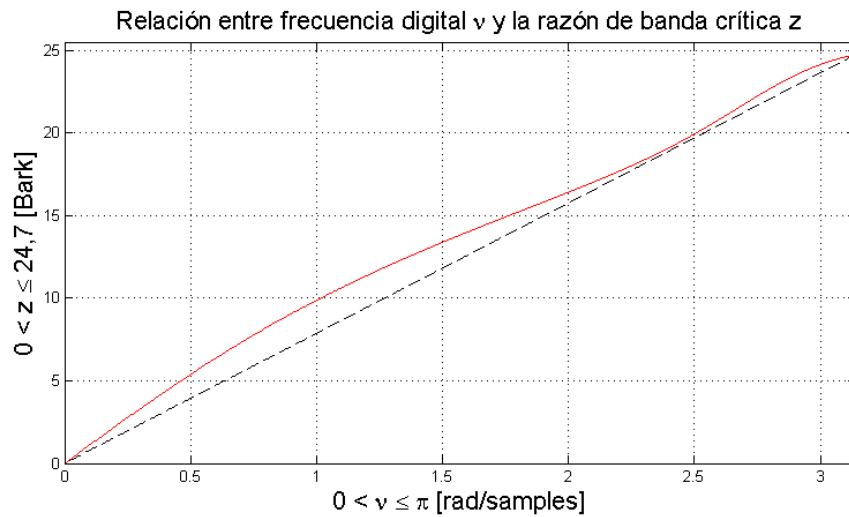


Figura 3.3: La curva roja representa la relación aproximadamente lineal entre la razón de banda crítica z y la frecuencia en dominio warp ν

Se definirá, entonces, la aproximación a la razón de banda crítica usando la técnica de *Warping* como z_ν , de acuerdo a la Ecuación 3.5.

$$z \approx z_\nu = \frac{\nu}{\pi} \cdot z_s \quad (3.5)$$

Donde z_s corresponde a la máxima frecuencia representable en un sistema digital expresada en Barks.

La Ecuación 3.3 sólo tiene validez para frecuencias menores a 22.050 Hz, por lo que la aproximación de la Ecuación 3.5 es válida para su uso con frecuencias de muestreo menores o iguales a 44.100 Hz. Para hacerla extensible a frecuencias de muestreo mayores, basta con omitir toda aquella información vinculada a las frecuencias superiores a 22.050 Hz. Las frecuencias por sobre este límite no representan un aporte significativo en el fenómeno de sensación auditiva.

El diseño de cada filtro pasa banda del banco de bandas críticas se hará partiendo de las especificaciones digitales que nos entregue la variable z o, lo que es aproximadamente equivalente, la variable z_ν .

Los filtros pasabanda se formarán al disponer un filtro pasa bajos y un filtro pasa altos en cascada (en serie), por lo que la frecuencia de corte del LPF coincidirá con la frecuencia de corte superior del pasa banda, mientras que la frecuencia de corte del HPF coincidirá con la frecuencia de corte inferior del pasa banda. De acuerdo a la tabla 2.3, estas frecuencias de corte coinciden con los valores enteros desde 0 a 24 Barks, sin embargo, para el proceso de diseño deben ser traducidos a una frecuencia digital ν entre 0 y π . Estos valores se muestran en la Tabla 3.2.

Las columnas de frecuencias digitales ν_i y ν_u fueron obtenidas considerando la Ecuación 3.5, por lo que:

$$\nu = \frac{z_\nu}{z_s} \cdot \pi \quad (3.6)$$

3.2.2. Diseño de los filtros pasa-bajos

Los parámetros de diseño requeridos para obtener un filtro pasa bajos son las frecuencias angulares digitales de corte ν_u (Tabla 3.2) y la atenuación en dB/Barks. Tal como se mencionó en el apartado 2.2.3, para este último parámetro, se considerará el modelo de Terhardt, el que postula que la pendiente superior S_2 de atenuación para una banda crítica es dependiente de su frecuencia central. Si reescribimos la Ecuación 2.13 se obtiene:

Banda Nº	Frecuencias			Frecuencias		
	f_i [Hz]	z_i [Bark]	ν_i [rad/samples]	f_u [Hz]	z_u [Bark]	ν_u [rad/samples]
0	20	0,2	0,0127	100	1	0,1270
1	100	1	0,1270	200	2	0,2540
2	200	2	0,2540	300	3	0,3810
3	300	3	0,3810	400	4	0,5070
4	400	4	0,5070	510	5	0,6349
5	510	5	0,6349	630	6	0,7619
6	630	6	0,7619	770	7	0,8889
7	770	7	0,8889	920	8	1,0159
8	920	8	1,0159	1.080	9	1,1429
9	1.080	9	1,1429	1.270	10	1,2698
10	1.270	10	1,2698	1.480	11	1,3968
11	1.480	11	1,3968	1.720	12	1,5238
12	1.720	12	1,5238	2.000	13	1,6508
13	2.000	13	1,6508	2.320	14	1,7778
14	2.320	14	1,7778	2.700	15	1,9048
15	2.700	15	1,9048	3.150	16	2,0317
16	3.150	16	2,0317	3.700	17	2,1587
17	3.700	17	2,1587	4.400	18	2,2857
18	4.400	18	2,2857	5.300	19	2,4127
19	5.300	19	2,4127	6.400	20	2,5397
20	6.400	20	2,5397	7.700	21	2,6667
21	7.700	21	2,6667	9.500	22	2,7936
22	9.500	22	2,7936	12.000	23	2,9206
23	12.000	23	2,9206	15.500	24	3,0476

Tabla 3.2: Frecuencias de corte en Hz, en Barks y en rad/samples (considerando un dominio *Warp*) para una $f_s = 44.100$ Hz. Las frecuencias ν_i corresponderán a las frecuencias de corte de cada HPF del banco a diseñar, mientras que las frecuencias ν_u corresponderán a las frecuencias de los LPF.

$$S_2 = -24 - 230/f_c + 0,2 \cdot L_p \quad (3.7)$$

El parámetro L_p corresponde al nivel de presión sonora, en dB, de un tono puro centrado en f_c que produce la banda de enmascaramiento o banda crítica cuya pendiente de atenuación, para frecuencias superiores a f_c está dada por la Ecuación 3.7. Se consideró un L_p de 70 dB, por corresponder a un nivel de presión sonora RMS común en la reproducción de piezas musicales y cine (asociado a un valor *peak* aproximado de 85 dB).

Banda N°	f_c [Hz]	S_2 [dB/Bark]	Banda N°	f_c [Hz]	S_2 [dB/Bark]
0	50	14,60	12	1.850	10,12
1	150	11,53	13	2.150	10,11
2	250	10,92	14	2.500	10,09
3	350	10,66	15	2.900	10,08
4	450	10,51	16	3.400	10,07
5	570	10,40	17	4.000	10,06
6	700	10,33	18	4.800	10,05
7	840	10,27	19	5.800	10,04
8	1.000	10,23	20	7.000	10,03
9	1.170	10,20	21	8.500	10,03
10	1.370	10,17	22	10.500	10,02
11	1.600	10,14	23	13.500	10,02

Tabla 3.3: Valores de la pendiente de atenuación S_2 para el diseño de filtros pasa bajos Butterworth para las 24 primeras bandas críticas. Se usó la Ecuación 3.7 con $L_p = 70$ dB.

Si ahora consideramos la Ecuación B.11 y la reexpresamos en términos de nuestros parámetros de diseño:

$$N = \frac{\log_{10}(10^{(A/10)} - 1)}{2 \log_{10}(\tan(\nu_s/2)/\tan(\nu_p/2))} \quad (3.8)$$

Donde A corresponde a la atenuación requerida entre la región de pasa banda y el inicio de la banda de rechazo, por lo que $A = S_2 + 3$ dB².

La función de transferencia en dominio analógico $H(s)$ del LPF Butterworth diseñado a partir de las especificaciones en el dominio discreto está dada por:

$$H(s) = \frac{1}{\prod_{k=1}^N (s - p_k)} \quad (3.9)$$

Donde los valores p_k son los polos del sistema, definidos por la Ecuación 2.43.

² S_2 representa la atenuación del filtro por cada Bark de frecuencia. Se consideró como referencia la frecuencias de corte del LPF. Como inicio de la zona de rechaza banda se consideró la frecuencia de corte del LPF de la banda adyacente superior (que dista en 1 Bark aproximadamente). En estas condiciones, la atenuación A requerida para el diseño del LPF Butterworth corresponde a 3 dB por sobre S_2 , ya que A corresponde a la atenuación entre la zona de pasa banda y de rechaza banda y no respecto de la frecuencia de corte del filtro como representa la atenuación S_2 .

Finalmente, para obtener la función de transferencia digital $H(z^{-1})$ debe aplicarse una transformación bilineal a la ya conocida función $H(s)$ de acuerdo a la Ecuación 2.45. En el apéndice A se muestra un ejemplo de cómo aplicar esta transformación bilineal, de manera de conseguir el mapeamiento del plano analógico s a la circunferencia unitaria en el dominio z .

3.2.3. Diseño de los filtros pasa-altos

Al igual que en el caso de los filtros pasa bajos (LPF), para el diseño de los filtros pasa altos (HPF) se requiere conocer las frecuencias angulares digitales de corte ν_i (Tabla 3.2) además de la atenuación en dB/Barks. El modelo de Terhardt plantea, esta vez, una atenuación constante para cada banda de $S_1 = 27$ dB/Bark. De igual manera, se considerará una atenuación A que es 3 dB mayor que S_1 .

$$N = \frac{\log_{10}(10^{(A/10)} - 1)}{2 \log_{10}(\tan((\pi - \nu_s)/2) / \tan((\pi - \nu_p)/2))} \quad (3.10)$$

En este caso, la frecuencia digital de corte (o de paso) ν_p está 1 Bark por sobre la frecuencia digital de rechaza banda ν_s . Debido a la simetría que supone la técnica de reversión espectral para el diseño de HPF a partir de LPF, en la Ecuación 3.10 se usan las frecuencias digitales *complementarias* a las frecuencias de diseño de un HPF.

Con la aproximación entera de N , usando la Ecuación 3.10, se realiza el diseño de un LPF con frecuencia de corte $\omega_p = \pi - \nu_p$. En la función de transferencia digital obtenida se debe reemplazar cada variable z^{-1} por su inverso aditivo $-z^{-1}$. En el Apéndice B se presenta un ejemplo de diseño aplicando este procedimiento.

3.2.4. Construcción de los filtros pasa-banda

Los filtros pasa banda del banco de filtros auditivo que se desea implementar tienen características altamente no simétricas, vale decir, las bandas no presentan las mismas pendientes de atenuación para las frecuencias superiores a la frecuencia de corte que para las frecuencias inferiores a ella. Debido a que se requiere tener absoluto control del comportamiento de cada pasa banda es que se optó por

disponer en cascada un filtro pasa bajos (LPF) y un filtro pasa altos (HPF) para obtener cada una de las bandas del banco de filtros.

En la Tabla 3.4 se muestran los órdenes de cada LPF y BPF Butterworth utilizados en la construcción de las bandas críticas aproximadas por la técnica de *Warping* [28]. Por efectos prácticos no se muestran los coeficientes α_k y β_k ³ de cada una de las funciones de transferencia asociadas, sin embargo, se adjunta a este trabajo un disco compacto que contiene almacenados estos valores en formato de MATLAB®(extensión *.mat).

Banda Nº	Orden		Banda Nº	Orden	
	LPF	HPF		LPF	HPF
0	3	1	12	10	24
1	4	4	13	9	24
2	5	8	14	9	24
3	6	10	15	8	23
4	7	13	16	8	21
5	7	16	17	7	19
6	8	18	18	6	17
7	9	20	19	5	15
8	9	22	20	4	12
9	9	23	21	3	10
10	10	24	22	2	7
11	10	24	23	1	4

Tabla 3.4: Orden de los filtros pasa bajos (LPF) y pasa altos (HPF) usados en la construcción de los filtros auditivos usando la técnica de *Warping* de frecuencias. En general, se puede decir que los filtros pasa banda resultantes son de un orden elevado.

En la Figura 3.4 se muestran 3 de las 24 bandas del banco de filtros propuesto. Aquí, el eje de frecuencias es presentado en términos de la frecuencia en dominio *Warp* ν , que se asemeja a la escala psicoacústica Bark.

Cabe mencionar que las especificaciones del banco de filtros obtenido se aproximan a los valores tomados como referencia (Tabla 3.2) principalmente en los

³La denominación α_k y β_k para los coeficientes de las funciones de transferencia asociadas a cada una de las bandas indica que el banco de filtros fue diseñado en el dominio *warp* de frecuencias. La denominación a_k y b_k sólo será utilizada en caso de trabajar con un eje lineal de frecuencias.

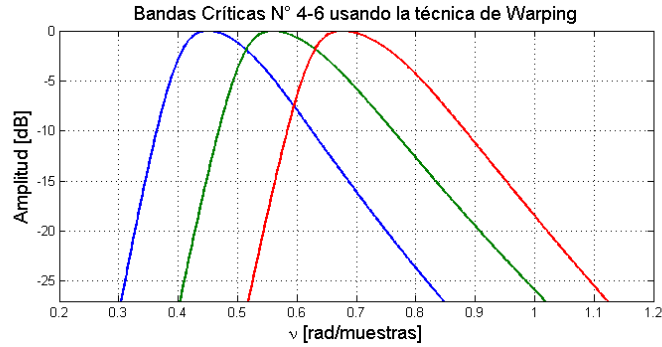


Figura 3.4: Bandas N° 4, 5 y 6 del banco de filtros en bandas críticas aproximados usando la técnica de *Warping* de frecuencias.

respectivos anchos de banda, así como en la selectividad de cada filtro (pendientes de atenuación del banco). Las frecuencias de corte, entretanto, presentan un corrimiento considerable respecto de las referencias mayormente en la región de alta frecuencia. Las frecuencias centrales y de corte –expresadas en Hz– de cada banda del banco de filtros diseñado considerando la técnica de *Warping* son mostradas en la Tabla 3.5.

Banda N°	Frecuencia[Hz]			Banda N°	Frecuencia[Hz]		
	f_i	f_c	f_u		f_i	f_c	f_u
0	20	70	130	12	1.820	1.940	2.140
1	130	180	260	13	2.080	2.210	2.450
2	250	300	390	14	2.360	2.510	2.780
3	380	430	520	15	2.680	2.850	3.190
4	510	560	660	16	3.070	3.280	3.680
5	640	700	810	17	3.530	3.790	4.300
6	780	840	960	18	4.100	4.440	5.120
7	920	990	1.110	19	4.840	5.290	6.220
8	1.080	1.150	1.290	20	5.810	6.460	7.780
9	1.240	1.320	1.460	21	7.190	8.110	10.080
10	1.420	1.510	1.660	22	9.190	10.660	13.630
11	1.620	1.710	1.890	23	12.250	14.750	18.660

Tabla 3.5: Características del banco de 24 filtros en bandas críticas usando la técnica de *Warping*. Las frecuencias mostradas se han redondeado al múltiplo de 10 más cercano. Debido al corrimiento de las frecuencias de corte como consecuencia del procedimiento de diseño, la última banda queda en su mayor parte por sobre los 12.500 Hz.

Tal como se muestra en la Tabla 3.5 gran parte de la última banda crítica (Banda N° 23) se encuentra por sobre los 12.500 Hz. Todo ese rango deberá ser omitido al efectuar una ponderación de sonoridad, debido a que las curvas de sonoridad no están definidas sobre este límite. La frecuencia de corte superior f_u de la Banda N° 22 también se encuentra por sobre los 12.500 Hz por lo que se debe hacer la misma consideración. Estas situaciones se traducen en una especie de truncamiento de bandas al usar una ponderación en Fones, por lo que podría plantearse una especie de factor de corrección para compensar el porcentaje de banda omitido.

3.3. Filtrado multibanda de señales de audio

Las bandas críticas aproximadas, tal como son presentadas en la Figura 3.4 suponen el filtrado de una señal de audio a la cual ya se haya efectuado la técnica de *Warping*.

La aplicación del *warping* de frecuencias a una señal temporal $x[n]$ se obtiene al leer la salida $w[n]$ del sistema de la Figura 2.26⁴ cuando a él ingresa la señal $x[n]$. La cantidad M de filtros *AllPass* dispuestos en cascada es un parámetro que influye en la señal post-*warping* $w[n]$. Como los filtros *AllPass* son filtros de respuesta impulsiva infinita (IIR), una baja cantidad M de elementos produciría un truncamiento en la respuesta del sistema. En un procedimiento empírico, se determinó que para obtener una señal $w[n]$ apropiada⁵, se requiere una cadena de filtros *AllPass* de $M = 7 \cdot N$ elementos, donde N es la longitud de la señal original $x[n]$.

La situación anterior hace que la técnica de *warping* de frecuencias no sea viable para una implementación en tiempo real, ya que la sola conversión de una señal de audio al dominio *warp* requiere un alto costo computacional. A eso se debe agregar que el proceso de filtrado multibanda considerando un banco

⁴Debe incluirse una etapa de pre-énfasis.

⁵Se requiere que la señal $w[n]$ asegure una respuesta en frecuencia unitaria desde 0 a 12.500 Hz.

de 24 filtros de orden N^6 , con N entre 4 y 34, también requiere un alto costo computacional.

3.3.1. Banco de filtros en el dominio lineal de frecuencias

Se aplicó una transformación bilineal inversa al banco de filtros diseñado. Lo que se desea es trabajar el banco de filtros en un eje lineal de frecuencias. Recordemos que la técnica de *warping* considera el reemplazo de cada retardo unitario z^{-1} por un filtro *AllPass* de orden 1 dado por:

$$D(z^{-1}) = \frac{-\lambda + z^{-1}}{1 - \lambda z^{-1}} \quad (3.11)$$

Por comparación con la Ecuación A.4 (Apéndice A), la transformación bilineal inversa asociada está dada por:

$$D^{-1}(z^{-1}) = \frac{\lambda + z^{-1}}{1 + \lambda z^{-1}} \quad (3.12)$$

Si consideramos una función de transferencia $G(z^{-1})$, en dominio *warp*:

$$G(z^{-1}) = \frac{\beta_0 + \beta_1 z^{-1} + \dots + \beta_N z^{-N}}{\alpha_0 + \alpha_1 z^{-1} + \dots + \alpha_N z^{-N}} \quad (3.13)$$

Los coeficientes β_k y α_k de la función de transferencia $G(z^{-1})$ denotan que su filtro característico está en dominio *warp*. Al reemplazar z^{-1} por $D^{-1}(z^{-1})$ (Ecuación 3.12) en la Ecuación 3.13 se llega a una función de transferencia equivalente, pero en dominio lineal:

$$G(z^{-1}) = \frac{b_0 + b_1 z^{-1} + \dots + b_N z^{-N}}{a_0 + a_1 z^{-1} + \dots + a_N z^{-N}} \quad (3.14)$$

3.3.2. Implementación Recursiva del Banco de Filtros

Si se conocen los valores a_k y b_k que caracterizan la función de transferencia $G_i(z^{-1})$, que se asocia a la banda i de un banco de filtros, entonces se cumple

⁶ $N_{\text{total}} = N_{LFF} + N_{HFF}$

con la ecuación en diferencias lineales de la Ecuación 2.33. Si en ella se despeja la muestra $y[n]$, se obtiene:

$$y[n] = b_0x[n] + b_1x[n-1] + \dots + b_Nx[n-N] - a_1y[n-1] - \dots - a_Ny[n-N] \quad (3.15)$$

Donde la serie temporal y corresponde a la señal filtrada por la banda i , al ingresar una señal de entrada x . La m -ésima muestra de salida $y[m]$ se determina a partir de la muestra actual de entrada, $x[m]$, así como de las muestras retardadas tanto de entrada como de salida $x[m-k]$ e $y[m-k]$, con $k = 1, 2, \dots, m$.

A mayor orden N del sistema digital $G_i(z^{-1})$, mayor cantidad de sumas ponderadas se requieren para el cálculo de la señal filtrada $y[n]$.

3.3.3. Filtrado en el dominio de la Frecuencia

Para analizar una señal de audio, sólo se requiere extraer la información necesaria para concretar la suma energética de cada banda. Gracias al teorema de Parseval, presentado en la Ecuación 2.30, esta suma de energías puede hacerse indistintamente en el dominio del tiempo (usando la señal filtrada o en el dominio de la frecuencia.

Para el cálculo del valor cuadrático medio $y_{i_{RMS}}$ asociado a la señal $x[n]$ filtrada por la banda i :

$$y_{i_{RMS}} = 10 \log_{10} \left(\frac{1}{N} \sum_{i=1}^N y_i^2[n] \right) \quad (3.16)$$

$$y_{i_{RMS}} = 10 \log_{10} \left[\frac{1}{N} \left(\frac{1}{K} \sum_{k=1}^K Y_i^2[k] \right) \right] \quad (3.17)$$

Donde $K = N/2$ y la señal $Y_i[k]$ corresponde a la transformada de Fourier de la señal de salida $y_i[n]$. Además, se cumple que:

$$Y_i[k] = X[k] \cdot H_i[k] \quad (3.18)$$

Donde $X[k]$ con k entre 1 y K corresponde a la transformada de Fourier de la señal de entrada $x[n]$. Como consecuencia de esto, conociendo $H_i[k]$ se puede obtener $Y_i[k]$ y por lo tanto se puede calcular $y_{i_{RMS}}$, usando la Ecuación 3.17. En este contexto, el almacenamiento de la respuesta $H_i[k]$ nos permitiría prescindir de la implementación recursiva del la filtro i y, probablemente, también de los coeficientes a_k y b_k de la función de transferencia.

En [29] se propone el almacenamiento de la respuesta en frecuencia $H_i(\omega)$ utilizando Redes Neuronales Artificiales del tipo Radial Base, lo que permite una posterior reconstrucción de cada banda de frecuencia.

3.3.4. Filtrado usando Redes Neuronales Artificiales de Funciones Radial Base

La utilización de Redes Neuronales Artificiales (RNA) para efectuar el filtrado de señales gira en torno a la idea de almacenar la respuesta en frecuencia (o de amplitud) de un filtro determinado. En nuestro caso se utilizaron RNA de funciones radial base (RBFN) para almacenar las respuestas en amplitud $H_i(\omega)$ de cada una de las bandas críticas diseñadas usando la técnica de *Warping* de frecuencias [29].

El rango de frecuencias a interpolar va desde 20 a 22.050 Hz, mientras que el rango dinámico se aproximará para las amplitudes de la respuesta en frecuencia entre -100 y 0 dBFS.

En la Figura 3.5 se muestra la interpolación de 10 de las 24 bandas críticas mediante la implementación de las RBFN de la Tabla 3.6 al considerar una frecuencia de muestreo de 64.000 Hz. Cada filtro pasa banda fue implementado usando una red diferente.

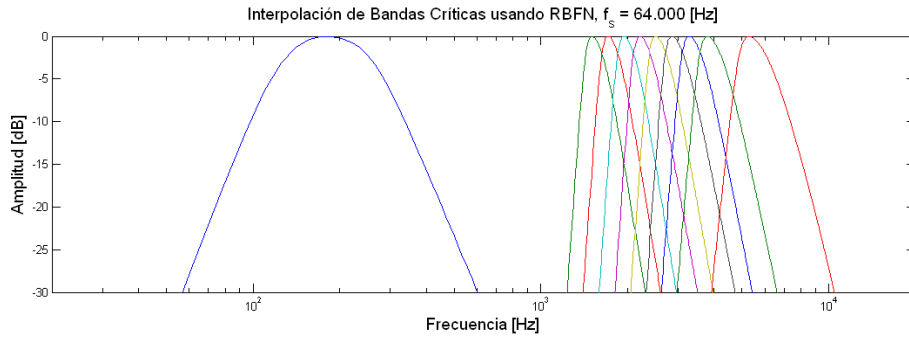


Figura 3.5: Algunas de las Bandas Críticas interpoladas usando las RNA del tipo RBFN, considerando una frecuencia de muestreo $f_s = 64.000$ Hz. El entrenamiento fue realizado considerando un $f_s = 44.100$ Hz.

3.3.5. Filtrado en tiempo real: Comparación entre metodologías

En esta sección se presenta un análisis cuantitativo del consumo computacional de la metodología de filtrado temporal efectuando una **implementación recursiva** y de la metodología de filtrado en el dominio de la frecuencia usando **Redes Neuronales Artificiales RNA**. Se implementó un prototipo para cada metodología en un plug-in VST 2.4 usando como herramienta de monitoreo el medidor de carga de CPU, *VST Performance*, que poseen las aplicaciones de audio de Steinberg.

En la Tabla 3.6 se muestran las características de los sistemas digitales que posibilitan la implementación de los filtros auditivos en bandas críticas usando la técnica de *Warping*.

En el caso del filtrado temporal el factor más influyente en el costo computacional del algoritmo es el orden de los pasa bajos y pasa altos. En este contexto, las bandas más costosas son las centradas en 1.510, 1.710 y 1.940 Hz, que corresponden a filtros de orden 34. Por otro lado, considerando un filtrado en el dominio de frecuencias usando RBFN el factor más influyente en el costo computacional del algoritmo de filtrado es el número de neuronas en la capa oculta, donde las 2 primeras bandas llevan la delantera, con 276 neuronas.

Al implementar las 24 bandas críticas, fue posible apreciar una notable supe-

Frecuencia Central f_c [Hz]	Orden de Filtro IIR		Configuración Óptima para para la RBFN	Frecuencia Central f_c [Hz]	Orden de Filtro IIR		Configuración Óptima para para la RBFN
	LPF	HPF			LPF	HPF	
70	3	1	1-276-1	1.940	10	24	1-195-1
180	4	4	1-276-1	2.210	9	24	1-208-1
300	5	8	1-250-1	2.510	9	24	1-221-1
430	6	10	1-247-1	2.850	8	23	1-175-1
560	7	13	1-248-1	3.280	8	21	1-171-1
700	7	16	1-251-1	3.790	7	19	1-143-1
840	8	18	1-248-1	4.440	6	17	1-102-1
990	9	20	1-208-1	5.290	5	15	1-112-1
1.150	9	22	1-244-1	6.460	4	12	1-126-1
1.320	9	23	1-210-1	8.110	3	10	1-134-1
1.510	10	24	1-189-1	10.660	2	7	1-155-1
1.710	10	24	1-225-1	14.750	1	4	1-177-1

Tabla 3.6: Características de los filtros auditivos utilizados en el entrenamiento de las RBFN. Se muestran las configuraciones óptimas encontradas, que aseguran un error despreciable en la interpolación de cada banda entre 20 y 22.050 Hz a niveles por sobre -100 dBFS. Para la implementación recursiva de cada banda debe considerarse 2 filtros IIR en cascada con los órdenes especificados en las columnas *Orden LPF* y *Orden HPF*.

rrioridad en la implementación del banco de filtros al usar RBFN, situación que fue imposible de dimensionar debido a que la carga de CPU del prototipo de filtrado temporal superaba el 100%. Por este motivo, se repitió la implementación de los prototipos de filtrado temporal y en el dominio de frecuencias, pero sólo considerando 10 de las 24 bandas críticas⁷.

En la Figura 3.6 se muestra el consumo de CPU devuelto por el medidor *VST performance* para los plug-ins que implementan los 2 algoritmos de filtrado de señales. Para todas las mediciones se utilizó un Laptop con procesador AMD Turion® de 64 bits, doble núcleo, 1,90 GHz y una memoria RAM de 1 GB.

El bajo costo computacional del algoritmo de filtrado que utiliza RBFN se atribuye a que la interpolación de la respuesta en amplitud de cada banda crítica se realiza una sola vez, al cargar el plug-in o al cambiar la frecuencia de muestreo

⁷Las bandas escogidas corresponden a las bandas 1, 10-17 y 19, las cuales incluyen las mencionadas bandas "más costosas" para cada uno de los métodos. Las 10 bandas consideradas son resaltadas en negrillas en la Tabla 3.6



Figura 3.6: Carga de CPU devuelto por el medidor *VST performance* para el filtrado de señales considerando 10 filtros auditivos. (a) Consumo de CPU de un 48,00 % para el algoritmo filtrado temporal (implementación recursiva). (b) Consumo de CPU de un 2,89 % para el algoritmo de filtrado usando RBFN.

de la sesión de audio. Luego de eso, el filtrado de señales se traduce en una serie de multiplicaciones, de acuerdo a la Ecuación 3.18, siendo este proceso el que produce una carga de CPU de un 2,89 % para las 10 bandas implementadas.

Como consecuencia del análisis anterior, si una implementación de las bandas críticas usando filtros IIR es poco viable para un análisis en tiempo real, la utilización de filtros FIR para el mismo fin entregará un consumo de CPU aún mayor. Por este motivo se descartó la implementación de prototipos para el análisis espectral usando filtros FIR.

3.3.6. Precisión del algoritmo de filtrado usando RBFN

El algoritmo de filtrado usando RBFN resultó ser considerablemente mejor en cuanto a tiempo de ejecución, sin embargo, como esta metodología trabaja en el dominio de la frecuencia, resulta ser apenas una aproximación al filtrado temporal. Debido a esto es que a continuación se presenta un análisis de la precisión en el cálculo de los valores RMS al usar RBFN en comparación con su implementación recursiva.

Se utilizaron 2 bloques de audio de 4.096 muestras extraídos de 2 archivos de audio digital (`Audio1.wav` y `Audio2.wav`) muestreados a $f_s = 44.100$ Hz. Los valores RMS de `Audio1` y `Audio2` son mostrados en la Figura 3.7.

Al utilizar una transformada rápida de Fourier (FFT) de 4.096 puntos, la máxima diferencia en el cálculo energético por banda de los archivos analizados es de 1,5 dB (en la región de alta frecuencia), mientras que el promedio aritmético de la desviación de amplitudes es de 1,1 dB. Si consideramos que la mínima

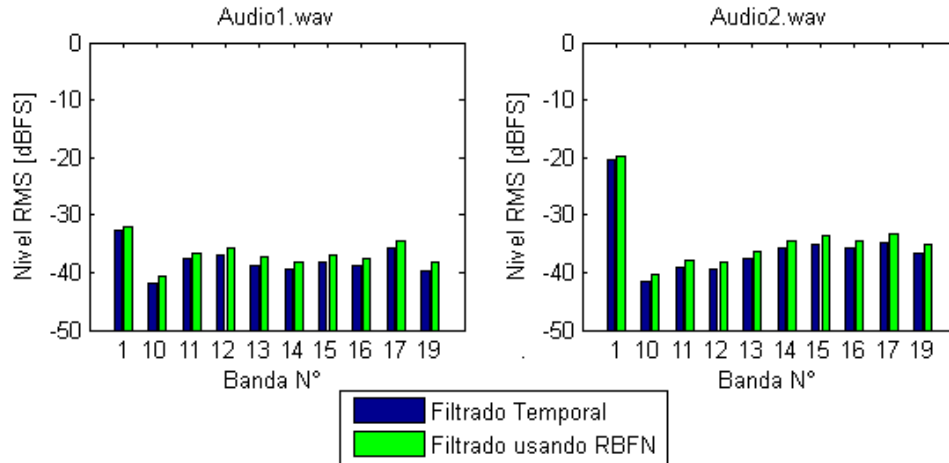


Figura 3.7: Valores RMS correspondientes a algunas bandas de frecuencia aplicados a 2 fragmentos de audio de 4.096 muestras a una $f_s = 44.100$ Hz. Se muestran los valores obtenidos al aplicar pasa bandas de respuesta impulsiva infinita (IIR), así como sus respectivas aproximaciones determinadas en el dominio de la frecuencia.

diferencia percibida por el oído humano⁸ es aceptada como 1 dB a 1 kHz [9], la variación entre los resultados al filtrar usando RBFN y filtros IIR puede ser considerada despreciable.



Figura 3.8: Valores RMS en los plug-ins VST programados. El cálculo en el dominio de la frecuencia (derecha) muestra una tendencia a presentar valores RMS mayores que la implementación temporal (izquierda), tal como ocurre en su versión *off-line* programada en MATLAB® (Figura 3.7).

En la Figura 3.8 se muestran los 2 plug-ins VST implementados tanto para la evaluación de eficacia como para la de precisión de las metodologías usadas para el filtrado temporal y en el dominio de la frecuencia. Tal como ocurrió en

⁸Del inglés: *Just-Noticeable Difference* (JND).

el prototipo *off-line* efectuado en MATLAB, los valores de amplitud RMS por banda tienden a ser mayores en la implementación en el dominio de frecuencias.

3.3.7. Ponderación psicoacústica de frecuencias considerando la técnica de *Warping*

Lo que se ha logrado hasta el momento es la implementación de un banco de filtros en bandas críticas con características altamente selectivas que efectúa un cálculo energético en el dominio de la frecuencia, sin embargo, en el eje de frecuencias no se ha introducido ninguna ponderación psicoacústica.

Para efectuar una ponderación psicoacústica de cada componente de frecuencia, de acuerdo a la resolución espectral del oído humano⁹ se debe introducir una modificación a la Ecuación 3.19 propuesta para el cálculo RMS de la señal de audio digital $x[n]$ filtrada por la banda i del banco de filtro psicoacústico, $y_{i_{RMS}}$.

$$y_{i_{RMS}} = 10 \log_{10} \left[\frac{1}{N} \left(\frac{1}{K} \sum_{k=1}^K Y_i^2[k] \right) \right] \quad (3.19)$$

Esta Ecuación es equivalente a la Ecuación 3.20:

$$y_{i_{RMS}} = 10 \log_{10} \left[\frac{1}{N} \left(\frac{1}{K} \sum_{k=1}^K Y_i^2[k] \Delta\omega_k \right) \right] \quad (3.20)$$

Donde $\Delta\omega_k = 1$ es un parámetro relacionado con el espaciado normalizado entre frecuencias de manera que $\sum_k \Delta\omega_k = K$, lo que se traduce en una ponderación equitativa para cada componente de frecuencia, en otras palabras, la amplitud $Y_i[k_1]$ para una componente de baja frecuencia tiene la misma importancia que una componente de alta frecuencia $Y_i[k_2]$.

Si el parámetro $\Delta\omega_k$ lo cambiamos por $\Delta\nu_k$, el cual representa un espaciado de frecuencias considerando la técnica de *Warping*, también en un eje normalizado, entonces se le brindará una mayor ponderación de frecuencias a la región de baja frecuencia de manera similar a la resolución espectral auditiva.

⁹Vale decir, una mayor importancia a las frecuencias bajas que a las frecuencias altas.

Esto se produce debido a que el factor $\Delta\nu_k$ decrece con la frecuencia, tal como es posible apreciar en la Figura 3.9(b).

$$y_{i_{RMS}} = 10 \log_{10} \left[\frac{1}{N} \left(\frac{1}{K} \sum_{k=1}^K Y_i^2[k] \Delta\nu_k \right) \right] \quad (3.21)$$

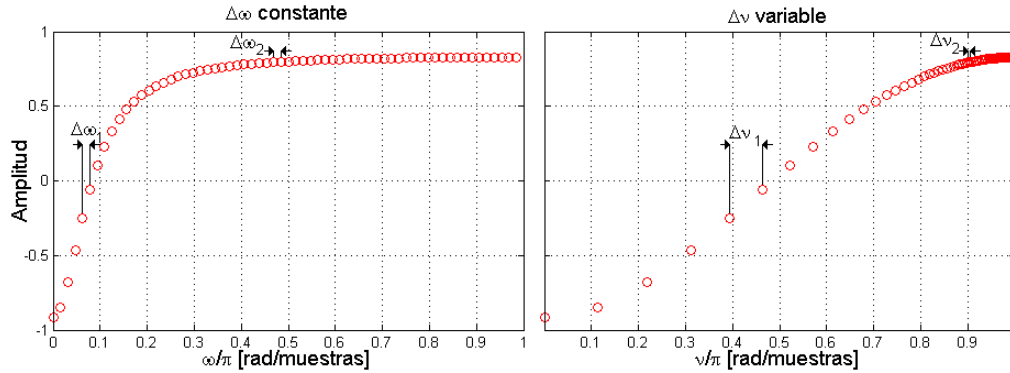


Figura 3.9: Se muestra una función $Y[k]$ (a) en el dominio lineal de frecuencias y (b) en el dominio *Warp* de frecuencias. En el caso lineal, $\Delta\omega_1 = \Delta\omega_2$, por lo que se otorga una ponderación equitativa a cada componente de frecuencia cuando se calcula un valor RMS. Por otro lado, al usar la frecuencia en dominio *Warp* $\nu = \nu(\omega)$ (Ecuación 2.58) se produce que $\Delta\nu_1 > \Delta\nu_2$, por lo que las amplitudes vinculadas a componentes de baja frecuencia serán más influyentes en la determinación de valores RMS que las de alta frecuencia.

A esta ponderación introducida en el eje de frecuencias se debe añadir la ponderación que se considerará en el eje de la amplitud. La Ecuación 3.21 será aplicada directamente cuando se desee aplicar en conjunto con las curvas de ponderación A, C, B, D o Z. Para el caso del cálculo de sonoridad, la suma energética se hace determinando la Sonoridad Específica considerando que $z \approx z_\nu = \nu \cdot z_s / \pi$ (Ecuación 3.5).

3.4. Comparación entre modelos de Sonoridad: ISO 532B y Método Propuesto

El análisis comparativo entre el modelo de sonoridad propuesto en la norma ISO 532:1975-B (basado en un análisis en bandas de 1/3 de octava) y el método propuesto (basado en un análisis aproximado en bandas críticas) permitió el

desprendimiento de una publicación científica a fines de 2009 [27]¹⁰. En este trabajo, se mostró cómo la metodología propuesta arroja resultados comparables respecto de los otros métodos, detectando existosamente las regiones de frecuencia en donde existe un mayor o menor énfasis de amplitud de una señal de audio. La comparación fue hecha considerando una condición de campo libre.

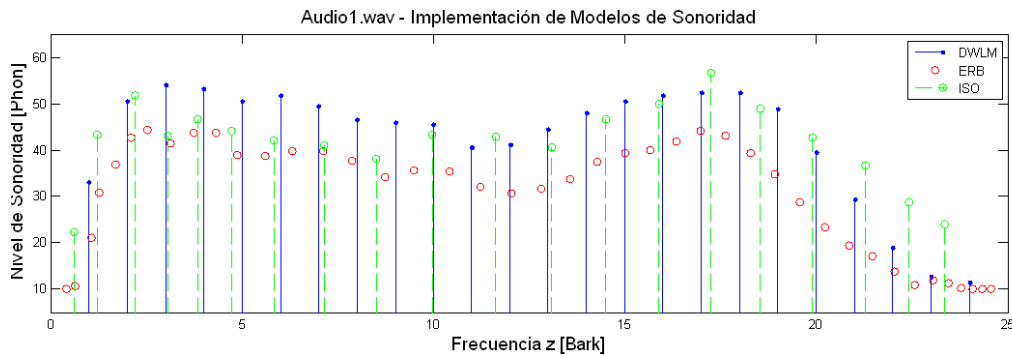


Figura 3.10: En azul se presenta el método propuesto, en rojo la implementación del modelo coclear de Patterson-Holdsworth y en verde el Método B (ISO 532:1975). Todos los modelos muestran que el oído humano presenta una mayor sensibilidad a este fragmento de audio en las frecuencias de 3 a 4 Barks (300 a 400 Hz) y de 17 a 18 Barks (en torno a 4 kHz).

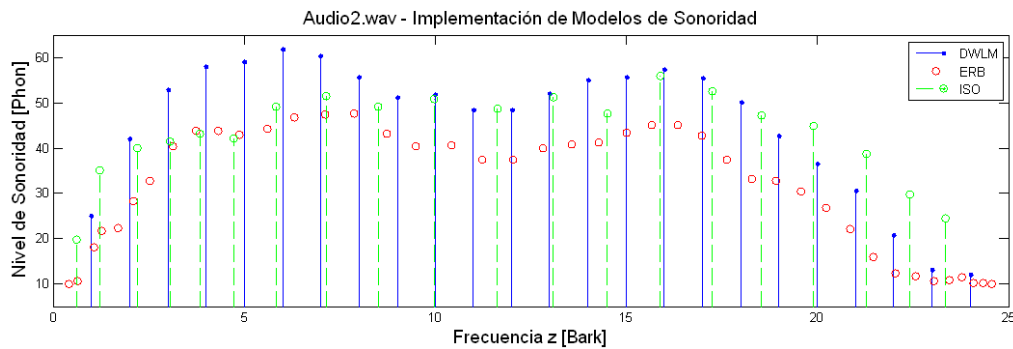


Figura 3.11: En azul se presenta el método propuesto, en rojo la implementación del modelo coclear de Patterson-Holdsworth y en verde el Método B (ISO 532:1975). Todos los modelos muestran que el oído humano presenta una mayor sensibilidad a este fragmento de audio en las frecuencias de 6 a 7 Barks (600 a 900 Hz) y de 16 a 17 Barks (en torno a 3,5 kHz).

¹⁰En este trabajo, es comparado también el modelo auditivo de Patterson-Holdsworth que se basa en un banco de filtros ERB, el cual corresponde a un modelo en bandas críticas ampliamente aceptado actualmente.

En las Figuras 3.10 y 3.11, puede observarse que el modelo propuesto, presentado en azul como DWLM¹¹, entrega una mayor ponderación a las bajas frecuencias que el Método B, presentado en verde, esencialmente en el rango de frecuencias de 3 a 7 Barks (300 a 900 Hz). En cuanto a la suma de sonoridad utilizando ambos métodos, para el fragmento Audio1 se obtuvo 33,1 Sones usando el DWLM y 22,6 Sones (GF) usando el Método B, mientras que para Audio2 los valores respectivos son 41,1 Sones y 25,1 Sones (GF).

La gran diferencia en los valores de sonoridad obtenidos entre ambos métodos (que bordean los 6 Fones), puede atribuirse a que el método propuesto considera los nuevos contornos de igual sonoridad, los cuales presentan grandes diferencias en la región de baja frecuencia respecto a las curvas de Robinson-Dadson, ampliamente aceptadas por más de 4 décadas, llegando a diferencias de hasta 14 dB. Además, el ancho de banda obtenido al usar la técnica de *Warping* durante el proceso de diseño de las bandas críticas, tiende a ser mayor que los valores experimentales publicados en [40] (de 1 a 1,2 Barks).

Pese a lo anterior, el método propuesto permite detectar en qué bandas críticas existe una mayor concentración de energías y, en este aspecto, no difiere de manera considerable con los resultados arrojados por la ISO 532-B.

3.5. Prototipo para el Analizador de Espectro Psicoacústico en Tiempo Real

Se elaboró un prototipo que permite analizar las componentes de frecuencia de una señal de audio en tiempo real con la posibilidad de ponderar parámetros psicoacústicos de amplitud y frecuencia. El prototipo ha sido construido a partir de la interacción de los conceptos revisados en las secciones precedentes. En la Figura 3.12 se muestra el diagrama en bloques del sistema a ser implementado en el plug-in VST, el cual fue diseñado tanto en MATLAB® como en Microsoft Visual Studio 2008.

¹¹Modelo de Sonoridad de Ponderación Dinámica: (*Dynamic Weighting Loudness Model*).

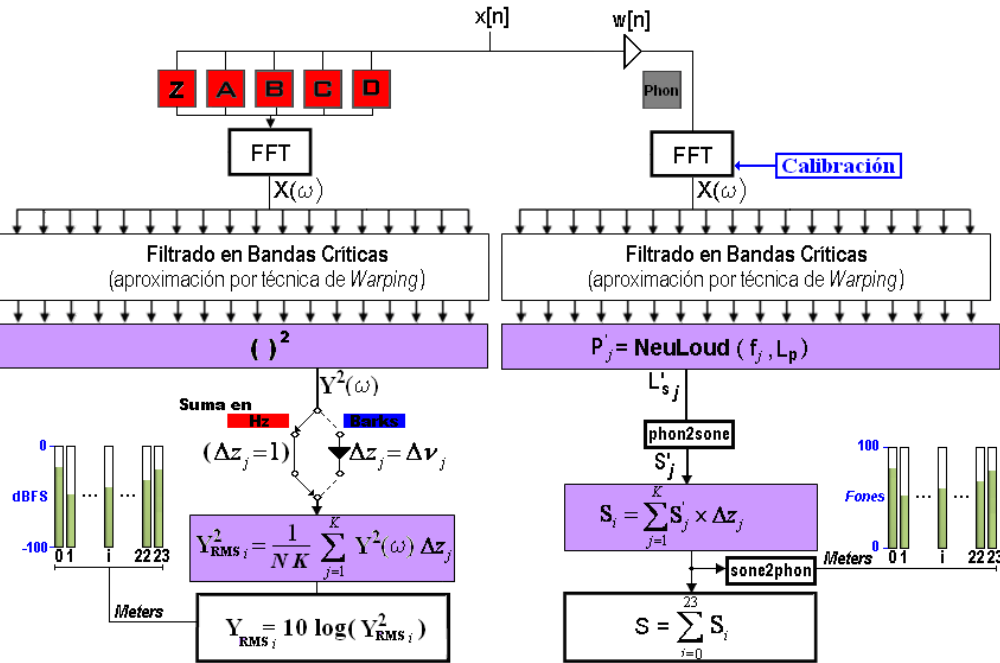


Figura 3.12: Prototipo a ser implementado en un Plug-in de audio VST para el análisis de señales de audio considerando parámetros psicoacústicos.

La señal filtrada en el bloque de *Bandas Críticas* de la Figura 3.12 se determina mediante la multiplicación entre las respuestas en frecuencia de la señal de entrada $x[n]$ y la de cada pasa banda, es decir, $Y_i(\omega) = X(\omega) \cdot H_{i_{RBF}}(\omega)$, donde el subíndice RBF indica que la respuesta en frecuencia de la banda i es interpolada usando Redes Neuronales Artificiales de Funciones Radial Base, RBFN.

La sección de análisis mediante el cálculo de sonoridad está inspirada directamente en el modelo de ponderación dinámica presentado en la sección 3.1.2. La ventana temporal $w[n]$ incluida en el primer *release* es del tipo rectangular.

Cuando se usan las curvas de ponderación en frecuencia, se incluye la posibilidad de efectuar la suma energética en dominio lineal (suma energética tradicional) o con una ponderación psicoacústica del eje de frecuencias.

Tal como se muestra en la Figura 3.12, se incluyó una etapa de calibración cuando se utiliza la ponderación de amplitud en Fones, debido a que el rótulo

presentado en la norma ISO 226:2003 [19] para la obtención del nivel de sonoridad P , en Fones, se aplica sobre un Nivel de Presión Sonora medido en (dB re. $2 \mu\text{Pa}$). En el caso de las curvas de ponderación A, B, C y D, independiente del nivel de presión sonora asociado a las respectivas amplitudes de una señal $x[n]$, se le aplica una atenuación fija que depende sólo de la frecuencia f , en Hz, asociada a cada componente de frecuencia del espectro $X(\omega)$ ¹²

3.5.1. Implementación en un plug-in VST

Durante la fase de diseño del prototipo del analizador de espectro psicoacústico se trabajó mayormente en MATLAB®, sin embargo, nuestro objetivo es concretar un análisis en tiempo real, por lo que en esta segunda fase se escogió implementar un plug-in de audio VST en su versión 2.4, sobre la base del paquete de desarrollo de software, SDK, producido por **Steinberg Technologies**.

La programación de plug-ins VST ha sido estudiada en diversas tesis [6], [2], [22], por lo que se recomienda su revisión. Por este motivo, sólo se revisarán conceptos generales vinculados principalmente a la elaboración de interfaces gráficas.

La programación de plug-ins VST, en general, se efectúa sobre 2 líneas paralelas: implementación del algoritmo de procesamiento e implementación de una interfaz gráfica (GUI).

El prototipo final implementado para el análisis en bandas críticas propuesto cuenta con aproximadamente 1.500 líneas de programación: 1.000 para el algoritmo de análisis (implementación) y 500 para la elaboración de la interfaz gráfica. La sesión de trabajo de Visual Studio 2008 puede ser encontrada en el CD anexo a esta investigación.

¹²Debemos tener en consideración que el objetivo de esta investigación se centra en el análisis de la distribución energética en bandas de frecuencia de una señal de audio determinada. Al convertir dBFS en dB re. $20 \mu\text{Pa}$ sólo se efectúa un cambio de escala de los valores a ser mostrados en los *VU Meters*, pero mantiene constante la relación entre bandas (ΔdB).

Interfaz gráfica

La Interfaz Gráfica de Usuario, GUI¹³, es parte fundamental del plug-in de audio VST que se desea implementar ya que un análisis espectral no introduce en sí un proceso de audio, por lo que su resultado no es audible, sino que se entrega de manera visual. Si el programador no especifica una GUI, el *Host* de audio suministra una pero a base de *faders* o potenciómetros y no de *Meters* (indicadores de nivel) como es lo que se requiere en nuestro caso.

A la presencia de *Meters* debe sumarse la posibilidad de escoger las ponderaciones de amplitud y de frecuencia. Estos parámetros de entrada tampoco pueden ser escogidos de manera intuitiva usando los controles de una GUI genérica. Se requiere trabajar, entonces, una GUI personalizada. La interfaz gráfica diseñada se muestra en la Figura 3.13.

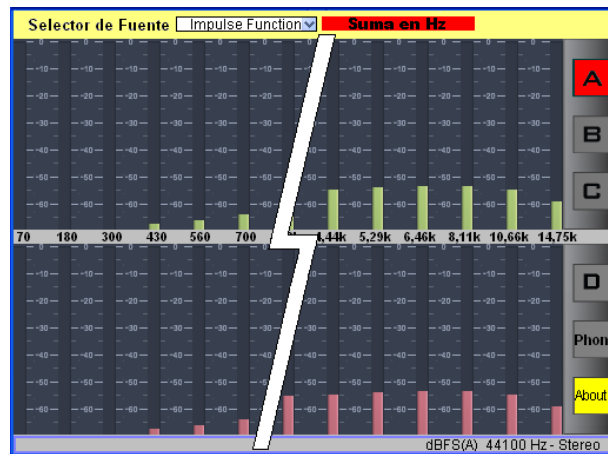


Figura 3.13: Implementación del Plug-in analizador de espectro en formato *Stereo*. Se muestran sólo algunas de las 24 bandas críticas implementadas. En la configuración mostrada, se está aplicando la curva de ponderación A y se está considerando el eje de frecuencias en Hz, por lo que este análisis entrega resultados similares a un análisis en bandas de 1/3 de octava.

Los controles incluidos en la interfaz gráfica son los siguientes: selector de fuentes, selector del tipo de ponderación de frecuencias (Suma en Bark o Suma en Hz) y Selector de las curvas de ponderación de amplitud.

¹³Del inglés: *Graphical User Interface*.

La opción **Selector de Fuentes** debe ser seteada a *Audio Inputs* para efectuar el análisis en bandas críticas de la señal de audio. Al escoger *Impulse function*, se desactiva el análisis de señales de audio, suministrándose repetitivamente un impulso unitario (respuesta en frecuencia unitaria), permitiendo la visualización de información vinculada a la respuesta al impulso de cada una de las bandas críticas implementadas.

La elección del dominio de integración se efectúa al escoger **Suma en Hz** o **Suma en Barks**. En el primero de los casos se asume un span de frecuencias equiespaciado, en Hz, mientras que en la opción **Suma en Barks**, se asume un dominio de frecuencias en Barks (aproximados usando la técnica de *warping* de frecuencias) otorgando una importancia mayor a las bajas frecuencias que a las frecuencias medias y medias altas.

En el sector derecho de la GUI del plug-in se encuentra el panel de elección de la ponderación en amplitud a utilizar. Este panel es mostrado en la Figura 3.14.

La elección de las ponderaciones a utilizar, tanto en frecuencia como en amplitud, se basan en el concepto de variables *booleanas* (Activado/Desactivado), de manera que existe una variable que controla el tipo de ponderación en frecuencia (`bWarpOn`, por omisión en *False* lo que corresponde a la opción **Suma en Hz**) y otra variable que controla la curva de ponderación en amplitud que se encuentra activa (`iCurve`, que va desde 0 a 5, donde cada número indica qué ponderación está activa. Por ejemplo `iCurve=0,1,2,3` o 4 implica que la curva Z, A, B, C o D está activa, mientras que `iCurve=5` implica que la ponderación en Fones está activa).

Un parámetro de gran importancia que también es posible de visualizar en la interfaz gráfica del plug-in es la frecuencia de muestreo f_s correspondiente a la sesión de audio que se está utilizando. Para la entrega de resultados fiables, se recomienda que este valor se encuentre por encima de $f_s = 32.000$ Hz.

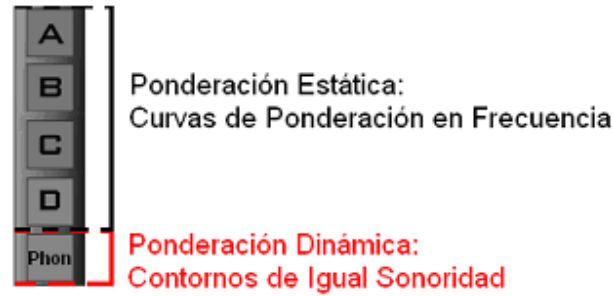


Figura 3.14: Panel para la selección de la ponderación de amplitud.

Otro concepto importante durante el diseño de la GUI del plug-in fue la utilización de coordenadas X e Y, según lo define el SDK 2.4 de Steinberg. El tamaño de la ventana en que se desplegará el plug-in es definida al crear un componente del tipo `CFrame` y se mide en píxeles. Dentro de este componente (por convención, llamado `frame`) es donde se ubicarán cada uno de los componentes de la interfaz gráfica. El origen de `frame` se encuentra en el extremo superior izquierdo, de manera que la coordenada X aumenta hacia la derecha y la coordenada Y aumenta hacia abajo.

Análisis de Desempeño en Tiempo Real

En la sección 3.3.5 se implementaron 2 plug-ins de audio VST. De acuerdo a la comparación hecha allí, se determinó que el filtrado de señales óptimo se logra aproximando las bandas de frecuencia con Redes Neuronales Artificiales de Funciones Radial Base (RBFN). Las 10 bandas que se habían implementado consumían un 2,89% de CPU. Al implementar las 24 bandas de frecuencia, el valor devuelto por el medidor *VST Performance* de Steinberg se incrementó a un 5,80% de carga de CPU.

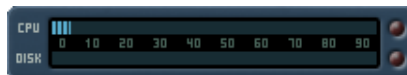


Figura 3.15: Carga de CPU de 5,80%

Tal como se mencionó en la sección 3.3.5, el porcentaje de carga de CPU

mostrado en la Figura 3.15 no corresponde a la interpolación de las bandas de frecuencia, sino a la multiplicación de cada una de las componentes de frecuencia de cada banda $H_i[k]$ que posibilitan el cálculo de la salida $Y_i[k]$. La interpolación, sin embargo, es un proceso que se lleva a cabo cuando el plug-in es cargado y cuando la frecuencia de muestreo de la sesión de audio es cambiada, debido a que en estas circunstancias la frecuencia f ajusta sus valores equiespaciadamente entre 0 y $f_s/2$, por lo que deben recalcularse las amplitudes $H_i[k]$ correspondiéndose a los nuevos valores de cada componente de frecuencia. El mencionado proceso de interpolación es un proceso que lleva menos de 3 segundos.

Con todos los antecedentes anteriores, podemos decir que el filtrado de señales implementado de esta manera es altamente eficaz, al menos para efectos del análisis de señales digitales.

Capítulo 4

ANÁLISIS Y RESULTADOS

Una vez que hemos diseñado el plug-in VST que permite efectuar un análisis psicoacústico de las señales de audio digital, debemos poner a prueba su capacidad de servir como herramienta de monitoreo a las etapas de mezcla y/o masterización de piezas musicales. Para validar la utilización del analizador de espectro psicoacústico debemos buscar y encontrar algún tipo de patrón en la lectura de valores RMS, los cuales pueden hacerse visibles al considerar algún tipo de ponderación psicoacústica. Esta premisa es parte de las hipótesis de la presente investigación.

El plug-in será utilizado en canciones pertenecientes a álbumes musicales reconocidos por su buena sonoridad. Como criterio objetivo en la elección de las piezas musicales se escogieron 7 temas dentro de los ganadores de los premios Grammy de los últimos 30 años, particularmente en la categoría “Grabación del Año” (*Record of the Year*), la cual es una distinción entregada a una única canción en un álbum que, desde el año 1999, reconoce no sólo al artista y a su productor, sino que también al ingeniero de grabación, de mezcla y/o de masterización partícipes de la producción musical.

Los 7 temas escogidos son resumidos en la Tabla 4.1. De manera adicional, se eligieron 2 temas correspondientes a artistas sudamericanos de manera de comparar nuestra realidad a la de las grandes industrias musicales (estadounidense y europea). Se escogió, además, una última canción dentro de los nominados al Grammy de “Grabación del año”. Estos 3 temas son mostrados en la Tabla 4.2.

Año	Artista	Canción
2008	Amy Winehouse	<i>Rehab</i>
2004	Norah Jones	<i>Don't know why</i>
2002	U2	<i>Walk on</i>
2001	U2	<i>Beautiful day</i>
1996	Seal	<i>Kiss from a Rose</i>
1991	Phil Collins	<i>Another day in paradise</i>
1983	Toto	<i>Rosanna</i>

Tabla 4.1: Temas seleccionados de entre los ganadores del Grammy en la categoría *Record of the Year*.

Año	Artista	Canción
2004	La Ley	<i>Ámate y sálvate</i>
2004	Divididos	<i>Dame un limón</i>
2003	Outkast	<i>Hey ya</i>

Tabla 4.2: Otros temas seleccionados. La canción *Hey ya* resultó nominada pero no ganadora en la categoría *Record of the year*. Por otro lado, los discos *Libertad* y *Vivo Acá* de donde se extrajeron respectivamente el tema de La Ley y el de Divididos fueron nominados al Grammy Latino en la categoría “Mejor Album de Rock Vocal de Dúo o Grupo”, resultando ganador el disco de La Ley.

4.1. Análisis de los archivos de audio: consideraciones

Para el análisis efectuado por parte del plug-in analizador de espectro, deben tenerse ciertas consideraciones, de manera de asegurar la repetitividad de la metodología aplicada, cuyos resultados se presentan en esta sección.

Una consideración importante y hasta el momento no tan abordada es la **calibración** de la señal de audio $x[n]$ a ser analizada. Como se mencionó en la sección 3.5, se consideró la calibración de $x[n]$ solamente cuando se utiliza una ponderación de amplitud en Fones. Por simplicidad, se asume que la señal de audio $x[n]$ posee un promedio energético igual a -12 dBFS, valor que fue ajustado a 70 dB. Debido a esta *calibración estática* es que se requiere que cualquier señal de audio a ser analizada deba poseer un valor RMS equivalente a -12 dBFS.

Otras consideraciones de carácter sistemático están relacionadas, por ejemplo, al formato de almacenamiento de las pistas de audio que serán analizadas.

Fueron extraídos 90 segundos de cada una de las canciones siendo almacenadas en formato *wav* a una frecuencia de muestreo de 44,1 kHz con una resolución de amplitud de 16 bits.

El análisis efectuado por el plug-in analizador de espectro utiliza un *buffer* de $N = 4.096$ muestras. En fases de prototipaje, se comprobó que los análisis espectrales del canal izquierdo y derecho de una misma canción entregaban información casi idéntica, de manera que al efectuar una suma energética promediada cada 5 segundos, los resultados entregados mostraban diferencias inferiores a 1 dB (para el caso de las curvas de ponderación A, B, C y D) y 2 Fones (para el caso de ponderación en Fones). Por este motivo se decidió omitir el canal derecho de cada fragmento escogido, asumiendo que éste tiene un comportamiento espectral representado por el canal izquierdo. De esta forma, al considerar 90 segundos de música, en formato monofónico, a una frecuencia de muestreo de 44.100 Hz y usando un *buffer* de 4.096 muestras, se obtienen aproximadamente 954 bloques de análisis.

Como el análisis introducido se basa en el concepto de bandas críticas, a cada uno de los 954 bloques de análisis (por cada fragmento de audio) se asociará un valor de amplitud por cada una de las 24 bandas críticas propuestas. Sin embargo, esta matriz de 954 x 24 valores está asociada a tan solo 1 descriptor acústico, el cual puede corresponder a dB(A), dB(B), dB(C), dB(D), dB(Z) o Fon. Además, debe considerarse que la suma energética realizada puede efectuar una ponderación lineal de frecuencias o bien una ponderación psicoacústica de frecuencias (ver sección 3.3.7), determinando más variantes a los descriptores anteriormente mencionados. Son 8 las combinaciones de descriptores evaluadas, las que son mostradas en la Tabla 4.3, motivo por el cual, el análisis aplicado a cada fragmento de audio es almacenado en una matriz de 954 x 24 x 8 elementos.

Ponderación de amplitud	Ponderación de Frecuencia	
	Lineal (lin)	Psicoacústica (warp)
Z	Sí	Sí
A	Sí	Sí
B	-	Sí
C	-	Sí
D	-	Sí
Fon	-	Sí

Tabla 4.3: Ponderación de amplitud y ponderación de frecuencia considerada por el plug-in analizador de espectro. Son considerados 8 descriptores acústicos, considerando las variantes Z_{lin} y Z_{warp} , A_{lin} y A_{warp} , B_{warp} , C_{warp} , D_{warp} y Fon.

4.2. Metodología para el análisis de temas musicales

La comparación entre fragmentos de audio se hará en base a los valores medios de amplitud para cada una de las 8 variantes de ponderación. En este contexto, se optó por la utilización de la moda estadística como estimador del valor medio de amplitud en cada caso de ponderación. La elección de este estimador, está basada en la gran importancia que poseen los **histogramas** en el campo del procesamiento de señales. Un histograma indica el número de ocurrencias de un evento –en nuestro caso de un rango de amplitudes– determinando la moda estadística al leer el rango de amplitud (evento) con mayor cantidad de ocurrencias. De esta forma, se obtendrá la moda estadística para cada banda de frecuencia considerando los 954 bloques de análisis. Por cada fragmento de audio y por cada ponderación considerada, se obtendrán 24 valores medios (modas).

En la Figura 4.1 se explica la obtención del valor medio de amplitud para el descriptor $\text{dBFS}(A)_{lin}$ aplicados al fragmento de audio rotulado como *Número 1* (Rehab de Amy Winehouse) para la banda crítica 9¹. Los -40 dBFS(A) obtenidos de esta lectura representan el intervalo con mayor número de ocurrencias (256 de 954 veces) y corresponde a uno de los 24 valores medios que resumirán el análisis del fragmento al usar la curva de ponderación A.

El análisis de los fragmentos de audio estará basado en el entrecruzamiento de

¹Se debe recordar que las Bandas Críticas fueron rotuladas entre 0 y 23 en la Tabla 2.3.

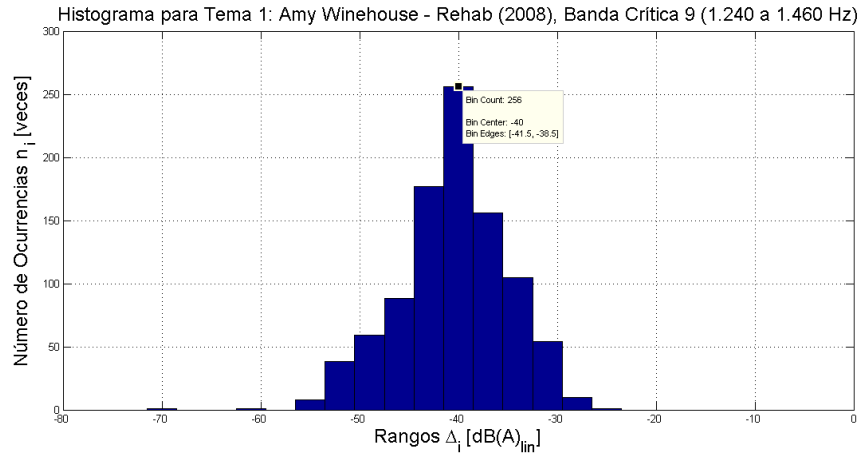


Figura 4.1: Histograma que considera un número total de 954 valores de amplitud (1 a cada valor filtrado de la canción Rehab en la banda crítica 9) y que han sido agrupados en intervalos de 3 en 3 decibeles ponderados en A, considerando una ponderación lineal de frecuencias. Se puede apreciar que la Moda estadística corresponde al intervalo de amplitud entre $-41,5$ y $-38,5 \text{ dBFS(A)}_{\text{lin}}$.

las medias de amplitud asociadas a cada una de las bandas críticas considerando cada uno de los descriptores acústicos de la Tabla 4.3, procurando su clasificación en 2 o más grupos de acuerdo a la coloración (ecualización) presentada por cada tema al usar cada uno de los 8 descriptores acústicos. Se analizará qué descriptor facilita más la discriminación de estos grupos de ecualización o balance espectral.

4.3. Análisis de los fragmentos musicales

En la Tabla 4.4 se enumeran las 10 producciones musicales que serán analizadas usando los criterios psicoacústicos abordados en esta investigación.

En primera instancia se analizó cuán plana es la respuesta en amplitud de cada fragmento. Si se consideran los 24 valores de amplitud para cada tipo de ponderación, la respuesta en frecuencia (o balance espectral) más plana es aquella que presenta una menor desviación entre sus valores. Se utilizó como estimador la desviación estándar S de los datos de amplitud.

En la Tabla 4.5 son presentadas las desviaciones estándar S asociadas a los

Tema N°	Artista	Canción
1	Amy Winehouse	<i>Rehab</i>
2	Norah Jones	<i>Don't know why</i>
3	U2	<i>Walk on</i>
4	U2	<i>Beautiful day</i>
5	Seal	<i>Kiss from a Rose</i>
6	Phil Collins	<i>Another day in paradise</i>
7	Toto	<i>Rosanna</i>
8	La Ley	<i>Ámate y sálvate</i>
9	Divididos	<i>Dame un limón</i>
10	Outkast	<i>Hey ya</i>

Tabla 4.4: Temas seleccionados. En esta oportunidad se presenta la rotulación utilizada para cada canción: tema 1 a tema 10.

10 fragmentos de 90 segundos analizados.

Tema	dB(Z)	dB(A)	dB(Z) _w	dB(A) _w	dB(B) _w	dB(C) _w	dB(D) _w	Fones
1	5,7658	5,9782	9,9170	9,4005	11,3865	11,9891	9,4446	18,4044
2	8,5274	7,2401	12,3845	11,6854	13,0161	13,7525	10,9239	16,0822
3	4,7486	4,7640	9,0856	8,4766	10,1639	10,7918	7,8036	17,4181
4	8,1230	6,5595	11,3951	8,8677	12,2554	12,9104	9,0758	17,8673
5	10,0445	7,5929	13,4692	11,8358	13,8541	14,7628	10,1951	15,5675
6	7,53146	6,9082	12,0914	10,7592	13,0461	13,7139	10,4041	15,4561
7	6,2228	4,1788	9,6605	8,4352	10,5781	11,3865	8,0585	14,7053
8	4,7365	5,4474	9,2069	9,4317	10,5720	11,3800	8,6894	19,6620
9	6,9611	6,1171	11,5909	9,5203	11,6798	12,6811	10,0736	16,5720
10	6,9270	6,8904	11,2133	10,9753	12,4423	12,7682	9,8104	20,7830
S_{prom}	6,9588	6,1677	11,0014	9,9388	11,8994	12,6136	9,4479	17,2518

Tabla 4.5: Desviación estándar S para los 10 fragmentos musicales analizados. El estimador acústico que presentó una respuesta en frecuencia más plana correspondió a los decibeles ponderados en A, dB(A), mientras que la respuesta menos plana fue registrada por el propuesta ponderación en Fones.

La ponderación que presentó una respuesta en frecuencia más plana, utilizando el criterio de búsqueda de la mínima desviación estándar S , fue la curva de ponderación A. De los estimadores que consideran una aproximación a la razón de banda crítica, el que menor desviación presentó fue la curva de ponderación D_w .

La propuesta ponderación en Fones presentó una mayor variabilidad en los

valores de amplitud de las 24 bandas, por lo que corresponde a la ponderación en amplitud menos plana de entre los descriptores acústicos evaluados.

Ante los resultados anteriores, se efectuó un entrecruzamiento de los datos en gráficos de amplitud en función de cada banda crítica, de manera de detectar visualmente un patrón de balance espectral. En este contexto, las curvas de ponderación A y D_w , que habían entregado las distribuciones más planas, no presentaron un patrón definido, mientras que la ponderación en Fones permitió la clara discriminación de 2 tendencias en el balance espectral de los fragmentos: (a) énfasis en baja frecuencia de aproximadamente 10 fones respecto de la banda crítica 10 y una desviación menor a 5 fones hasta la banda crítica 18 y (b) énfasis en baja frecuencia de aproximadamente 15 fones respecto de la banda crítica 9 y énfasis menor a 10 fones en frecuencias medias altas (banda 15 o 16). Los temas musicales clasificados en el **Grupo A** fueron los temas 1 (*Rehab*), 3 (*Walk on*), 8 (*Ámate y Sálvate*), 9 (*Dame un Limón*) y 10 (*Hey ya!*), mientras que en el **Grupo B** se clasificaron los temas 2 (*Don't know why*), 4 (*Beautiful day*), 5 (*Kiss from a Rose*), 6 (*Another day in paradise*) y 7 (*Rosanna*).

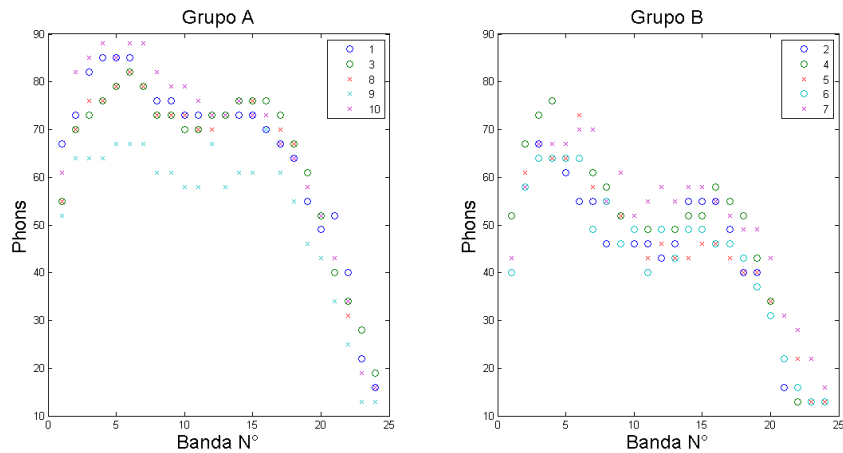


Figura 4.2: Grupo A y Grupo B. Ambas nubes de puntos responden a 2 patrones de balance espectral claramente diferenciados.

Al observar la Figura 4.2 es posible aseverar que utilizando una ponderación psicoacústica (en este caso en Fones) es posible discriminar patrones de balance

espectral que no son visibles de manera clara utilizando los descriptores acústicos convencionales.

Una observación relevante tiene relación con los años de publicación de los temas analizados. La mayor parte de las canciones pertenecientes al **Grupo A** corresponden a grabaciones efectuadas durante la década 2000-2010, mientras que los 3 temas más antiguos (1983-1996) están clasificados en el **Grupo B**.

No se descarta la aparición de más grupos de balance espectral si se analizasen más canciones y más estilos, sin embargo, ya es importante notar que es posible detectar un número finito de grupos de ecualización, lo que permitiría masterizar piezas musicales de manera automática al ser catalogadas dentro de un grupo de balance espectral (Grupo A o B, por ejemplo).

Capítulo 5

CONCLUSIONES Y DISCUSIÓN

Uno de los aportes importantes de la presente investigación es la utilización de las Redes Neuronales Artificiales del tipo Radial Base en la implementación de bancos de filtros de alta selectividad, haciendo viable el filtrado en tiempo real de señales de audio digital. Sin embargo, esta implementación no entrega ninguna información de fase, motivo por el cual, se hace útil sólo en procesos de filtraje con fines de análisis.

Luego del análisis de resultados, es posible aseverar que considerando criterios psicoacústicos la búsqueda de la respuesta más plana no se aproxima al proceso efectuado por los profesionales del *mastering* durante el proceso de producción musical. Pese a lo anterior, fue posible detectar balances de amplitud en función de las bandas críticas (patrones de ecualización) al usar una ponderación de amplitud en Fones, en vez de usar las convencionales curvas de ponderación A y C y las derogadas curvas B y D. En los fragmentos de audio analizados se pudo visualizar una tendencia de coloración en la región de baja frecuencia, con un énfasis de entre 10 y 15 fones, y en las frecuencias medias altas, con un énfasis menor a 10 fones en torno a la banda crítica 17 (4 kHz). El rango de frecuencias de las bandas 10 y 11 (en torno a 1,4 kHz) son las que presentan un valor mínimo de amplitud.

Respecto de las técnicas utilizadas en la realización de este análisis psicoacústico, la aplicación de la técnica de *warping* no es apta para efectuar un análisis en

tiempo real de datos en el rango completo de frecuencias (20-12.500 Hz¹), pero sí resulta apropiada para una aproximación a la razón de banda crítica z , en Barks. La técnica de *warping* puede aplicarse en tiempo real sólo consiguiendo fidelidad en un rango de frecuencias por debajo de los 4 kHz, lo que no nos resulta de utilidad para los alcances de la presente investigación.

Considerando la efectividad en la aproximación a la razón de banda crítica, se buscó aprovechar teóricamente la técnica de *Warping* de frecuencias, a través de las propiedades de la transformación bilineal asociada, de manera de poder efectuar una suma energética en bandas críticas, usando un eje lineal de frecuencias, pero determinando un Δz_w variable, introduciendo una mayor resolución en baja frecuencia.

Dentro de los tópicos relacionados al tema en desarrollo que no fueron fuertemente explorados están la generalización del análisis propuesto corrigiendo la condición de campo libre, para lograr una condición de campo difuso, la cual se aproxima más al ambiente de reproducción musical que un auditor destina usualmente para estos fines.

Un gran problema en la consideración de parámetros psicoacústicos es la dificultad para establecer normativas internacionales que objetivicen el procedimiento de medición de sonoridad. La referenciada norma ISO 532 –aún vigente– contiene métodos antiguos que consideran los contornos de igual sonoridad obsoletos, además de establecer su aplicación para señales acústicas de características no tonales (ruidos). Con los años otros modelos se han formulado, que han superado las limitaciones principalmente del Método B, pero ellos no han sido propuestos en la elaboración de nuevas normativas que evalúen la sensación de sonoridad. Incluso han aparecido nuevos modelos de bandas críticas como lo son los Filtros de Área Rectangular Equivalente (ERB), los cuales desestiman la caracterización del oído humano en 24 bandas críticas. Pese a esta situación, se

¹El rango de frecuencias entre 20 y 12.500 Hz corresponde al rango de frecuencias en que están definidos los contornos de igual sonoridad (ISO 226:2003), por lo que se considera este rango de frecuencias como el rango completo de análisis.

mostró que de igual manera es posible obtener información relevante a partir de piezas musicales utilizando estimadores acústicos alternativos a las populares curvas de ponderación en frecuencia.

Bibliografía

- [1] R. Aarts. A comparison of some loudness measures for loudspeaker listening tests. *Audio Engineering Society*, 40(3):60–74, Marzo 1992. [cita en pag. 3]
- [2] J. P. Bello. Diseño e implementación de un plug-in vst multicanal (5.1) que incluya las etapas de downmix, bass management y meters. Universidad Tecnológica de Chile, 2007. [cita en pag. 80]
- [3] R. Bladon and B. Lindblom. Modeling the judgment of vowel quality differences. *Journal of the Acoustical Society of America*, 69(5):1414–1422, Mayo 1981. [cita en pag. 15]
- [4] R. Cassidy and J. Smith. Auditory filter bank lab, 2008. [cita en pag. 4]
- [5] R. Churchill, J. Brown, and R. Verhey. *Complex variables and applications*. Editorial McGraw-hill, 3 edition, 1960. [cita en pag. 100]
- [6] C. Espinoza. Diseño e implementación de plug-ins de audio vst. Universidad Tecnológica de Chile, 2003. [cita en pag. 80]
- [7] V. Espinoza and A. Osses. Ponderación psicoacústica de la distorsión no lineal de dispositivos de audio: una propuesta para su medición. Primeras Jornadas Regionales de Acústica AdAA, Rosario, Argentina, Noviembre 2009. [cita en pag. 57]
- [8] V. Espinoza, R. Venegas, and S. Floody. Modelo de sonoridad usando redes neuronales artificiales. V Congreso Iberoamericano de Acústica FIA, Santiago, Chile, Octubre 2006. [cita en pag. 19, 54, 56, 59]
- [9] H. Fastl and E. Zwicker. *Psychoacoustics: Facts & models*. Editorial Springer, 3 edition, 2007. [cita en pag. 12, 22, 74, 116]
- [10] H. Fletcher and W. Munson. Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America*, 5:82–108, Octubre 1933. [cita en pag. 2]

- [11] D. D. Greenwood. A cochlear frequency mapping-position function for several species—29 years later. *Journal of the Acoustical Society of America*, 87:2592–2605, Junio 1990. [cita en pag. 10]
- [12] K. Gurney. *An introduction to neural networks*. UCL Press, 1997. [cita en pag. 45]
- [13] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U. K. Laine, and J. Huopaniemi. Frequency-warped signal processing for audio applications. *Journal of the Audio Engineering Society*, 48:1011–31, Noviembre 2000. [cita en pag. 41]
- [14] International Electrotechnical Commission, IEC. *IEC 60537. Frequency weighting for the measurement of aircraft noise (D-weighting)*, 1976. [cita en pag. 19]
- [15] International Electrotechnical Commission, IEC. *IEC 60651. Sound Level Meters*, 1979. [cita en pag. 18, 122]
- [16] International Electrotechnical Commission, IEC. *IEC 61260: Electroacoustics. Octave-band and Fractional-Octave-Band Filters*, 1995. [cita en pag. 5, 7, 122]
- [17] International Electrotechnical Commission, IEC. *IEC 61672-1: Electroacoustics. Sound Level Meters*, 2002. [cita en pag. 3, 18, 122]
- [18] International Organization for Standardization, ISO. *ISO 532:1975. Acoustics. Method for calculating loudness level*, 1975. [cita en pag. 24]
- [19] International Organization for Standardization, ISO. *ISO 226:2003. Acoustics. Normal equal-loudness-level contours*, 2003. [cita en pag. 15, 16, 20, 56, 80]
- [20] M. Kahrs and K. Brandenburg. *Applications of Digital Signal Processing to Audio and Acoustics*. Kluwer Academic Publishers, 2002. [cita en pag. 13, 14, 116]
- [21] B. Katz. Integrated approach to metering, monitoring, and leveling practices, part 1: Two-channel metering. *J. Audio Eng. Soc*, 48(9):800–809, 2000. [cita en pag. 55]
- [22] I. Klepac and O. Muñoz. Programación y aplicación de plug-in vst multimeter. Universidad Tecnológica de Chile, 2007. [cita en pag. 80]
- [23] S. Mitra. *Digital Signal Processing. A computer based approach*. Editorial McGraw-hill, 2 edition, 2001. [cita en pag. 32]
- [24] T. Oliveira e Silva. Laguerre filters, an introduction. *Revista do Detua*, 1(3), Enero 1995. [cita en pag. 44]

- [25] A. Oppenheim. *Discrete-Time Signal Processing*. Editorial Prentice Hall, 2 edition, 1999. [cita en pag. 32, 33, 35, 37]
- [26] M. Orr. Introduction to radial basis function networks. University of Edinburgh, 1996. [cita en pag. 49, 119]
- [27] A. Osses and V. Espinoza. Análisis comparativo entre modelos de sonoridad: Propuesta de un modelo de ponderación dinámica. Primeras Jornadas Regionales de Acústica AdAA, Rosario, Argentina, Noviembre 2009. [cita en pag. 58, 77, 119]
- [28] A. Osses and V. Espinoza. Análisis espectral en bandas críticas usando la técnica de warping de frecuencias. 7 Congresso de Áudio da AES Brasil, São Paulo, Brasil, Mayo 2009. [cita en pag. 65]
- [29] A. Osses, R. Retamal, and H. Ángel. Análisis mulibanda en dominio de la frecuencia usando redes neuronales artificiales. 7 Conferencia AES Chile, Santiago, Chile, Noviembre 2009. [cita en pag. 70]
- [30] D. Robinson and R. Dadson. A re-determination of the equal-loudness relations for pure tones. *Br. J. Appl. Phys.*, 7:166–181, 1956. [cita en pag. 17, 117]
- [31] S. Samper and L. Herrera. Optimización de sistemas de audio profesional de acuerdo a factores psicoacústicos. Universidad Pérez Rosales, 2003. [cita en pag. 24]
- [32] B. Shenoi. *Introduction to Signal Processing and Filter Design*. John Wiley & Sons Inc., 2006. [cita en pag. 36, 104]
- [33] J. Smith and J. S. Abel. Bark and erb bilinear transforms. *IEEE Tr. Speech and Audio Processing*, pages 697–708, Noviembre 1999. [cita en pag. 43]
- [34] S. Smith. *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Publishing, 1997-1998. [cita en pag. 29, 30, 38]
- [35] Y. Suzuki and H. Takeshima. Equal-loudness-level contours for pure tones. *Journal of the Acoustical Society of America*, 116(2):918–933, 2004. [cita en pag. 16, 117]
- [36] J. Tebelskis. *Speech recognition using Neural Networks*. PhD thesis, Carnegie Mellon University, 1995. [cita en pag. 45, 118]
- [37] E. Terhardt. Calculating virtual pitch. *Hearing Research*, 1:155–182, 1979. [cita en pag. 13]

- [38] E. Terhardt, G. Stoll, and M. Sweewann. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America*, 71(3):679–688, 1981. [cita en pag. 13]
- [39] R. Venegas. Diseño e implementación de un modelo de localización sonora espacial utilizando técnicas de inteligencia computacional. Universidad Tecnológica de Chile, 2006. [cita en pag. 44]
- [40] E. Zwicker. Subdivision of the audible frequency range into critical bands. *Journal of the Acoustical Society of America*, 33(2):1961, 1961. [cita en pag. 11, 14, 78]
- [41] E. Zwicker, H. Fastl, U. Widmann, K. Hurakata, S. Kuwano, and S. Namba. Program for calculating loudness according to din 45631 (iso 532b). *Journal of the Acoustical Society of Japan*, 12(1):39–42, 1991. [cita en pag. 28]
- [42] E. Zwicker and E. Terhardt. Analytical expressions for critical band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America*, 68(6):1523–1525, 1980. [cita en pag. 11]

Apéndice A

TRANSFORMACIÓN BILINEAL

La técnica de Transformación Bilineal, también conocida como transformación lineal fraccionaria, permite el mapeamiento o conversión de funciones pasando de un dominio asociado a una variable independiente z a otro dominio definido por una variable independiente w [5].

$$w = T(z) = \frac{az + b}{cz + d} \quad (\text{A.1})$$

El mapeamiento está definido por las constantes complejas (generalmente reales) a , b , c y d que deben cumplir $(ad - bc) \neq 0$. Si efectuamos la multiplicación cruzada en la Ecuación A.1 y agrupamos todos los términos algebraicos al lado izquierdo de la ecuación, se obtiene una expresión de la forma:

$$Awz + Bw + Cz + D = 0 \quad (\text{A.2})$$

Lo cual es lineal en w y lineal en z o *bilineal* en w y z .

Si la Ecuación es desarrollada para la variable independiente z se obtiene:

$$z = \frac{-dw + b}{cw - a} \quad (\text{A.3})$$

Esto puede interpretarse como que la transformación lineal fraccionaria de la Ecuación A.1 representa un mapeamiento uno-a-uno entre los dominios de las variables z y w . Si cambiamos arbitrariamente las variables de la Ecuación A.3, obtendremos la expresión para la transformación bilineal inversa $T^{-1}(z)$:

$$T^{-1}(z) = \frac{-dz + b}{cz - a} \quad (\text{A.4})$$

Esto quiere decir que podremos pasar tanto del dominio z al dominio w , como del dominio w al dominio z .

A.1. *Warping* de frecuencias

La técnica de *Warping* de frecuencias consiste en un mapeamiento entre dominios digitales, introduciendo una distorsión *bucleada* (o curvada) al eje de frecuencias. Esta distorsión resulta del reemplazo de cada retardo unitario z^{-1} de una estructura digital por un filtro IIR de orden 1 determinado por $D(z^{-1})$.

$$D(z^{-1}) = \frac{-\lambda + z^{-1}}{1 - \lambda z^{-1}} \quad (\text{A.5})$$

La técnica de *Warping* es un caso particular de la transformación bilineal con parámetros $a = d = -\lambda$ y $b = c = 1$. La transformación bilineal inversa $D^{-1}(z^{-1})$ se obtiene al intercambiar a y d con los signos contrarios, tal como se muestra en la Ecuación A.6.

$$D^{-1}(z^{-1}) = \frac{\lambda + z^{-1}}{1 + \lambda z^{-1}} \quad (\text{A.6})$$

Tal como se presentó en la sección 2.3.3, la distorsión entre las frecuencias digitales ω y ν está dada por:

$$\nu(\omega) = \omega + 2 \cdot \arctan \left(\frac{\lambda \sin(\omega)}{1 - \lambda \cos(\omega)} \right) \quad (\text{A.7})$$

Sin embargo, como consecuencia de la transformación inversa de la Expresión A.6, también se puede aseverar que:

$$\omega(\nu) = \nu + 2 \cdot \arctan \left(\frac{-\lambda \sin(\nu)}{1 + \lambda \cos(\nu)} \right) \quad (\text{A.8})$$

A.2. Mapeamiento del dominio analógico al dominio digital

Tal como se mencionó en la sección 2.3.2, para utilizar las técnicas de diseño de filtros en dominio analógico y aplicarlas al dominio digital se requiere efectuar una transformación bilineal a la función de transferencia analógica $H(s)$ de la Ecuación A.9.

$$H(s) = \frac{b_n s^n + b_{n-1} s^{n-1} + \dots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0} \quad (\text{A.9})$$

La transformación se efectúa al reemplazar la variable independiente s por la expresión de la Ecuación A.10.

$$s = \frac{2}{T} \cdot \left(\frac{1 - z^{-1}}{1 + z^{-1}} \right) \quad (\text{A.10})$$

Donde T es el período de muestreo de una señal analógica dado por $1/f_s$ [s].

El reemplazo a efectuar está dado, entonces, por:

$$H(z^{-1}) = H(s) \Big|_{s=\frac{2}{T} \left(\frac{1-z^{-1}}{1+z^{-1}} \right)} \quad (\text{A.11})$$

La función de transferencia digital $H(z^{-1})$ representa un sistema de respuesta impulsiva infinita (IIR).

Si despejamos la variable z de la Ecuación A.10, se obtiene:

$$z = \frac{1 + (T/2)s}{1 - (T/2)s} \quad (\text{A.12})$$

Para conocer la respuesta en frecuencia de un sistema analógico, comúnmente se reemplaza el parámetro s por $j\Omega$. Además, si se determina el módulo de la expresión evaluada, se puede inferir el rango de valores que pueden adquirir las frecuencias del nuevo dominio digital z .

$$z = \frac{1 + j\Omega T/2}{1 - j\Omega T/2} \quad (\text{A.13})$$

$$|z| = \frac{\sqrt{1^2 + (\Omega T/2)^2}}{\sqrt{1^2 + (-\Omega T/2)^2}} = 1 \quad (\text{A.14})$$

La Ecuación A.14 muestra que la transformación bilineal de la Ecuación A.10 realiza, efectivamente, un mapeamiento de la frecuencia analógica Ω dispuesta en el eje imaginario (eje vertical) del plano s hacia una circunferencia de radio unitario en el plano z , permitiendo una interpretación digital del nuevo dominio de trabajo.

Para analizar la relación exacta entre la frecuencia angular analógica Ω y la frecuencia digital ω , debe reemplazarse la variable z^{-1} por $\exp(-j\omega)$ en la Ecuación A.10:

$$s = \frac{2}{T} \cdot \left(\frac{1 - e^{-j\omega}}{1 + e^{-j\omega}} \right) \quad (\text{A.15})$$

Si multiplicamos convenientemente en el numerador y denominador por $e^{j\omega/2}$:

$$\begin{aligned} s &= \frac{2}{T} \cdot \left(\frac{e^{j\omega/2}(1 - e^{-j\omega})}{e^{j\omega/2}(1 + e^{-j\omega})} \right) \\ s &= \frac{2}{T} \cdot \left(\frac{e^{j\omega/2} - e^{-j\omega/2}}{e^{j\omega/2} + e^{-j\omega/2}} \right) \end{aligned} \quad (\text{A.16})$$

Recordando las identidades trigonométricas de Euler:

$$\begin{aligned} \sin(\alpha) &= (e^{j\alpha} - e^{-j\alpha})/(2j) \\ \cos(\alpha) &= (e^{j\alpha} + e^{-j\alpha})/2 \end{aligned}$$

Es posible observar que el numerador de la Ecuación A.16 está asociado a la función seno del ángulo $\omega/2$, mientras que el denominador de la misma expresión se asocia a la función coseno.

$$s = \frac{2}{T} \cdot \left(\frac{2j \sin\left(\frac{\omega}{2}\right)}{2 \cos\left(\frac{\omega}{2}\right)} \right) \quad (\text{A.17})$$

Si ahora reemplazamos s por $j\Omega$ y ordenamos la expresión se obtiene la expresión final que relaciona las frecuencias Ω y ω :

$$\Omega = \frac{2}{T} \cdot \tan(\omega/2) \quad (\text{A.18})$$

A.2.1. Conversión de filtros pasa bajos a filtros pasa altos

En la literatura asociada a las técnicas de diseño de filtros (tanto digitales como analógicos) generalmente se trabajan los filtros pasa bajos (LPF), debido a que a partir de ellos se pueden diseñar los filtros pasa altos (HPF), pasa banda y rechaza banda. Lo que se requiere, es aplicar una transformación bilineal apropiada a la función de transferencia de un LPF ya diseñado, de manera de obtener una nueva función $H(z^{-1})$.

En esta sección sólo se abordarán los detalles de la transformación bilineal que permite pasar del diseño de un LPF al de un HPF. Según [32] esta transformación está dada por las Ecuaciones A.19 y A.20.

$$z^{-1} \rightarrow -\frac{a + z^{-1}}{1 + az^{-1}} \quad (\text{A.19})$$

Con:

$$a = -\frac{\cos [(\omega_p + \omega'_p)/2]}{\cos [(\omega_p - \omega'_p)/2]} \quad (\text{A.20})$$

Donde ω_p corresponde a la frecuencia de corte del LPF y ω'_p la frecuencia deseada para el HPF, ambas expresadas en [rad/muestra].

Pese a que las frecuencias ω_p y ω'_p pueden corresponder a cualquier frecuencia digital entre 0 y π , vamos a revisar qué ocurre si existe una relación particular de dependencia entre ellas: $\omega'_p = \pi - \omega_p$. Si esta relación es considerada en la Ecuación A.20:

$$a = -\frac{\cos [(\omega_p + (\pi - \omega_p))/2]}{\cos [(\omega_p - (\pi - \omega_p))/2]} = -\frac{\cos [\pi/2]}{\cos [\omega_p]} = 0$$

Es decir, que si se considera la conversión de un filtro pasa bajos a un filtro pasa altos cuyas frecuencias de corte están relacionadas por $\omega'_p = \pi - \omega_p$, hace que el parámetro a de la Expresión A.19 se anule. Esto simplifica considerablemente la transformación bilineal requerida, reduciéndola a:

$$z^{-1} \rightarrow -z^{-1} \quad (\text{A.21})$$

Apéndice B

DISEÑO DE FILTROS PASA BANDA

En esta sección se diseñará un filtro pasabanda a partir de un filtro pasa bajos (LPF) y un filtro pasa altos en cascada (HPF).

Se desea obtener un filtro pasa banda de respuesta impulsiva infinita (IIR) con banda de paso entre 2.000 y 3.000 Hz. Para el filtro pasa bajos (LPF) considere el sistema RC presentado en la sección 2.3.1. Además, se requiere una atenuación de al menos 15 dB para las frecuencias inferiores a 1.500 Hz.

B.1. Filtro Pasa bajos: Transformación Bilineal de un Sistema Analógico $H(s)$

La función de transferencia de un sistema RC es conocida y se comporta como un filtro pasa bajos. Si recordamos su función de transferencia:

$$H(s) = \frac{1}{1 + RCs} \quad (\text{B.1})$$

Donde R y C son los valores de resistencia y capacitancia de los elementos del circuito RC, que determinan la frecuencia de corte del sistema, tal como se manifiesta en la Ecuación B.2.

$$f_c = \frac{1}{2\pi RC} \quad (\text{B.2})$$

Si usamos una resistencia y un condensador con valores $R = 50\Omega$ y $C = 1 \cdot 10^{-6}F$, se obtiene una frecuencia de corte f_c de 3.180 Hz (aproximadamente

3.000 Hz). Observemos que un circuito RC corresponde a un sistema con un polo único (*Single Pole*) y es, por tanto, un sistema de orden 1.

La función de transferencia digital $H(z^{-1})$ se obtiene aplicando la transformación bilineal de la Ecuación A.10. En este caso no podemos omitir la constante T , ya que el sistema RC no fue diseñado a partir de especificaciones en dominio digital de frecuencias. La constante T debe asociarse a la frecuencia de muestreo como: $T = 1/f_s$. Se considerará una frecuencia de muestreo $f_s = 44.100$ Hz. Si efectuamos la transformación bilineal:

$$H(z^{-1}) = \frac{1}{1 + RC \frac{2}{T} \left(\frac{1-z^{-1}}{1+z^{-1}} \right)}$$

Desarrollando se obtiene:

$$H(z^{-1}) = \frac{T + Tz^{-1}}{(T + 2RC) + (T - 2RC)z^{-1}} \quad (\text{B.3})$$

Si evaluamos los valores numéricos de nuestro caso:

$$H(z^{-1}) = \frac{2,27 \cdot 10^{-5} + 2,27 \cdot 10^{-5} z^{-1}}{(2,27 \cdot 10^{-5} + 10 \cdot 10^{-5}) + (2,27 \cdot 10^{-5} - 10 \cdot 10^{-5}) z^{-1}}$$

Finalmente, simplificamos, obteniéndose la función de transferencia de la Ecuación B.4:

$$H(z^{-1}) = \frac{2,27 + 2,27z^{-1}}{12,27 - 7,73z^{-1}} \quad (\text{B.4})$$

B.2. Filtro Pasa altos: Diseño de un Filtro Butterworth

Se diseñará un filtro pasa altos Butterworth con frecuencia de corte $f_p = 2.000$ Hz y frecuencia límite de la zona de rechaza banda es de $f_r = 1.500$ Hz. Además, se debe asegurar una atenuación mínima de 15 dB en toda la banda de rechazo. Se considerará una frecuencia de muestreo f_s igual a 44.100 Hz.

El primer paso es convertir las especificaciones al dominio digital de frecuencias usando: $\omega = 2\pi f/f_s$. De esta forma:

$$\omega_c = 2\pi \frac{2.000}{44.100} = 0,091\pi \quad [\text{rad/muestras}] \quad (\text{B.5})$$

$$\omega_r = 2\pi \frac{1.500}{44.100} = 0,068\pi \quad [\text{rad/muestras}] \quad (\text{B.6})$$

El diseño del filtro pasa altos que posee los parámetros anteriores se efectúa partiendo de un filtro pasa bajos (LPF) con parámetros *complementarios* de manera tal que:

$$\nu_c = \pi - \omega_c = \pi - 0,091\pi = 0,909\pi \quad [\text{rad/muestras}] \quad (\text{B.7})$$

$$\nu_r = \pi - \omega_r = \pi - 0,068\pi = 0,932\pi \quad [\text{rad/muestras}] \quad (\text{B.8})$$

Para el diseño del LPF con frecuencia de corte ν_c debe aplicarse el procedimiento descrito en la sección 2.3.2. Para conseguir ello, primero debemos obtener las frecuencias pre-warpeadas Ω_c y Ω_r asociadas que, como consecuencia de la transformación bilineal son:

$$\Omega_c = 2 \tan\left(\frac{\nu_c}{2}\right) = 2 \tan\left(\frac{0,909\pi}{2}\right) = 14,05 \quad [\text{rad/s}] \quad (\text{B.9})$$

$$\Omega_r = 2 \tan\left(\frac{\nu_r}{2}\right) = 2 \tan\left(\frac{0,932\pi}{2}\right) = 18,65 \quad [\text{rad/s}] \quad (\text{B.10})$$

Todos los parámetros calculados y que se considerarán en el diseño del HPF requerido se resumen en la Tabla B.1.

	f [Hz]	ω [rad/muestra]	ν [rad/muestra]	Ω [rad/s]
Frecuencia de corte	2.000	0,091 π	0,909 π	14,05
Frecuencia de rechazo	1.500	0,068 π	0,932 π	18,65

Tabla B.1: Parámetros de diseño para un HPF con corte en 2.000 Hz y banda de rechazo por debajo de los 1.500 Hz. El LPF asociado posee una frecuencia de corte de 20.050 Hz e inicio de la zona de rechazo en 20.550 Hz. Las frecuencias ν y Ω se muestran en las 2 últimas columnas.

Calculamos el orden del filtro pasa bajos en función de las frecuencias pre-warpeadas Ω_p y Ω_r :

$$N = \frac{\log_{10}(10^{(A/10)} - 1)}{2 \log_{10}\left(\frac{\Omega_r}{\Omega_p}\right)} = \frac{\log_{10}(10^{(A/10)} - 1)}{2 \log_{10}\left(\frac{18,65}{14,05}\right)} = 6,012 \approx 6 \quad (\text{B.11})$$

El orden N obtenido es aproximadamente igual a 6. Ahora se buscarán los polos asociados a un LPF con corte en 1 [rad/s]. La expresión genérica para determinar los polos de un LPF Butterworth de orden N es:

$$p_k = \exp\left[j \frac{2k + N - 1}{2N} \pi\right] \quad (\text{B.12})$$

Donde k es un entero que va desde 1 a N . Al aplicar esta ecuación usando $N = 6$ se obtienen los polos resumidos en la Tabla B.2.

k	p_k	k	p_k
1	$\exp(j \frac{7}{12} \pi)$	4	$\exp(j \frac{13}{12} \pi)$
2	$\exp(j \frac{9}{12} \pi)$	5	$\exp(j \frac{15}{12} \pi)$
3	$\exp(j \frac{11}{12} \pi)$	6	$\exp(j \frac{17}{12} \pi)$

Tabla B.2: Polos de la función de transferencia de un LPF de orden 6 y frecuencia de corte $\Omega_p = 1$ [rad/s].

Los polos p_1 y p_6 , p_2 y p_5 y p_3 y p_4 corresponden respectivamente a pares complejos conjugados.

La función de transferencia está dada por:

$$H(s) = \prod_{k=1}^6 \frac{1}{(s - p_k)} \quad (\text{B.13})$$

Si agrupamos convenientemente las multiplicaciones de 2 en 2:

$$H(s) = \frac{1}{(s - p_1)(s - p_6)} \cdot \frac{1}{(s - p_2)(s - p_5)} \cdot \frac{1}{(s - p_3)(s - p_4)} \quad (\text{B.14})$$

Si reemplazamos los valores p_k de la Tabla B.2, se obtiene:

$$H(s) = \frac{1}{(s^2 + 0,518s + 1)} \cdot \frac{1}{(s^2 - 1,414s - p_5)} \cdot \frac{1}{(s^2 + 1,932s + 1)} \quad (\text{B.15})$$

Si reemplazamos s por $s/\Omega_c = s/14,05$ se desplaza la frecuencia de corte del LPF diseñado de 1 [rad/s] a 14 [rad/s]:

$$H(s) = \frac{1}{(0,0051s^2 + 0,037s + 1)} \cdot \frac{1}{(0,0051s^2 + 0,1s + 1)} \cdot \frac{1}{(0,0051s^2 + 0,138s + 1)} \quad (\text{B.16})$$

Debemos ahora convertir la función de transferencia obtenida al dominio digital. Para ello debe aplicarse la transformación bilineal de la Ecuación A.10. Se considerará un $T = 1$.

$$H(z^{-1}) = \frac{1}{\left(0,0051 \left[2 \left(\frac{1-z^{-1}}{1+z^{-1}}\right)\right]^2 + 0,037 \left[2 \left(\frac{1-z^{-1}}{1+z^{-1}}\right)\right] + 1\right)} \cdot \frac{1}{\left(0,0051 \left[2 \left(\frac{1-z^{-1}}{1+z^{-1}}\right)\right]^2 + 0,1 \left[2 \left(\frac{1-z^{-1}}{1+z^{-1}}\right)\right] + 1\right)} \cdot \frac{1}{\left(0,0051 \left[2 \left(\frac{1-z^{-1}}{1+z^{-1}}\right)\right]^2 + 0,138 \left[2 \left(\frac{1-z^{-1}}{1+z^{-1}}\right)\right] + 1\right)} \quad (\text{B.17})$$

Si desarrollamos y simplificamos:

$$H(z^{-1}) = \left(\frac{1+2z^{-1}+z^{-2}}{1,094+1,96z^{-1}+0,946z^{-2}} \right) \cdot \left(\frac{1+2z^{-1}+z^{-2}}{1,22+1,96z^{-1}+0,82z^{-2}} \right) \cdot \left(\frac{1+2z^{-1}+z^{-2}}{1,296+1,96z^{-1}+0,744z^{-2}} \right) \quad (\text{B.18})$$

La igualdad B.18 representa la función de transferencia $H(z^{-1})$ de un LPF que guarda una estrecha relación con nuestro HPF objetivo. Sólo nos queda reemplazar la variable z^{-1} por su inverso aditivo $-z^{-1}$. Este reemplazo concreta un proceso de reversión espectral al LPF de la Expresión B.18, obteniendo el HPF final, cuya función de transferencia es la presentada en la Expresión B.19.

$$H(z^{-1}) = \left(\frac{1-2z^{-1}+z^{-2}}{1,094-1,96z^{-1}+0,946z^{-2}} \right) \cdot \left(\frac{1-2z^{-1}+z^{-2}}{1,22-1,96z^{-1}+0,82z^{-2}} \right) \cdot \left(\frac{1-2z^{-1}+z^{-2}}{1,296-1,96z^{-1}+0,744z^{-2}} \right) \quad (\text{B.19})$$

B.3. Filtro Pasa banda

Para implementar el filtro pasa banda (BPF) con banda de paso entre 2.000 y 3.000 Hz (aproximadamente) se implementará el filtro pasa bajos (LPF) representado por la Ecuación B.4 seguido por el filtro pasa altos (HPF) de la Ecuación B.19.

La Figura B.1 muestra la respuesta en frecuencia del BPF obtenido a partir de los filtros LPF y HPF diseñados.

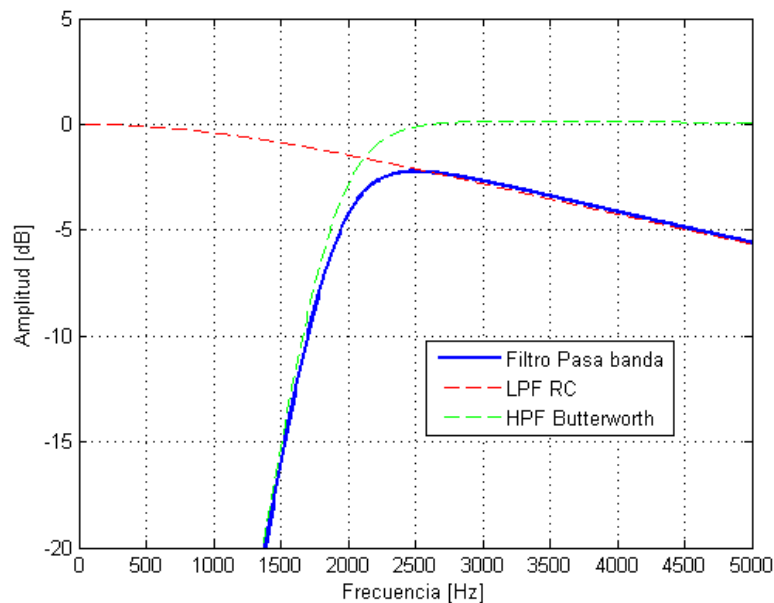


Figura B.1: Se muestra la respuesta en frecuencia del filtro pasa banda que se obtuvo al disponer en cascada un filtro pasa bajos (curva roja) y un filtro pasa altos (curva verde).

Los filtros Butterworth tienen por propiedad mantener una respuesta en frecuencia bastante plana en la banda de paso, al contrario de otros tipos de filtros como el sistema RC implementado en el LPF del ejemplo, en donde el filtro comienza a atenuar gradualmente desde una frecuencia muy inferior a la frecuencia de corte, como se puede apreciar en la Figura B.1. Por otro lado, los filtros Butterworth no presentan el fenómeno de rizado o *Ripple* en la banda de paso común en otros tipos de filtros.

Apéndice C

PLUG-INS DE AUDIO VST

De manera general, un **plug-in** es un *software* que interactúa con una aplicación anfitriona o *Host*, extendiendo sus funcionalidades en una o más tareas específicas.

En el caso particular de los plug-ins de audio, en 1996, la compañía Steinberg Media Technologies lanzó el protocolo VST¹, siendo incluido en su conocida aplicación de audio Cubase, versión 3.0. Este concepto, además de extender las funcionalidades del *Host* de audio, se presentó como un protocolo de plataforma abierta de manera que cualquier programador que quisiese desarrollar un plug-in VST, lo hiciera utilizando apropiadamente las herramientas que Steinberg disponibilizó a través de su Kit de Desarrollo de Software, SDK.

Entre los *Hosts* de audio más populares para los plug-ins VST se encuentran Cubase, Nuendo, WaveLab y Adobe Audition. En esta investigación se utilizó Steinberg Nuendo en su versión 3.0.2. La versión más reciente del SDK VST soportada por Nuendo 3 es la versión 2.4².

A continuación se presenta una breve revisión de las principales funciones del SDK VST 2.4 que se requieren para la implementación del plug-in propuesto en la presente investigación.

¹Del inglés: *Virtual Studio Technology*

²La versión más reciente del SDK VST es el *release* 3.0, sin embargo, él no presenta compatibilidad ni con Nuendo 3 ni con Cubase SE, que corresponde a la versión de Cubase licenciada por la universidad.

C.1. SDK VST

El Kit de Desarrollo de Software, SDK, VST 2.4 contiene un conjunto de funciones pre-programadas orientadas a la implementación de algoritmos para (a) el procesamiento de audio y para (b) la implementación de la Interfaz Gráfica de Usuario, GUI. Se hace la distinción entre el proceso de audio y la interfaz gráfica de usuario (GUI), ya que ambos procesos se ejecutan de manera paralela en el sistema operativo, manteniendo una independencia entre las líneas de programación de ambas implementaciones.

C.1.1. Principales funciones para la implementación del plug-in VST

En esta sección detallaremos las funciones extraídas del SDK VST tanto para la manipulación de audio como para la implementación de la GUI, requeridas para la implementación del plug-in analizador de espectro.

Procesamiento de Audio

El plug-in de Análisis Espectral propuesto no introduce alteraciones en la señal de salida. Debido a esto, las funciones que se requieren del SDK deben asegurar una correcta asignación de entrada-salida y debe permitirnos el acceso a la información de entrada para efectuar los cálculos energéticos asociados. En este contexto, sólo se requieren las funciones siguientes:

setNumInputs/setNumOutputs: Define el número de entradas/salidas del plug-in VST. Para el análisis de una señal de audio Stereo tanto las entradas como las salidas deben ser ajustadas a 2.

processReplacing: En esta función se aplica el proceso de audio. Toma las entradas, las procesa y entrega el resultado sobrescribiendo el Buffer de salida. En nuestro caso, las entradas deben ser copiadas directamente a las salidas, almacenando en una variable cada Buffer de audio trabajado.

Debido a la naturaleza de nuestro analizador no fue necesaria la implementación de algunas típicas funciones para la implementación del procesamiento de audio como son las funciones `setParameter` y `getParameter`, las cuales son llamadas cada vez que el usuario modifica algún parámetro del plug-in.

GUI

1. `open/close`: La función `open` es llamada por el *Host* cuando se muestra en pantalla la Interfaz Gráfica de Usuario (GUI), mientras que la función `close` se llama cada vez que la ventana del editor es cerrada.
2. `idle`: Tiene una función similar a la de `processReplacing`. Se llama periódicamente mientras el editor está abierto, realizando la actualización de los parámetros de la GUI.

Otra función que ha sido omitida es la función `valueChanged`, llamada cada vez que se modifica un parámetro de la GUI.

C.2. Creando un proyecto en Microsoft Visual Studio 2008

El SDK VST 2.4 trae dentro de su directorio algunos plug-ins de ejemplo, a partir de los cuales se puede armar una sesión de trabajo en el compilador de C++ escogido, reutilizando la mayor cantidad de líneas de código a partir de estos demos.

Otra alternativa es la utilización de asistentes de ejecución de nuevos proyectos para plug-ins VST (*VST Wizards*).

C.3. VST Wizard

Disponible en <http://ag-works.net/source.wizards.htm>. Está programado para su utilización en Microsoft Visual Studio 2005, sin embargo, haciendo unas pequeñas modificaciones se hace extensible al Microsoft Visual Studio 2008, que es la versión con la que estamos trabajando.

Una vez instalado el *VST Wizard*, basta con abrir Microsoft Visual Studio 2008 y escoger **new project**. Dentro de los tipos de proyecto disponibles para la creación, figurará VST Wizard, tal como lo muestra la Figura C.1. Debe escogerse el nombre del proyecto y la ubicación que tendrá en el computador.

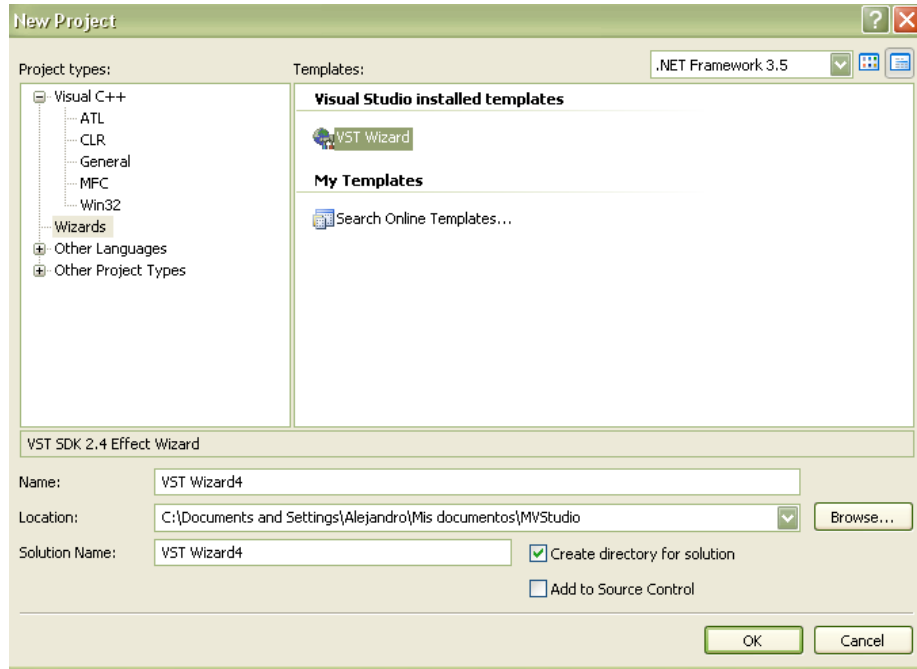
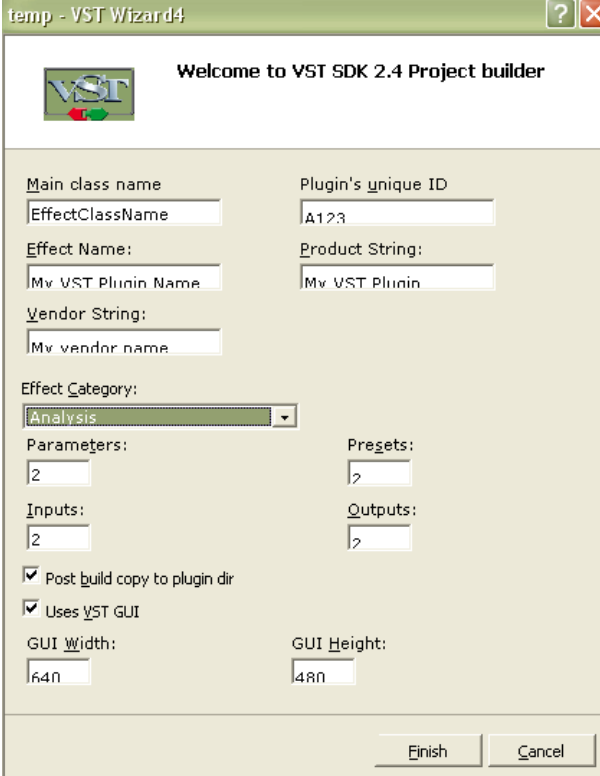


Figura C.1: Pantalla para la creación de un nuevo proyecto, en Microsoft Visual Studio 2008.

Luego aparece una pantalla con una serie de campos para completar (Figura C.2), todos para la definición del nuevo plug-in VST. Inicialmente, de lo único que debemos preocuparnos es de la activación del *checkbox Uses VST GUI* y deben escogerse 2 entradas y 2 salidas (para un plug-in en configuración Stereo). Al presionar Finalizar (*Finish*) se creará automáticamente un proyecto conteniendo toda la estructura de archivos y la estructura base del código a programar, incluyendo tanto para la implementación del algoritmo como para la programación de la interfaz gráfica.



The image shows a Windows-style dialog box titled "temp - VST Wizard4". The window has a green title bar with a question mark icon and a close button. The main content area is titled "Welcome to VST SDK 2.4 Project builder" and features the VST logo. Below the title, there are several input fields and checkboxes for configuring a new plugin:

- Main class name:** Input field containing "EffectClassName".
- Plugin's unique ID:** Input field containing "A123".
- Effect Name:** Input field containing "My VST Plugin Name".
- Product String:** Input field containing "My VST Plugin".
- Vendor String:** Input field containing "My vendor name".
- Effect Category:** A dropdown menu currently set to "Analysis".
- Parameters:** Input field containing "2".
- Presets:** Input field containing "2".
- Inputs:** Input field containing "2".
- Outputs:** Input field containing "2".
- Post build copy to plugin dir:** A checked checkbox.
- Uses VST GUI:** A checked checkbox.
- GUI Width:** Input field containing "640".
- GUI Height:** Input field containing "480".

At the bottom right of the dialog, there are two buttons: "Finish" and "Cancel".

Figura C.2: Configuración de las especificaciones del nuevo plug-in (número de entradas/salidas, nombre, etc.)

Índice de figuras

2.1. Banco de filtros ideal con 3 bandas. Las bandas poseen una pendiente de atenuación infinita en las frecuencias de corte inferior f_i y superior f_s de cada banda, por lo que idealmente las bandas no están solapadas.	5
2.2. La línea continua (azul) corresponde a la respuesta en frecuencia de una banda centrada en 5.000 Hz, la cual cumple con los requerimientos de atenuación mínima exigidos por la IEC 61260 (cruces en rojo).	9
2.3. Oído interno. Está compuesto principalmente por los conductos semi-circulares y la coclea o caracol.	10
2.4. Escalas de alturas en relación a la longitud de la coclea, sin daños [9]. Las escalas de frecuencia y posición (longitud) mostradas se relacionan mediante la Ecuación 2.9.	12
2.5. Razón de banda crítica Z, medida en Barks, en función de la frecuencia [kHz]. El parámetro Z se asocia a las bandas críticas adyacentes. Sólo se muestran los primeros 24 valores.	13
2.6. (a) Todas aquellas señales cuya amplitud caiga bajo la curva de enmascaramiento (curva negra) no será percibida ante la presencia del tono puro de 1 kHz a 80 dB (curva roja). (b) Aproximación a las pendientes de una banda crítica mediante líneas rectas.	14
2.7. Modelo de Terhardt para el umbral de enmascaramiento de señales sinusoidales. Se puede apreciar la variación de las pendientes superiores en función de la intensidad del tono enmascarador [20].	14

2.8. Curvas de igual sonoridad, según Suzuki y Takeshima[35]. Se tiene el nivel de presión sonora en función de la frecuencia. Cada curva es caracterizada por un valor en Fones que es igual al nivel de presión sonora a 1 kHz. 16

2.9. Contornos de igual sonoridad según Robinson y Dadson [30]. Son destacadas las curvas de 40, 70 y 100 Fon, a partir de las cuales se infirieron las curvas de ponderación A, B y C: curvas inversas a la desviación en dB de la respectiva isofónica usando como referencia la frecuencia de 1.000 Hz. 17

2.10. (a) Curva isofónica 40, normalizada a 0 dB@ 1 kHz. (b) Curva isofónica 40 complementaria, comparada con la curva de ponderación A. . . . 18

2.11. Curvas de igual nivel de presión sonora, determinados usando la Red Neuronal Artificial *feed-forward* 2-11-9-1. En el eje de las ordenadas se lee el Nivel de Sonoridad P asociado a cada punto de frecuencia y nivel de presión sonora L_p (rotulado en las leyendas). 20

2.12. Transformada de Fourier para 2 señales de audio digital (arriba) y descomposición en un banco de 24 bandas de frecuencia (abajo). . . . 24

2.13. Gráfico escogido para condiciones de campo libre, considerando un nivel máximo $L = 69,5$ dB @ 63 Hz. 28

2.14. Se muestra el rectángulo de área equivalente (región coloreada) al área bajo la curva de sonoridad específica (área bajo curva azul), su altura leída en Sones entrega la sonoridad total S . En este caso, $S = 16,3$ Sones (GF). 29

2.15. (a) Filtros pasa bajos (LPF) y pasa altos (HPF); (b) Filtro pasa banda (BPF) formado de la disposición en cadena de un HPF y un LPF, donde $f_{\text{corte HPF}} < f_{\text{corte LPF}}$ 30

2.16. Circuito Analógico RC, compuesto por una resistencia de R Ohms y un condensador de C Faradios. 31

2.17. Estructura digital que permite la implementación recursiva de los filtros de respuesta impulsiva infinita (IIR). Si los coeficientes a_k fueran nulos, entonces la estructura digital carecería de realimentación y su implementación sería más sencilla. Se hablaría, entonces, de un filtro de respuesta impulsiva finita (FIR). 33

2.18. (a) Especificaciones para la respuesta en frecuencia analógica de un filtros pasa bajos (LPF) y (b) sus respectivas especificaciones en dominio digital. 34

2.19. Respuesta de amplitud para filtros Butterworth de diferentes órdenes. 35

2.20. Sistema digital temporal que produce un HPF sobre la base de un LPF. 39

2.21. Respuesta en frecuencia de un LPF (a) original y (b) con reversión espectral (HPF). 39

2.22. Esquema de implementación de un sistema caracterizado por la función de transferencia $H(z^{-1})$ como (arriba) un sistema de orden N y, (abajo) como $N/2$ sistemas de orden 2. La implementación en cascadas de filtros de orden 2 es más robusta ante la aparición de errores de redondeo. 41

2.23. Sistema digital con Función de Transferencia $D(z^{-1})$ 42

2.24. Warping de frecuencias para diferentes valores de λ , normalizadas $(\omega/\pi, \nu/\pi)$ 42

2.25. Razón de banda crítica z y frecuencia digital ν como función de la frecuencia ω . *Warping* de frecuencias utilizando el factor λ apropiado. Las frecuencias están normalizadas. 43

2.26. Cadena de filtros *AllPass*. Hace posible aplicar la técnica de *warping* a una señal de audio cualquiera. 44

2.27. Cadena de filtros que implementa el *warping* de frecuencias a una señal $x[n]$. Esta estructura *AllPass* es conocida como filtro de Laguerre. La etapa de pre-énfasis está caracterizada por la función de transferencia $A(z^{-1})$ de la Ecuación 2.60 44

2.28. Topologías de redes [36]: (a) No estructuradas, (b) en capas, (c) recurrentes, (d) modulares. 45

2.29. Red de funciones Radial Base, RBFN, con n parámetros de entrada, m neuronas en la capa oculta y 1 salida. Imagen extraída de [26] . . . 49

2.30. Funciones radial base de Cauchy y Gauss, para el caso unidimensional (1 sola entrada) con centros en $c = 0$ y $r = 1$ 50

2.31. Ejemplo de una red de funciones radial base (RBFN), con 1 parámetro de entrada y $m = 3$ neuronas en la capa oculta. Las funciones radiales $h_j(x)$ 51

2.32. Salida $f(x)$ de la RBFN de la Figura 2.31. Se muestran las funciones radiales que generan la salida $f(x)$. La curva en negro representa la salida de la red. Se destaca el valor adquirido cuando $x = 0,6$ 52

3.1. Curva de enmascaramiento global usando el modelo de Terhardt. De los 4 tonos puros presentados, el segundo tono puro es enmascarado por el umbral generado por el primer tono, por lo que el tono de amplitud 55 dB no será percibido por el oído humano. 57

3.2. Modelo para el cálculo de sonoridad propuesto en [27]. Se debe considerar una etapa de calibración para la señal $x[n]$ que permita la conversión de dBFS a dB (re.20 μ Pa) 58

3.3. La curva roja representa la relación aproximadamente lineal entre la razón de banda crítica z y la frecuencia en dominio warp ν 60

3.4. Bandas N° 4, 5 y 6 del banco de filtros en bandas críticas aproximados usando la técnica de *Warping* de frecuencias. 66

3.5. Algunas de las Bandas Críticas interpoladas usando las RNA del tipo RBFN, considerando una frecuencia de muestreo $f_s = 64.000$ Hz. El entrenamiento fue realizado considerando un $f_s = 44.100$ Hz. 71

3.6. Carga de CPU devuelto por el medidor *VST performance* para el filtrado de señales considerando 10 filtros auditivos. (a) Consumo de CPU de un 48,00 % para el algoritmo filtrado temporal (implementación recursiva). (b) Consumo de CPU de un 2,89 % para el algoritmo de filtrado usando RBFN. 73

3.7. Valores RMS correspondientes a algunas bandas de frecuencia aplicados a 2 fragmentos de audio de 4.096 muestras a una $f_s = 44.100$ Hz. Se muestran los valores obtenidos al aplicar pasa bandas de respuesta impulsiva infinita (IIR), así como sus respectivas aproximaciones determinadas en el dominio de la frecuencia. 74

3.8. Valores RMS en los plug-ins VST programados. El cálculo en el dominio de la frecuencia (derecha) muestra una tendencia a presentar valores RMS mayores que la implementación temporal (izquierda), tal como ocurre en su versión *off-line* programada en MATLAB ® (Figura 3.7). 74

3.9. Se muestra una función $Y[k]$ (a) en el dominio lineal de frecuencias y (b) en el dominio *Warp* de frecuencias. En el caso lineal, $\Delta\omega_1 = \Delta\omega_2$, por lo que se otorga una ponderación equitativa a cada componente de frecuencia cuando se calcula un valor RMS. Por otro lado, al usar la frecuencia en dominio *Warp* $\nu = \nu(\omega)$ (Ecuación 2.58) se produce que $\Delta\nu_1 > \Delta\nu_2$, por lo que las amplitudes vinculadas a componentes de baja frecuencia serán más influyentes en la determinación de valores RMS que las de alta frecuencia. 76

3.10. En azul se presenta el método propuesto, en rojo la implementación del modelo coclear de Patterson-Holdsworth y en verde el Método B (ISO 532:1975). Todos los modelos muestran que el oído humano presenta una mayor sensibilidad a este fragmento de audio en las frecuencias de 3 a 4 Barks (300 a 400 Hz) y de 17 a 18 Barks (en torno a 4 kHz). 77

3.11. En azul se presenta el método propuesto, en rojo la implementación del modelo coclear de Patterson-Holdsworth y en verde el Método B (ISO 532:1975). Todos los modelos muestran que el oído humano presenta una mayor sensibilidad a este fragmento de audio en las frecuencias de 6 a 7 Barks (600 a 900 Hz) y de 16 a 17 Barks (en torno a 3,5 kHz). 77

3.12. Prototipo a ser implementado en un Plug-in de audio VST para el análisis de señales de audio considerando parámetros psicoacústicos. . 79

3.13. Implementación del Plug-in analizador de espectro en formato *Stereo*.
 Se muestran sólo algunas de las 24 bandas críticas implementadas. En la configuración mostrada, se está aplicando la curva de ponderación A y se está considerando el eje de frecuencias en Hz, por lo que este análisis entrega resultados similares a un análisis en bandas de 1/3 de octava. 81

3.14. Panel para la selección de la ponderación de amplitud. 83

3.15. Carga de CPU de 5,80% 83

4.1. Histograma que considera un número total de 954 valores de amplitud (1 a cada valor filtrado de la canción Rehab en la banda crítica 9) y que han sido agrupados en intervalos de 3 en 3 decibeles ponderados en A, considerando una ponderación lineal de frecuencias. Se puede apreciar que la Moda estadística corresponde al intervalo de amplitud entre -41,5 y -38,5 dBFS(A)_{lin}. 89

4.2. Grupo A y Grupo B. Ambas nubes de puntos responden a 2 patrones de balance espectral claramente diferenciados. 91

B.1. Se muestra la respuesta en frecuencia del filtro pasa banda que se obtuvo al disponer en cascada un filtro pasa bajos (curva roja) y un filtro pasa altos (curva verde). 110

C.1. Pantalla para la creación de un nuevo proyecto, en Microsoft Visual Studio 2008. 114

C.2. Configuración de las especificaciones del nuevo plug-in (número de entradas/salidas, nombre, etc.) 115

Índice de tablas

2.1. Bandas de 1/3 de octava. Tanto los números de banda k como las frecuencias centrales nominales aquí mostradas se encuentran estandarizadas en la norma IEC 61260:1995 [16].	7
2.2. Límites en la atenuación relativa para filtros de 1/3 de octava, según IEC 61260.	8
2.3. Bandas críticas. Son mostradas las especificaciones de 25 filtros pasabanda. El último de ellos es una extensión de los valores definidos por Zwicker.	12
2.4. Polos considerados en la definición de las curvas de ponderación en frecuencia A, B y C. Los estándares considerados corresponden a la norma IEC 60651:1979 [15] la cual fue reemplazada por la norma IEC 61672 [17]. En este caso se hace referencia a una norma derogada debido a que en ella se definió la curva de ponderación B.	18
2.5. Se especifica la nomenclatura que debe utilizarse al usar el Método A o el Método B en el cálculo de sonoridad. El Método A es recomendable en conjunto a un análisis en bandas de 1 octava y sólo está disponible para condiciones de campo difuso. El método B considera bandas de 1/3 de octava y entrega la posibilidad de cálculo en condiciones de campo libre.	25
2.6. Niveles de Presión Sonora L_{pi} en bandas de 1/3 de octava asociados a una fuente de ruido estacionaria. Se determinará el nivel de sonoridad total P usando el procedimiento ISO 532-B.	26
2.7. <i>Niveles de Presión Sonora en Bandas Críticas L_i.</i>	27
2.8. Funciones de agregación a una neurona artificial k	46

2.9. Algunas funciones de activación, también llamadas funciones de transferencia. 47

2.10. Parámetros asociados a una red de funciones radial base (RBFN). . . 50

3.1. Coeficientes b_k y a_k de las funciones de transferencia digital $H(z^{-1})$ para las curvas de ponderación en frecuencia A, B, C y D. 54

3.2. Frecuencias de corte en Hz, en Barks y en rad/samples (considerando un dominio *Warp*) para una $f_s = 44.100$ Hz. Las frecuencias ν_i corresponderán a las frecuencias de corte de cada HPF del banco a diseñar, mientras que las frecuencias ν_u corresponderán a las frecuencias de los LPF. 62

3.3. Valores de la pendiente de atenuación S_2 para el diseño de filtros pasa bajos Butterworth para las 24 primeras bandas críticas. Se usó la Ecuación 3.7 con $L_p = 70$ dB. 63

3.4. Orden de los filtros pasa bajos (LPF) y pasa altos (HPF) usados en la construcción de los filtros auditivos usando la técnica de *Warping* de frecuencias. En general, se puede decir que los filtros pasa banda resultantes son de un orden elevado. 65

3.5. Características del banco de 24 filtros en bandas críticas usando la técnica de *Warping*. Las frecuencias mostradas se han redondeado al múltiplo de 10 más cercano. Debido al corrimiento de las frecuencias de corte como consecuencia del procedimiento de diseño, la última banda queda en su mayor parte por sobre los 12.500 Hz. 66

3.6. Características de los filtros auditivos utilizados en el entrenamiento de las RBFN. Se muestran las configuraciones óptimas encontradas, que aseguran un error despreciable en la interpolación de cada banda entre 20 y 22.050 Hz a niveles por sobre -100 dBFS. Para la implementación recursiva de cada banda debe considerarse 2 filtros IIR en cascada con los órdenes especificados en las columnas *Orden LPF* y *Orden HPF*. . 72

4.1. Temas seleccionados de entre los ganadores del Grammy en la categoría *Record of the Year*. 86

4.2. Otros temas seleccionados. La canción *Hey* ya resultó nominada pero no ganadora en la categoría *Record of the year*. Por otro lado, los discos *Libertad* y *Vivo Acá* de donde se extrajeron respectivamente el tema de La Ley y el de Divididos fueron nominados al Grammy Latino en la categoría “Mejor Album de Rock Vocal de Dúo o Grupo”, resultando ganador el disco de La Ley. 86

4.3. Ponderación de amplitud y ponderación de frecuencia considerada por el plug-in analizador de espectro. Son considerados 8 descriptores acústicos, considerando las variantes Z_{lin} y Z_{warp} , A_{lin} y A_{warp} , B_{warp} , C_{warp} , D_{warp} y Fon. 88

4.4. Temas seleccionados. En esta oportunidad se presenta la rotulación utilizada para cada canción: tema 1 a tema 10. 90

4.5. Desviación estándar S para los 10 fragmentos musicales analizados. El estimador acústico que presentó una respuesta en frecuencia más plana correspondió a los decibeles ponderados en A, dB(A), mientras que la respuesta menos plana fue registrada por el propuesta ponderación en Fones. 90

B.1. Parámetros de diseño para un HPF con corte en 2.000 Hz y banda de rechazo por debajo de los 1.500 Hz. El LPF asociado posee una frecuencia de corte de 20.050 Hz e inicio de la zona de rechazo en 20.550 Hz. Las frecuencias ν y Ω se muestran en las 2 últimas columnas. 107

B.2. Polos de la función de transferencia de un LPF de orden 6 y frecuencia de corte $\Omega_p = 1$ [rad/s]. 108