



UNIVERSIDAD
POLITECNICA
DE VALENCIA

TESIS DOCTORAL

Procesado Tiempo-Frecuencia: Aplicación a la percepción humana del sonido.

Autor: Andrés Camacho García.
Directora: Gema Piñero Sipán.

Agradecimientos

Quisiera agradecer a Gema, mi directora de tesis, todo el tiempo que me ha dedicado, sus consejos y sus correcciones a este trabajo.

También quisiera mostrar mi gratitud al departamento de Maquinas y Motores Térmicos de la Universidad Politécnica de Valencia por proporcionarnos las señales de motor utilizadas en esta tesis.

A Raúl, Jorge y el resto de compañeros de sección gracias por vuestro apoyo y compañerismo.

A Laura, el equipo del ITEAM y a mis compañeros y alumnos de la EPSA que han colaborado en los test de audición, gracias por vuestra colaboración desinteresada.

A mi Familia, gracias por estar a mi lado.

Resumen

Síntesis.

Esta tesis está dedicada al estudio de las distintas herramientas que se disponen en el campo del Procesado Digital de Señales para analizar la percepción humana del sonido. Para ello, en la primera parte de la tesis se exponen los fundamentos fisiológicos de la percepción humana del sonido, así como los métodos y parámetros fundamentales que permiten evaluar la percepción subjetiva producida, lo que también se conoce como estudio psicoacústico de los sonidos. El análisis psicoacústico permite calcular el valor de una serie de parámetros del sonido para combinarlos de una forma adecuada con el fin de obtener una calificación de su calidad percibida (Sound Quality), lo más parecida posible a la que percibiría una persona. Según los parámetros psicoacústicos combinados, podremos obtener la calificación del grado de molestia, bienestar, stress, disonancia, etc. que produce el sonido en una persona.

Respecto a las herramientas disponibles en el campo del Procesado Digital de Señales que puedan ayudar a mejorar el análisis psicoacústico, en esta tesis se realiza un estudio exhaustivo de las posibilidades del procesado Tiempo-Frecuencia. Además de realizar este estudio, el trabajo de investigación se ha centrado en la aplicación del análisis psicoacústico para la calificación de la molestia que producen ciertos tipos de ruido como son los ruidos de motor, ya sea el ruido de combustión producido por el motor, ya sea el ruido de escape. Se sabe que la molestia producida por estos ruidos se puede modelar utilizando una combinación de valores de Loudness (Sonoridad), Sharpness (Agudeza o Tonalidad) y Roughness (Aspereza). El cálculo de los dos primeros parámetros está estandarizado, mientras que existen varias aproximaciones para el cálculo del Roughness, presentando todas ellas serias limitaciones al analizar sonidos de motor, puesto que fueron desarrolladas y probadas para aplicarlas a señales simples.

En esta tesis se ha desarrollado un modelo de síntesis de señales de ruido basado en técnicas Tiempo-Frecuencia cuyo objetivo es sintetizar señales de motor lo más realistas posibles, pero buscando un modelo suficientemente sencillo para que en él se puedan identificar parámetros que influyan directamente sobre el Roughness percibido. Para ello, ambos ruidos de combustión y de escape se han modelado básicamente como la convolución entre una señal patrón y un tren de deltas modulado. Gracias a este modelo parametrizado, se ha comprobado que el Roughness de las señales sintetizadas viene determinado por la amplitud y la frecuencia de modulación del tren de deltas, mientras que se mantiene independiente de la forma que tenga la señal patrón. Estos resultados permitirán diseñar sonidos que produzcan menos molestia incidiendo en las señales de

excitación que los producen, más que en los patrones de repetición que presentan.

Objetivos.

El objetivo general de esta tesis es realizar un estudio de las herramientas Tiempo-Frecuencia del Procesado Digital de Señales que pueden ser utilizadas en el análisis psicoacústico de cualquier sonido. Como objetivo particular, el trabajo aquí presentado pretende aportar nuevos métodos para mejorar la estimación de la molestia percibida en los ruidos de combustión y escape de un automóvil. Para ello, la investigación se ha centrado en el análisis de uno de los parámetros psicoacústicos que más influyen en la percepción de la molestia, el Roughness. Dado que los métodos clásicos de cálculo del Roughness han tomado siempre como referencia señales sencillas, el objetivo de esta tesis ha sido encontrar una adecuada metodología para el cálculo del Roughness aplicado a señales más complejas como son las señales de ruido de motor.

Elementos de la metodología a destacar.

La metodología utilizada se ha basado en primer lugar en una exhaustiva recopilación bibliográfica (continuamente actualizada durante todo el desarrollo de la investigación) sobre las técnicas y metodologías propuestas para el análisis psicoacústico de señales complejas, como son las señales de motor. Cabe destacar en este punto la multidisciplinariedad del tema tratado, ya que este estudio ha abarcado: 1) Revisión de la fisiología del oído humano y su influencia en la percepción del sonido; 2) Revisión de todos los parámetros acústicos y psicoacústicos, y su metodología de cálculo, que pueden ser utilizados para describir la percepción subjetiva que produce un sonido; 3) Revisión de las características mecánicas y físicas de un motor de combustión que han permitido sentar las bases de un modelo realista para sintetizar señales de ruido de motores. En segundo lugar, para el estudio del ruido de motor se disponía de una gran cantidad y variación de señales de motor grabadas en muy buenas condiciones acústicas (cámara anecoica), lo que ha permitido realizar un estudio sobre el modelo de generación de estas señales y obtener un modelo fiable basado en la convolución entre una señal patrón y un tren de deltas modulado. Por último, se ha podido validar el modelo propuesto como un modelo adecuado para la obtención del Roughness percibido gracias a la realización de experimentos con jurado que han proporcionado valoraciones fiables de los parámetros psicoacústicos de las señales. Cabe destacar que el test de jurado es prácticamente la única metodología usada en la literatura que permite confirmar la validez de un modelo psicoacústico, y como tal se ha usado a lo largo del trabajo presentado en esta tesis.

Resultados logrados.

En el entorno de trabajo descrito, las aportaciones originales de esta tesis consisten en:

-
- Un exhaustivo estudio del estado del arte de las técnicas de Procesado Tiempo-Frecuencia con aplicación a la percepción subjetiva del sonido.
 - El desarrollo de un modelo de síntesis de ruido de motor que permite relacionar los principales parámetros psicoacústicos que modelan la molestia (Loudness, Sharpness y Roughness) con parámetros sencillos usados en el modelo.
 - Análisis de las prestaciones de los métodos clásicos de cálculo del Roughness y propuesta de un nuevo método adecuado a las señales de ruido de motor.
 - Desarrollo de un nuevo tipo de señal con interés en psicoacústica, la chirp de barrido uniforme en la escala Bark, que permite investigar los efectos de correlación entre bandas críticas a la hora de calcular los principales parámetros psicoacústicos.

Resum

Síntesi.

Esta tesi està dedicada a l'estudi de les distintes ferramentes que es disposen en el camp del Processat Digital de Senyals per a analitzar la percepció humana del so. Per a això, en la primera part de la tesi s'exposen els fonaments fisiològics de la percepció humana del so, així com els mètodes i paràmetres fonamentals que permeten avaluar la percepció subjectiva produïda, la qual cosa també es coneix com a estudi psicoacústic dels sons. L'anàlisi psicoacústic permet calcular el valor d'una sèrie de paràmetres del so per a combinar-los d'una forma adequada a fi d'obtindre una qualificació de la qualitat percebuda (Sound Quality) el més pareguda possible a la que percebria una persona. Segons els paràmetres psicoacústics combinats, podrem obtindre la qualificació del grau de molèstia, benestar, estrés, dissonància, etc. que produïx el so en una persona.

Respecte a les ferramentes disponibles en el camp del Processat Digital de Senyals que puguen ajudar a millorar l'anàlisi psicoacústic, en esta tesi es realitza un estudi exhaustiu de les possibilitats del processat Temps-Freqüència. A més de realitzar este estudi, el treball d'investigació s'ha centrat en l'aplicació de l'anàlisi psicoacústic per a la qualificació de la molèstia que produïxen certs tipus de soroll com són els sorolls de motor, ja siga el soroll de combustió produït pel motor, ja siga el soroll d'escapament. Se sap que la molèstia produïda per estos sorolls es pot modelar utilitzant una combinació de valors de Loudness (Sonoritat), Sharpness (Agudesa o Tonalitat) i Roughness (Aspror). El càlcul dels dos primers paràmetres està estandarditzat, mentres que hi ha diverses aproximacions per al càlcul del Roughness, presentant totes elles serioses limitacions a l'analitzar sons de motor, ja que van ser desenvolupades i provades per a aplicar-les a senyals simples.

En esta tesi s'ha desenvolupat un model de síntesi de senyals de soroll basat en tècniques Temps-Freqüència l'objectiu del qual és sintetitzar senyals de motor el més realistes possible, però buscant un model prou senzill perquè en ell es puguen identificar paràmetres que influïsquen directament sobre el Roughness percebut. Per a això, ambdós sorolls de combustió i de fuga s'han modelat bàsicament com la convolució entre un senyal patró i un tren de deltes modulats. Gràcies a este model parametritzat, s'ha comprovat que el Roughness dels senyals sintetitzats ve determinat per l'amplitud i la freqüència de modulació del tren de deltes, mentres que es manté independent de la forma que tinga el senyal patró. Estos resultats permetran dissenyar sons que produïsquen menys molèstia incidint en els senyals d'excitació que els produïxen, més que en els patrons de repetició que presenten.

Objectius.

L'objectiu general d'esta tesi és realitzar un estudi de les ferramentes Temps-Freqüència del Processat Digital de Senyals que poden ser utilitzades en l'anàlisi psicoacústic de qualsevol so. Com a objectiu particular, el treball ací presentat pretén aportar nous mètodes per a millorar l'estimació de la molèstia percebuda en els sorolls de combustió i escapament d'un automòbil. Per a això, la investigació s'ha centrat en l'anàlisi d'un dels paràmetres psicoacústics que més influïxen en la percepció de la molèstia, el Roughness. Atés que els mètodes clàssics de càlcul del Roughness han pres sempre com a referència senyals senzills, l'objectiu d'esta tesi ha sigut trobar una metodologia adequada per al càlcul del Roughness aplicat a senyals més complexos com són els senyals de soroll de motor.

Elements de la metodologia a destacar.

La metodologia utilitzada s'ha basat en primer lloc en una exhaustiva recopilació bibliogràfica (continuament actualitzada durant tot el desenvolupament de la investigació) sobre les tècniques i metodologies proposades per a l'anàlisi psicoacústic de senyals complexos, com són els senyals de motor. Cal destacar en este punt la multidisciplinarietat del tema tractat, ja que este estudi ha comprés: 1) Revisió de la fisiologia del sistema auditiu humà i la seua influència en la percepció del so; 2) Revisió de tots els paràmetres acústics i psicoacústics, i la seua metodologia de càlcul, que poden ser utilitzats per a descriure la percepció subjectiva que produïx un so; 3) Revisió de les característiques mecàniques i físiques d'un motor de combustió que han permés assentar les bases d'un model realista per a sintetitzar senyals de soroll de motors. En segon lloc, per a l'estudi del soroll de motor es disposava d'una gran quantitat i variació de senyals de motor gravades en molt bones condicions acústiques (cambra anecoica), la qual cosa ha permés realitzar un estudi sobre el model de generació d'estos senyals i obtindre un model fiable basat en la convolució entre un senyal patró i un tren de deltes modulats. Finalment, s'ha pogut validar el model proposat com un model adequat per a l'obtenció del Roughness percebut gràcies a la realització d'experiments amb jurat que han proporcionat valoracions fiables dels paràmetres psicoacústics dels senyals. Cal destacar que el test de jurat és pràcticament l'única metodologia usada en la literatura que permet confirmar la validesa d'un model psicoacústic, i com a tal s'ha usat al llarg del treball presentat en esta tesi.

Resultats aconseguits.

En l'entorn de treball descrit, les aportacions originals d'esta tesi consistixen en:

- Un exhaustiu estudi de l'estat de l'art de les tècniques de Processat Temps-Freqüència amb aplicació a la percepció subjectiva del so.
- El desenvolupament d'un model de síntesi de soroll de motor que permet relacionar els principals paràmetres psicoacústics que modelen la molèstia (Loudness, Sharpness i Roughness) amb paràmetres senzills usats en el model.

-
- Anàlisi de les prestacions dels mètodes clàssics de càlcul del Roughness i proposta d'un nou mètode adequat als senyals de soroll de motor.
 - Desenvolupament d'un nou tipus de senyal amb interès en psicoacústica, la chirp d'exploració uniforme en l'escala Bark, que permet investigar els efectes de correlació entre bandes crítiques a l'hora de calcular els principals paràmetres psicoacústics.

Abstract

Synthesis.

This thesis is dedicated to the study of the different tools that are available in the field of Digital Signal Processing to analyze the human perception of sound. To do this the first part of the thesis presents the physiological basis of human perception of sound, and the methods and basic parameters concerning evaluation of subjective perception, which are also known as psychoacoustic studies of sounds. Psychoacoustic analysis allows the calculation of a set of parameters of sound, to combine them in an appropriate way to obtain a rating of perceived quality (Sound Quality), as close as possible to that perceived by a person. From a combination of psychoacoustic parameters, we obtain the characterization of the degree of discomfort, wellness, stress, disharmony, etc. that a sound produces in a person.

Regarding the tools available in the field of Digital Signal Processing that can help to improve the psychoacoustic analysis, this thesis presents a comprehensive study of the possibilities of time-frequency processing. In addition to this study, the research has focused on the implementation of psychoacoustic analysis for the classification of the discomfort produced by certain types of noise such as engine noise; either the noise produced by the engine combustion or the exhaust noise. It is known that the discomfort caused by these noises can be modeled using a combination of values of Loudness, Sharpness and Roughness. The calculation of the first two parameters is standardized, while there are several approaches to the calculation of Roughness, all of which have serious limitations when analyzing engine sounds, since they were developed and tested for application to simple signals.

This thesis has developed a synthesis model of sound signals based on time-frequency techniques whose aim is to synthesize engine signals as realistic as possible, but looking for a model simple enough to allow those parameters that directly influence the perceived Roughness to be identified. To this end, both exhaust noise and combustion noise are modeled essentially as the convolution between a signal pattern and a modulated train of deltas. With this parameterized model, it was found that the Roughness of the synthesized signals is determined by the amplitude and modulation frequency of the train of deltas, while being independent of the shape of the pattern signal. These results will allow us to design sounds that produce less discomfort working with the excitation signals that produce them, rather than the repetition patterns.

Objectives.

The general aim of this thesis is to study the time-frequency tools of Digital Signal Processing that can be used in the psychoacoustic analysis of any sound. As a particular objective, the work presented here aims to provide new methods to improve the estimation of the perceived discomfort of combustion and exhaust noises of a car. To do this, this research has focused on the analysis of one of the psychoacoustic parameters that most influences the perception of discomfort, Roughness. Since traditional methods of Roughness calculation have always worked with simple signals, the objective of this thesis has been to find an appropriate methodology for calculating the Roughness applied to more complex signals such as the engine noise signals.

Elements of the methodology to emphasize.

The methodology is based firstly on an exhaustive bibliographic compilation (continuously updated throughout the development of research) on the techniques and methodologies proposed for the psychoacoustic analysis of complex signals, such as engine signals. It should be noted at this point that a multi-discipline approach has been taken to the topic, as this study has covered: 1) Review of the physiology of the human ear and its influence on the perception of sound; 2) Review of all acoustic and psychoacoustic parameters and its calculation methodology, which can be used to describe the subjective perception that produces a sound; 3) Review of mechanical and physical properties of a combustion engine that provided the basis for a realistic model to synthesize other noise signals. Secondly, to study the engine noise we had a large number and variety of engine signals recorded in very good acoustic conditions (anechoic chamber), which allowed a study of the process of generation of these signals and a reliable model based on the convolution between a signal pattern and a train of deltas modulation. Finally, the proposed model has been validated as an appropriate model to obtain the perceived Roughness. All this has been made possible by conducting experiments with a jury who have provided reliable assessments of the psychoacoustic parameters of the signals. We must note that the jury test is practically the only method used in the literature to confirm the validity of a psychoacoustic model, and as such has been used throughout the work presented in this thesis.

Results achieved.

In the work environment described, the original contributions of this thesis are:

- An exhaustive study of the state of the art of the techniques for time-frequency processing with application to the subjective perception of sound.
- The development of a synthesis model of engine noise that relates the main parameters that model psychoacoustic discomfort (Loudness, Sharpness and Roughness) with simple parameters used in the model.

-
- Analysis of the performance of the classical methods of Roughness calculation and the proposal of a new method which is suitable for engine noise signals.
 - Development of a new type of signal that will prove to be very interesting in the field of psychoacoustics, the even scanning chirp in the Bark scale, which allows us to investigate the effects of correlation between critical bands when calculating the main parameters of psychoacoustics.

Índice general

1. Introducción	1
2. El oído humano: acústica y psicoacústica	5
2.1. Introducción.	5
2.2. Generación y propagación de sonidos.	5
2.2.1. Velocidad.	6
2.2.2. Presión.	7
2.2.3. Intensidad.	7
2.2.4. Potencia.	8
2.3. Anatomía del oído humano	9
2.3.1. El oído externo.	9
2.3.2. El oído medio.	13
2.3.3. El oído interno.	14
2.4. Fundamentos de psicoacústica	17
2.4.1. No linealidad en la percepción del nivel.	17
2.4.2. Enmascaramiento frecuencial.	18
2.4.3. Enmascaramiento temporal.	19
2.4.4. No linealidad en la percepción de la frecuencia.	22
2.5. Parámetros acústicos.	24
2.5.1. Ponderación frecuencial.	24
2.5.2. Ponderación Temporal.	25
2.5.3. Análisis en bandas de octava y tercio de octava.	26
2.5.4. Parámetros estadísticos.	28
2.5.5. Parámetros espectrales.	31
2.6. Parámetros psicoacústicos.	32
2.6.1. Escalas de frecuencia.	32
2.6.2. Pitch.	34
2.6.3. Loudness	35
2.6.4. Sharpness.	41
2.6.5. Fluctuation Strength.	43
2.6.6. Roughness.	45
2.6.7. Duración subjetiva.	49
2.7. Índices globales de calidad de sonidos.	49

3. Análisis TF.	53
3.1. Introducción.	53
3.1.1. Representación de señales que varían en el tiempo.	54
3.1.2. Dominio del tiempo y dominio de la frecuencia.	55
3.1.3. Frecuencia instantánea.	56
3.2. Representaciones Tiempo-Frecuencia de señales continuas.	57
3.2.1. Representaciones tiempo frecuencia lineales.	59
3.2.2. Representaciones Tiempo-Frecuencia cuadráticas.	68
3.3. Distribuciones Tiempo-Frecuencia discretas.	68
3.3.1. Espectrograma y STFT.	69
3.3.2. Escalograma y Wavelets.	73
3.4. Procesado Tiempo Frecuencia en Acústica y Psicoacústica.	74
3.4.1. Wavelet Packet.	75
3.4.2. Warped Spectrum.	77
3.4.3. Cepstrum.	80
3.5. Comparativa Técnicas TF	82
4. Modelado de ruido de motor	85
4.1. Introducción.	85
4.2. Ruido producido por un motor	86
4.2.1. Presión en el interior de un cilindro.	86
4.2.2. Ruido de combustión.	94
4.2.3. Ruido de escape.	99
4.3. Modelo de ruido de motor periódico.	101
4.4. Nuevo modelo de ruido de motor: modelo cuasiperiódico.	102
4.4.1. Descripción del modelo.	102
4.4.2. Obtención de los parámetros del modelo propuesto.	108
4.4.3. Validación subjetiva del modelo propuesto: test de realismo	111
4.5. Análisis frecuencial de los distintos modelos.	114
4.5.1. Señal excitación $c(t)$: tren de deltas periódico.	114
4.5.2. Señal excitación $c(t)$: tren de deltas modulado en AM.	117
4.5.3. Señal excitación $c(t)$: tren de deltas con amplitud aleatoria.	119
4.5.4. Comparación entre modelos.	121
5. Parámetros Psicoacústicos en el modelo de señales Cuasiperiódicas.	123
5.1. Introducción.	123
5.2. Envoltente en los canales de banda crítica.	124
5.2.1. Señal excitación $c(t)$: tren de deltas periódico.	127
5.2.2. Modelo de Feng.	130
5.2.3. Señal excitación $c(t)$: tren de deltas modulado en AM.	133
5.2.4. Convolución de patrón con tren de deltas aleatorias	133
5.3. Evaluación de la calidad de los estimadores de Roughness.	134
5.3.1. Los jueces.	134
5.3.2. Estímulos utilizados.	134
5.3.3. Igualación del Loudness de las señales.	134

5.3.4.	Instrumentación utilizada en el experimento.	135
5.3.5.	Procedimiento.	135
5.3.6.	Resultados del test.	135
5.4.	Evaluación psicoacústica de las señales sintéticas de escape	136
5.4.1.	Estímulos utilizados.	136
5.4.2.	Procedimiento.	138
5.4.3.	Resultados del test para ruido sintético de 1000 r.p.m.	139
5.4.4.	Resultados del test para ruidos de 2000, 3000 y 4000 r.p.m.	142
5.5.	Nuevos patrones sintéticos.	145
5.5.1.	La señal chirp de barrido uniforme en la escala Bark.	147
5.5.2.	Combinación de varias chirp.	149
5.6.	Evaluación psicoacústica de señales sintéticas con patrón chirp.	151
5.6.1.	Dependencia de la forma del patrón (una chirp).	151
5.6.2.	Dependencia de la duración del patrón (dos chirps).	152
5.6.3.	Dependencia de la excitación instantánea del pat. (dos chirps). . . .	155
5.7.	Conclusiones.	155
6.	Conclusiones y líneas futuras	157
	Apéndice 1: Wavelets y filtros de reconstrucción perfecta	159
	Apéndice 2: Publicaciones Relacionadas.	169
	Bibliografía	171

Capítulo 1

Introducción

El ruido generado por productos de consumo se está convirtiendo en un factor relevante de su diseño. Los esfuerzos en las tareas de ingeniería de sonido no están únicamente centrados en reducir el nivel de presión sonora, sino también en reducir la sensación de molestia percibida por el usuario. Además, el sonido de algunos productos se diseña para conseguir ciertas características deseables, es decir se fabrican para que suenen de la forma que deben de sonar. Basta pensar en el timbre de un teléfono móvil o en el sonido del motor de un coche deportivo.

No obstante, cuando se evalúan sensaciones producidas por sonidos, la utilización del Nivel de Presión Sonora (SPL) u otros parámetros físicos basados en un simple análisis espectral son inadecuados para explicar las sensaciones percibidas. Por esta razón, se han propuesto nuevos parámetros relacionados con la percepción humana del sonido, conocidos como parámetros psicoacústicos.

Los parámetros psicoacústicos se pueden calcular mediante procedimientos que evalúan numéricamente algunas sensaciones asociadas a los sonidos escuchados. Los parámetros psicoacústicos más importantes son:

- Loudness: mide la sensación de potencia o nivel sonoro. Su cálculo fue estandarizado por el International Standard Organization (ISO532) en 1975 [1].
- Sharpness: evalúa la sensación subjetiva del timbre de los sonidos y está basado principalmente en el contenido frecuencial. También ha sido objeto de estandarización por la norma Alemana DIN 45692 en el año 2009 [2].
- Roughness: cuantifica la sensación de vibración producida por variaciones temporales rápidas en los sonidos.
- Fluctuation Strength: cuantifica la sensación de vibración producida por variaciones lentas en los sonidos.

Zwicker y Fastl [3, 4], han descrito todas estas sensaciones y han llevado a cabo una serie de experimentos exhaustivos para analizarlas, incluyendo propuestas de útiles mo-

delos matemáticos para su cálculo. Estos modelos matemáticos, así como otras versiones propuestas por diversos autores como Aures [5], Moore [6], Blauert [7] o Daniel y Weber [8] son ampliamente utilizados en experimentos de calidad de sonido y han sido programados en varios paquetes comerciales [9, 10, 11].

La industria, con el fin de mejorar el sonido de sus productos, necesita valorar la molestia que producen, y para esto existen dos opciones: realizar una evaluación de los sonidos por parte de un grupo de expertos o consumidores (test de jurado) o desarrollar sistemas automáticos que proporcionen una medida de esta molestia [12]. Dentro del segundo grupo encontramos medidas calculadas en función de los parámetros psicoacústicos anteriores [3, 13, 14, 15, 16, 17, 18], pero cada una de estas medidas debe de desarrollarse para un tipo de producto en concreto. En el caso de la industria automovilística, el Roughness es uno de los parámetros más influyentes en la valoración de la molestia producida el ruido asociado al motor y otras partes de los vehículos [19, 20, 21, 22, 23, 24]. Sin embargo los métodos de cálculo de Roughness utilizados habitualmente, presentan serias limitaciones cuando se utilizan para analizar sonidos complejos [19, 22], puesto que fueron desarrollados y probados utilizando señales simples tales como modulaciones de amplitud y frecuencia sobre señales senoidales o ruido blanco.

Para poder evaluar correctamente la molestia que produce el sonido de un motor, hace falta primero desarrollar un método que permita estimar correctamente su Roughness. Y para esto se necesita disponer de un gran número de ruidos de motor diferente con los que experimentar. Pero al pedir a un jurado que evalúe el Roughness de varios sonidos comparándolos entre sí, las diferencias que puedan existir en el resto de parámetros psicoacústicos producen sesgo en la estimación del Roughness. Para evitar este sesgo, es necesario utilizar en los test de jurado, grupos de señales de distinto Roughness pero valores idénticos en el resto de parámetros psicoacústicos. Obtener grupos de señales con estas características es imposible si se trabaja con sonidos producidos por motores reales puesto que una modificación en alguno de sus elementos o condiciones de funcionamiento, genera sonidos donde todos los parámetros psicoacústicos han variado.

Con el fin de encontrar una solución a la evaluación del Roughness en este tipo de ruidos complejos, en esta tesis se ha desarrollado un nuevo método para sintetizar el ruido de banda ancha producido por una señal de motor y estimar la sensación de Roughness que este produce. Este modelo se basa en el hecho de que los ruidos producidos por el motor de un coche se ajustan muy bien a un modelo de señal patrón que se repite de forma periódica, con pequeñas variaciones en su amplitud. La extracción de este patrón a partir de sonidos reales ha permitido generar ruidos sintéticos sobre los que analizar los parámetros psicoacústicos. La estructura de la tesis es la siguiente:

- El capítulo 2 es una revisión del estado del arte en materia de Psicoacústica.
- El capítulo 3 se es una revisión del estado del arte en cuanto a las herramientas de análisis tiempo frecuencia necesarias en la evaluación de parámetros psicoacústicos y de las señales de motor.
- El capítulo 4 analiza dos tipos de ruido producidos por motores de combustión

interna: el ruido de combustión y el ruido de escape. A continuación se diseña un modelo que genera señales sintéticas, y se evalúan las prestaciones del modelo.

- El capítulo 5 está dedicado al estudio de los parámetros psicoacústicos de las señales sintéticas generadas. En este capítulo, también se describe como modificar los parámetros del generador de señales descrito en el capítulo 4 para variar los parámetros psicoacústicos de una señal sintética y se valida con test de pares la bondad del modelo propuesto.
- El capítulo 6 recoge las principales conclusiones y plantea algunas líneas futuras asociadas al trabajo planteado

Capítulo 2

El oído humano: acústica y psicoacústica

2.1. Introducción.

Para analizar el grado de molestia o desagrado que produce un sonido es necesario estudiar sus características o, más concretamente, las características que las personas somos capaces de percibir de este sonido. La ciencia que estudia cómo las personas percibimos los sonidos recibe el nombre de Psicoacústica, y su desarrollo se ha realizado sobre los cimientos de la acústica (dedicada a analizar los parámetros físicos de los sonidos). Cabe destacar, sin embargo, que los avances que en las últimas décadas se han producido en materia de Psicoacústica no habrían sido posibles sin el impulso proporcionado por el vertiginoso desarrollo experimentado por la electrónica y el procesado de señal. Estos avances han hecho posible la implementación de modelos digitales del complejo funcionamiento del sistema auditivo humano.

Este capítulo es una revisión del estado del arte de la Psicoacústica y su utilización en la evaluación de la calidad de los sonidos. El capítulo comienza con una revisión de los conceptos básicos de acústica relacionados con la generación y propagación de sonidos. A continuación, se analiza la anatomía y el funcionamiento del oído humano para posteriormente introducir los conceptos más importantes en los que se basa la Psicoacústica. Una vez revisados los fundamentos de la acústica y la Psicoacústica, se describen las magnitudes utilizadas en estas ciencias para la medida de parámetros esenciales de los sonidos. Por último, se analiza cómo combinar los parámetros básicos de los sonidos para evaluar características más complejas como el grado de molestia que producen.

2.2. Generación y propagación de sonidos.

Según el diccionario de la Real Academia Española [25], el sonido es la sensación producida en el órgano del oído por el movimiento vibratorio de los cuerpos, transmitido

por un medio elástico, como el aire.

La Acústica es la ciencia que estudia la producción, transmisión y percepción del sonido tanto en el intervalo de la audición humana como en las frecuencias ultrasónicas e infrasónicas.

Para que exista un sonido en un determinado punto es necesario que exista una variación de la presión en ese punto. Pero, las variaciones de presión en un punto de un medio hacen que se produzcan también variaciones de presión en sus proximidades dando lugar a lo que se conoce como ondas de presión.

2.2.1. Velocidad.

De acuerdo con las leyes de la física, debe existir una relación entre la presión y la velocidad de una partícula dentro de un gas para que se conserve el momentum [26]:

$$\text{grad}(p) = -\rho_0 \frac{\partial v}{\partial t} \quad (2.1)$$

donde:

p = Presión sonora.

v = Vector de velocidad de partícula.

t = Tiempo.

ρ = Densidad estática del gas.

Cuando en un gas ideal las variaciones de presión y densidad son pequeñas comparadas con los valores estáticos y las velocidades de las partículas son mucho menores que la velocidad del sonido, la ecuación anterior se transforma en una ecuación de onda:

$$c^2 \Delta p = \frac{\partial^2 p}{\partial t^2} \quad (2.2)$$

Donde la velocidad de propagación de la presión (c) se puede expresar como [26, 27, 28]:

$$c = \sqrt{\frac{\gamma RT}{m}} \quad (2.3)$$

donde:

R es la constante de los gases,

m es el peso molecular promedio del aire ($R/m = 287 \text{ J/kgK}$),

γ es la razón de los calores específicos $\gamma = cp/cv$, siendo igual a 1,4 para el aire, y

T es la temperatura absoluta en grados kelvin.

En una atmósfera estándar se considera que T es 293,15 K, dando un valor de 343 m/s ó 1.235 km/hora. Esta fórmula supone que la transmisión del sonido se realiza sin pérdidas de energía en el medio, aproximación muy cercana a la realidad.

Una velocidad aproximada (en m/s) puede ser calculada mediante la siguiente fórmula [26]:

$$c = (331,5 + 0,6t) \quad (2.4)$$

donde t es la temperatura en grados Celsius:

$$t = T - 273\text{kelvins} \quad (2.5)$$

2.2.2. Presión.

El valor normal de la presión atmosférica en un punto es de 1 atm. (101325 Pascales); pero cuando en ese punto existe un sonido se producen ligeras variaciones de la presión en torno a este valor. El valor cuadrático medio de dichas variaciones de presión p_{rms} es lo que se conoce como presión sonora. Lo habitual es trabajar con este valor expresado en decibelios, que es lo que se conoce como Nivel de Presión Sonora (Sound Pressure Level o SPL).

Al expresar el Nivel de Presión Sonora (en decibelios) se toma como referencia el valor de $20 \mu Pa$ (valor próximo a las presiones mínimas que el oído es capaz de detectar). Y como la potencia del sonido es proporcional al cuadrado de la presión sonora, la expresión del Nivel de Presión Sonora es:

$$SPL(dB) = 20 \log \frac{p_{rms}}{20\mu Pa} \quad (2.6)$$

2.2.3. Intensidad.

Otra medida física aplicada a los sonidos es la Intensidad Sonora que también se suele utilizar en su versión logarítmica (Nivel de Intensidad Sonora).

La intensidad sonora existente en un punto i es la potencia que atraviesa una unidad de superficie perpendicular a la dirección de propagación de la onda y sus unidades son w/m^2 . Al expresar esta magnitud en decibelios para obtener el nivel de intensidad sonora se normaliza respecto a un valor de referencia de $10^{-12}w/m^2$ quedando:

$$IL(dB) = 10 \log \frac{i}{10^{-12}w/m^2} \quad (2.7)$$

En campo libre, tanto para ondas planas como esféricas, la magnitud de la intensidad en la dirección de propagación de las ondas acústicas está relacionada con la presión acústica por [29, 30]:

$$i = \frac{p^2}{\rho_0 c} \quad (2.8)$$

donde:

ρ_0 es la densidad del aire y
 c la velocidad del ruido en el aire.

A la cantidad $\rho_0 c$ se le llama resistencia característica del medio a través del cual viaja la onda, que en el caso del aire, para una presión barométrica de 1 atm (101.325 Pa) y a una temperatura de 20° C tiene un valor de 415 N.s/m³. Por lo que [30]:

$$IL(dB) = 10 \log \frac{i}{10^{-12} w/m^2} = 10 \log \frac{p^2/\rho_0 c}{10^{-12} w/m^2} = 20 \log \frac{p_{rms}}{20 \mu Pa} + 10 \log \frac{(20 \mu Pa)^2/\rho_0 c}{10^{-12} w/m^2} \quad (2.9)$$

$$IL(dB) = SPL(dB) - 0,16 \quad (2.10)$$

Es decir, que el nivel de intensidad acústica y el nivel de presión sonora son prácticamente idénticos, lo que puede llevar a confusión puesto que esta identidad sólo es aplicable en campo libre para ondas planas o esféricas, no siendo válida para posiciones muy cercanas a la fuente. En campos acústicos más complicados, no se puede aplicar la ecuación anterior.

2.2.4. Potencia.

Se define como potencia acústica a la cantidad de energía emitida por una fuente por unidad de tiempo, expresándose en vatios (W) o en picovatios ($10^{-12}W$).

Frecuentemente, es conveniente expresar la potencia acústica de una fuente en una escala logarítmica, empleando entonces el “nivel de potencia sonora” L_W en dB, dado por:

$$LW(dB) = 10 \log \frac{W}{10^{-12} \text{vatios}} \quad (2.11)$$

Donde

W es la potencia de la fuente en vatios y
 10^{-12}vatios (o 1 picovatio) es la potencia de referencia.

Si consideramos una fuente puntual, es decir, aquella fuente acústica cuyas dimensiones son pequeñas comparadas con la longitud de onda del sonido radiado, emitiendo una potencia W vatios uniformemente en todas las direcciones, la intensidad acústica producida por ésta en la dirección radial será [29, 30]:

$$I = \frac{W}{4\pi r^2} \quad (2.12)$$

Donde:

W es la potencia de la fuente y

r la distancia a la que se mide la intensidad de potencia acústica.

Es decir, que:

$$IL(dB) = LW(dB) - 20 \log(r) - 10 \log(4\pi) = LW(dB) - 20 \log(r) - 11 \quad (2.13)$$

Por lo tanto, las medidas logarítmicas de estas tres magnitudes están relacionadas en campo libre mediante la ecuación:

$$IL(dB) = SPL(dB) - 0,16 = LW(dB) - 20 \log(r) - 11 \quad (2.14)$$

2.3. Anatomía del oído humano

La capacidad auditiva de los animales varía de unos a otros dependiendo de la estructura de sus oídos. Uno de los parámetros importantes en el análisis del fenómeno de la audición es el margen de frecuencias audibles, que en el caso de los humanos (con una audición normal) se extiende entre los valores de 20 Hz y 20.000 Hz, aunque se pueden encontrar animales, como el murciélago, cuyo margen de frecuencias audibles es muy superior.

Para que podamos escuchar e interpretar un sonido, las variaciones de presión existentes en el aire deben ser transformadas en impulsos nerviosos, que las neuronas se encargarán de transportar hasta el cerebro. En este proceso de transducción de energía interviene el oído, que se encuentra dividido en tres partes: oído externo, oído medio y oído interno [6, 31, 32, 33, 34].

2.3.1. El oído externo.

El oído externo está formado por el pabellón auricular y el conducto auditivo externo. El pabellón auricular u oreja, se encarga de proporcionar directividad al sistema de recepción de sonidos humano y de convertir una onda radiada en una onda guiada que se

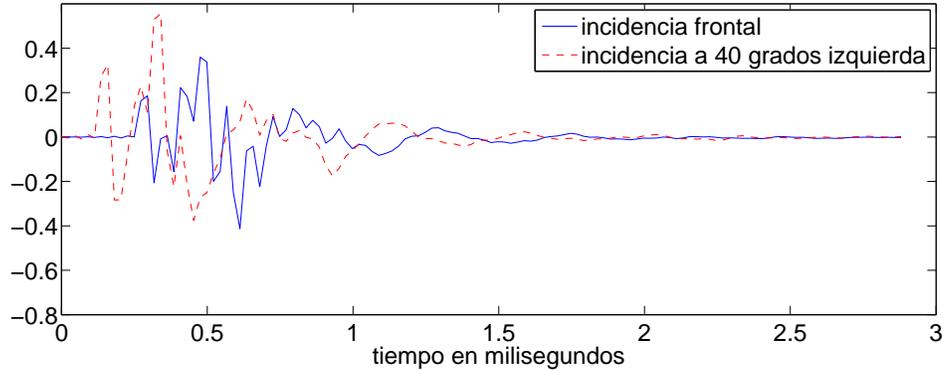


Figura 2.1: Respuesta impulsional de la oreja izquierda para sonidos provenientes de dos direcciones diferentes (0 grados de acimut o dirección frontal y 40 grados de acimut respectivamente) en el plano horizontal (elevación=0 grados). Se toma como dirección de referencia la frontal y como ángulos positivos aquéllos asociados a la mitad izquierda de la cabeza (datos tomados de [36]).

desplazará a través del conducto auditivo externo. El conducto auditivo externo se comporta como una línea de transmisión con una respuesta dependiente de la frecuencia. Al final del conducto se encuentra el tímpano, que es la membrana que comunica el oído externo con el oído medio [35].

Son características fundamentales del pabellón auditivo la ganancia, el diagrama de radiación, la respuesta en frecuencia y la impedancia característica; son características fundamentales del canal auditivo su atenuación y su impedancia característica.

El pabellón auricular, junto a la cabeza y el torso confieren a nuestro sistema auditivo una directividad (los sonidos procedentes de unas direcciones se perciban con mayor intensidad que los que proceden de otras) que depende de la frecuencia. Para caracterizar este efecto se realizan mediciones de sonidos procedentes de varias direcciones dando lugar a las funciones HRTF (Head Related Transfer Function) que dependen de la anatomía de cada persona. La medida de las mismas se realiza registrando, mediante unos diminutos micrófonos introducidos en el canal auditivo externo, las ondas sonoras que llegan a ambos oídos en respuesta a una señal de prueba que emiten los altavoces situados en las posiciones del espacio que se desean caracterizar, obteniéndose una respuesta impulsional para cada una de las distintas direcciones ensayadas [36, 37]. La utilización de estas funciones HRTF es fundamental para simular entornos tridimensionales utilizando auriculares [38, 39], ya que la diferencia de niveles y fases recibidas en los dos oídos es lo que permite determinar la dirección de llegada de los sonidos [40]. Estas funciones, que pueden ser medidas sobre personas o maniqués, dan origen a gran cantidad de datos (como se puede apreciar en las figuras 2.1, 2.2 y 2.3, que en ocasiones son modeladas mediante expresiones matemáticas que permiten trabajar de manera más cómoda [41, 42]).

La transformada de Fourier de estas funciones permite observar que el pabellón auditivo se comporta de forma selectiva en frecuencia, es decir, atenúa unas frecuencias mientras que resalta otras. Y este comportamiento en frecuencia es diferente para cada

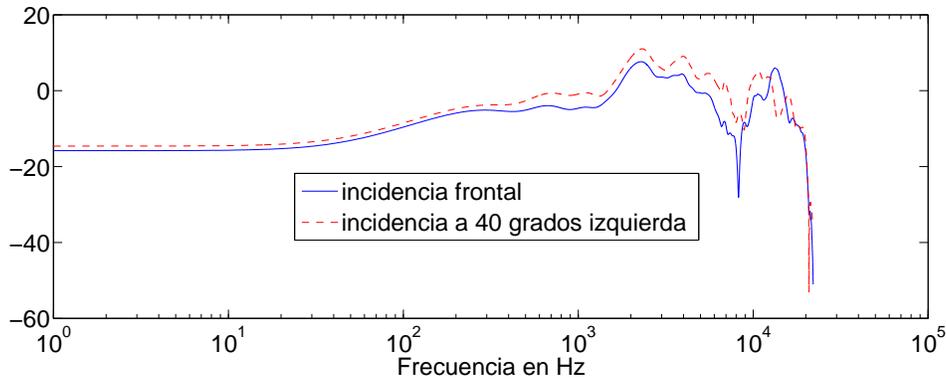


Figura 2.2: Respuesta frecuencial de la oreja izquierda para sonidos provenientes de dos direcciones diferentes (0 grados o dirección frontal y 40 grados de acimut respectivamente) en el plano horizontal (elevación=0 grados). Se toma como dirección de referencia la frontal y como ángulos positivos aquéllos asociados a la mitad izquierda de la cabeza. La representación corresponde a la transformada de Fourier de la figura 2.1.

dirección de recepción de sonido, como se puede apreciar en la figura 2.2.

Otra información que se puede obtener a partir de las funciones HRTF son los diagramas de radiación que el pabellón auditivo presenta en diferentes direcciones. La figura 2.2 contiene representaciones de estos diagramas de radiación para dos frecuencias distintas, y en ella se puede observar que: mientras que en bajas frecuencias el comportamiento del pabellón auditivo es poco dependiente del ángulo de incidencia del sonido, a medida que la frecuencia aumenta la respuesta es altamente dependiente de la dirección de llegada (figura 2.3).

El canal auditivo, que une la oreja con el tímpano, se comporta como una línea de transmisión de sonido que se puede modelar de forma física como un tubo de aproximadamente 25 milímetros con uno de sus extremos cerrado, que producirá por tanto reflexiones y ondas estacionarias. El análisis de canales auditivos reales ha llevado a la conclusión de que la función de transferencia de este elemento presenta resonancias (mínimos de atenuación) en valores de frecuencia próximos a 4 kHz y a 8 kHz [35, 43, 44].

La función de transferencia correspondiente a la relación entre la presión acústica (o nivel de presión acústica) que se medirá en un punto en ausencia de obstáculos y la que se recibe realmente en el tímpano (al final del oído externo), es la combinación de la función de transferencia de la oreja (que dependerá de la dirección de incidencia del sonido) y la función de transferencia del canal auditivo.

La combinación de estas dos funciones de transferencia da como resultado un máximo de recepción en torno a los 2.5 kHz, y la dirección en la que se produce esta ganancia está en torno a los 60° en el plano horizontal (midiendo los ángulos respecto a la frontal). Para esta frecuencia, si se representa la directividad en distintos ángulos, se puede observar que el ancho de haz a 3dB es aproximadamente de 60° y la relación delante detrás de 16dB [35], de forma similar a lo que ocurre con la frecuencia de 6000 Hz de la figura 2.3.

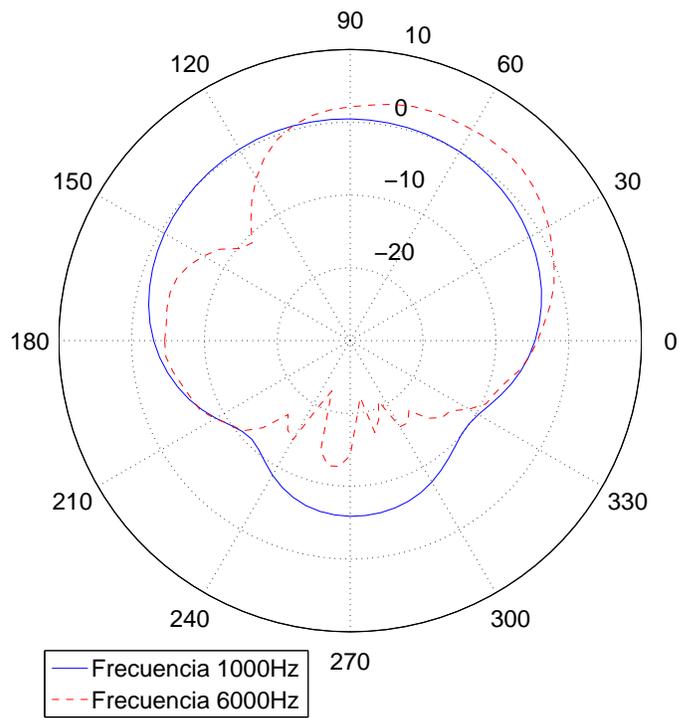


Figura 2.3: Directividad de la oreja izquierda medida en el plano horizontal (elevación=0) en función del ángulo de incidencia acimutal medida en dB (datos tomados de [36]).

Como se puede comprobar, la señal recibida es altamente dependiente de la dirección de llegada por lo que cuando se realizan experimentos de audición se suelen considerar dos casos extremos:

- *Campo libre*: se supone que las señales proceden de una sola dirección de llegada que se corresponde con la frontal. Este caso es habitual en espacio libre, cámaras anecoicas o propagación en espacios donde no existen obstáculos ni paredes en las proximidades del emisor y el receptor y estos se encuentran próximos.

- *Campo difuso*: se supone que el sonido proviene simultáneamente de todas las direcciones del espacio o que existen distintas ondas incidentes desde direcciones indeterminadas. Este tipo de propagación es más similar a las condiciones reales que se producen en locales con paredes en los que el emisor y el receptor se encuentran alejados entre sí.

Otro caso importante a considerar es la *recepción de sonidos a través de auriculares*. En este caso dejan de ser observables las características de directividad del oído y aparecen atenuaciones debidas a pérdidas producidas por fenómenos de desadaptación de impedancias [45, 46]. La recepción de los sonidos será diferente dependiendo del tipo de auricular que se utilice (supra-aurales, circumaurales o intraauriculares) y de si estos son abiertos o cerrados debido a las diferentes impedancias y respuestas frecuenciales que éstos presentan.

2.3.2. El oído medio.

El oído medio está formado por una cavidad llena de aire donde se alojan tres huesecillos (unidos entre si formando una cadena) que reciben el nombre de martillo, yunque y estribo. El martillo se encuentra unido al tímpano mientras que el estribo se encuentra unido a la ventana oval (que forma parte del oído interno). La misión del oído medio es la de adaptar impedancias, puesto que la recepción de las señales sonoras se va a producir en el oído interno dentro de un medio líquido mientras que la señal captada por la oreja procede de un medio gaseoso (cuya impedancia es mucho menor). Además los tres huesecillos forman un sistema de palancas que amplifican la señal de presión recibida. Esta amplificación se puede considerar lineal cuando las presiones de entradas no sean excesivas, ya que si se sobrepasa un cierto límite ($90dB_{SPL}$) los músculos que sostienen los huesecillos se tensan disminuyendo notablemente la amplificación. Este efecto recibe el nombre de reflejo timpánico [47].

El oído medio tiene como misión la de adaptar impedancias y amplificar la señal de presión existente en el oído externo. El modelo equivalente de este sistema será el de un transformador que presentará distintas ganancias e impedancias de entrada y salida para diferentes frecuencias.

La superficie del tímpano es aproximadamente 17 veces superior a la superficie de la ventana oval. Este hecho unido al efecto de palanca que se produce en los huesecillos del oído medio hace que la relación entre la presión existente en el tímpano y la presión existente en la ventana oval sea aproximadamente 22:1. Pero la función del oído medio no es solamente acústica, puesto que también es el encargado de proteger el oído frente a

niveles de presión elevados y proporcionar una barrera mecánica frente a golpes y sonidos propagados por la estructura ósea [48, 49, 50].

Hay que tener en cuenta que las características del oído medio pueden variar dependiendo principalmente de dos factores:

- *Diferencias de presión estática entre el oído medio y el oído externo*: pueden provocar una considerable atenuación de la señal de sonido afectando sobre todo a las frecuencias bajas.

- *Contracción de los músculos del oído medio*: los músculos del oído medio se pueden contraer debido a múltiples causas (sonidos de alta intensidad, movimientos de los músculos de la boca, movimientos del cuerpo en general ...) y esto produce también una disminución de la señal recibida.

2.3.3. El oído interno.

El oído interno es la parte del oído donde se convierten las ondas sonoras en señales eléctricas (impulsos nerviosos). La parte fundamental del oído interno es la cóclea o caracol, que es un tubo de aproximadamente 32 milímetros de longitud enrollado en forma de espiral y relleno con distintos líquidos por el que se propaga una onda de presión. Dicho tubo está dividido longitudinalmente en mitades separadas por una membrana (Membrana Basilar), y ambas mitades se encuentran rellenas de líquido. El funcionamiento de la cóclea corresponde al de un sistema hidráulico de parámetros distribuidos que va filtrando la señal de presión con distintos filtros a medida que avanza por su interior. Dentro de este conducto, se disponen longitudinalmente los transeptores (llamados células ciliares) encargados de convertir la señal acústica en eléctrica [51]. Gran parte de los avances en el estudio de la cóclea se deben a von Gerog Békésy [52] que recibió el premio Nobel de medicina en 1961 por sus estudios de este órgano.

Conocer el comportamiento físico de la cóclea es esencial para entender los fenómenos relacionados con la audición; ya que el comportamiento de este órgano determina la respuesta en frecuencia del oído, los umbrales de percepción y las características no lineales del proceso de transducción entre otras cosas. Uno de los extremos de la cóclea está en contacto con el estribo (oído medio) y presenta una apertura que recibe el nombre de ventana oval; y el otro extremo (que recibe el nombre de helicotema) se encuentra en la parte más interna de la espiral. La onda de presión entra en la cóclea a través del estribo y se propaga hacia el helicotema, pero las características de propagación no homogéneas de la cóclea provocan que a medida que la onda avanza se vayan eliminando las altas frecuencias, por lo que no se forman ondas estacionarias dentro del tubo, sino que aparecen un tipo particular de ondas llamadas ondas viajeras (*traveling waves*). Esta forma de propagación hace que cada punto de la cóclea se comporte como un filtro hidráulico. De este modo, las altas frecuencias estarán presentes únicamente en el extremo en que se encuentra la ventana oval; mientras que las bajas frecuencias se propagarán hasta el final de la cóclea, apareciendo en cada punto de la cóclea una frecuencia de resonancia distinta. Al excitar el oído interno con una señal senoidal de frecuencia f_0 , se producirán vibracio-

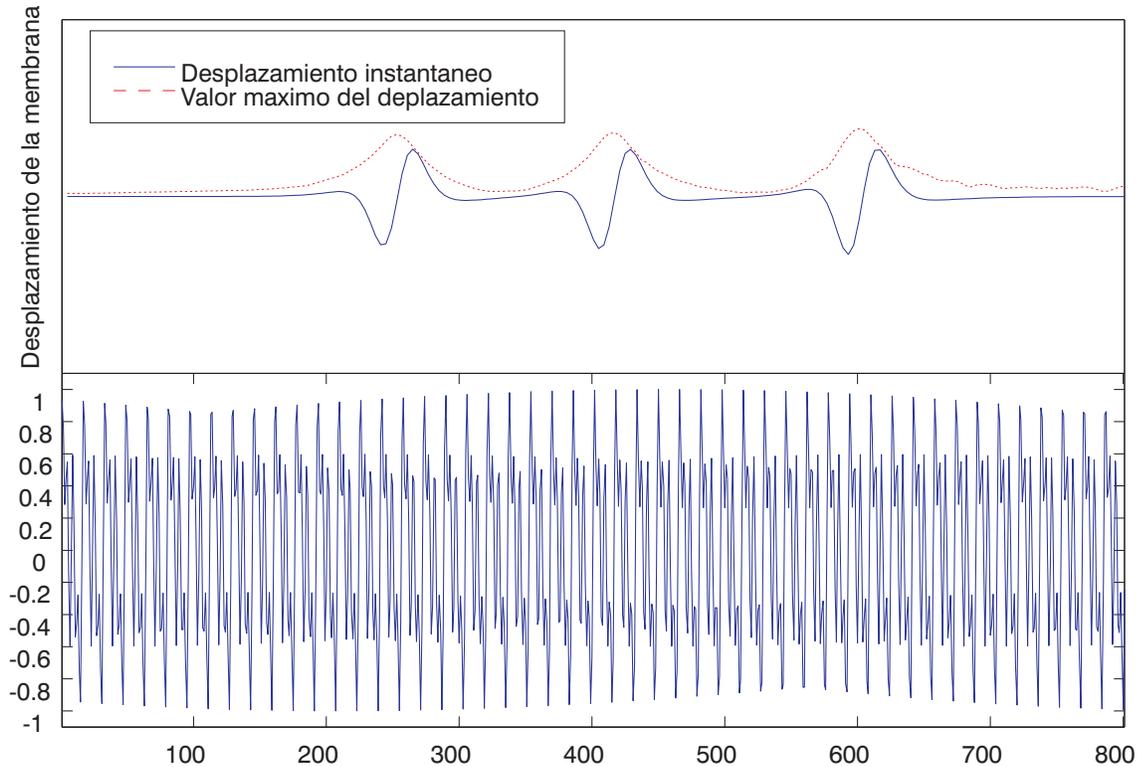


Figura 2.4: La parte superior de la figura representa los valores máximos de desplazamiento producidos en la membrana basilar cuando se excita el oído interno con una señal compuesta por tres tonos de frecuencias 500Hz, 1000Hz y 2000Hz de igual amplitud y sumados en fase. El eje de abscisas corresponde a distancias en la membrana basilar de la cóclea medidas desde el helicotema, mientras que el eje de ordenadas corresponde a desplazamientos (la línea continua representa el desplazamiento medido en un instante de tiempo y la discontinua los máximos alcanzados en cada punto al reproducir la señal). La parte inferior de la figura corresponde a la representación temporal de la señal excitación correspondiente a tres señales cosenos sumadas en fase.

nes a lo largo de la cóclea, que serán mayores en la posición cuya frecuencia de resonancia coincida con f_0 . Si la cóclea es excitada simultáneamente con varias señales senoidales, se producirán oscilaciones que presentarán máximos locales en los puntos cuya frecuencia de resonancia coincida con cada una de las frecuencias de excitación. Las frecuencias más bajas, penetran más en la cóclea y generan máximos de vibración en lugares próximos al helicotema, mientras que las frecuencias altas penetran menos en la cóclea y los puntos de máxima vibración asociados a ellas se encuentran más próximos a la ventana oval. La figura 2.4 corresponde a una simulación por ordenador (basada en el modelo de Cooke [53]) de los máximos niveles de desplazamiento registrados en la membrana basilar cuando se excita el oído interno con una señal compuesta por tres tonos simultáneos (500Hz, 1000Hz y 2000Hz).

Como se puede observar en la figura 2.4, la frecuencia de resonancia no se distribuye de forma lineal a lo largo de toda la cóclea sino que la zona de resonancia correspondiente

a las frecuencias bajas es mucho mayor que la correspondiente a frecuencias altas: la distancia entre el punto con frecuencia de resonancia 500Hz y el punto de frecuencia de resonancia 1000Hz es aproximadamente igual a la distancia entre el punto con frecuencia de resonancia 1000Hz y el punto de frecuencia de resonancia 2000Hz.

A lo largo de los 32 milímetros de longitud de la cóclea se encuentran las células ciliares, que son las encargadas de reaccionar a las vibraciones produciendo señales nerviosas que serán enviadas al cerebro mediante el nervio auditivo. Las células ciliares se clasifican en dos grupos: las células ciliares internas (que son las que realmente se encargan de generar las señales nerviosas) y las células ciliares externas (que interactúan con las células ciliares internas para formar sistemas amplificadores realimentados no lineales). Cada una de las células ciliares generan, de forma estadística, unos impulsos eléctricos que tendrán mayor probabilidad de aparecer cuando mayor sea la oscilación mecánica recibida por la célula.

Al nacer, los humanos disponemos aproximadamente de 3500 células ciliares internas, que van muriendo con la edad al ser excitadas con sonidos muy intensos, lo que ocasiona pérdida de audición en personas ancianas muy expuestas a ruidos.

A la vista de lo anterior, el oído interno se puede modelar como un banco de filtros lineales llamados filtros auditivos (3500 filtros con bandas de paso solapadas) donde la salida de cada uno de ellos se encuentra unida a un detector no lineal.

Al analizar las partes del oído humano, se puede observar que la complejidad va en aumento desde sus partes más externas hasta las más internas. El oído externo se modela fácilmente como un sistema mecánico lineal, el oído medio también puede ser modelado en la mayoría de las veces como un sistema lineal, aunque en este caso intervienen sistemas biológicos (los músculos que controlan la cadena de huesecillos) que pueden variar sus características. El modelado del oído interno es más complejo puesto que, aunque una primera parte es un sistema hidráulico con un modelo mecánico conocido, otra parte (la receptora) está formada por células ciliares cuya respuesta es no lineal y responde a comportamientos biológicos [54]. Actualmente, el grado de conocimiento del funcionamiento de la cóclea permite realizar implantes en su interior cuando se encuentra dañado, reemplazando el sistema hidráulico por circuitos electrónicos que simulan su comportamiento [55]. La señal nerviosa generada por las células ciliares es transportada al cerebro por el nervio auditivo [56, 57], que también se comporta de forma no lineal. Existen modelos informáticos que representan las señales acústicas o eléctricas presentes en cada uno de los elementos tratados hasta llegar al nervio auditivo [58, 59]. Otra dificultad en el avance de la ciencia en el conocimiento de las partes más profundas del oído estriba en que la realización de medidas precisa de la colocación de sensores en distintas partes del oído, lo cual no presenta problemas en el oído externo, pero en el resto de partes del oído la utilización de técnicas invasivas hace imposible la experimentación con seres humanos vivos, siendo sólo posible la experimentación con tejidos muertos o con animales [60, 61], cuyo comportamiento puede diferir del comportamiento del oído humano vivo.

Finalmente, además de la señal física captada por el sentido del oído, está su procesado o interpretación, y ahí el modelado es mucho más complejo puesto que a un mismo sonido recibido por distintos individuos se le pueden asignar características distintas: lo que a

uno le parece agradable a otro le puede parecer molesto, lo que a uno le parece relajante a otro le puede parecer estresante... Mientras que la captación de sonidos es algo similar en todos los individuos, su procesado e interpretación es algo muy particular de cada uno pues ahí interviene la psicología, educación, estado de ánimo, etc., de cada individuo.

2.4. Fundamentos de psicoacústica

La estructura del sistema auditivo humano [2.3](#) permite explicar gran parte de los efectos auditivos producidos por los sonidos. A continuación se resumen los efectos y sensaciones auditivos más relevantes en cuanto a la percepción humana del sonido, que han permitido el desarrollo de modelos y medidas psicoacústicas.

2.4.1. No linealidad en la percepción del nivel.

Aunque parte del sistema auditivo se ha modelado como un sistema lineal, el conjunto completo se comporta de forma no lineal debido, por un lado, a la conversión de señal acústica a impulsos nerviosos y, por otro lado, el procesado que hace el cerebro de los sonidos.

Este funcionamiento no lineal supone, por ejemplo, que cuando oímos un sonido de amplitud doble a la de otro, la sensación que percibimos no es la de un sonido el doble de intenso sino que simplemente se aprecia una sensación de nivel sonoro un poco más alto [[6](#), [3](#)]. Por otra parte, dos sonidos con la misma potencia, pero de diferente frecuencia, no se perciben igualmente intensos sino que la sensación que produce depende de la frecuencia que tienen. La figura [2.5](#) es una representación de lo que se conoce como curvas isofónicas, es decir, curvas de sonidos que producen la misma sensación de sonoridad (que se mide en sones). Las curvas isofónicas se obtienen experimentalmente pidiendo a un grupo de personas (jurado) que ajuste la intensidad de un sonido hasta que le cause la misma sensación que otro, normalmente se toma como referencia una señal de 1kHz y se pide al jurado que ajuste los niveles de señales de frecuencia distinta a ésta obteniendo cada una de las líneas representadas en la figura [2.5](#).

Una primera aproximación de la sensación de sonoridad la proporciona la ley de Stevens [[63](#)]:

$$L = k \cdot I^{0,3} \tag{2.15}$$

donde L mide la “sensación” de intensidad sonora (Loudness) k es una constante que depende de las unidades utilizadas e I representa la “magnitud física” de intensidad sonora.

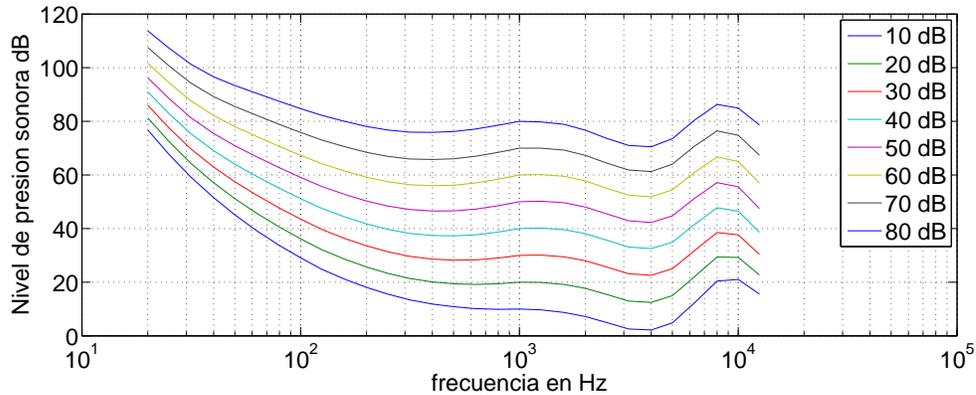


Figura 2.5: Curvas isofónicas representadas a partir de las tablas contenidas en [62]. Cada una de las curvas representa el nivel de presión sonora que debe tener un sonido para que cause la misma sensación de sonoridad que otro de frecuencia distinta. En el eje de abscisas se representa la frecuencia en Hz y en el de ordenadas el nivel de intensidad sonora en dB_{SPL} .

2.4.2. Enmascaramiento frecuencial.

El enmascaramiento consiste en la no percepción de un sonido (enmascarado) a causa de la existencia de otro (enmascarante). Esta es una propiedad muy importante del sistema auditivo que debe estar presente en cualquier modelo de percepción del oído. Existen dos tipos de enmascaramiento: el enmascaramiento temporal y el enmascaramiento frecuencial [3, 6, 64].

El enmascaramiento frecuencial es un efecto producido en el oído humano que hace que el sonido enmascarado sea imperceptible cuando se escucha *simultáneamente* con el sonido enmascarante. La capacidad de un sonido para enmascarar frecuencialmente a otro depende del nivel sonoro y de la frecuencia de ambos. Como ejemplo de este fenómeno, la figura 2.6 presenta la capacidad enmascarante de una frecuencia de 1kHz de distintos niveles sonoros.

En la figura 2.6 se presentan 4 líneas correspondientes al enmascaramiento producido por sonido enmascarante correspondiente a una señal de banda estrecha (160Hz de ancho de banda) centrada en la frecuencia de 1kHz con distintos niveles de intensidad (100dB, 80dB, 60dB, 40dB). Se puede observar que un sonido enmascarante puede enmascarar otros sonidos con niveles próximos al primero cuando la diferencia entre las frecuencias de los dos sonidos es pequeña. Sin embargo la capacidad enmascarante de un sonido decrece a medida que la diferencia entre las frecuencias de los sonidos enmascarado y enmascarante se hace mayor, disminuyendo la capacidad de enmascaramiento a razón de 100dB por octava cuando el sonido enmascarado tiene frecuencias inferiores que el enmascarante. También se observa en la figura 2.6 que el patrón de enmascaramiento presenta una forma simétrica cuando el sonido enmascarante es de un nivel bajo, mientras que cuando posee niveles altos esta simetría desaparece.

Desde el punto de vista fisiológico, el efecto de enmascaramiento frecuencial se puede

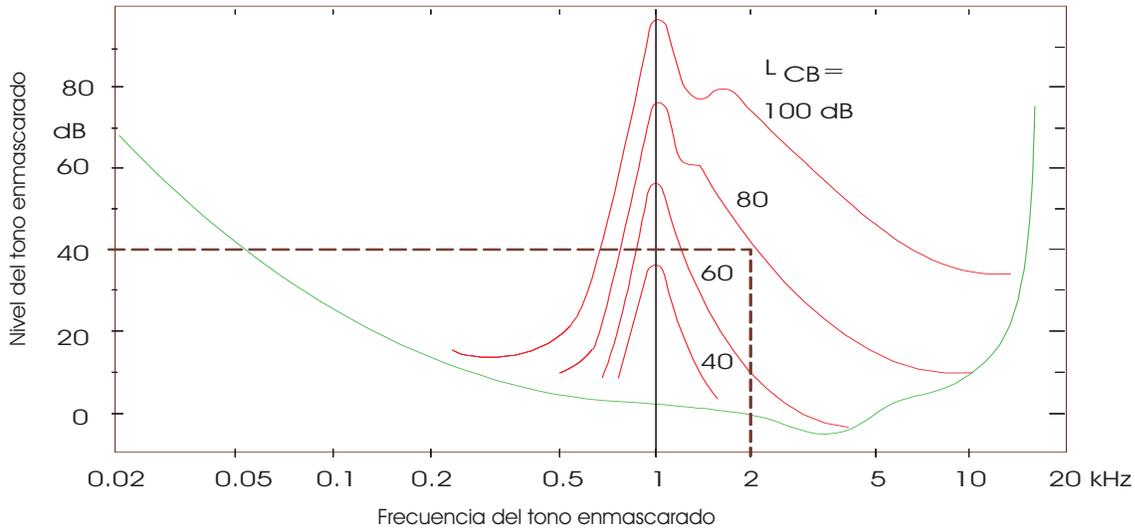


Figura 2.6: Capacidad enmascarante de una frecuencia de 1kHz [4]. Cada una de las líneas representa el nivel mínimo que debe de tener un tono para que, escuchado simultáneamente con el sonido enmascarante, resulte perceptible. Las distintas líneas corresponden a distintos niveles de sonidos enmascarantes. En este experimento el sonido enmascarante es una señal de banda estrecha (160Hz de ancho de banda) en lugar de un tono puro.

explicar mediante la utilización de la respuesta frecuencial de los filtros auditivos asociados a cada posición coclear.

2.4.3. Enmascaramiento temporal.

El enmascaramiento temporal es un efecto producido en el oído humano que hace que un sonido enmascarado sea imperceptible cuando se produce antes o después de otro sonido enmascarante. La capacidad de un sonido para enmascarar temporalmente a otro depende del nivel sonoro de ambos del tiempo transcurrido entre ellos y de la duración del sonido enmascarante. Cuando el sonido enmascarado se produce después del enmascarante el efecto se conoce como post-enmascaramiento o enmascaramiento hacia delante (post-masking o forward masking), mientras que si el orden es el inverso se habla de pre-enmascaramiento o enmascaramiento hacia atrás (pre-masking o backward masking). El efecto es similar al que se produce en el sistema visual humano, cuando después de ver una luz muy intensa el ojo queda cegado durante unos instantes. La dependencia del post-enmascaramiento del nivel de la señal enmascarante se puede apreciar en la figura 2.7, donde se representa en función del tiempo el enmascaramiento producido por tres señales.

En la figura 2.7 se puede apreciar la gran dependencia que el efecto de post-enmascaramiento tiene respecto al nivel del sonido enmascarante, puesto que la velocidad de caída de la curva de enmascaramiento es mucho más abrupta para los sonidos fuertes (80dB) que

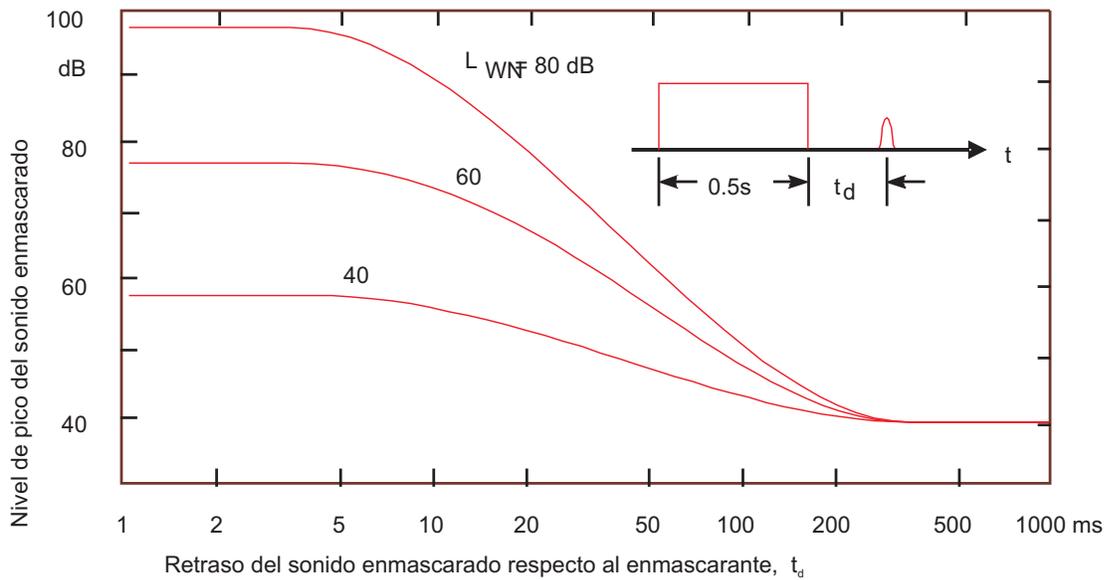


Figura 2.7: post-enmascaramiento dependiente del nivel de la señal enmascarante [4]. Cada una de las líneas representa el nivel mínimo que debe de tener un pulso gaussiano de $20 \mu\text{s}$ de duración, para que resulte perceptible escuchado posteriormente al de un sonido enmascarante de 0.5 segundos de duración compuesto por ruido blanco. Las distintas líneas corresponden a distintos niveles de sonidos enmascarantes (80dB, 60dB y 40dB). Las representaciones se realizan en función del tiempo transcurrido entre el final del sonido enmascarante y el ruido enmascarado.

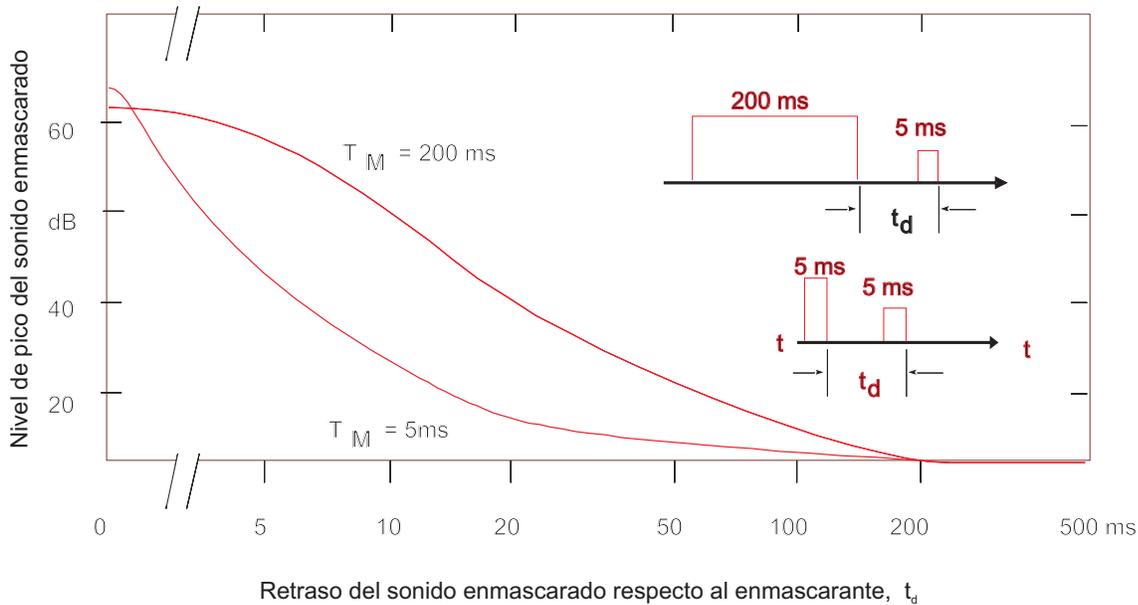


Figura 2.8: post-enmascaramiento dependiente de la duración de la señal enmascarante [4]. Cada una de las líneas representa el nivel mínimo que debe de tener un pulso de frecuencia 2 kHz y 5ms de duración, para que resulte perceptible escuchado posteriormente al de un sonido enmascarante. Las distintas líneas corresponden a distintas duraciones de sonidos (5ms y 200ms). Las representaciones se realizan en función del tiempo transcurrido entre el final del sonido enmascarante y el final del ruido enmascarado.

para los sonidos débiles (40dB). También se puede observar que el efecto de enmascaramiento desaparece aproximadamente 200 μ s después de finalizar el sonido enmascarante, independientemente del nivel que este tenga.

Por otro lado, el enmascaramiento temporal no depende únicamente de nivel de la señal enmascarante, sino también de la duración de ésta como se puede apreciar en la figura 2.8, donde se representa en función del tiempo el enmascaramiento producido por dos señales de distinta duración.

De la figura 2.8 se desprende que la relación entre el enmascaramiento y la duración del sonido enmascarante es altamente no lineal (al igual que ocurría con la dependencia del nivel) puesto que las dos señales enmascarantes de distinta duración analizadas presentan pendientes de decaimiento muy diferentes.

Desde el punto de vista fisiológico, el enmascaramiento temporal se debe al tiempo de respuesta de las células ciliares y a la saturación que sufren al ser sometidas a excitaciones intensas.

Los sistemas de compresión de audio [65, 66] utilizan estas características del oído para codificar señales con el menor número posible de bits. Esto se debe por un lado a que los sonidos enmascarados no se codifican, y por otro, a que se puede reducir el número de bits empleados en la cuantificación de tal modo que el ruido de cuantificación quede

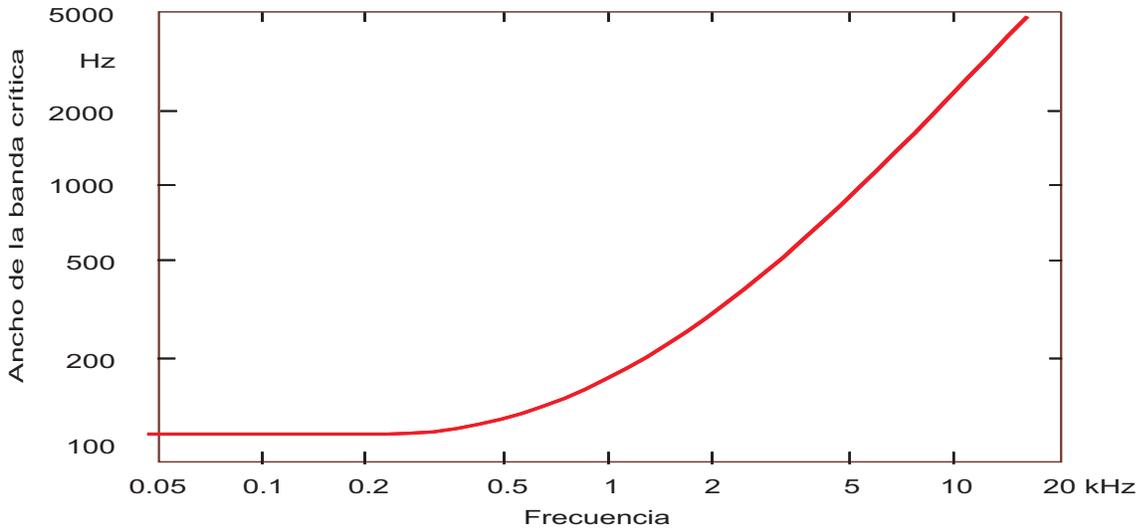


Figura 2.9: Ancho de banda de las bandas críticas en función de la frecuencia.

enmascarado por la señal deseada.

2.4.4. No linealidad en la percepción de la frecuencia.

Puesto que las células ciliares existentes en el oído interno son las encargadas de transformar las vibraciones de la membrana basilar en impulsos nerviosos, y cada una de ellas responde a un rango de frecuencias dependiendo de su posición en la cóclea comportándose como un filtro paso banda, al aplicar al oído sonidos de banda estrecha con diferente ancho de banda, serán percibidos de forma distinta según exciten a una o varias de estas células.

Partiendo de este modelo de banco de filtros y con el fin de determinar la anchura espectral de cada uno de estos filtros, se han realizado experimentos psicoacústicos basados en el hecho de que los sonidos del mismo nivel cuya anchura espectral es superior a la anchura del filtro se perciben de forma distinta que si la anchura es inferior, puesto que cuando el sonido tiene mayor ancho de banda que el filtro, una parte de su potencia pasará a través del filtro adyacente. Estos experimentos han permitido obtener unos anchos espectrales dependientes de la frecuencia (figura 2.9) que se han denominado bandas críticas (ya que al sobrepasar este ancho de banda el sonido se percibe diferente).

Las bandas críticas mostradas en la figura 2.9 se pueden aproximar en función de la frecuencia mediante la expresión:

$$\Delta f_G(\text{Hz}) = 25 + 75 \left(1 + 1/4 f^2(\text{kHz})\right)^{0.69} \quad (2.16)$$

Un modelo muy extendido de percepción del oído consiste en descomponer los sonidos haciéndolos pasar a través un banco de filtros ideales (cuya respuesta frecuencial es 1 en la banda de paso y 0 en la banda atenuada) de ancho de banda una banda crítica; y con

Filtro	1	2	3	4	5	6	7	8	9	10	11	12
límite inferior	20	100	200	300	400	510	630	770	920	1080	1270	1480
límite superior	100	200	300	400	510	630	770	920	1080	1270	1480	1720
Filtro	13	14	15	16	17	18	19	20	21	22	23	24
límite inferior	1720	2000	2320	2700	3150	3700	4400	5300	6400	7700	9500	12000
límite superior	2000	2320	2700	3150	3700	4400	5300	6400	7700	9500	12000	15500

Tabla 2.1: banco de filtros de banda crítica

el final de la banda de paso coincidente con el principio de la banda de paso del filtro siguiente. Este banco de filtros abarca las frecuencias comprendidas entre 0Hz y 15500Hz y las frecuencias de corte de cada banda son las que se muestran en la tabla 2.1:

De esta forma, en lugar de trabajar con la cantidad de potencia o energía que contiene una determinada frecuencia se suele trabajar con la cantidad de potencia o energía contenida en cada una de estas bandas (puesto que representa mejor el funcionamiento del oído humano), es decir, en lugar de trabajar en una escala de frecuencias se suele trabajar con esta escala de bandas llamada escala Bark donde la frecuencia 100 corresponde a 1 Bark, la frecuencia 200 a 2 Bark, etc. El término que se utiliza en honor a Barkhausen, que fue el primer científico que utilizó el fonio para medir el nivel de sonoridad (Loudness Level).

La relación entre estas dos escalas (escala de frecuencias y escala Bark) se puede aproximar mediante la expresión:

$$z(\text{Bark}) = 13 \arctan(0'76f(\text{kHz})) + 3'5 \arctan(f(\text{kHz})/7'5)^2 \quad (2.17)$$

Zwicker [3] comprobó que la sensación de frecuencia producida por los sonidos medida en mel, guarda una relación lineal con la medida de la frecuencia del sonido en la escala Bark.

A pesar de que esta escala se emplea ampliamente en psicoacústica, resulta imprecisa para algunos autores por suponer que los filtros que modelan el comportamiento del oído tienen forma rectangular y por la forma en que se obtiene esta escala [6, 67].

La ecuación (2.17) fue modificada por H. Traunmüller [68] en 1990 mediante una fórmula que se ajusta a los resultados experimentales con mayor exactitud:

$$z'(\text{Bark}) = \frac{26,81}{1 + \frac{1960}{f}} - 0,53 \quad (2.18)$$

Pero esta expresión presenta cierto sesgo en las frecuencias situadas en los extremos superior e inferior del rango de frecuencias audibles que se corrige añadiendo al término anterior una pequeña corrección:

$$z = \begin{cases} z' + 0,15 * (2 - z') & \text{Si } z' < 2 \\ z' + 0,22 * (z' - 20,1) & \text{Si } z' > 20,1 \\ z' & \text{En cualquier otro caso} \end{cases} \quad (2.19)$$

2.5. Parámetros acústicos.

En esta sección y la siguiente, se van a analizar los principales parámetros utilizados en la medida de sonidos. Estos parámetros se van a dividir en dos grupos: parámetros acústicos y parámetros psicoacústicos.

Al hablar de parámetros psicoacústicos nos referiremos a aquellas características de los sonidos que dependen de la interpretación humana y el procesado realizado por el cerebro, mientras que al referirnos a parámetros acústicos nos referiremos a aquéllos en los que se miden características físicas de los sonidos, sin tratar de buscar interpretaciones psicológicas de éstos.

La medida más habitual realizada sobre un sonido es el nivel de presión sonora (SPL) cuya unidad es el decibelio y está medido tomando como referencia la presión acústica de 20 micropascales, aunque también es habitual aplicar a estas medidas algún tipo de ponderación temporal y frecuencial. El instrumento empleado para realizar mediciones de presión sonora es el sonómetro, del que existen dos versiones: un sonómetro básico (normalmente analógico) cuyo funcionamiento y características han sido normalizados en la norma UNE-EN606051 [69], y un sonómetro avanzado (sonómetro integrador) realizado mediante técnicas digitales cuyo funcionamiento y características han sido normalizados en la norma UNE-EN60804 [70]. A continuación se exponen algunos de los procedimientos y medidas realizadas comúnmente por los *sonómetros*.

2.5.1. Ponderación frecuencial.

A pesar de que las medidas acústicas no tratan de medir las sensaciones producidas por el sonido, la palabra sonido implica frecuencias audibles por el oído humano normal, por lo que no sería lógico dar como medida de sonido en un punto el valor cuadrático medio del nivel de presión sonora en dicho punto, puesto que algunas frecuencias no son captadas por el oído y no pueden ser consideradas sonido. Del mismo modo frecuencias próximas a los límites de audición (20Hz y 20000Hz), que son muy atenuadas por el sistema auditivo, no pueden ser tratadas del mismo modo que frecuencias próximas a 1kHz, donde el oído es más sensible.

Por esta razón suele ser habitual filtrar los sonidos con filtros que se asemejen a la respuesta frecuencial del oído humano antes de medirlas. Pero como el oído se comporta de forma no lineal y presenta características diferentes según responda a niveles altos o bajos, fundamentalmente existen tres formas distintas de caracterizar el filtrado realizado por el oído (figura 2.10): ponderación A, ponderación B y ponderación C. La ponderación A modela la curva isofónica (figura 2.5) correspondiente a 40 dB, la ponderación B la

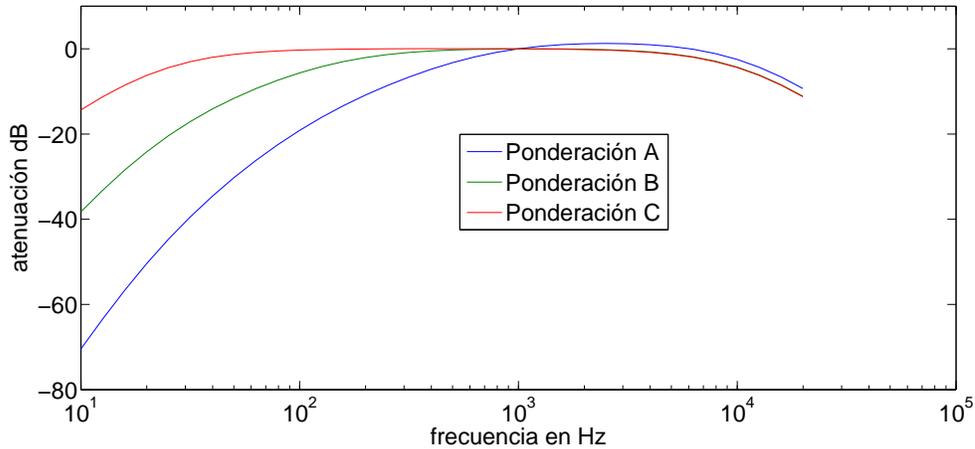


Figura 2.10: Atenuación correspondiente a distintas ponderaciones del sonido.

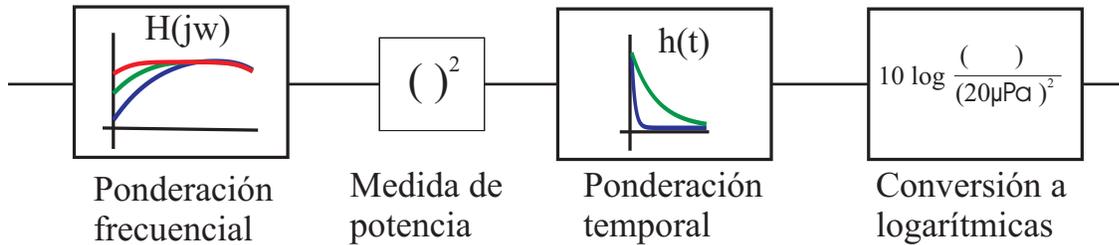


Figura 2.11: Esquema de funcionamiento de un sonómetro.

correspondiente a 70 y la ponderación C a 100 [71]. Puesto que los niveles más habituales en el entorno cotidiano están más próximos a los 40dB, la ponderación A es la más utilizada.

2.5.2. Ponderación Temporal.

Al realizar medidas sobre un sonido, este puede variar muy rápidamente por lo que a veces, las medidas instantáneas no son demasiado útiles y se procede a ponderarlas temporalmente. De esta forma el esquema habitual de la medida nivel de presión acústica realizada por un sonómetro consiste en: primero aplicar una ponderación frecuencial al sonido captado si se estima oportuno, segundo medir su potencia y tercero ponderar temporalmente la medida obtenida. Por lo que el esquema general de un sonómetro queda como se indica en la figura 2.11.

Habitualmente se emplean dos tipos distintos de ponderaciones temporales en función de la velocidad a la que varía el parámetro que se está midiendo: la ponderación rápida (Fast), que está indicada para aquellos sonidos que presentan pocas fluctuaciones; y la ponderación lenta (Slow) que está indicada para la medida de aquellos sonidos que presentan mayor variabilidad.

Los circuitos encargados de realizar las ponderaciones temporales deben ser sistemas

lineales e invariantes, al igual que los ponderadores frecuenciales, pero así como aquéllos se modelaban en frecuencia mediante su respuesta frecuencial, éstos se modelan en el tiempo mediante su respuesta temporal, que es:

$$h(t) = e^{-\frac{t}{\tau}} \quad (2.20)$$

donde

t es el tiempo y

τ es la constante de tiempo, que vale 125 milisegundos para la ponderación fast y 1000 milisegundos para la ponderación slow.

Además de las ponderaciones fast y slow muchos sonómetros disponen de la función impulsive, que permite medir variaciones mucho más rápidas de los sonidos.

2.5.3. Análisis en bandas de octava y tercio de octava.

El modelo de sonómetro representado en el apartado anterior proporciona una medida global del nivel de intensidad sonora que puede estar ponderada de forma similar a la respuesta del oído humano. Esta medida global en ocasiones no es suficiente información puesto que a menudo resulta interesante conocer las características que los sonidos presentan en distintas frecuencias. En estas situaciones se descompone el sonido en bandas de frecuencia que han sido normalizadas en acústica según la norma UNE-EN61260 [72] y que se obtienen a partir de una frecuencia fundamental que es 1000Hz.

Según la precisión frecuencial que se desee obtener de las señales, se utilizan filtros de banda de octava o de tercio de octava (que proporcionan mayor resolución frecuencial puesto que los filtros son más estrechos).

La norma permite trabajar con los filtros obtenidos en dos bases distintas:

-Base 10: la frecuencia central de un filtro de octava se obtiene multiplicando la frecuencia central del filtro de octava anterior por el factor $10^{3/10} = 1,9953$. La frecuencia central de un filtro de tercio de octava se obtiene multiplicando la frecuencia central del filtro de tercio de octava anterior por el factor $10^{1/10} = 1,2589$.

-Base 2: la frecuencia central de un filtros de octava se obtiene multiplicando la frecuencia central del filtro de octava anterior por el factor 2. La frecuencia central de un filtros de tercio de octava se obtiene multiplicando la frecuencia central del filtro de tercio de octava anterior por el factor $2^{1/3} = 1,2599$.

En ambos casos todos los filtros que componen el banco de filtros se obtienen a partir de la frecuencia 1000Hz multiplicando o dividiendo por el factor correspondiente. Los resultados son muy similares (aunque se recomienda utilizar la base 10) y los valores obtenidos son los indicados en las tablas 2.2 y 2.3.

Índice	frecuencia exacta en base 10	frecuencia exacta en base 2	frecuencia nominal central
-16	25,119	24,803	25
-15	31,623	31,250	31,5
-14	39,811	39,373	40
-13	50,119	49,606	50
-12	63,096	62,500	63
-11	79,433	78,745	80
-10	100,000	99,213	100
-9	125,89	125,00	125
-8	158,49	157,49	160
-7	199,53	198,43	200
-6	251,19	250,00	250
-5	316,23	314,98	315
-4	398,11	396,85	400
-3	501,19	500,00	500
-2	630,96	629,96	630
-1	794,33	793,70	800
0	1000,0	1000,0	1000
1	1258,9	1259,9	1250
2	1584,9	1587,4	1600
3	1995,3	2000,0	2000
4	2511,9	2519,8	2500
5	3162,3	3174,8	3150
6	3981,1	4000,0	4000
7	5011,9	5039,7	5000
8	6309,6	6349,6	6300
9	7943,3	8000,0	8000
10	10000,0	10079,4	10000
11	12589	12699	12500
12	15849	16000	16000
13	19953	20159	20000

Tabla 2.2: frecuencias centrales de los filtros de tercio de octava utilizados en acústica según la norma UNE-EN61260 [72].

Índice	frecuencia exacta en base 10	frecuencia exacta en base 2	frecuencia nominal central
-5	31,623	31,250	31,5
-4	63,096	62,500	63
-3	125,89	125,00	125
-2	251,19	250,00	250
-1	501,19	500,00	500
0	1000,0	1000,0	1000
1	1995,3	2000,0	2000
2	3981,1	4000,0	4000
3	7943,3	8000,0	8000
4	15849	16000	16000

Tabla 2.3: frecuencias centrales de los filtros de octava utilizados en acústica según la norma UNE-EN61260 [72].

2.5.4. Parámetros estadísticos.

Cuando se realizan medidas de señales durante periodos largos de tiempo y se registran los resultados, a menudo se cuenta con una gran cantidad de datos que pueden ser resumidos mediante sus percentiles, deciles o cuartiles.

Percentiles son 99 valores que distribuyen la serie de datos, ordenada de forma creciente o decreciente, en cien tramos iguales, en los que cada uno de ellos concentra el 1% de los resultados.

Deciles son 9 valores que distribuyen la serie de datos, ordenada de forma creciente o decreciente, en diez tramos iguales, en los que cada uno de ellos concentra el 10% de los resultados. Los deciles están relacionados con los percentiles, puesto que el decil 1 coincide con el percentil 10, el decil 2 coincide con el percentil 20, etc.

Cuartiles son 3 valores que distribuyen la serie de datos, ordenada de forma creciente o decreciente, en cuatro tramos iguales, en los que cada uno de ellos concentra el 25% de los resultados. Los cuartiles están relacionados con los percentiles y deciles, así pues el cuartil 2 coincide con el percentil 50 o el decil 5.

Cuando se mide el ruido de tráfico se trabaja con los percentiles 10, 50 y 90 de los niveles sonoros ofrecidos por un sonómetro [73].

También es habitual tratar los datos proporcionados por los sonómetros como variables aleatorias y hacer medidas sobre ellas:

Medidas de localización

Estas medidas indican el valor central de una variable alrededor del cual se concentra su función densidad de probabilidad.

La media de una variable aleatoria (o esperanza matemática) coincide con su momento de orden 1 alrededor del origen [74].

Se define momento de orden k alrededor del origen de una variable aleatoria X a la esperanza matemática de la variable elevada a k :

$$\mu'_k = E[(X)^k] \quad (2.21)$$

Cuando no se dispone de la función de distribución de la variable aleatoria, su momento de orden k alrededor del origen se puede estimar a partir de una serie correspondiente a n observaciones mediante la expresión:

$$\widetilde{\mu}'_k = \frac{1}{n} \sum_{i=1}^n (X_i)^k \quad (2.22)$$

Siendo X_i el valor obtenido en la observación i de la variable aleatoria.

El apóstrofe del símbolo μ'_k utilizado para designar el momento alrededor del origen y de su estimación $\widetilde{\mu}'_k$, se ha empleado para distinguirlo de μ_k y $\widetilde{\mu}_k$, que representa el momento centrado de orden k y su estimación respectivamente. El momento centrado de orden k se define como la esperanza matemática de la variable menos su media elevada a k :

$$\mu_k = E[(X - E[X])^k] \quad (2.23)$$

El momento centrado de orden 1 de cualquier variable aleatoria es 0. Al igual que el momento alrededor del origen, el momento centrado puede ser estimado a partir de una serie de n observaciones. En este caso se emplea la ecuación:

$$\widetilde{\mu}_k = \frac{1}{n} \sum_{i=1}^n (X_i - \widetilde{\mu}_1)^k \quad (2.24)$$

La mediana de una serie de n valores previamente ordenados de forma creciente se define como el valor que deja a cada lado (por encima y por debajo) la mitad de los valores de la muestra (y coincide con su percentil 50).

La moda de una serie de n valores es el valor que más se repite.

Medidas de dispersión

Estas medidas indican la forma en la que la variable se distribuye alrededor de un valor central midiendo si su función densidad de probabilidad está concentrada o no.

La varianza de una variable aleatoria se define como su momento centrado de orden 2 (2.23) y se designa con el símbolo σ^2 [74]:

$$\sigma^2 = \mu_2 \quad (2.25)$$

La varianza mide la distancia existente entre los valores de la serie y la media, y su unidad de medida es la unidad de medida de la variable aleatoria al cuadrado.

Cuando no se dispone de la función de distribución de la variable aleatoria, su varianza se puede estimar a partir de una serie correspondiente a una cantidad finita de observaciones realizando el sumatorio de las diferencias al cuadrado entre cada observación y la media y dividiendo por el tamaño de la muestra:

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \tilde{\mu}_1)^2 \quad (2.26)$$

La varianza siempre será mayor que cero. Cuanto más bajo sea su valor, más concentrados están los valores de la serie alrededor de la media. Por el contrario, mientras mayor sea la varianza, más dispersos están.

La desviación típica es la raíz cuadrada de la varianza y se designa con el símbolo σ :

$$\sigma = \sqrt{\sigma^2} = \sqrt{\mu_2} \quad (2.27)$$

la desviación típica de una variable aleatoria tiene las mismas unidades que la variable aleatoria, y proporciona la misma información que la varianza.

El rango de una serie de valores se define como la diferencia entre el valor máximo y el mínimo de la serie.

Medidas de forma

Estas medidas proporcionan información sobre la forma de una función densidad de probabilidad indicando si ésta presenta asimetrías o picos y si éstos son pronunciados.

La asimetría o skewness es un parámetro adimensional que se define como el tercer momento estandarizado, se calcula dividiendo el tercer momento entre el cubo de la desviación típica, y se representa mediante el símbolo γ_1 :

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \quad (2.28)$$

El concepto de asimetría de una función densidad de probabilidad o de una serie se refiere a si la curva que forman sus valores presenta la misma forma a izquierda y derecha de un valor central (media):

- si $\gamma_1 = 0$ se dice que la distribución es simétrica: existe la misma concentración de valores a la derecha y a la izquierda de la media
- si $\gamma_1 > 0$ se dice que la distribución es asimétrica positiva: la falda de la función densidad de probabilidad a la derecha del centro es más larga que a la izquierda por lo que esta función está más concentrada a la izquierda y más dispersa a la derecha.
- si $\gamma_1 < 0$ se dice que la distribución es asimétrica negativa: la falda de la función densidad de probabilidad a la izquierda del centro es más larga que a la derecha por lo que esta función está más concentrada a la derecha y más dispersa a la izquierda.

La Curtosis (o apuntamiento) es un parámetro adimensional que se calcula restando 3 al cuarto momento estandarizado (cociente entre el momento de orden cuatro y la desviación típica elevada a cuatro), y se representa mediante el símbolo γ_2 :

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 \quad (2.29)$$

Se definen 3 tipos de distribuciones según su grado de curtosis:

- si $\gamma_2 = 0$ se dice que la distribución es mesocúrtica: presenta un grado de concentración medio alrededor de los valores centrales de la variable (el mismo que presenta una distribución normal).
- si $\gamma_2 > 0$ se dice que la distribución es leptocúrtica: presenta un elevado grado de concentración alrededor de los valores centrales de la variable.
- si $\gamma_2 < 0$ se dice que la distribución es platicúrtica: presenta un reducido grado de concentración alrededor de los valores centrales de la variable.

2.5.5. Parámetros espectrales.

Cuando se analiza el ruido producido por el motor de un coche, el tipo de señal es muy característico, pues se puede modelar como la suma de una señal periódica y un ruido de banda ancha. La parte periódica del ruido de motor está formada por armónicos de la frecuencia de rotación del motor. En algunas ocasiones aparecen armónicos de la mitad de esta frecuencia que reciben el nombre de armónicos de medio orden. Parámetros asociados a la distribución espectral de los armónicos de motor como su contenido en armónicos de medio orden o su relación con otros parámetros están relacionados con efectos auditivos asociados a estas señales como el retumbar (Rumble noise) o la turbiedad (muddiness) [18]. Al trabajar con este tipo de señales se utilizan parámetros asociados a la distribución de armónicos como: suma de la potencia total de los armónicos de medio orden, suma de la potencia total de los armónicos de orden completo, potencia de los armónicos de orden par e impar y la relación entre ellos.

2.6. Parámetros psicoacústicos.

En los apartados siguientes se presentarán los parámetros psicoacústicos más utilizados, sus definiciones, unidades y la forma de calcularlos cuando exista un modelo para su cálculo. Todos estos parámetros fueron estudiados a partir de la experimentación en multitud de tests psicológicos por el profesor Eberhard Zwicker [3], quien propuso métodos para calcularlos, por lo que para cada método se dedica un apartado al desarrollo del método de cálculo propuesto por Zwicker. Existen otros trabajos que explican muchos de los parámetros desarrollados por Zwicker desde la perspectiva del análisis de la fisiología del oído [6].

El Loudness ha sido normalizado para señales estacionarias mediante normas nacionales o internacionales (DIN 45631/A1 [75], ISO 532/UNE 74014:1978 [1], ANSI S3.4 [76]), por lo que existe una forma estándar de calcularlo (propuesto por Zwicker), mientras que el resto de parámetros no han sido normalizados y han aparecido nuevos métodos de cálculo que tratan de mejorar las estimaciones respecto a las obtenidas por los métodos propuestos por Zwicker para obtener valores que se ajusten más a los resultados obtenidos experimentalmente. El cálculo del Sharpness también ha sido estandarizado mediante la norma alemana DIN 45692 [2], y el cálculo del Roughness está actualmente siendo considerado por DIN [77] para establecer una futura norma. La existencia de parámetros no normalizados todavía hace que los resultados proporcionados por los distintos fabricantes de software sean diferentes [11].

Todos estos parámetros psicoacústicos son ampliamente utilizados en el diseño de sonidos [78, 79, 15, 80] o el análisis de las sensaciones que los sonidos provocan en los humanos [81, 82, 83, 84, 73, 85, 86].

2.6.1. Escalas de frecuencia.

La unidad utilizada en física para la medida de la frecuencia de una señal periódica es el hercio, pero la sensación subjetiva de frecuencia no es proporcional a la frecuencia física (sec 2.4.4), por lo que han aparecido nuevas escalas de frecuencias más acordes a la sensación subjetiva de frecuencia (y que no son proporcionales a las escalas de frecuencias en Hercios).

Escalas logarítmicas.

En el estudio de la música, es habitual hablar de octavas. Esta forma de medir las frecuencias deriva del cálculo del logaritmo en base 2 de la frecuencia física de un sonido (sec. 2.5.3). También es habitual hablar de décadas, que es una medida de frecuencias basada en logaritmos decimales. Tanto la medida de frecuencias en décadas como en octavas, supone una buena aproximación de las frecuencias subjetivas cuando se trabaja con frecuencias superiores a 500Hz. De ahí que en acústica se utilicen las bandas de octava y de tercio de octava (sec. 2.5.3).

Posición coclear.

La parte del oído interno que descompone los sonidos en distintas bandas de frecuencia es la cóclea. Al analizar este órgano, se puede observar que cada posición presenta su propia frecuencia de resonancia (la del filtro auditivo asociado a esa posición). Estudios realizados sobre el oído han permitido obtener la frecuencia de resonancia de cada punto de la cóclea:

$$f = 165'4(10^{0'06 \cdot x} - 1) \quad (2.30)$$

donde:

x es la posición de la cóclea en mm.

f es la frecuencia de resonancia en Hz.

Y reordenando la expresión anterior se deduce que la posición coclear asociada a una frecuencia dada es:

$$x = \frac{1}{0'06} \log_{10} \left(\frac{f - 165'4}{165'4} \right) \quad (2.31)$$

A partir de este punto se puede establecer una escala de posiciones cocleares.

Escala Bark.

Esta escala, muy utilizada en psicoacústica, se ha obtenido mediante experimentos subjetivos, donde los sujetos han respondido de forma distinta a los estímulos en función de su ancho de banda. De esta forma se ha modelado parte del oído como un banco de filtros rectangulares de ancho dependiente de la frecuencia central del filtro como se describió en la sección 2.4.4.

La escala Bark mide sencillamente el número de filtro auditivo asociado a cada frecuencia. La escala Bark se analizó en la sección 2.4.4 y su relación con la escala de frecuencias medida en hercios corresponde a las ecuaciones (2.17), (2.18) y (2.19).

Escala ERB.

La escala Bark modela el oído como un banco de filtros rectangulares, cosa que no es del todo cierta y que no coincide con el funcionamiento físico del oído. Por otra parte la forma de obtener experimentalmente la escala Bark presenta algunos problemas, por lo que se ha desarrollado otra escala, similar a la escala Bark, basada en suposiciones más acordes con la fisiología de la cóclea [6, 87, 88, 28].

La respuesta en frecuencia de un filtro auditivo se puede aproximar mediante filtros Gammatone [53, 89, 90, 91], cuya respuesta impulsional es:

$$h(t) = t^{\gamma-1} \cdot \tilde{\alpha}^t \quad (2.32)$$

Siendo γ el orden del filtro (que se toma de orden 4 [91]) y $\tilde{\alpha}$ una constante compleja que permite fijar la frecuencia central y el ancho de banda del filtro (y cuyo valor depende del filtro auditivo que se desee modelar). Una versión mejorada de este modelo consiste utilizar la función gammachirp [92] consistente en una modificación de la función gammatone mediante el uso de la función chirp.

En este modelo, el oído se representa mediante un banco de filtros auditivos (no rectangulares) del que se ha calculado, de forma experimental, su ancho de banda rectangular equivalente (ERB).

El ERB depende de la frecuencia de los sonidos. Si partiendo de la frecuencia 0 se divide el espectro en segmentos de anchura igual al ancho de banda del filtro rectangular equivalente asociado a cada frecuencia, se obtiene la escala ERB que está relacionado con la escala en Hz [88] mediante la expresión:

$$x = 21'4 \log_{10}(4,37f + 1) \quad (2.33)$$

donde

f es la frecuencia en kilohercios y

x es la frecuencia medida en escala ERB.

2.6.2. Pitch.

Este parámetro sólo está definido para sonidos periódicos y representa la estimación subjetiva de su frecuencia fundamental [3, 6, 87].

A frecuencias bajas, se observa que un sonido correspondiente a una señal senoidal pura de doble frecuencia que otro, produce una sensación de frecuencia subjetiva del doble que la primera. Pero, esta relación no se conserva al aumentar la frecuencia de los sonidos, puesto que al tratar con sonidos de frecuencias muy altas, duplicar la frecuencia de éstos supone tan sólo un pequeño incremento de la sensación subjetiva de frecuencia o Pitch.

El Pitch o frecuencia subjetiva de un sonido periódico tiene como unidad el mel y el cálculo de este parámetro es de gran interés cuando se trabaja con instrumentos musicales, ya que estos generan señales periódicas o cuasiperiódicas.

El **símbolo** utilizado para este parámetro es H_F y se mide en **mel**.

Referencia: Un tono puro de 125 Hz tiene un Pitch de 125 mel.

Pitch de Zwicker.

Según Zwicker [3] el Pitch de una señal senoidal pura se calcula mediante la expresión:

$$H_F = (f_T/Hz)(1 + v) \quad (2.34)$$

donde:

f_T es la frecuencia del tono y

v es la desviación de pitch que se descompone en dos partes:

$$v = v_L + v_M \quad (2.35)$$

siendo:

v_L la componente debida a la dependencia del pitch con el nivel del sonido y

v_M la componente debida al enmascaramiento.

Para tonos complejos formados por varias señales senoidales puras, el pitch corresponde a la frecuencia fundamental del sonido (que coincide con la separación frecuencial existente entre las señales senoidales que lo forman).

2.6.3. Loudness

El Loudness (o sonoridad) de una señal acústica es una característica determinante a la hora de evaluar su calidad subjetiva. La sonoridad de una señal de banda estrecha está en función de su frecuencia, siendo las frecuencias próximas a 4 kHz las que mayor sonoridad tienen para un mismo nivel de presión; y las frecuencias más altas y más bajas dentro del margen de frecuencias perceptibles las que menos. Los métodos de cálculo del Loudness tienen en cuenta el modelo del oído como analizador espectral y por tanto también los enmascaramientos temporales y frecuenciales que se producen [3, 87, 93].

Los sonómetros que se encuentran en el mercado no miden loudness sino SPL con distintas ponderaciones. Pero el loudness se ha introducido como un parámetro superior al nivel de presión sonora con ponderación A porque muestra mayor correspondencia con la sensación subjetiva de volumen. En [94] se estudia como combinar los distintos parámetros medidos por un sonómetro convencional para aproximar el valor de Loudness.

El modelo de Loudness desarrollado por Zwicker [3] ha sido trasladado a la norma ISO-532-B [1] y ha sido objeto de mejoras propuestas por Moore [88] (que utiliza la escala ERB en lugar de la Bark) y por Chalupper [95] (que permite simular el funcionamiento de oídos dañados).

El **símbolo** utilizado para el Loudness es N y se mide en **sones (o sonios)**.

Referencia: Un tono puro de 40 dB SPL y frecuencia de 1kHz tiene Loudness de 1 sone.

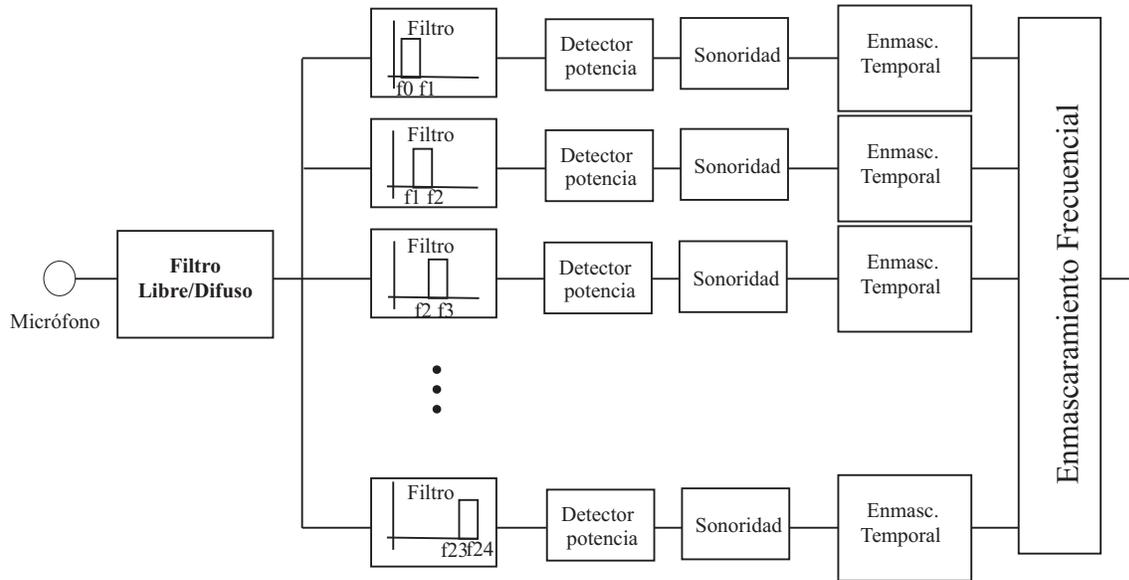


Figura 2.12: Esquema del método de cálculo de loudness propuesto por Zwicker.

Loudness específico.

El modelo de cálculo de Loudness propuesto por Zwicker [3] representado en la figura 2.12, calcula el Loudness de una señal como la suma del Loudness asociado a cada una de las bandas de frecuencia en que se descompone la señal. El Loudness asociado a cada banda recibe el nombre de *Loudness Específico*, se mide en sones/bark y se representa por el símbolo N' , donde el apóstrofe significa específico (es decir particularizado en una banda). El adjetivo *Específico* se utiliza también para otros parámetros psicoacústicos cuando se hace referencia a la distribución del parámetro entre las bandas críticas y se representa añadiendo un apóstrofe al símbolo del parámetro respectivo.

La figura 2.12 describe el procedimiento de cálculo del Loudness mediante circuitos electrónicos o programas informáticos: Después de filtrar la señal con un filtro de campo libre o campo difuso en función de la dirección de llegada de la señal, se divide ésta en 24 bandas (Bandas críticas) de 1 Bark de anchura.

La señal obtenida a la salida de cada uno de los 24 filtros anteriores se introduce en un estimador de potencia (formado por un rectificador de señal y un filtro paso bajo). La potencia detectada en cada canal se convierte a nivel de sonoridad, mediante una función no lineal que depende de la banda de trabajo.

Al nivel de sonoridad de cada banda se le realiza un proceso que modela el enmascaramiento temporal. En este proceso se calcula el nivel de sonoridad de un determinado instante a partir de la salida del bloque anterior en el instante actual y en los anteriores.

Por último se modela el enmascaramiento frecuencial mediante un bloque que calcula el nivel de sonoridad en una banda para un instante determinado, a partir de los valores de sonoridad en ese instante, de la banda calculada y sus bandas contiguas. De esta forma

se obtiene un nivel de sonoridad para cada banda en un instante de tiempo (*Loudness específico*).

Si se suman, para un instante de tiempo los valores correspondientes al Loudness específico de todas las bandas se obtiene el Loudness instantáneo y al promediar el Loudness instantáneo para un determinado intervalo de tiempo se obtiene el Loudness total en dicho intervalo.

Este modelo ha sido la base de un programa normalizado para calcular el Loudness publicado en la norma DIN45631 [75] y también ha sido la base para otro método que permite el cálculo del loudness de los sonidos estacionarios de forma gráfica que se ha publicado en la norma ISO 532-B [1].

Fundamentos del Loudness de Zwicker

El modelo propuesto por Zwicker ha sido adoptado por la norma ISO-532-B [1] que es de aplicación para señales con cualquier tipo de distribución espectral pero con características temporales estacionarias debido a la utilización de filtros estrechos cuyo tiempo de respuesta es relativamente largo. Este estándar está basado en los siguientes hechos empíricos:

1. La existencia de *bandas críticas* que modelan mejor el análisis humano de las frecuencias que una escala en Hercios, ya que tienen en cuenta que la sensación de frecuencia percibida por el ser humano no está en relación lineal con la frecuencia física.
2. El cálculo del Loudness total de una señal corresponde a la suma del aportado por cada banda crítica (*Loudness específico*).
3. Un sonido en una banda puede provocar excitación (y por lo tanto Loudness) en las bandas adyacentes, y esto produce el fenómeno conocido como *enmascaramiento frecuencial*.
4. La sensación de sonoridad de un sonido que llegue al oído desde *todas las direcciones* difiere de la que se tendría si este se aproxima en *dirección frontal* (aunque ambos tengan el mismo nivel de presión sonora).
5. Existe una relación entre nivel sonoro y sensación de sonoridad que depende de la intensidad del sonido y de la banda crítica en la que se encuentre.
6. El cálculo del Loudness se realiza de forma gráfica utilizando una *plantilla normalizada*. El estándar dispone de distintas plantillas dependiendo del nivel máximo de la señal que se va a analizar y de si se recibe en campo libre o difuso. En la figura 2.13 se representa un ejemplo de plantilla contenida en [1]

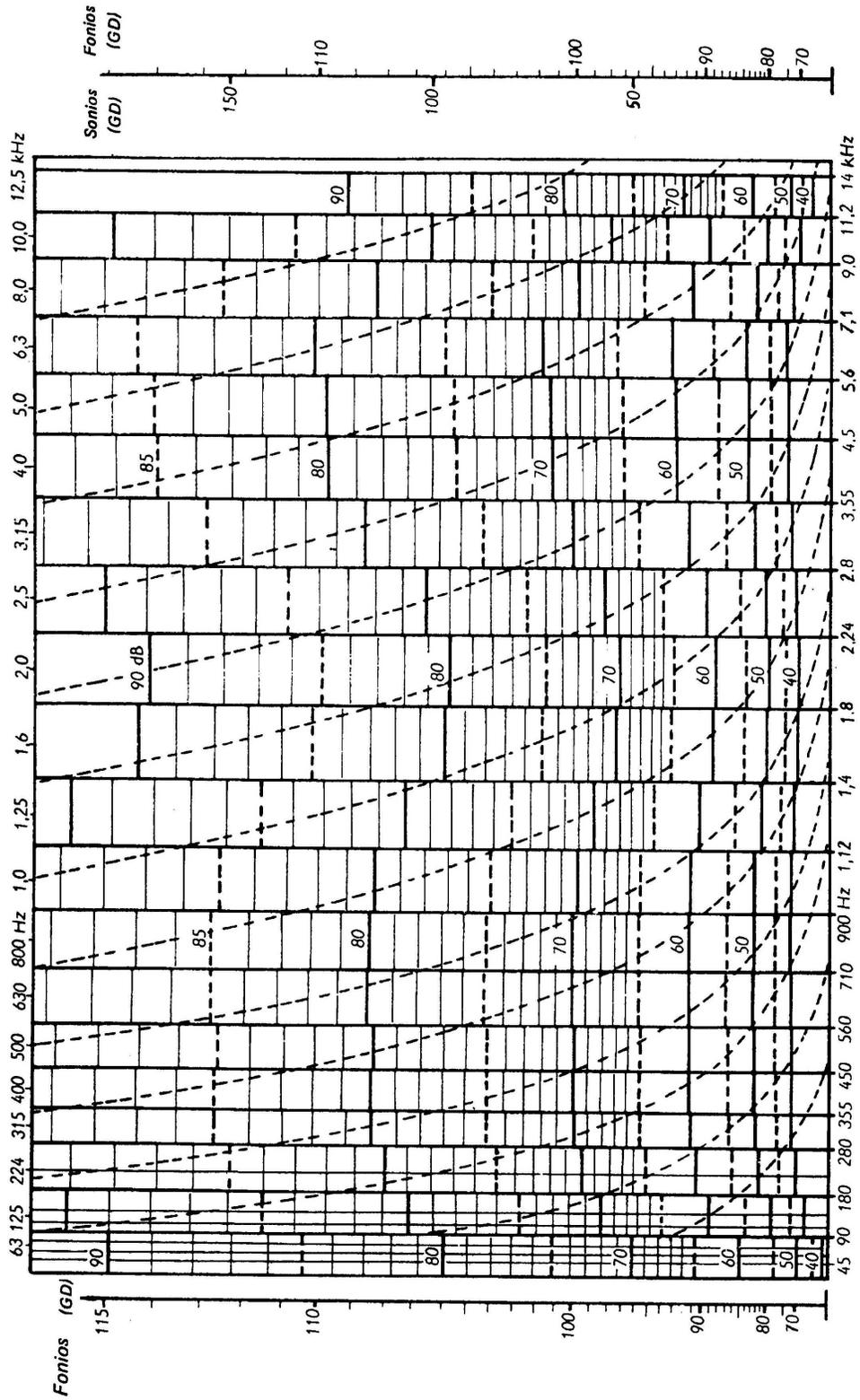


Figura 2.13: Plantilla para cálculo de Loudness según la norma ISO-532-B.

Cálculo del Loudness de Zwicker (ISO 532-B [1]).

El punto de partida de este método es el nivel de señal contenido en cada una de las bandas de tercio de octava que aparecen en la figura 2.13, por lo que se deberá hacer pasar la señal a analizar a través de un banco de filtros y a continuación medir la potencia existente en la salida de cada uno de ellos. Una vez hecho esto, el procedimiento para el cálculo del Loudness se descompone en los siguientes pasos:

- Paso 1:

Se empezará *seleccionando la plantilla* adecuada en función de la dirección de llegada del sonido y del nivel de la señal que se vaya a utilizar. A continuación, se debe obtener el valor del nivel de intensidad sonora en cada uno de los filtros de tercio de octava definidos en el estándar. Después, se deberán combinar todas las bandas por debajo de 90Hz para agruparlas en una sola medida, también se deberán agrupar en una sola medida las tres bandas comprendidas entre 90 y 180 Hz y las dos comprendidas entre 180 y 280Hz. Todos los valores así calculados se trasladarán a la plantilla mediante líneas horizontales situadas en la banda correspondiente y cuya altura indicará el nivel de éstas. La gráfica obtenida de esta forma recibe el nombre de Nivel de banda crítica.

- Paso 2:

Se deben incluir las *pendientes de excitación* del siguiente modo: si el nivel de la banda inmediatamente superior es mayor que el de la banda analizada, la línea de transición entre las bandas corresponderá a una vertical, pero si la banda inmediatamente superior tiene menor nivel, la transición se realizará mediante una línea curva paralela a las que aparecen en la tabla. De esta forma se habrá construido una línea continua que corresponde al nivel de excitación específico (si se lee en la escala de dB) o al Loudness específico (si se lee en la escala de sones).

- Paso 3:

La última fase en la obtención del Loudness consiste en la *integración de la gráfica* de Loudness específico, que según el estándar se debe realizar de forma gráfica dibujando una recta horizontal cuya área sea la misma que la que hay bajo la línea dibujada en el paso anterior y leyendo la altura de la gráfica en la escala graduada, para obtener el valor del Loudness.

Aunque en la actualidad existen sistemas de procesado digital que permiten implementar fácilmente cualquier tipo de filtro asociado a la escala Bark, paradójicamente cuando se quieren realizar cálculos de Loudness normalizados se deben de diseñar sistemas digitales que reproduzcan el funcionamiento de los antiguos sistemas analógicos y los filtros de tercio de octava, puesto que la norma se redactó teniendo en cuenta la tecnología analógica de la época.

Loudness Level.

Este parámetro está estrechamente ligado al Loudness, proporcionando la misma información. Indica el nivel que debe de tener un tono puro de 1 kHz para que produzca la misma sensación de sonoridad.

Su **símbolo** es L_N , se mide en **phones (o fonios)** y se relaciona con el Loudness del siguiente modo:

si $N \geq 1$ sone:

$$L_N = 40 + 10 \log_2(N) \quad (2.36)$$

si $N < 1$ sone

$$L_N = 40 + (N + 0'0005)^{0'35} \quad (2.37)$$

Referencia: Un tono puro de 40 dB SPL y frecuencia de 1kHz tiene Loudness Level de 40 phones.

Loudness de Señales no estacionarias.

El Loudness de una señal estacionaria [67] permanece constante a lo largo del tiempo, mientras que el Loudness de una señal no estacionaria varía a lo largo del tiempo, por lo que es habitual realizar medidas instantáneas y estadísticas de estos valores.

Las medidas estadísticas más habituales realizadas sobre el Loudness son:

N_{max} Loudness máximo.

N_{min} Loudness mínimo.

N_{mean} Loudness medio.

σ_N Desviación típica del Loudness.

$N(n)$ Loudness excedido durante el $n\%$ del tiempo.

Las medidas instantáneas más habituales realizadas sobre el Loudness son:

N_{imax} Loudness instantáneo máximo,

N_{imin} Loudness instantáneo mínimo,

N_{imean} Loudness instantáneo medio,

σ_{Ni} Desviación instantáneo típica del Loudness y

$N_i(n)$ Loudness instantáneo excedido durante el $n\%$ del tiempo.

2.6.4. Sharpness.

Este parámetro se corresponde con la sensación subjetiva de Frecuencia media del sonido. Un sonido agudo en el que predominen las frecuencias altas tendrá un Sharpness elevado mientras que un sonido grave en el que predominan las frecuencias bajas tendrá un Sharpness reducido.

Cálculo del Sharpness: Los métodos utilizados para el cálculo del Sharpness son principalmente dos: uno propuesto por E. Zwicker y otro propuesto por Aures posteriormente para corregir los problemas detectados en las estimaciones realizadas mediante el método anterior.

El **símbolo** utilizado para el Sharpness es S y las unidades en que se mide son el **acum**.

Referencia: Un ruido de ancho espectral un Bark centrado en la frecuencia de 1kHz 60 dB SPL tiene, por definición, un Sharpness de 1 acum.

Sharpness de Zwicker.

El método propuesto por Zwicker [3] para el cálculo del Sharpness se corresponde con el cálculo del momento (centro de gravedad) del Loudness específico (en escala Bark). En los experimentos realizados por Zwicker observó que las bandas críticas altas producen un mayor Sharpness que las que les corresponde en el cálculo del momento, por lo que introdujo un factor de ponderación $g(z)$ para valores altos de la escala:

$$S = \frac{\int_0^{24} N'(z) \cdot g(z) \cdot z \cdot dz}{\int_0^{24} N'(z) \cdot dz} \text{acum} \quad (2.38)$$

donde:

z es la frecuencia medida en escala Bark,

$N'(z)$ es el loudness específico y

$g(z)$ es una función ponderadora en función de la frecuencia que vale:

$$g(z) = 1, \quad \text{si } z \leq 16 \quad (2.39)$$

$$g(z) = 0'066 \cdot e^{0'171 \cdot z}, \quad \text{si } z > 16 \quad (2.40)$$

La gráfica de esta función ponderadora es la que se puede ver en la figura 2.14.

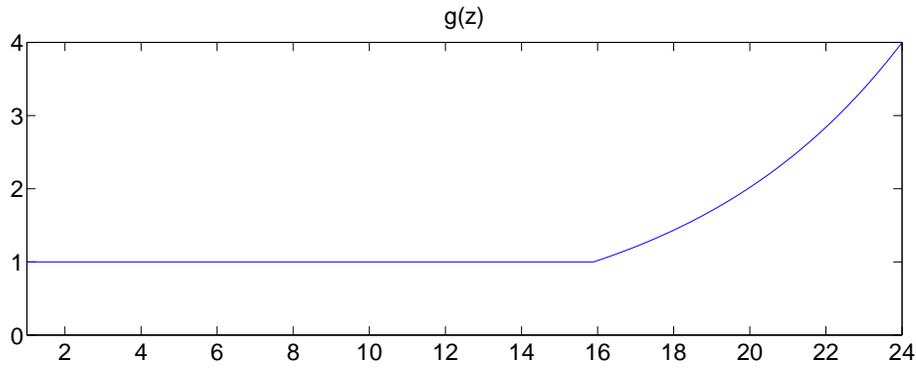


Figura 2.14: Función de ponderación del sharpness en función de la frecuencia (en escala Bark).

Sharpness de Aures.

El método propuesto por Zwicker para el cálculo del Sharpness es independiente del nivel sonoro de la señal que se analiza. No obstante, los experimentos realizados demuestran cierta dependencia entre el Sharpness y el nivel sonoro de la señal. Por eso, Aures propone calcular el Sharpness mediante un método derivado del anterior que tenga en cuenta esta dependencia:

$$S = \frac{\int_0^{24} N'(z) \cdot g(z) \cdot z \cdot dz}{\ln(0'05N + 1)} \text{acum} \quad (2.41)$$

Siendo

$N'(z)$ el loudness específico y

N el loudness total de la señal.

Sharpness de Señales no estacionarias.

Al igual que ocurría con el Loudness, es habitual realizar medidas instantáneas y estadísticas del Sharpness de las señales que no son estacionarias, siendo las principales:

S_{max} Sharpness máximo,

S_{min} Sharpness mínimo,

S_{mean} Sharpness medio,

σ_S Desviación típica del Loudness y

$S(n)$ Sharpness excedido durante el $n\%$ del tiempo.

2.6.5. Fluctuation Strength.

Distintos tipos de sonidos tienen asociado distintos procedimientos para calcular la molestia que producen. Por ejemplo, en el caso de golpes contra el suelo de objetos pesados la molestia va asociada a fluctuación [96].

Una señal coseno modulada en amplitud con otro coseno corresponde a la expresión:

$$s(t) = (1 + m \cos(2\pi f_m t)) \cos(2\pi f_c t) \quad (2.42)$$

Siendo m el índice de modulación, f_m la frecuencia de modulación y f_c la frecuencia de portadora.

Los sonidos representados por la ecuación (2.42) producen sensaciones diferentes cuando se escuchan, dependiendo de que la frecuencia de modulación sea alta o baja. Cuando un sonido es modulado con una frecuencia baja, la sensación percibida es la de Fluctuation Strength (fuerza de las fluctuaciones); mientras que si la frecuencia moduladora aumenta, el efecto producido es el de Roughness (discutido en la sección 2.6.6).

La sensación de Fluctuation Strength es la de un sonido cuya amplitud varía de forma perceptible en el tiempo. El oído percibe esta sensación cuando las frecuencias moduladoras toman valores desde 0 hasta 20 Hz, siendo los valores en torno a 4 Hz los que mayor sensación de fluctuación producen.

El **símbolo** utilizado para Fluctuation Strength es **F** y su unidad de medida el **vacil**.

Referencia: Un tono de 60 dB SPL y frecuencia de 1kHz modulado en AM con profundidad 100 % y 4 Hz de frecuencia de modulación tiene, por definición una Fluctuation Strength de 1 vacil.

Fluctuation Strength de Zwicker.

Zwicker [3] propone un método para el cálculo de este parámetro aplicable a señales moduladas en amplitud sinusoidalmente, de las que se conoce o se puede calcular la frecuencia de modulación:

$$F = \alpha \frac{\Delta L}{(f_{mod}/4Hz) + (4Hz/f_{mod})} \quad (2.43)$$

donde:

α es un factor de proporcionalidad que dependerá de la señal que se va a modular,

ΔL representa la variación temporal del patrón de enmascaramiento, que es la diferencia entre el máximo y el mínimo en el patrón de enmascaramiento temporal. Es importante hacer notar que este parámetro no es el

índice de modulación m utilizado en (2.42), puesto que debido a efectos de enmascaramiento las fluctuaciones de señal realmente percibidas son inferiores a las representadas con este índice,

f_{mod} es la frecuencia de la señal moduladora.

Después de numeros experimentos, Zwicker, consiguió determinar el valor de la constante α y refinar la fórmula anterior para dos casos particulares de señales.

1. Ruidos de banda ancha modulados en amplitud por una señal senoidal:

$$F_{BBN} = \frac{5,8(1,25 \cdot m - 0,25) \cdot [0,05 \cdot (L_{BBN}/dB) - 1]}{(f_{mod}/5Hz)^2 + (4Hz/f_{mod}) + 1,5} vacil \quad (2.44)$$

donde:

m es el índice de modulación, (coincide con el que aparece en (2.42) aunque en este caso la portadora es de banda ancha),

L_{BBN}/dB el nivel del sonido de banda ancha expresado en decibelios y

f_{mod} es la frecuencia de la señal moduladora.

2. Tonos modulados en amplitud o en frecuencia:

$$F = \frac{0,008 \int_0^{24Bark} (\Delta L / (dB/Bark)) \cdot dz}{(f_{mod}/4Hz) + (4Hz/f_{mod})} \quad (2.45)$$

En este caso, se calcula la fluctuación en cada Bark y a continuación se obtiene el total por integración.

De forma práctica, para aquellas señales en las que la variación temporal del patrón de enmascaramiento en cada banda (ΔL) es difícil de obtener, se recurre a estimaciones realizadas a través de percentiles del Loudness en la banda analizada. Una aproximación de este parámetro es:

$$\Delta L = 4 \cdot 20 \log \left(\frac{N'(1)}{N'(99)} \right) \quad (2.46)$$

Siendo $N'(1)$ el Loudness excedido durante el 1 % del tiempo en la banda y $N'(99)$ el excedido durante el 99 % del tiempo.

2.6.6. Roughness.

Conceptualmente, el Roughness es un parámetro similar al Fluctuation Strength (analizado en el párrafo anterior), ya que mide variaciones de frecuencia o amplitud que se produce en una señal. Sin embargo, la sensación de Roughness es muy diferente a la de Fluctuation Strength, puesto que en el parámetro anterior se consideraban las variaciones lentas de las señales y en éste se consideran las variaciones rápidas. El parámetro Roughness toma un valor máximo cuando las oscilaciones que se producen en la amplitud o la frecuencia de una señal está en torno a los 70Hz. El Roughness de los sonidos está relacionado con la capacidad del oído para separar sonidos que proceden de distintas fuentes [97].

Los primeros estudios acerca del Roughness proceden de la teoría de la música cuando Helmholtz [98] definió el término disonancia como un “fenómeno desagradable producido cuando se suman dos tonos de frecuencias similares” en contraste con la consonancia (asociada a sonidos agradables), aunque esta asociación requiere cierto entrenamiento en teoría de la música [99]. El Roughness fue más tarde analizado por Terhardt [100] y varios métodos de cálculo de este parámetro fueron propuestos por Aures [5], Zwicker y Fastl [3] y Daniel y Weber [8, 101].

Todos estos métodos se basan en la descomposición frecuencial llevada a cabo por el oído humano: los sonidos se descomponen en bandas (bandas críticas habitualmente) y, mediante diferentes cálculos, se estima el Roughness en cada una de estas bandas, obteniendo lo que se conoce como Roughness específico. Finalmente se suma el Roughness específico de cada banda para obtener el valor del Roughness total.

Otra característica relevante de la percepción humana del sonido considerada en todos los cálculos de Roughness es el efecto de enmascaramiento temporal y frecuencial. Este efecto se ha estudiado en tonos modulados en amplitud (AM), tonos modulados en frecuencia (FM) y señales de banda estrecha moduladas, obteniendo patrones de enmascaramiento para diferentes índices y frecuencias de modulación. La variación temporal del patrón de enmascaramiento temporal se puede describir mediante la profundidad de enmascaramiento temporal (ΔL), que representa la diferencia de niveles entre el máximo y el mínimo en el patrón de enmascaramiento temporal. En términos generales, esta profundidad de enmascaramiento temporal es menor que el índice de modulación de la envolvente del sonido enmascarante debido al post-enmascaramiento.

Si consideramos el método de cálculo de Roughness propuesto por Zwicker et al. [3], la obtención del Roughness Específico en cada banda crítica es proporcional al producto de la profundidad de enmascaramiento temporal ΔL y la frecuencia de modulación f_{mod} , y el Roughness total se obtiene sumando los valores de Roughness Específico.

Aures [5] desarrolló otro método donde el cálculo del Roughness total se realizaba de un modo más complejo que el propuesto por Zwicker en [3], pero manteniendo el propósito de calcular el Roughness Total a partir de la evaluación previa del Roughness Específico. Ambos métodos son adecuados cuando se trabaja con tonos modulados en AM y en FM, pero proporcionan resultados no adecuados al trabajar con ruidos de banda ancha.

Con el fin de proporcionar una medida más adecuada del Roughness, Daniel y Weber [8] propusieron un nuevo método que ponderaba el Roughness específico en función de la correlación cruzada de los valores obtenidos en bandas adyacentes. Trabajos posteriores [102] han mostrado que este procedimiento caracteriza el Roughness bastante mejor que el propuesto por Zwicker en [3], no sólo para ruidos blancos y paso banda sino también cuando se utilizan tonos modulados en AM y en FM.

Este método de cálculo de Roughness se ha convertido en un standard de facto del cálculo de Roughness utilizado en multitud de experimentos: El modelo de cálculo de Roughness de Daniel y Weber se ha utilizado para el estudio del Roughness producido por varias fuentes distribuidas espacialmente [103], ruidos de motor [22, 23] o el cálculo de roughness de otros sonidos: bolas que se mueven [104], bocinas [79] y otros estudios que analizan distintos tipos de sonidos [105, 13, 106].

Una versión fisiológica del Roughness corresponde con el índice de sincronización [107, 108, 109] y está relacionado con la sincronización de los disparos de las neuronas del oído. Para calcular el índice de sincronización que produce un sonido se analizan las señales eléctricas generadas por las neuronas asociadas a distintas bandas críticas del oído y el grado de correlación entre ellas.

El **símbolo** utilizado para el Roughness es **R** y las unidad en que se mide es el **asper**.

Referencia: Un tono de 60 dB SPL y frecuencia de 1kHz modulado en AM con profundidad 100 % y 70 Hz de frecuencia de modulación tiene, por definición, un Roughness de 1 asper.

La mayor parte de los métodos de cálculo del Roughness están basados en la descomposición de los sonidos en bandas que modelan la descomposición frecuencial realizada por el oído. Una vez descompuesto el sonido en bandas se aplican diversos métodos para analizar el Roughness en cada una de estas bandas (Roughness específico) y por último se utiliza este parámetro para el cálculo del Roughness total del sonido analizado.

Roughness de Zwicker.

El primer método de cálculo de Roughness fue desarrollado por Zwicker [3], quien propuso que el Loudness debía ser proporcional al producto entre la profundidad de enmascaramiento temporal (ΔL) y la frecuencia de modulación de la señal (f_{mod}):

$$R \propto f_{mod}\Delta L \quad (2.47)$$

Como la profundidad de enmascaramiento temporal no coincide con la profundidad de modulación de una señal, y puede variar según la banda de frecuencia analizada, se han desarrollado procedimientos para el cálculo de dicho parámetro a partir de los percentiles de la sonoridad que presenta la señal en distintas bandas.

De forma práctica [10], se calcula el Roughness específico (en cada Bark) mediante la expresión:

$$R'(z) = 0,0003 \cdot (f_{mod}(z)/Hz) \cdot (\Delta L_E/(dB/Bark))_{asper}. \quad (2.48)$$

Donde

$$\Delta L_E = 20 \log \left(\frac{N'(1)}{N'(99)} \right) \quad (2.49)$$

Siendo

$N'(1)$ el Loudness excedido durante el 1 % del tiempo en la banda,

$N'(99)$ el excedido durante el 99 % del tiempo y

$f_{mod}(z)$ la frecuencia de modulación en el Bark z .

A partir del Roughness específico, se calcula el Roughness total sumando el Roughness específico de cada banda:

$$R = \sum_1^{24} R'(z) \quad (2.50)$$

Roughness de Daniel y Weber.

En 1997 Daniel y Weber [8] publicaron un nuevo método para el cálculo del Roughness basado en los resultados de experimentos anteriores [110, 111, 112, 113, 114, 115, 116] que estudiaban la dependencia del Roughness con:

- El índice de modulación, en modulaciones de amplitud y frecuencia.
- La relación entre amplitudes de la suma de dos tonos de distintas frecuencias.
- Frecuencia de la portadora y tipo de modulación.
- Nivel y distribución espectral de la señal

El proceso de cálculo del Roughness se corresponde con el esquema de la figura 2.15 y se puede resumir en los siguientes pasos:

Enventanado de la señal: El método de cálculo comienza con el enventanado de la señal a analizar utilizando ventanas de Blackman de 200 ms de duración para obtener la señal $x(t)$.

Análisis de 47 bandas críticas: Las frecuencias de análisis se extienden desde el Bark 0 hasta el bark 24. Las 47 bandas se obtienen al dividir este rango frecuencial en bandas críticas de ancho un bark con solapamiento del 50 %.

Cálculo de la excitación específica en la banda i ($e_i(t)$): Consiste en replicar la señal $x(t)$ atenuando cada una de sus componentes frecuenciales pero sin modificar su fase según el siguiente criterio [117]:

- Las frecuencias asociadas a la banda i no se atenúan.
- Las frecuencias superiores a la banda i se atenúan en función de su distancia a la frecuencia de corte superior de la banda a razón de:

$$S_1 = -27 \frac{dB}{Bark} \quad (2.51)$$

- Las frecuencias inferiores a la banda i se atenúan en función de su distancia a la frecuencia de corte inferior de la banda (y teniendo en cuenta su nivel y frecuencia) a razón de:

$$S_2 = \left[-24 - \frac{0,23kHz}{f} + \frac{0,2L}{dB} \right] \frac{dB}{Bark} \quad (2.52)$$

Obtención de parámetros de la envolvente de $e_i(t)$: La excitación específica de la banda i es una señal paso banda cuyo Roughness está asociado a su envolvente, de la que es necesario obtener dos parámetros:

Media de la envolvente ($h_{0,i}$), que es una constante.

Envolvente filtrada paso banda ($h_{BP,i}(t)$). Puesto que el Roughness de una señal modulada depende de la frecuencia de modulación, el filtrado aplicado a la envolvente tiene la misión de ponderar de forma adecuada cada frecuencia de modulación. Dado que el roughness de una señal depende de la frecuencia central de ésta, el filtro aplicado a cada banda es diferente.

Para extraer estos dos parámetros se trabaja con el módulo de la señal $e_i(t)$.

Cálculo del índice de modulación generalizado m_i^* . Es el cociente de dividir la desviación típica de la envolvente filtrada entre la media de la envolvente. El valor máximo de este parámetro es 1 por lo que si el resultado de la operación anterior es superior a este valor, se toma $m_i^* = 1$. Esta limitación evita valores excesivamente altos en el Roughness cuando aparecen pulsos en la envolvente:

$$m_i^* = \min \left\{ 1, \frac{std(h_{BP,i}(t))}{h_{0,i}} \right\} \quad (2.53)$$

Cálculo del Roughness específico en la banda i r_i^* . Para esto se tiene en cuenta, además del índice de modulación generalizado de la banda i , un factor de ponderación (que depende de i) y la correlación de la envolvente en la banda i con las envolventes de la banda $i + 2$ e $i - 2$, según la expresión:

$$r_i = (g(z_i) \cdot m_i^* \cdot k_{i-2} \cdot k_i)^2 \quad (2.54)$$

Donde $g(z_i)$ es un factor que pondera el peso de la banda en el Roughness total de la señal y toma valores entre 0,6 y 1,1; k_{i-2} es la correlación entre las envolventes de la banda $i - 2$ y la banda i , y k_i es la correlación entre las envolventes de la banda i y la banda $i + 2$.

Cálculo del Roughness total R . Se suma el Roughness específico de cada banda normalizando por un factor de 0.25 para conseguir que un tono de 60 dB SPL y frecuencia de 1kHz modulado en AM con profundidad 100 % y 70 Hz de frecuencia de modulación tenga un Roughness de 1 asper:

$$R = \sum_{i=1}^{47} r_i(\text{asper}) \quad (2.55)$$

De esta forma, en el proceso de cálculo de la profundidad de modulación de cada canal, se tiene en cuenta que hay frecuencias moduladoras que contribuyen más que otras al Roughness percibido y que la ponderación de estas frecuencias es diferente según la banda que se analice.

2.6.7. Duración subjetiva.

La duración de un sonido es una propiedad muy interesante para ciertas aplicaciones como el análisis y síntesis de señales de voz o de música. Aunque se podría pensar que la duración no es algo subjetivo y que la duración de un sonido es el tiempo (medido en segundos) durante el que se escucha dicho sonido, esto no es cierto para sonidos de corta duración. Así pues, se debe distinguir entre la duración física de un sonido y la sensación de duración que produce éste. Mediante experimentos se ha podido comprobar que para sonidos de duración inferior a 100ms la duración física y subjetiva de los sonidos no coincide, siendo mayores las duraciones subjetivas que las duraciones físicas.

El **símbolo** utilizado para el Duración subjetiva es **D** y las unidad en que se mide es el **dura**.

Referencia: Un tono de 60 dB SPL y frecuencia de 1kHz con una duración de un segundo, tiene una duración subjetiva de 1 dura.

2.7. Índices globales de calidad de sonidos.

El sonido juega un papel fundamental en la evaluación de un producto [80] y esto hace que los fabricantes tengan cada vez más interés en el “diseño del sonido” de estos. El carácter desagradable o molesto del sonido de un producto repercute negativamente en su precio. Takada [118] comprobó en un estudio de mercado hecho con aspiradoras, que aquellas que producían ruidos más molestos tenían un precio inferior en un 12 %

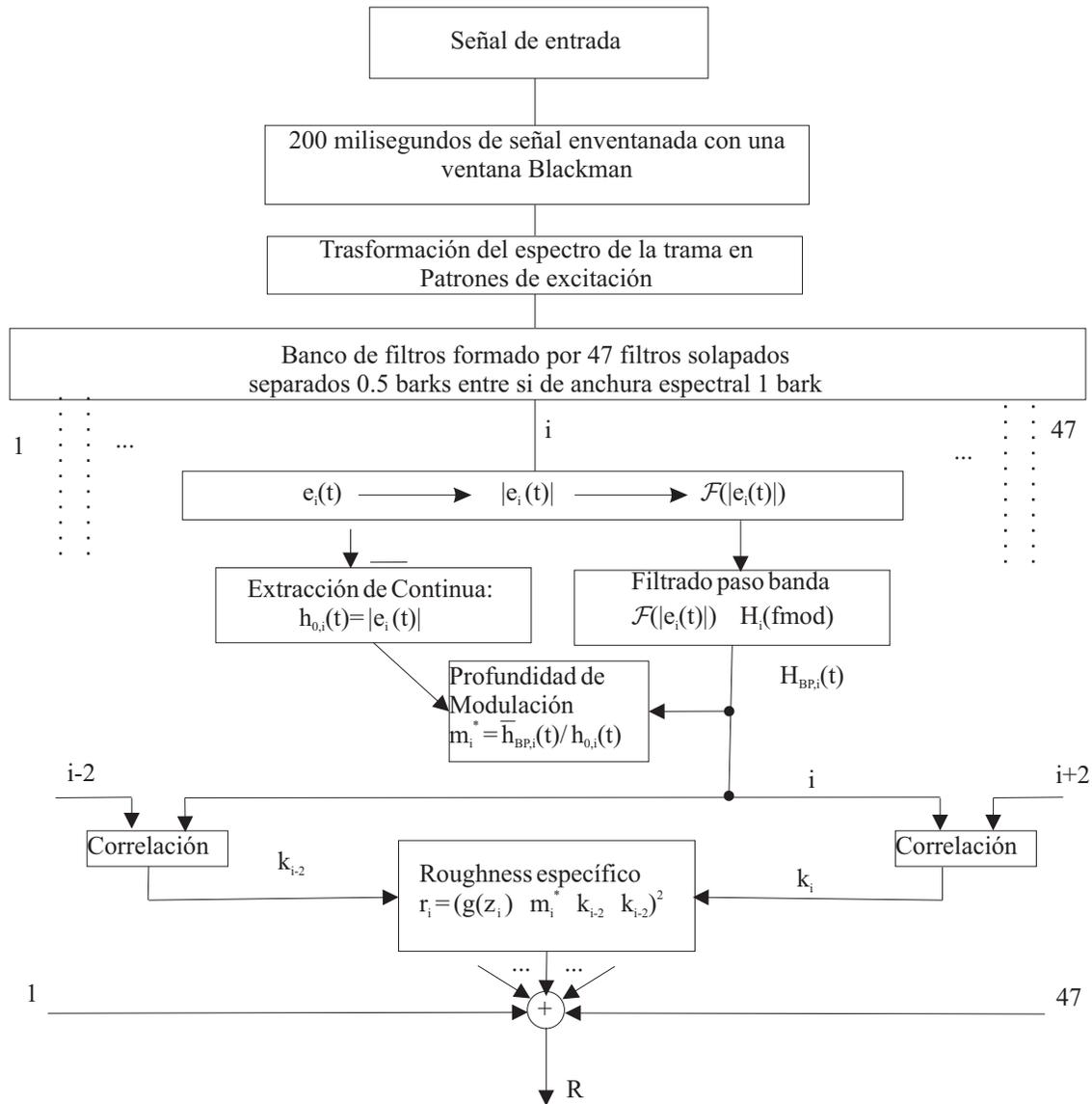


Figura 2.15: Cálculo del Roughness de una señal utilizando el método de Daniel y Weber.

a las menos molestas y también comprobó que el grado de molestia estaba asociado al sharpness.

Los parámetros Loudness, Sharpness, Fluctuation Strength y Roughness se agrupan bajo la denominación *medidas de calidad de sonido* (Sound Quality Metrics). La utilización de estos parámetros ofrece información psicoacústica de los sonidos prescindiendo de la necesidad de un jurado, y esto permite reducir los errores en las medidas asociados a factores humanos y proporciona datos con menor coste económico [12, 119].

Las *medidas de calidad de sonido* se han utilizado para analizar el grado de molestia que producen diversos sonidos como una nevera [81], el aire acondicionado en el interior de un vehículo [82], el elevallas de un automóvil [83], el tráfico urbano producido por los automóviles [84, 73], o el ruido asociado a tranvías, autobuses [85] y trenes [86], entre otros.

Como consecuencia de este tipo de estudios se han desarrollado expresiones matemáticas que proporcionan (en función las *medidas de calidad de sonido*) una “calificación global del grado de molestia o agrado” como: el “Psychoacoustic Annoyance” o el “Sensory Pleasantness” desarrollados por Zwicker [3]; el “perceived annoyance” descrito por Fastl [14]; o el “Unpleasantness Scores” calculado por Ellermeier [13]. Pero la combinación del Loudness, Sharpness, Fluctuation Strength y Roughness para conformar esta “calificación global del grado de molestia o agrado” puede variar entre distintas familias de productos [14, 15], por lo que también se han desarrollado calificaciones globales asociadas a un determinado tipo de producto. En este sentido podemos encontrar: un índice de molestia (annoyance index) [16] y un valor de agrado (pleasantness) [16] desarrollado para aspiradoras, un índice de molestia desarrollado por la compañía AVL para ruidos de motor [17], o un índice de molestia específicamente diseñado para motores diesel [18].

En el entorno automovilístico, las *medidas de calidad de sonido* se han utilizado para analizar parámetros que afectan a la comodidad del conductor y los pasajeros como el booming [120] (que se produce cuando señales con reducido ancho de banda y sus armónicos procedentes del motor se introducen en el habitáculo del vehículo) o el ruido de impacto [121] (que se produce al atravesar irregularidades en el terreno).

Capítulo 3

Análisis TF.

3.1. Introducción.

El análisis en frecuencia fue introducido por Fourier al estudiar la física del calor en 1807 (aunque sus trabajos no fueron publicados hasta años más tarde [122]). La transformada de Fourier permite descomponer una señal en una suma de funciones coseno (o exponenciales complejas en el caso más genérico) de distintas frecuencias, y es especialmente útil cuando se trabaja con señales estacionarias. Si se trabaja con señales no estacionarias, la información proporcionada por la transformada de Fourier es insuficiente ya que indica las componentes frecuenciales de la señal pero no el instante en el que estas aparecen. En este caso se hace necesario un análisis bidimensional de la señal llamado análisis Tiempo-Frecuencia (TF).

El análisis TF se puede implementar realizando un análisis frecuencial sobre una ventana de la señal; y para ello existen dos alternativas dependiendo de la resolución espectral necesaria en cada banda de frecuencias. Si se desea la misma resolución en altas que en bajas frecuencias se emplean técnicas derivadas de la transformada de Fourier como la Transformada Corta de Fourier o STFT (siglas de Short Time Fourier Transform) [123, 124]. Si se desea mayor resolución frecuencial en las bandas de frecuencias bajas que en las altas se emplean técnicas derivadas de la Transformada Wavelet (WT).

La primera función wavelet conocida (la wavelet de Haar) fue desarrollada por el matemático Húngaro Alfred Haar en el año 1909 en su tesis doctoral dedicada al estudio de sistemas de funciones ortogonales [125]. En estos primeros momentos de desarrollo de la teoría de wavelets no se utilizó la palabra “wavelet”, la función base utilizada era muy simple y no tenía muchas de las propiedades recomendables de las wavelets. Las primeras aplicaciones prácticas de las wavelets corresponden al campo de la mecánica cuántica y datan de la década de 1980 [126]. A partir de este momento la teoría de wavelets se ha desarrollado de forma muy rápida y han aparecido multitud de aplicaciones. Cuando aparecieron las wavelets algunos pensaron en ellas como una potente herramienta que podía dar mucho recorrido en el futuro, pero otros no vieron más que un cambio de nombre a técnicas que se habían utilizado anteriormente [127]: análisis con Q constante [128],

funciones de ambigüedad cruzada y banda ancha [129], banco de filtros de banda de octava con reconstrucción perfecta [130, 131] o una variación de la descomposición piramidal Laplaciana [132]. En cualquier caso, las wavelets aportan una unificación de todos estos trabajos en una única teoría.

Las representaciones obtenidas mediante análisis wavelet o STFT son representaciones lineales puesto que son proporcionales a los valores que toma la señal, pero existen otro tipo de representaciones proporcionales a la energía de la señal que reciben el nombre de representaciones cuadráticas. Ambos tipos de representaciones están limitados por el “Principio de incertidumbre” [133]: para mejorar la resolución temporal de un análisis TF, se debe hacer a costa de empeorar la resolución frecuencial (y viceversa).

El análisis tiempo frecuencia que proporciona una mayor resolución en los dominios tiempo y frecuencia es el realizado mediante la transformación de Wigner-Ville, pero presenta el inconveniente de tener términos de interferencia elevados [134], por lo que la mayoría de los métodos de análisis tiempo-frecuencia consisten en un “suavizado” de éste. La transformación de Wigner-Ville pertenece al grupo de transformaciones cuadráticas y el suavizado de esta representación se puede realizar de dos formas [135]: realizando un filtrado bidimensional (para obtener la misma resolución en bajas que en altas frecuencias, de forma similar a la STFT), o realizando un filtrado bidimensional afín (para obtener mayor resolución en frecuencias bajas que en altas de forma similar a la WT).

En este capítulo se describirán en primer lugar los parámetros más importantes en el procesado tiempo-frecuencia de señales mediante el uso de técnicas digitales como: periodo de muestreo, longitud de la ventana, número de puntos de la transformada o solape.

El presente capítulo también examina las características de las distintas técnicas de análisis tiempo frecuencia y desarrolla aquéllas que han tenido mayor aplicación en el campo de la Psicoacústica. Se estudia de forma separada la problemática asociada al análisis tiempo frecuencia discreto, debido a su gran relevancia desde el punto de vista práctico, puesto que a menudo no es trivial trasladar conceptos desarrollados para el análisis continuo (como sucede con la operación de escalado que es la base del análisis wavelet). El capítulo está estructurado en cuatro secciones: la introducción presenta algunos conceptos relativos al análisis TF, la siguiente sección está dedicada al análisis de señales continuas cuyas características varían con el tiempo, la sección tercera trata sobre la descomposición de señales discretas y la última sección se centra en técnicas de procesado TF aplicadas a Psicoacústica.

3.1.1. Representación de señales que varían en el tiempo.

Una determinada señal se puede representar de múltiples formas dependiendo de la aplicación considerada. En la mayor parte de aplicaciones de ingeniería se suelen representar las señales en función del tiempo, aunque a la hora de analizar o diseñar sistemas es habitual realizar representaciones en el dominio de la frecuencia, puesto que algunos parámetros de las señales y los sistemas se caracterizan mejor en este dominio. Así pues, a pesar de la infinidad de posibilidades existentes a la hora de representar una señal, las

variables más importantes que aparecen en la naturaleza son el tiempo y la frecuencia: el dominio temporal nos dice cómo varía la amplitud de una señal a lo largo del tiempo mientras que el dominio de la frecuencia nos indica con qué frecuencia suceden las variaciones. El puente que sirve de unión entre el dominio del tiempo y el dominio de la frecuencia es la transformada de Fourier (TF).

En acústica y psicoacústica, el análisis frecuencial es fundamental puesto que el oído realiza una descomposición de los sonidos en bandas de frecuencia. Como se verá en este capítulo, al realizar un análisis en el dominio del tiempo no se dispone de información frecuencial, mientras que si se trabaja en el dominio de la frecuencia se pierde la información temporal de la señal. Cuando se trabaja con señales no estacionarias (cuyas características varían a lo largo del tiempo), ni el análisis temporal ni el análisis frecuencial son suficientes para caracterizarlas completamente siendo necesario un análisis mixto llamado análisis tiempo-frecuencia.

3.1.2. Dominio del tiempo y dominio de la frecuencia.

La Transformada de Fourier (TF) (correspondiente a la ecuación (3.1)) y la transformada de Fourier Inversa (TFI) (correspondiente a la ecuación (3.2)) establecen una relación entre el dominio del tiempo (señal $x(t)$) y el dominio de la frecuencia. A la representación de una señal en el dominio de la frecuencia se le suele denominar espectro.

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft} df \quad (3.1)$$

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt \quad (3.2)$$

Por lo general la transformada de Fourier es una función que toma valores complejos (aunque en el dominio del tiempo la señal sólo tome valores reales), puesto que al observar la ecuación (3.2) se puede comprobar que en el cálculo de la transformada de Fourier intervienen valores complejos. Debido a este hecho, en ocasiones, las representaciones de los espectros se deben realizar mediante dos gráficas: una de ellas representando el módulo de la transformada de Fourier y la otra representando la fase de ésta (aunque es habitual prescindir de la representación de la fase puesto que la información que proporciona es difícil de interpretar).

Además, se puede demostrar [136] a partir de la ecuación (3.2) que cuando se analizan señales reales (como la mayor parte de las señales que aparecen en la naturaleza) el módulo de la transformada de Fourier tiene simetría par y la fase tiene simetría impar respecto a la frecuencia 0, por lo que en las representaciones frecuenciales suelen aparecer únicamente las frecuencias positivas y reciben el nombre de espectros unilaterales.

El dominio del tiempo y de la frecuencia son dos formas alternativas de representar una misma señal, y la transformada de Fourier permite pasar de uno a otro, aunque no permite combinaciones de los dos dominios. Dicho de otro modo, la información temporal se pierde en el dominio de la frecuencia, es decir, el espectro nos proporciona información de las frecuencias contenidas en una señal y de la amplitud de cada una de ellas en la señal pero no nos permite conocer en qué momento estas frecuencias están presentes en la señal.

3.1.3. Frecuencia instantánea.

En una modulación FM compleja, la señal se ajusta a la expresión:

$$y(t) = Ae^{j\theta(t)} \quad (3.3)$$

consistente en un fasor de amplitud constante y frecuencia variable. En este caso, se define frecuencia instantánea de la señal $x(t)$ como:

$$f_x(t) = \frac{1}{2\pi} \frac{d\theta(t)}{dt} \quad (3.4)$$

Este concepto de frecuencia instantánea, y la forma de calcularlo, se puede generalizar para cualquier función compleja, puesto que puede ser escrita descomponiendo su módulo y fase de la siguiente forma:

$$x(t) = A(t)e^{j\theta(t)} \quad (3.5)$$

y por tanto la fórmula general para calcular la frecuencia instantánea de una señal compleja se obtiene mediante la generalización de la expresión(3.4) consistente en:

$$f_x(t) = \frac{1}{2\pi} \frac{d}{dt} \arg \{x(t)\} \quad (3.6)$$

Pero este modelo únicamente es válido cuando la señal de trabajo se corresponde con el modelo de un solo fasor cuya frecuencia y amplitud varían con el tiempo. Si la señal de trabajo está compuesta por dos frecuencias simultaneas (f_1 y f_2), la frecuencia instantánea calculada mediante la expresión (3.6) se corresponderá a un valor intermedio entre las dos frecuencias. Esto hace que si se aplica este cálculo señales reales en lugar de complejas, la frecuencia instantánea resulte ser siempre 0 (puesto que las señales reales siempre tienen un espectro simétrico compuesto por señales de frecuencias instantáneas positivas y negativas igualmente intensas). Por lo tanto, si se desea calcular la frecuencia instantánea de una señal real, no se puede aplicar directamente la expresión(3.6) a la señal sino que hay que trabajar con una versión compleja de la señal real en la que se hayan eliminado las frecuencias negativas y que recibe el nombre de versión analítica de la señal $x(t)$. La versión analítica de una señal real se puede calcular mediante la expresión:

$$z(t) = x(t) + jHT \{x(t)\} \quad (3.7)$$

donde $z(t)$ representa la versión analítica de la señal $x(t)$ y HT representa la transformada de Hilbert (o red de desplazamiento de fase de 90° [136]).

Esta técnica ofrece muy buenos resultados para el análisis de modulaciones y ha sido utilizada en aplicaciones como la demodulación de señales y análisis de señal de voz [137, 138, 139].

La frecuencia instantánea definida mediante (3.4) se aplica a la expresión de la señal en el dominio del tiempo. Pero, puesto que las señales se pueden describir en el dominio del tiempo y de la frecuencia, podemos preguntarnos en qué instante de tiempo se produce una determinada frecuencia f y esto se llama retardo de grupo para la frecuencia f definida por:

$$t_x(f) = -\frac{1}{2\pi} \frac{d}{df} \arg \{X(f)\} \quad (3.8)$$

En definitiva, puesto que la ecuación (3.1) y la ecuación (3.2), que permiten pasar del dominio del tiempo al de la frecuencia y viceversa son muy similares, cualquier magnitud definida en el dominio del tiempo será similar a otra definida en el dominio de la frecuencia, y esta pareja recibe el nombre de magnitudes duales. El retardo de grupo es, por tanto, la magnitud dual de la frecuencia instantánea.

3.2. Representaciones Tiempo-Frecuencia de señales continuas.

El concepto de frecuencia instantánea es únicamente útil en los casos en los que la señal esté formada por una sola componente espectral cuya frecuencia varía con el tiempo, y las representaciones de frecuencia instantánea corresponden a una curva en el plano tiempo frecuencia. Para poder trabajar con cualquier tipo de señal, donde pueden estar presentes de forma simultánea distintas frecuencias con distintas amplitudes, es necesario sustituir la curva asociada a las representaciones de frecuencia instantánea por superficies en el plano tiempo frecuencia (figura 3.1), lo que matemáticamente corresponde a una función de dos variables ($T_x(t, f)$) asociada a una determinada función x , que recibe el nombre de “representación tiempo frecuencia” de la señal x (o “TFR” del inglés time-frequency representation) .

Existe una gran variedad de representaciones tiempo frecuencia y se pueden clasificar en dos grupos:

Representaciones tiempo frecuencia lineales: son aquellas representaciones en las que a una combinación lineal de dos señales le corresponde a una representación

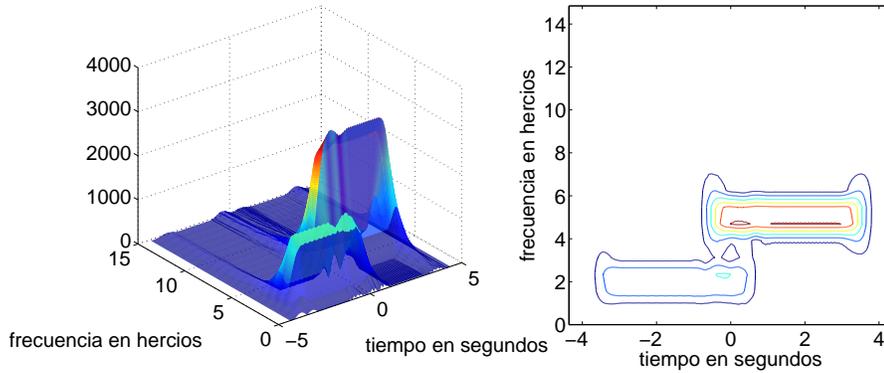


Figura 3.1: Representación tiempo frecuencia de la señal formada por dos cosenos de amplitudes 50 y 100, frecuencias 2 y 5 Hz que se extienden desde $t = -2,5$ a $t = 0,5$ (la primera) y desde $t = -0,5$ a $t = 2,5$ (la segunda). En la parte izquierda aparece una representación tridimensional de la amplitud para cada punto en el eje tiempo-frecuencia. En la parte derecha aparece la misma información representada mediante curvas de nivel.

tiempo frecuencia igual a la misma combinación lineal aplicada a las representaciones tiempo-frecuencia de las señales que se combinan. O expresado matemáticamente:

$$x(t) = c_1x_1(t) + c_2x_2(t) \Rightarrow T_x(t, f) = c_1T_{x_1}(t, f) + c_2T_{x_2}(t, f) \quad (3.9)$$

Representaciones tiempo frecuencia no lineales son aquellas que no cumplen las condiciones anteriores. Entre este grupo de representaciones son de especial relevancia las denominadas cuadráticas por lo que el grupo se subdivide en:

Representaciones tiempo frecuencia cuadráticas son las que cumplen el principio de superposición cuadrática:

$$\begin{aligned} x(t) &= c_1x_1(t) + c_2x_2(t) \Rightarrow \\ \Rightarrow T_x(t, f) &= |c_1|^2 T_{x_1}(t, f) + |c_2|^2 T_{x_2}(t, f) + c_1c_2^* T_{x_1, x_2}(t, f) + c_2c_1^* T_{x_2, x_1}(t, f) \end{aligned} \quad (3.10)$$

Donde T_{x_1} y T_{x_2} son las representaciones tiempo frecuencia de cada una de las dos señales que se combinan linealmente y T_{x_1, x_2} y T_{x_2, x_1} son representaciones conjuntas de las dos señales. Estas representaciones conjuntas deben cumplir:

$$T_{x,x}(t, f) = T_x(t, f) \quad (3.11)$$

Representaciones tiempo frecuencia no lineales y no cuadráticas son aquellas que no entran en ninguno de los grupos anteriores.

De todas las representaciones anteriores, las más utilizadas son las lineales y las cuadráticas.



Figura 3.2: Implementación de la STFT utilizando un filtro paso banda.

3.2.1. Representaciones tiempo frecuencia lineales.

Transformada corta de Fourier (STFT).

Aunque al realizar la transformada de Fourier se pierde la información temporal y por tanto la posibilidad de saber en qué momento se produce cada frecuencia, es posible conocer el contenido frecuencial de un determinado instante de tiempo analizando la transformada de Fourier en un pequeño margen de tiempo alrededor de ese instante. Para hacer esto, lo único que se necesita es “recortar” la señal con la que se va a trabajar para analizar únicamente un “trozo”, y esto se consigue realizando un enventanado (que consiste en multiplicar la señal a analizar por una señal llamada ventana), que atenúe o elimine la señal fuera del margen temporal que sea de interés. Si se realizan varios análisis desplazando esta ventana a lo largo de todo el eje de tiempos se obtiene una representación Tiempo-Frecuencia que recibe el nombre de Transformada corta de Fourier, también conocida como Transformada localizada de Fourier o Transformada de Fourier de ventana deslizante; aunque normalmente se abrevia mediante sus siglas en inglés STFT.

La transformada corta de Fourier de la señal $x(t)$ se define como:

$$STFT_x^\gamma(t, f) = \int_{t'=-\infty}^{\infty} x(t')\gamma^*(t' - t)e^{-j2\pi ft'} dt' \quad (3.12)$$

Siendo $\gamma^*(t' - t)$ la ventana de análisis (también llamada función de análisis) conjugada y desplazada al instante de tiempo $t' = t$. Esta ventana deberá de ser de una duración temporal corta para de esta forma eliminar la parte de la señal que no esté próxima al punto t .

Para aclarar el concepto de STFT, la figura 3.2 y la figura 3.3 representan dos implementaciones diferentes de esta transformada. En la figura 3.2 se puede observar como se puede obtener la STFT de una señal filtrándola mediante un filtro paso banda centrado en la frecuencia f y demodulando el resultado (desplazándolo en frecuencia desde la frecuencia central f hasta la frecuencia central 0). La figura 3.3 representa otra implementación de la STFT donde se ha invertido el orden de los procesos de filtrado y desplazamiento en frecuencia: la STFT se obtiene desplazando en frecuencia la señal una cantidad $-f$ y a continuación filtrando el resultado mediante un filtro paso bajo. En ambos casos la forma del filtro y su ancho de banda, corresponden a la ventana de análisis empleada en la STFT.

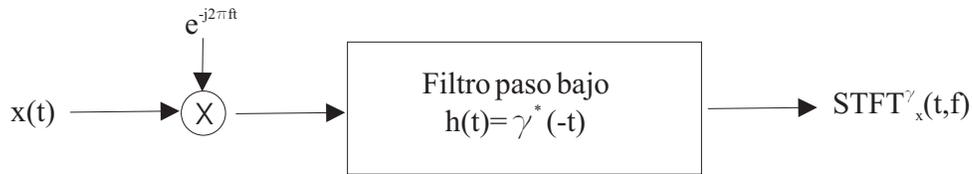


Figura 3.3: Implementación de la STFT utilizando un filtro paso bajo.

Así como la ecuación (3.12) muestra la forma de obtener la STFT partiendo de una señal $x(t)$ en el dominio del tiempo, también es posible obtener la STFT a partir del espectro de la señal, es decir a partir de $X(f)$. Para ello basta con realizar la siguiente operación:

$$STFT_x^\gamma(t, f) = e^{-j2\pi ft} \int_{f'=-\infty}^{\infty} X(f') \Gamma^*(f' - f) e^{j2\pi f' t} df' \quad (3.13)$$

Siendo $\Gamma(f)$, la transformada de Fourier de la ventana de análisis $\gamma(t)$ y $\Gamma^*(f' - f)$ una versión desplazada y conjugada de esta.

En la ecuación (3.12) se puede observar que la STFT de una señal $x(t)$ analizada en un instante de tiempo t , solamente tiene en cuenta la parte de la señal situada en las proximidades de t (puesto que al multiplicar por la ventana de análisis $\gamma^*(t' - t)$, se atenúa la señal que no esté próxima a t). A la vista de esto, interesa elegir ventanas cuya duración temporal sea pequeña para que la STFT tenga información precisa de cada instante de tiempo, es decir obtener buena resolución temporal.

De forma similar, en la ecuación (3.13) se puede observar que la STFT de una señal $X(f)$ analizada en una frecuencia dada f , solamente tiene en cuenta la parte de la señal situada en las proximidades de f (puesto que al multiplicar por la transformada de Fourier de la ventana de análisis $\Gamma^*(f' - f)$, se atenúa la parte de la señal que no esté próxima a f). A la vista de esto, interesa elegir ventanas cuyo ancho espectral sea pequeño para que la STFT indique exactamente la frecuencia existente en cada instante de tiempo, es decir obtener buena resolución frecuencial.

De los dos párrafos anteriores se deduce que sería deseable tener una buena resolución frecuencial a la vez que una buena resolución temporal y esto se conseguiría utilizando una función de análisis que fuera estrecha en el dominio del tiempo y también estrecha en el dominio de la frecuencia, pero el principio de incertidumbre [133] asegura que es imposible obtener funciones de análisis con anchos temporales y espectrales tan pequeños como se quiera: las ventanas con ancho temporal pequeño tienen el inconveniente de presentar un ancho espectral grande y viceversa.

Evidentemente la Transformada corta de Fourier de una señal va a depender en gran medida de la ventana de análisis que se elija. No obstante la STFT tiene, independientemente de la ventana de análisis elegida dos propiedades interesantes relacionadas con las representaciones tiempo frecuencia:

desplazamiento frecuencial: si una señal $x(t)$ se multiplica por un factor de frecuencia f_0 su STFT se desplaza f_0 en el eje de frecuencia. Lo cual se puede expresar matemáticamente mediante:

$$y(t) = x(t)e^{j2\pi f_0 t} \Rightarrow STFT_y^\gamma(t, f) = STFT_x^\gamma(t, f - f_0) \quad (3.14)$$

desplazamiento temporal: si una señal $x(t)$ se desplaza en el tiempo una cantidad t_0 , su STFT se desplaza en el tiempo la misma cantidad t_0 y se multiplica por un factor según la expresión:

$$y(t) = x(t - t_0) \Rightarrow STFT_y^\gamma(t, f) = STFT_x^\gamma(t - t_0, f)e^{-j2\pi f t_0} \quad (3.15)$$

transformada inversa (ISTFT): Una señal $x(t)$ se puede recuperar a partir de su transformada $STFT_x^\gamma(t, f)$ mediante la operación:

$$x(t) = \int_{t'=-\infty}^{\infty} \int_{f'=-\infty}^{\infty} STFT_x^\gamma(t', f')g(t - t')e^{j2\pi f' t} dt' df' \quad (3.16)$$

donde $g(t)$ se denomina “ventana de síntesis” y deberá mantener con la ventana de análisis ($\gamma(t)$) la relación:

$$\int_{t=-\infty}^{\infty} g(t)\gamma^*(t)dx = 1 \quad (3.17)$$

También es posible recomponer la señal $x(t)$ a partir de muestras equiespaciadas en tiempo y en frecuencia de la $STFT_x^\gamma(t, f)$ que reciben el nombre de coeficientes de Gabor [140]:

$$G_x(n, k) = STFT_x^\gamma(nT, kF) \quad (3.18)$$

En este caso $x(t)$ obtendrá mediante la expresión:

$$x(t) = \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} G_x(n, k)g_{nk}(t) \quad (3.19)$$

donde:

$$g_{nk}(t) = g(t - nT)e^{j2\pi(kF)t} \quad (3.20)$$

La transformada wavelet (WT).

Otra de las representaciones tiempo frecuencia lineal más importantes es la transformada wavelet en su versión tiempo frecuencia, que se define como [127]:

$$WT_x^\gamma(t, f) = \int_{t'=-\infty}^{\infty} x(t') \sqrt{|f/f_0|} \gamma^* \left(\frac{f}{f_0}(t' - t) \right) dt' \quad (3.21)$$

Donde $\gamma(t)$ es lo que se conoce como función de análisis; y juega un papel similar al de la ventana de análisis en la STFT. La expresión $\gamma^* \left(\frac{f}{f_0}(t' - t) \right)$ que aparece en la ecuación (3.21) es una versión de ésta conjugada y desplazada en tiempo y en frecuencia. Las funciones de análisis utilizadas en transformadas wavelet son señales paso banda (en lugar de las señales paso bajo que se utilizan como ventanas de análisis en STFT), que deben cumplir la condición:

$$\int_{t=-\infty}^{\infty} \gamma(t) dt = 0 \quad (3.22)$$

Además, si se compara la expresión (3.21), que se utiliza para definir la transformada Wavelet, con la expresión (3.12), que se utiliza para definir la STFT, se observa otra gran diferencia: en una transformada de Fourier (3.12) la ventana se desplaza en tiempo y frecuencia mediante la operación:

$$\gamma^*(t' - t) e^{-j2\pi f t'} \quad (3.23)$$

Si nos fijamos en el desplazamiento frecuencial (que se produce mediante la multiplicación por $e^{-j2\pi f t'}$), observamos que el espectro de la señal desplazada tiene exactamente la misma forma que la señal original (puesto que al multiplicar por un fasor, la representación frecuencial únicamente sufre un desplazamiento).

En un análisis wavelet (representado por la ecuación (3.21)), la función de análisis se desplaza en tiempo y frecuencia mediante la operación:

$$\sqrt{|f/f_0|} \gamma^* \left(\frac{f}{f_0}(t' - t) \right) \quad (3.24)$$

Si nos fijamos en el desplazamiento frecuencial (suponiendo $t = 0$), observamos que además del desplazamiento se produce un ensanchamiento (si $f < f_0$, o compresión en caso contrario) de la función de análisis; y esto no corresponde exactamente a un desplazamiento del espectro de la función sino a una compresión (si $f < f_0$, o ensanchamiento en caso contrario). Y como la función de análisis tiene un espectro paso banda, este escalado en la frecuencia supone también un desplazamiento de la frecuencia central del espectro

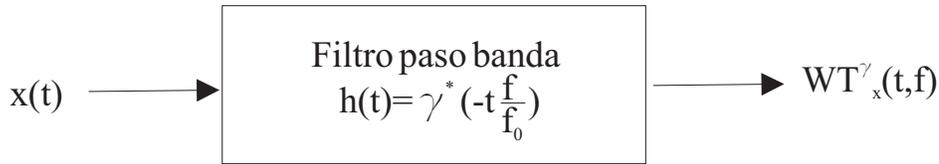


Figura 3.4: Implementación de la WT utilizando un filtro paso banda.

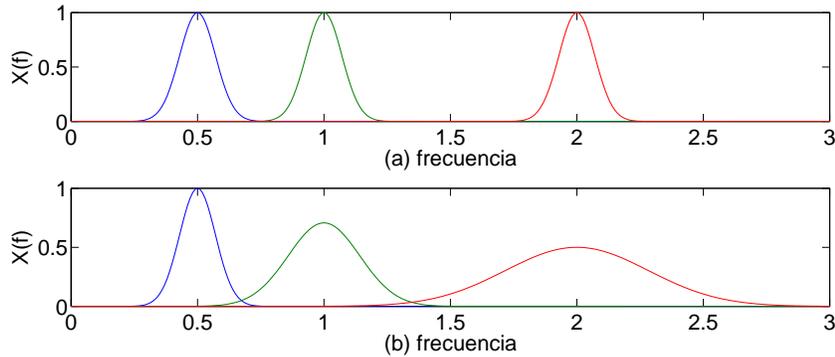


Figura 3.5: Ejemplos de filtros utilizados en la implementación de STFT (a) y WT (b) mediante bancos de filtros. En (a) se representan 3 filtros utilizados en la obtención de una STFT (según el esquema de la figura 3.2) de frecuencias centrales 0.5, 1 y 2 respectivamente. En (b) se representan 3 filtros utilizados en la obtención de una WT (según el esquema de la figura 3.4) de frecuencias centrales 0.5, 1 y 2 respectivamente. En la gráfica (a) se observa que todos los filtros tienen la misma forma mientras que en la gráfica (b) se observa que la forma del filtro depende de la frecuencia central de éste.

de la función de análisis, aunque en este caso (a diferencia de lo que ocurre en el análisis de Fourier) sí que se deforma el espectro de la función de análisis.

Para aclarar las similitudes y diferencias entre el análisis mediante WT y STFT, la figura 3.4 representa una implementación de la transformada wavelet mediante un filtro paso banda centrado en la frecuencia f . A diferencia de lo que ocurre en la figura 3.3 (que representa una implementación de la STFT) para obtener la WT mediante un filtrado no es necesario el uso de dos bloques (uno para seleccionar la frecuencia y otro para seleccionar la forma del filtro), y esto hace que al variar el parámetro f para ajustar la frecuencia se varía también la forma del filtro utilizado. De hecho, el filtro utilizado en la STFT tiene un ancho de banda constante e independiente de la frecuencia, mientras que el filtro utilizado en la WT tiene un ancho de banda proporcional a la frecuencia, o dicho de otro modo, el factor de calidad del filtro “Q” (cociente entre la frecuencia central del filtro y su ancho de banda) es constante. Por eso la WT se puede considerar como un análisis con factor de calidad constante.

En la figura 3.5 se muestran ejemplos de filtros utilizados en las implementaciones de STFT y WT de las figuras 3.2 y 3.4 respectivamente. En ambos casos se representan los filtros correspondientes a las frecuencias centrales 0.5, 1 y 2. Pero mientras que en (a) se observa que todos los filtros tienen la misma forma, en (b) la forma del filtro depende de

la frecuencia central de éste (factor de calidad constante).

Para que la transformada wavelet se pueda utilizar en el análisis de la variación temporal de las frecuencias contenidas en una señal, se requiere el uso de una función de análisis analítica compleja (que también separa los componentes de amplitud de los de fase) mientras que si lo que se desea es detectar discontinuidades o variaciones bruscas en señales se utilizan funciones de análisis reales [141]. Las funciones de análisis analíticas son funciones complejas que deben cumplir que su transformada de Fourier es cero para cualquier frecuencia negativa:

$$\Gamma_a(f) = 0 \quad \text{si } f < 0 \quad (3.25)$$

La transformada Wavelet tiene dos propiedades interesantes relacionadas con las representaciones tiempo frecuencia:

Desplazamiento temporal: si una señal $x(t)$ se desplaza en el tiempo una cantidad t_0 , su WT se desplaza en el tiempo la misma cantidad t_0 según la expresión:

$$y(t) = x(t - t_0) \Rightarrow WT_y^\gamma(t, f) = WT_x^\gamma(t - t_0, f) \quad (3.26)$$

Escalado: si una señal es escalada en el dominio del tiempo con un factor a , su transformada wavelet se escala con el inverso de ese factor ($1/a$) según indica la expresión:

$$y(t) = \sqrt{|a|}x(at) \Rightarrow WT_x^\gamma(t, f) = WT_x^\gamma(at, \frac{f}{a}) \quad (3.27)$$

Transformada Wavelet Inversa(IWT) y reconstrucción de señales.

En la sección (3.2.1) se definió la transformada wavelet en su versión tiempo frecuencia mediante la expresión (3.21) aunque la forma más habitual de expresarla es la versión tiempo escala de la transformada wavelet definida como:

$$WT_x^\gamma(t, a) = \int_{t'=-\infty}^{\infty} x(t') \frac{1}{\sqrt{a}} \gamma^* \left(\frac{t' - t}{a} \right) dt' \quad (3.28)$$

donde a recibe el nombre de escala.

Comparando la expresión (3.21) con la ecuación (3.28) se puede comprobar que la relación entre frecuencia y escala es:

$$a = \frac{f_0}{f} \quad (3.29)$$

y como se puede observar la escala y la frecuencia son inversamente proporcionales, por lo que al observar representaciones de transformadas wavelet se debe tener presente si son representaciones tiempo frecuencia o representaciones tiempo escala.

De forma similar a lo que ocurre con la STFT, se puede obtener una señal original a partir de su WT. Para señales reales o sus versiones analíticas, basta con considerar valores positivos de la variable escala según la expresión:

$$x(t) = c \int_{t'=-\infty}^{\infty} \int_{a=0}^{\infty} WT(t', a) \frac{1}{\sqrt{a}} \gamma\left(\frac{t-t'}{a}\right) dt' \frac{da}{a^2} \quad (3.30)$$

donde c es una constante que depende de la función de análisis utilizada.

Hay que tener en cuenta que la reconstrucción de señales realizada mediante la expresión anterior únicamente es posible con señales que no tengan componentes continuas, puesto que las funciones de análisis (o wavelets) son necesariamente señales paso banda con un nivel de continua nulo (ecuación (3.22)).

Al igual que ocurre con la STFT, la WT tiene gran cantidad de información redundante, y la reconstrucción de una señal se podría hacer mediante la utilización de “muestras” de esta información redundante de la WT de forma similar a la expansión de Gabor mostrada en la ecuación (3.19). Según esta expresión, se puede reconstruir una señal mediante el uso de unos coeficientes $G_x(n, k)$ que se pueden obtener tomando muestras equiespaciadas en tiempo (nT) y frecuencia (kF) de la STFT.

De forma similar, se puede reconstruir una señal a partir de su transformada wavelet. Aunque en este caso, las muestras no se deberán tomar de forma equiespaciada en el eje de frecuencias. Esto se debe a que la transformada wavelet presenta mayor resolución temporal y menor resolución frecuencial en frecuencias altas que en frecuencias bajas, como se puede observar en la figura 3.4 donde se interpreta la transformada wavelet como un filtrado mediante un banco de filtros de “Q” constante. Por lo que el muestreo del eje de frecuencias de la transformada wavelet se debe realizar según la expresión:

$$f_j = f_0 \alpha^j \quad (3.31)$$

donde α es el factor que determinará si el eje de frecuencias se muestreará con mayor o menor número de muestras (de forma similar a lo que ocurre con F en la expresión (3.36)).

El muestreo del eje de tiempo de la transformada wavelet tampoco se debe realizar con el mismo espaciado para todas las frecuencias, puesto que las frecuencias altas de la transformada wavelet corresponden a filtrados de la señal (figura 3.4) realizados con filtros de mayor ancho de banda que las frecuencias bajas. El teorema del muestreo de Nyquist indica que las señales de mayor ancho de banda deben ser muestreadas con intervalos de muestreos inferiores a las señales con menor ancho de banda. Por esta razón, cada una de las frecuencias f_j de la expresión (3.31) lleva asociado un tiempo de muestreo diferente y

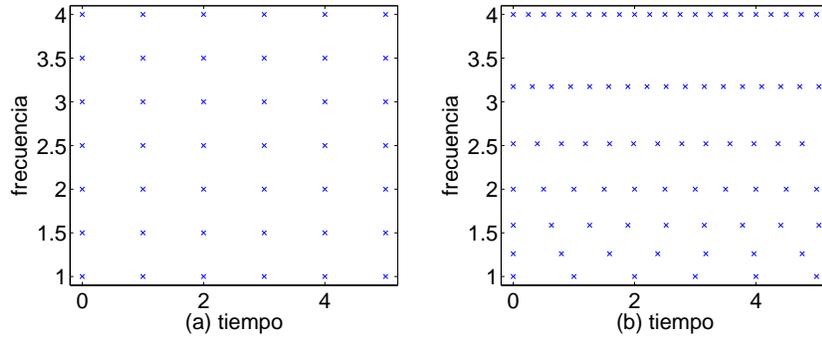


Figura 3.6: Puntos de muestreo del plano Tiempo frecuencia. Cada marca representa un punto de muestreo de la STFT (a) o la WT (b) para obtener los valores $G_x(n, k)$ de la ecuación (3.36) (a) y los valores $W(n, k)$ de la ecuación (3.34) (b). En (a) se ha tomado $F = 0,5$ y $T = 1$ y en (b) se ha tomado $\alpha = \sqrt[3]{2}$ y $T = 1$.

el eje de tiempos se deberá muestrear en los instantes:

$$t_k = \frac{kT}{\alpha^j} \quad (3.32)$$

Y teniendo en cuenta las condiciones anteriores, una señal se podrá reconstruir a partir de su transformada wavelet según la expresión:

$$x(t) = \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} W(n, k)w(t, f) \quad (3.33)$$

siendo:

$$W(n, k) = WT_x^\gamma\left(\frac{kT}{\alpha^j}, f_0\alpha^j\right) = \int_{t'=-\infty}^{\infty} x(t')\sqrt{\alpha^j}\gamma^*(\alpha^j t' - kT) dt' \quad (3.34)$$

y

$$w(t, f) = \alpha^j \tilde{\gamma}(\alpha^j t - kT) \quad (3.35)$$

La figura 3.6 representa dos formas de muestrear el plano tiempo-frecuencia. En (a) se representa el tipo de muestreo utilizado en las expansiones de Gabor para obtener $G_x(n, k)$ de la ecuación (3.19) con de $F = 0,5$ y $T = 1$. Y en (b) se representa el tipo de muestreo utilizado en WT para obtener los valores $W(n, k)$ de la ecuación (3.31) tomando $\alpha = \sqrt[3]{2}$ y $T = 1$.

Independencia lineal, ortogonalidad y completitud de bases.

Las ecuaciones (3.19) y (3.33) nos permiten descomponer señales mediante la combinación lineal de unas funciones base, donde los coeficientes de la combinación son muestras de la transformada corta de Fourier y de la transformada Wavelet, respectivamente. En este tipo de descomposiciones son muy importantes desde el punto de vista matemático las propiedades de independencia lineal, ortogonalidad y completitud de bases empleadas [135, 142, 143, 144].

La *completitud* de unas bases garantiza que cualquier señal de energía finita pueda ser expresada como combinación lineal de esas bases. En el caso de la expansión de Gabor, esta completitud se puede conseguir haciendo que los valores de T y F en la expresión (3.36) sean pequeños (valores positivos cercanos a 0). En el caso de las wavelets la completitud se consigue haciendo que los valores de T en la ecuación (3.36) sean pequeños (con valores positivos cercanos a 0) y los de α sean próximos a 1.

Si se seleccionan los valores de T , F y α como se ha indicado en el párrafo anterior, los coeficientes de las combinaciones lineales (obtenidos mediante el muestreo de la STFT y de la WT) contienen información redundante y suponen un sobremuestreo de la STFT y de la WT. En este caso las bases utilizadas en la combinación lineal no son *linealmente independientes* y por lo tanto los coeficientes $W(n, k)$ y $G_x(n, k)$ que aparecen en las expresiones (3.19) y (3.33) respectivamente no quedan unívocamente definidos.

Para que las bases de la expansión de Gabor sean completas, es necesario que los valores de T y F en la expresión (3.36) cumplan:

$$TF \leq 1 \quad (3.36)$$

y únicamente en el caso de que se cumpla la igualdad se consigue que los coeficientes $G(n, k)$ no contengan redundancia.

En cuanto a la *ortogonalidad* de las bases utilizadas en la expansión de Gabor cabe indicar que si se desea conseguir una buena localización en tiempo y en frecuencia las bases no serán ortogonales [143, 145]. Aunque se pueden realizar modificaciones en la expansión de Gabor para conseguir que las bases sean ortonormales [144].

En cuanto a la elección de los parámetros T y α en las transformadas wavelets ocurre algo similar a lo indicado para la expansión de Gabor: mientras α tome valores próximos a 1 y T valores pequeños, la completitud queda garantizada, aunque ello supone que las bases no sean ortogonales ni linealmente independientes. Pero en aplicaciones como codificación de señales o compresión, donde se busca un muestreo eficiente en el plano tiempo frecuencia, es típico utilizar un valor de $\alpha = 2$ (computación octava por octava) y un valor de T similar a la duración temporal de la wavelet [127, 146]. En estos casos, es difícil obtener unas bases linealmente independientes que sean a la vez completas y ortogonales. Estos aspectos fueron estudiados por Mayat [147, 148] y Meyer [149] que desarrollaron una completa teoría matemática basada en espacios de funciones multirresolución. A partir de esta teoría Daubichies [150, 151] y otros autores han conseguido desarrollar wavelets

que permitan la reconstrucción perfecta de señales.

3.2.2. Representaciones Tiempo-Frecuencia cuadráticas.

Tanto la STFT como la WT toman valores complejos, aunque en muchas aplicaciones la fase no aporta información relevante y se trabaja únicamente con sus módulos. En estos casos resulta interesante trabajar con el espectrograma y el escalograma, que son representaciones cuadráticas asociadas a la STFT y la WT que toman únicamente valores reales y positivos.

El espectrograma se define como:

$$SPEC_x^\gamma(t, f) = |STFT_x^\gamma(t, f)|^2 \quad (3.37)$$

Mientras que el escalograma se define como:

$$SCAL_x^\gamma(t, f) = |WT_x^\gamma(t, f)|^2 \quad (3.38)$$

Una de las aplicaciones más comunes del espectrograma es el tratamiento de la voz [152], aunque también se ha aplicado al análisis de otras señales no estacionarias como el estudio de las olas de los océanos [153] y también se está utilizando en aplicaciones novedosas como el estudio del genoma [154] mientras que entre las aplicaciones del escalograma se encuentra el estudio de ondas sísmicas [155] y principalmente la detección de roturas [156, 157, 158].

Las dos representaciones anteriores se pueden obtener mediante un filtrado bidimensional de otra representación cuadrática más general conocida como distribución de Wigner-Ville [159], que se define como:

$$W_x(t, f) = \int_{\tau=-\infty}^{\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau \quad (3.39)$$

Las aplicaciones de esta distribución han sido variadas: mejora en la detección de señales; análisis de voz, audio y acústica [160]; procesado y reconocimiento de imágenes; análisis de señales médicas, biológicas o sísmicas entre otras [161, 135, 162].

3.3. Distribuciones Tiempo-Frecuencia discretas.

Las señales procedentes de la naturaleza son generalmente continuas, aunque debido a los avances que se han producido en los sistemas electrónicos e informáticos de adquisición y procesado de datos, es cada vez más habitual convertir estas señales continuas

en discretas para procesarlas, representarlas o almacenarlas. Las representaciones tiempo frecuencia analizadas anteriormente en este capítulo serían difíciles de implementar (si no imposibles) mediante técnicas analógicas, puesto que requerirían del uso de filtros difícilmente realizables.

Así pues, la forma habitual de procesar y representar señales reales consiste en digitalizarlas primero (teniendo en cuenta las restricciones impuestas por el teorema del muestreo de Nyquist), para trabajar posteriormente con las señales discretas en lugar de las continuas. Por esta razón, los procedimientos y definiciones realizados en las secciones anteriores para señales continuas necesitan ser redefinidos para trabajar con señales discretas.

En esta sección se desarrollarán los conceptos de análisis tiempo-frecuencia asociados a señales discretas y los problemas de implementación asociados a este tipo de señales.

3.3.1. Espectrograma y STFT.

Definición de STFT.

En las figuras 3.2 y 3.3, se observa cómo la STFT de una señal continua se podría implementar mediante el uso de filtros paso banda o filtros paso bajo. En el caso del cálculo de STFT discretas, también existen dos posibilidades. La primera consiste en dividir una señal en pequeños trozos (mediante el enventanado) y aplicar posteriormente una transformada de Fourier a cada uno de esos “trozos” finalizando con un muestreo del resultado en el eje tiempo y en el eje frecuencia (si se considera necesario). La segunda posibilidad correspondería al filtrado de la señal mediante un banco de filtros y diezmar (en caso necesario) la señal de salida de cada uno de los filtros del banco de filtros. Las dos opciones son adecuadas para obtener representaciones de la STFT, pero la segunda opción no permite recuperar (salvo que se usen filtros especiales) la señal original a partir de la STFT. De ahora en adelante, nos referiremos a las dos formas expuestas para el cálculo de la STFT como método de la ventana y método del banco de filtros respectivamente.

Cuando se usa el método de la ventana, se define la STFT de una señal discreta $x[n]$ de la siguiente forma:

$$STFT_x^w[n, \lambda] = \sum_{m=-\infty}^{\infty} x[n+m]w[m]e^{-j2\pi\lambda m} \quad (3.40)$$

Siendo $w[n]$ la ventana aplicada.

En la expresión anterior, cuando se escribe $STFT_x^w[n, \lambda]$, el uso de un corchete para la variable n y un paréntesis para la variable λ indica que son una variable discreta y otra continua respectivamente.

A partir de esta transformada se puede recuperar la señal original mediante la operación:

$$x[n] = \frac{1}{w[0]} \int_{\lambda=0}^1 STFT_x^w[n, \lambda] d\lambda \quad (3.41)$$

Si, mediante el cambio de variable $m' = n + m$, se reescribe la ecuación (3.40), se obtiene:

$$STFT_x^w[n, \lambda] = \sum_{m=-\infty}^{\infty} x[m'] w[-(n - m')] e^{j2\pi\lambda(n-m')} \quad (3.42)$$

Y esto permite escribir la STFT como una convolución:

$$STFT_x^w[n, \lambda] = x[n] * h_\lambda[n] \quad (3.43)$$

siendo:

$$h_\lambda[n] = w[-n] e^{j2\pi\lambda n} \quad (3.44)$$

Si se toman ventanas (o funciones de análisis) simétricas respecto al origen de tiempos nos encontramos con problemas prácticos si se desea trabajar en tiempo real debido a la no causalidad, por lo que cuando se desea trabajar en tiempo real se seleccionan ventanas simétricas desplazadas convenientemente con el fin de evitar este tipo de problemas.

Cuando se usa el método del banco de filtros es habitual trabajar con una definición de la STFT distinta a (3.40):

$$\hat{X}^w[n, \lambda] = \sum_{m=-\infty}^{\infty} x[m] w[m - n] e^{-j2\pi\lambda m} \quad (3.45)$$

siendo la relación entre estas dos formas de definir la STFT [163]:

$$\hat{X}^w[n, \lambda] = e^{-j2\pi\lambda n} STFT_x^w[n, \lambda] \quad (3.46)$$

El propósito fundamental de la ventana en la STFT es limitar la duración de la señal a analizar, para que sus características espectrales permanezcan razonablemente constantes en el periodo analizado. Cuando más rápidamente cambian las características espectrales de la señal que se analiza, más corta deberá de ser la ventana utilizada. Pero al igual que ocurre con las STFT continua (como se discutió en la sección 3.2.1) cuando se disminuye el tamaño de una ventana de una STFT discreta, su resolución frecuencial también disminuye. La elección del tamaño de la ventana deberá ser un compromiso entre su resolución temporal y su resolución frecuencial.

Muestreo de la STFT en tiempo y en frecuencia.

En la ecuación (3.40) se definió la STFT de una señal discreta como una función de dos variables: el tiempo discreto (n), que se mide en muestras; y la frecuencia discreta (λ) que se mide en ciclos por muestra. De forma práctica, la STFT, solamente se puede analizar para un número finito de puntos en cada valor de n ; y este análisis se suele realizar para muestras equiespaciadas de la variable frecuencia.

La transformada de Fourier de una señal discreta se define mediante la expresión:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} \quad (3.47)$$

y la Transformada Discreta de Fourier (o DFT) de una señal $x[n]$ de duración N muestras, definida mediante la expresión:

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi}{N}kn} \quad k = 0, 1, \dots, N-1. \quad (3.48)$$

Se puede demostrar [163] que una señal de L muestras de duración, queda perfectamente definida mediante los N muestras equiespaciadas de su transformada de Fourier siempre que se cumpla que $N \geq L$.

$$X[k] = X(e^{j\omega})\Big|_{\omega=\frac{2\pi}{N}k} \quad k = 0, 1, \dots, N-1 \quad (3.49)$$

Estos N valores se pueden conseguir, de forma práctica, rellenando con 0's la señal $x[n]$ hasta conseguir que tenga una duración N y calculando a continuación su DFT, que se puede obtener de forma eficiente mediante el algoritmo FFT (Fast Fourier Transform).

Por lo tanto, cada una de las señales que se producen al inventanar una señal de duración infinita con una ventana de duración L muestras, deberá quedar perfectamente definida mediante las N muestras de su DFT (siempre que se cumpla que $N \geq L$.); y esto nos conduce a una versión muestreada en frecuencia de la ecuación (3.41):

$$X[n, k] = STFT_x^w[n, \lambda]_{\lambda=\frac{k}{N}} = STFT_x^w\left[n, \frac{k}{N}\right] = \sum_{m=0}^{L-1} x[n+m]w[m]e^{-j\frac{2\pi km}{N}} \quad (3.50)$$

La transformación representada en la ecuación (3.50) contiene información redundante, puesto que contiene un espectro para cada valor de n . También se puede demostrar [163] que es posible recuperar la señal $x[n]$ a partir de un muestreo en el tiempo de la ecuación (3.50) siempre que el intervalo de muestreo R sea menor o igual que el tamaño de la ventana L , quedando definida la versión muestreada de la STFT para señales discretas de la siguiente forma:

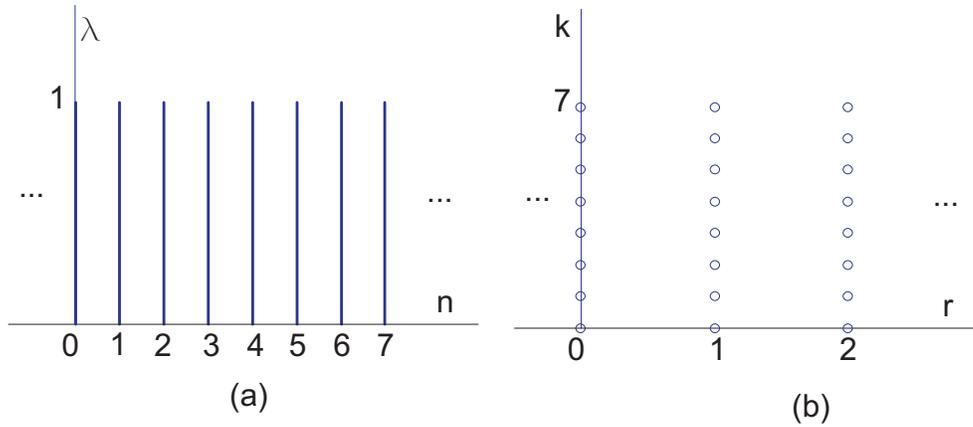


Figura 3.7: Dominio la STFT. En la parte izquierda (a) se representa el dominio de la STFT definida en la ecuación (3.40); y en la parte derecha (b) se representa el dominio de la STFT definida en la ecuación (3.51). La representación de la derecha se obtiene muestreando la de la izquierda con los parámetros $R=3$, $N=8$.

$$X[rR, k] = \sum_{m=0}^{L-1} x[rR + m]w[m]e^{-j\frac{2\pi km}{N}} \quad (3.51)$$

Debiendo cumplirse que el tamaño de cada una de las DFT (N) sea mayor o igual que el tamaño de la ventana utilizada (L) y este a su vez mayor o igual que el intervalo de muestreo de la variable tiempo (R):

$$N \geq L \geq R \quad (3.52)$$

Para aclarar estos conceptos, la figura 3.7 representa un ejemplo de los dominios de las STFT definidas en las ecuaciones (3.40) y su versión discreta (3.51) (zona del plano tiempo frecuencia en las que la STFT se encuentra definida). En la parte izquierda (a) se representa el dominio de la STFT definida en la ecuación (3.40), donde se puede apreciar que la función existe para cualquier valor de n natural y para valores de frecuencia correspondientes a los números reales entre el intervalo $[0, 1]$. En la parte derecha (b) se representa el dominio de la STFT definida en la ecuación (3.51): en este caso la función existe para valores de r naturales y valores de k naturales entre 0 y 7. La STFT asociada a la representación de la derecha (b) corresponde a un muestreo en los ejes tiempo y frecuencia de la STFT representada en la izquierda (a): el eje tiempo se ha muestreado de 3 en 3 muestras ($R = 3$) y el eje frecuencia se ha muestreado dividiendo el eje de frecuencias en 8 puntos ($N = 8$).

Dado que el eje de tiempo se muestrea con un periodo R y las ventanas utilizadas tienen una longitud $L (\geq R)$, las distintas ventanas tienen $(L - R)$ muestras comunes, y este valor $(L - R)$ se conoce como solape.

3.3.2. Escalograma y Wavelets.

La estrecha relación existente entre una STFT y una transformada wavelet continua queda de manifiesto cuando ambas se representan mediante bancos de filtros. La diferencia fundamental entre estas dos transformadas radica en que los filtros que componen el banco son de ancho de banda constante en el caso de STFT y de factor Q constante en el caso de la transformada Wavelet.

En el caso de las transformaciones discretas, la relación entre ambas es incluso mayor: además de la similitud que se observa cuando se comparan las estructuras de bancos de filtros, la teoría de wavelets discretas se puede desarrollar a partir de los conceptos asociados a la teoría de banco de filtros de reconstrucción perfecta desarrollada a principio de la década de 1980. De hecho, cuando se publicaron los primeros trabajos sobre wavelets, el gran interés que despertó en la comunidad científica no fue compartido por los diseñadores de bancos de filtros, puesto que desde su punto de vista esta teoría desarrollaba conceptos ya conocidos. Lo cierto es que la novedad de las wavelets radica en el proceso de iteración, puesto que los filtros utilizados en las wavelets deben de comportarse bien tras varias iteraciones que combinen diezmado y filtrado. Y esto último añade una nueva restricción que deben de cumplir los filtros empleados para la construcción de wavelets que no se tenía en cuenta cuando se diseñaban bancos de filtros: los filtros utilizados en wavelets deben tener un número “suficiente” de ceros en la frecuencia discreta 0.5 ($z=-1$).

En el Apéndice 1 se analizan los bancos de filtros de reconstrucción perfecta y se describe el modo en que se puede implementar un análisis wavelet mediante el uso de filtros. También se estudian las relaciones existentes entre wavelets continuas y discretas. En [164, 165] Strang realiza una revisión de la teoría de wavelets a partir de los conceptos desarrollados en el análisis mediante banco de filtros. En [166] se relacionan las wavelets con la codificación subbanda. A continuación se presentarán las familias de wavelets más comunes con las propiedades que las caracterizan.

El campo de aplicación de las wavelets es muy variado y existe una gran cantidad de formas de onda utilizadas como wavelets. Que un determinado tipo de wavelet pueda ser utilizado para una u otra aplicación depende de las propiedades que esta posea. Las familias de wavelets más utilizadas son Haar, Daubechies, Symlets, Coiflets, Biortogonales, Gaussianas, Sombrero Mejicano, Morlet [167], Meyer [168] y Shannon [169].

Algunas wavelets tienen “*soporte compacto*”, es decir, la wavelet toma valores no nulos solamente en un rango determinado de tiempos. Esto hace que la función de análisis esté muy localizada en el tiempo. Aunque cuando las wavelets son utilizadas para realizar análisis tiempo-frecuencia resulta interesante no solo su localización temporal sino también su localización frecuencial, que la energía de la wavelet esté contenida en un rango limitado de frecuencias.

Cuando se desea realizar un análisis mediante wavelets de una gran cantidad de datos, puede resultar interesante la existencia de un “*algoritmo rápido*” para el cálculo de la transformada wavelet, y esto no es posible para todas las transformadas wavelet.

Muchas propiedades del análisis wavelet se derivan de las propiedades de las funciones

wavelet utilizadas ($\Psi(t)$) y de las “*funciones de escalado*” ($\Phi(t)$) (aunque esta función no existe para todas las wavelets). Además, estas funciones, en la mayoría de los casos, no se pueden expresar mediante expresiones matemáticas analíticas. Una propiedad deseable de la función wavelet ($\Psi(t)$) es que sea “*derivable*” (pues esto garantiza que sea una función “suave”). También es deseable que los “*momentos*” de la función wavelet sean nulos hasta cierto orden.

En aplicaciones de compresión o transmisión de señales, resulta imprescindible que la transformada wavelet sea “*invertible*”, es decir que se pueda recuperar la señal original a partir de la transformada wavelet. Para ello basta con seleccionar los filtros H_0 y H_1 representados en la figura 6.2 de modo que sean “*biortogonales*” y así permitir la reconstrucción perfecta como se indica en el Apéndice 1. Si además se desea que los filtros de reconstrucción sean idénticos a los de análisis, la condición requerida será más estricta: se necesitarán filtros “*ortonormales*” (aunque en este caso los filtros no pueden ser simétricos).

Algunas familias de wavelets permiten analizar señales continuas y discretas, aunque otras solo pueden trabajar en uno de estos dos entornos.

3.4. Procesado Tiempo Frecuencia en Acústica y Psicoacústica.

Zwicker [3] definió un modelo de análisis espectral del oído basado en un banco de filtros analógicos de tercio de octava, que posteriormente se utilizó en el estándar internacional ISO532 [1] de cálculo de Loudness. Los filtros utilizados en este análisis tienen el mismo factor de calidad Q por lo que se podría decir que matemáticamente corresponden a un análisis wavelet continuo.

Con la aparición de las técnicas de procesado digital, la implementación práctica del banco de filtro de análisis utilizado en muchas aplicaciones psicoacústicas se ha realizado mediante la STFT debido a la eficiencia computacional conseguida mediante la FFT [170, 21]. La elección de la ventana de análisis es un factor crítico, pues para obtener filtros con suficiente selectividad frecuencial en las bandas críticas inferiores se necesitan ventanas con duraciones superiores a 10 ms. Pero esta duración es excesivamente alta para ciertas aplicaciones (por ejemplo para el cálculo del Roughness) en las bandas críticas altas; y por esta razón los productos comerciales de cálculo de parámetros psicoacústicos permiten la elección de la duración de la ventana de análisis según la finalidad [9, 10]. Por otro lado, las wavelets discretas presentan problemas con su selectividad frecuencial en las bandas superiores del espectro por lo que su aplicación en psicoacústica está muy limitada. No obstante, utilizando la teoría de wavelets discretas se han desarrollado bancos de filtros que proporcionan modelos aproximados del comportamiento del oído [171, 172, 173, 174].

En Psicoacústica han aparecido técnicas que mejoran la resolución de los análisis tiempo frecuencia derivadas de la compresión de voz y de audio. Entre estas técnicas cabe destacar *wavelet packet* (derivadas del análisis wavelet discreto) y *warped spectrum* (que

se obtienen mediante una modificación de la STFT que consigue deformar el eje frecuencia de forma apropiada).

Carnero [175] propuso en 1999 la utilización de *wavelet packet* en la codificación de voz que tuvo aplicaciones posteriores en la codificación de música y audio en general [176, 177] y que ha sido utilizado en la mejora de los algoritmos de compresión utilizados en MP3 [178].

El esquema de wavelet packet propuesto por Carnero, fue también empleado por Xing [179] para desarrollar un modelo psicoacústico del oído, utilizado entre otras aplicaciones para el análisis de la calidad de la voz [180].

El análisis wavelet packet está basado en la división del espectro de una señal en bandas frecuenciales mediante el uso de wavelets discretas: una señal se descompone en banda alta y banda baja dando lugar a dos nuevas señales que pueden ser descompuestas sucesivamente en distintos *niveles de descomposición*. Las señales aparecidas en cada nivel pueden ser descompuestas de nuevo o no, por lo que el análisis wavelet packet permite distintas representaciones de las señales en función del *árbol de descomposición* escogido. Partiendo del análisis de Carnero, Philippe [181] estudió la utilización de distintas familias de wavelets en diferentes puntos del árbol, Shao estudió la utilización del modelo psicoacústico de Carnero con distintas configuraciones del árbol de descomposición [182], Karmakar [183] diseñó un árbol de descomposición que se ajusta a la escala Bark y Abid [184] utilizó distintas wavelets para adaptarse a la escala Bark y a la escala ERB.

La idea del *warped spectrum* (o espectro combado), que permite deformar el eje frecuencia en una representación Tiempo-Frecuencia, fue introducida por Oppenheim [185] en 1971. Evangelista [171, 172] utilizó el warped spectrum para diseñar wavelets que permitieran una distribución del plano tiempo frecuencia que se ajustasen a los modelos cocleares y perceptuales utilizados en voz y música. Smith [186] calculó los parámetros que debía tener el warped spectrum para que se ajustase a la escala de frecuencias Bark y a la ERB; Härmä y Painter [187, 176] describen las aplicaciones de esta técnica en audio y acústica; y Parfieniuk y Petrovsky [188, 189] proponen un modelo del oído basado en la transformada discreta de Fourier warped.

La descomposición mediante cepstrum (u homomórfica) fue introducida por Bogert [190] en 1963. Se utiliza en aplicaciones de procesado digital de voz, por ejemplo para extraer el pitch de sonidos vocálicos o características espectrales de una señal [163, 191], ya que el cepstrum posee propiedades interesantes para el manejo de señales periódicas [192, 193].

A continuación se describe la utilización de todas estas técnicas en psicoacústica.

3.4.1. Wavelet Packet.

La transformada wavelet discreta es una transformación donde las escalas y los desplazamientos tan solo pueden tomar valores basados en las potencias de dos (escalas y posiciones diádicas). Una forma eficiente de implementar esta transformada es utilizando filtros que, mediante una estructura de codificación sub-banda bi-canal, permiten obte-

ner los coeficientes de la transformada wavelet dando origen a la Fast Wavelet Transform (FWT). En una transformación wavelet discreta, se descompone una señal en otras dos, filtrando la original paso bajo y paso alto. El filtrado paso bajo suele ser muy similar a la original excepto en algunos 'detalles' (de alta frecuencia) por lo que, en términos de wavelets, realizar esta operación de dividir una señal en sus componentes paso bajo y paso alto se llama obtener la aproximación (A) y el detalle (D) respectivamente.

Las aproximaciones y los detalles son diezmados dando origen al vector de coeficientes de la aproximación (cA) y al vector de coeficientes del detalle (cD). Es importante distinguir entre detalle o aproximación de nivel N (DN o AN) y coeficientes del detalle o la aproximación de nivel N (cDN o cAN) puesto que la aproximación tiene el mismo número de muestras que la señal original mientras que los coeficientes tienen un número menor de muestras porque en cada nivel se produce un diezmado.

Teóricamente, el proceso se podría repetir indefinidamente pero en realidad solamente se puede hacer hasta que únicamente quede una muestra. En la práctica se elige el número de iteraciones en función de la complejidad de la señal con la que se esté trabajando. La descomposición wavelet packet es una generalización de la descomposición wavelet que permite ampliar el estudio que se realiza mediante wavelets.

Mientras que en un análisis wavelet se descompone una señal en aproximación y detalle y a continuación se puede volver a descomponer únicamente la aproximación, con el uso de wavelet packet se puede descomponer en cada nivel tanto la aproximación como el detalle.

El análisis de señales mediante Wavelet Packet permite seleccionar la resolución temporal y frecuencial para distintas zonas del espectro. En cuanto a la función Wavelet a utilizar, se dispone de un amplio abanico de señales; pero entre todas ellas, la familia de wavelets que mejor preserva la selectividad en frecuencia a medida que aumenta el número de niveles es la de Daubechies [175]. El número de niveles necesarios en la Wavelet Packet depende de la resolución frecuencial que se desee alcanzar teniendo en cuenta que la resolución del análisis en el nivel p es:

$$\Delta f = \frac{f_s}{2^{p+1}} \quad (3.53)$$

La figura 3.8 representa el árbol de descomposición utilizado por Carnero [175] en la compresión de señales de voz. Al utilizar una frecuencia de muestreo de 16 kHz basta con un nivel de descomposición de 6 niveles para alcanzar una resolución en frecuencia de 125 Hz en las bandas más bajas.

Con este esquema se consigue dividir el espectro de la señal en 21 bandas, que corresponden aproximadamente a bandas críticas en la escala Bark.

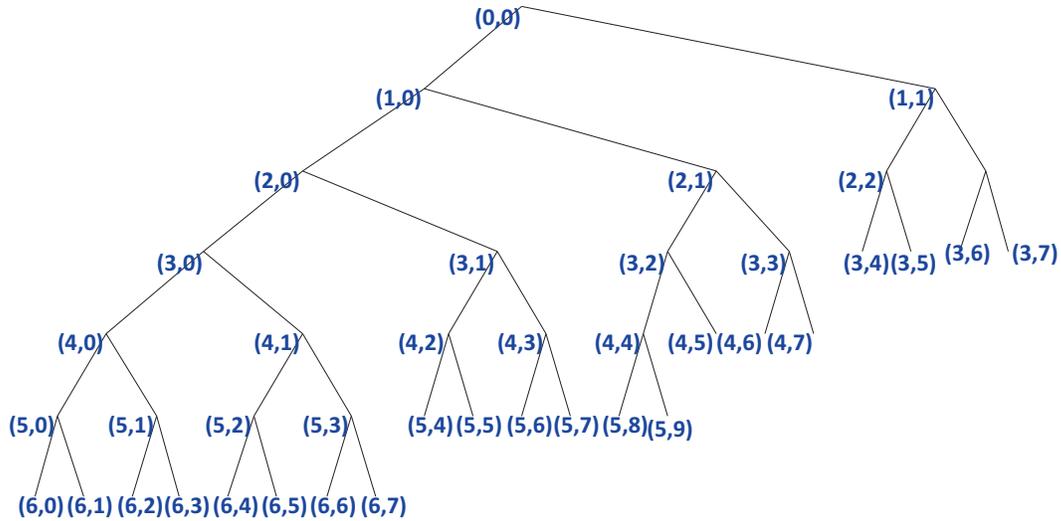


Figura 3.8: Descomposición wavelet packet. Los números entre paréntesis indican el nivel de la descomposición y el número asociado a la banda de frecuencias dentro de ese nivel.

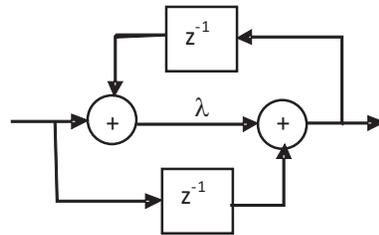


Figura 3.9: Bloque elemental utilizado en warped spectrum.

3.4.2. Warped Spectrum.

La escala frecuencial utilizada habitualmente en procesamiento digital de señales es lineal en relación con la escala Herciana, y por tanto la resolución frecuencial es uniforme en toda la escala de frecuencias comprendidas entre la frecuencia mínima representable (0Hz) y la máxima (la mitad de la frecuencia de muestreo).

Esto se debe a que el bloque fundamental en cualquier diseño digital es el retardador unitario z^{-1} , que retarda por igual cualquier frecuencia (una sola muestra). Por lo tanto, al hacer la transformada de Fourier se obtiene un conjunto de puntos que representan el contenido frecuencial de una señal en puntos equidistantes en frecuencia.

Una forma sencilla de trabajar con escalas que no estén relacionadas linealmente con la frecuencia Herciana, es sustituir el retardador de una muestra por un bloque diferente que permita transformar una escala lineal de frecuencias en otra no lineal, y este bloque puede ser un filtro paso todo de primer orden como el representado en la figura 3.9, cuya función de transferencia es:

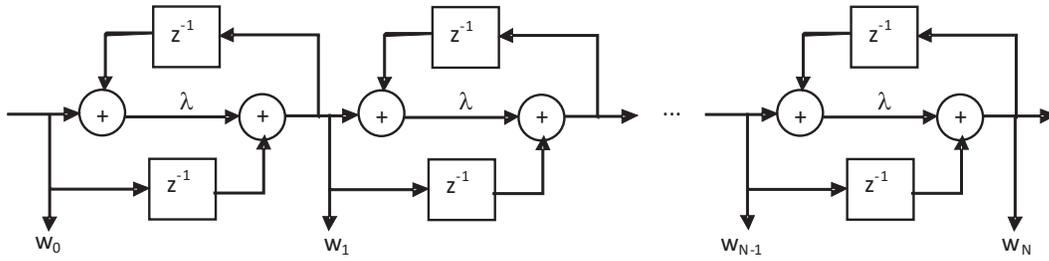


Figura 3.10: Filtrado mediante la técnica de warped spectrum.

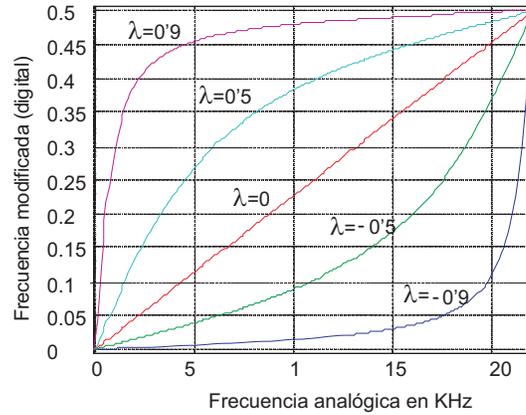


Figura 3.11: Conversión de frecuencias en warped spectrum.

$$D(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}} \quad (3.54)$$

Si se sustituye un retardador por este filtro paso todo, al realizar una operación como la de filtrado, en lugar de trabajar con $x[n], x[n-1] x[n-2] \dots x[n-N]$, se trabajará con la salida de N de estos sistemas situados en cascada $w_0, w_1, w_2, \dots, w_N$.

De modo que si se combinan los valores de $w_0, w_1, w_2, \dots, w_N$ de la misma forma que se combinarían los valores de $x[n], x[n-1] x[n-2] \dots x[n-N]$, para filtrar la pulsación ω tal y como se muestra en la figura 3.10, la frecuencia que en realidad se estará filtrando es:

$$\omega' = \arctan \frac{(1 - \lambda^2) \text{sen}(\omega)}{(1 + \lambda^2) \text{cos}(\omega) - 2\lambda} \quad (3.55)$$

donde $\omega = 2\pi f / f_s$ siendo f_s la frecuencia de muestreo.

Es decir, ω en escala Herciana, se transforma en ω' en la nueva escala Warped.

Se puede observar en la figura 3.11 que el valor que tome λ determinará la conversión de frecuencias que se pretende conseguir, de modo que si este parámetro es negativo la conversión consistirá en un ensanchamiento de las frecuencias altas y una compresión de las frecuencias bajas, mientras que si dicho parámetro es positivo, el proceso será el inverso, comprimiéndose las frecuencias altas y ensanchándose las frecuencias bajas, y

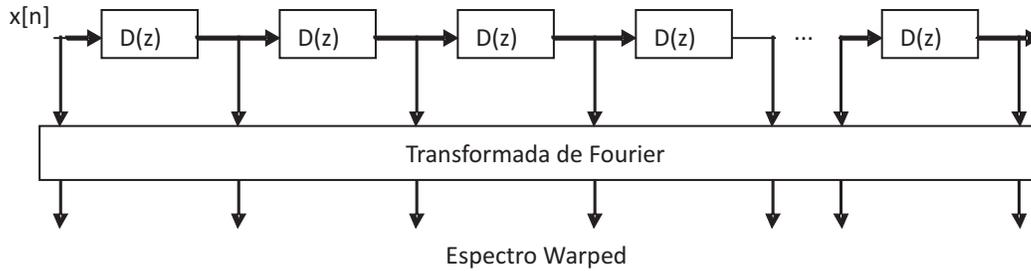


Figura 3.12: Obtención de un espectro combado.

sólo en el caso de $\lambda = 0$ no se realizará ninguna conversión de frecuencias, ya que eso corresponde a $D(z) = z^{-1}$.

Se puede obtener un espectro en el dominio Warped aplicando una FFT a la salida de una cadena de filtros paso todo, tal y como se muestra en la figura 3.12. El número de bloques paso todo que componen la cadena (N), determinará el número de puntos obtenidos al hacer una FFT ($N+1$).

Las aplicaciones de estos espectros Warped son muy variadas, aunque se debe de tener en cuenta que debido a la utilización de filtros IIR, no se puede utilizar en aplicaciones donde se necesite reconstruir las señales a partir de estos espectros.

Si se quiere convertir una escala lineal de frecuencias en una escala que se aproxime a la Bark, el parámetro λ se tendrá que elegir conforme a la expresión empírica (Smith y Abel) [186]:

$$\lambda = 1'0674 \left[\frac{2}{\pi} \arctan(0'06583 f_s) \right]^{1/2} - 0'1916 \quad (3.56)$$

Que ofrece el valor de λ en función de la frecuencia de muestreo utilizada en la digitalización de la señal analógica.

Al aplicar un filtro paso todo a una señal se aplican desfases diferentes a las distintas frecuencias que la componen. Los desfases asociados a cada frecuencia distinta se pueden obtener a partir de la respuesta frecuencial del filtro paso todo:

$$D(e^{j\omega}) = \frac{e^{-j\omega} - \lambda}{1 - \lambda e^{-j\omega}} \quad (3.57)$$

Como resultado, las frecuencias altas prácticamente no sufren ningún retraso comparadas con las bajas, que sufren un retraso de entre 0'5 y 1 ms por cada filtro paso todo. El efecto de estos retrasos desiguales se puede considerar despreciable cuando el número de filtros paso todo utilizados es pequeño, pero se puede convertir en un inconveniente importante cuando se utilizan una cantidad elevada de estos filtros.



Figura 3.13: Diagrama de bloques de la descomposición cepstrum de la señal $x(t)$.



Figura 3.14: Diagrama de bloques de la descomposición cepstrum de la señal $x[n]$ mediante la Transformada Discreta de Fourier.

3.4.3. Cepstrum.

El cepstrum complejo de una señal continua muestreada se puede calcular mediante el diagrama que aparece en la Figura 3.13: primero hay que calcular la transformada z ($X(z)$) de la señal $x[n]$ con la que se va a trabajar; a continuación se calcula el logaritmo de $X(z)$ para obtener $\hat{X}(z)$; y por último se calcula la transformada z inversa de $\hat{X}(z)$. El resultado, representado por $\hat{x}[n]$ recibe el nombre de cepstrum complejo de $x[n]$ (aunque habitualmente consiste en una secuencia real). El atributo “complejo” se utiliza para distinguirlo del cepstrum real, consistente básicamente en las mismas operaciones que el anterior, pero eliminando la parte imaginaria de $\hat{X}(z)$ para calcular $\hat{x}[n]$.

A pesar de que el cepstrum complejo de una señal se calcula utilizando la transformada z , las implementaciones prácticas de esta transformación utilizan la transformada discreta de Fourier (o su implementación más eficiente FFT) [163].

El diagrama de bloques asociado a la implementación práctica del cálculo del cepstrum mediante el uso de la FFT es el que se presenta en la Figura 3.14.

Cepstrum de una convolución.

“El cepstrum complejo de la convolución de dos señales es la suma de los cepstrum complejos de cada uno de ellas”. Esta propiedad se puede demostrar fácilmente a partir de la definición de cepstrum y partiendo de la convolución de dos señales:

$$x[n] = x_1[n] * x_2[n], \quad (3.58)$$

entonces su transformada de Fourier será:

$$X(z) = X_1(z) \cdot X_2(z), \quad (3.59)$$

y el logaritmo de la transformada:

$$\hat{X}(z) = \log\{X_1(z) \cdot X_2(z)\} = \log\{X_1(z)\} + \log\{X_2(z)\} = \hat{X}_1(z) + \hat{X}_2(z). \quad (3.60)$$

Con lo que la transformada “z” inversa queda:

$$\hat{x}[n] = \hat{x}_1[n] + \hat{x}_2[n], \quad (3.61)$$

Cepstrum de varias deltas equiespaciadas.

Si se considera el calculo del cepstrum de la señal:

$$c[n] = a_0\delta[n] + a_1\delta[n - N] + a_2\delta[n - 2N], \quad (3.62)$$

cuya transformada z es:

$$C(z) = a_0 + a_1z^{-N} + a_2z^{-2N} = a_0(1 - \alpha z^{-N})(1 - \beta z^{-N}), \quad (3.63)$$

y donde α y β pueden ser únicamente numeros reales o una pareja de números complejos conjugados.

Al calcular el logaritmo de $C(z)$ obtenemos:

$$\hat{C}(z) = \log(a_0) + \log(1 - \alpha z^{-N}) + \log(1 - \beta z^{-N}), \quad (3.64)$$

y desarrollando la serie de Taylor del logaritmo obtenemos:

$$\hat{C}(z) = \log(a_0) - \sum_{k=1}^{\infty} \frac{\alpha^k z^{-kN}}{k} - \sum_{k=1}^{\infty} \frac{\beta^k z^{-kN}}{k} = \log(a_0) - \sum_{k=1}^{\infty} \frac{(\alpha^k + \beta^k) z^{-kN}}{k}. \quad (3.65)$$

Por lo que el cepstrum complejo del tren de pulsos $c[n]$ es:

$$\hat{c}[n] = \begin{cases} \log(a_0) & , si \ n = 0 \\ - \sum_{k=1}^{\infty} \frac{(\alpha^k + \beta^k)}{k} \delta[n - kN] & , si \ n > 0 \end{cases} \quad (3.66)$$

y esto es un conjunto de deltas localizados en instantes de tiempo múltiplos del periodo de la señal.

La expresión anterior se puede generalizar diciendo que: “ una señal formada por un conjunto de varias deltas con distintas amplitudes y separadas entre si N muestras presenta un *cepstrum formado por* varias deltas separadas entres si N muestras”.

Cepstrum de una señal periódica.

Una señal periódica se puede escribir como la convolución:

$$y[n] = p[n] * c[n]. \quad (3.67)$$

Siendo $c[n]$. un tren de deltas y $p[n]$ la forma de onda asociada al patrón que se repite periódicamente. Por lo que el cepstrum de la señal periódica será:

$$\hat{y}[n] = \hat{p}[n] + \hat{c}[n]. \quad (3.68)$$

Y como el cepstrum de una señal periódica ($\hat{c}[n]$) está formado por deltas situadas en los múltiplos del periodo de repetición de la señal, resulta sencillo obtener $\hat{p}[n]$ a partir de $\hat{y}[n]$ simplemente eliminando aquellas deltas que aparecen en los múltiplos del periodo de la señal.

Además, puesto que el cepstrum de una señal ($c[n]$) compuesta por *deltas equiespaciadas y de distinta amplitud* también está formado por deltas situadas en los múltiplos del periodo de la señal, el método descrito también se puede utilizar para recuperar el patrón de repetición de señales donde éste se repita con amplitud variable. Este tipo de señales (que en esta tesis se han denominado “*señales Cuasiperiódicas*”) se pueden modelar mediante la expresión 3.67 siendo:

$$c[n] = \sum_{k=-\infty}^{\infty} a_k \delta[n - kN]. \quad (3.69)$$

A modo de ejemplo, la Figura 3.15 muestra un sonido formado mediante la convolución entre una señal patrón $p[n]$ y tres impulsos localizados en los instantes $n = 0$, $n = 36$ y $n = 72$ con amplitudes respectivas 1, 0,9 and 0,95. La figura muestra también el cepstrum de cada una de las señales utilizadas para formar el sonido.

Como se puede observar en la Figura 3.15, el cepstrum complejo de la convolución entre la señal patrón $p[n]$ y $c[n]$ corresponde a la suma de sus respectivos cepstrum complejos. De los apartados (d) y (f) del gráfico podemos afirmar que el cepstrum complejo de $p[n]$ ($\hat{p}[n]$), se puede extraer a partir del cepstrum complejo de la señal total $y[n]$ (ver ecuaciones (3.67)-(3.68)) simplemente eliminando los picos que aparecen en el lado derecho. Si se tiene en cuenta que el cepstrum complejo de la señal $c[n]$ decae al menos tan rápido como $1/|n|$, (ecuación 3.66), otro método por el que se puede conseguir una aproximación del cepstrum de $p[n]$ consiste en eliminar todas aquellas muestras de $\hat{y}[n]$ alejadas del origen.

3.5. Comparativa Técnicas TF

Las prestaciones de cada método de análisis TF dependen de la aplicación a que se destinen. En una representación gráfica Tiempo-Frecuencia bidimensional, los análisis TF cuadráticos proporcionan resultados de fase nula, lo que supone una ventaja en la representación de los resultados. Sin embargo estas representaciones cuadráticas no son adecuadas para el procesado o almacenamiento de señales no estacionarias puesto que al

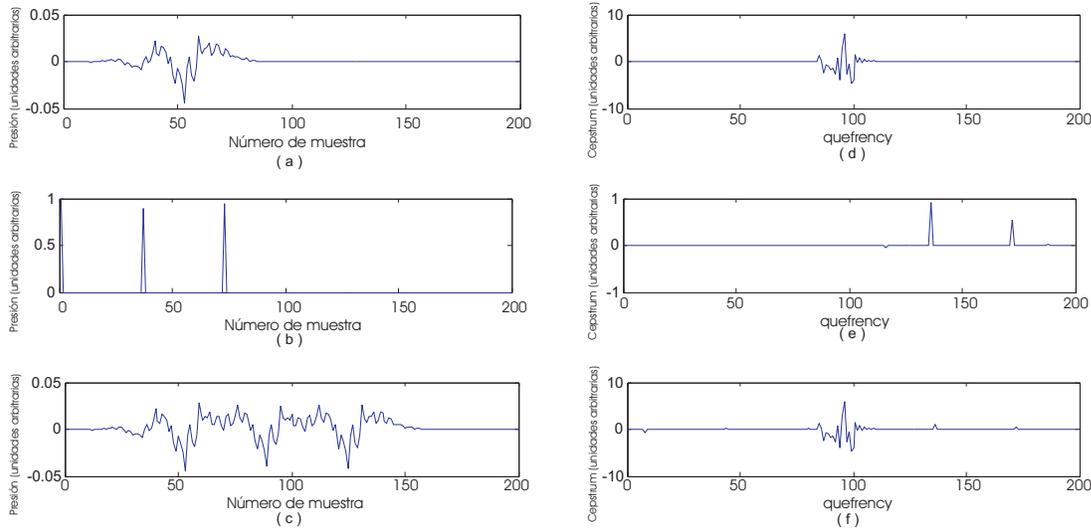


Figura 3.15: Sonido formado mediante la repetición de una señal patrón a intervalos regulares de tiempo. En la parte izquierda del gráfico y de arriba hacia abajo se representa: (a) patrón $p[n]$, (b) deltas equiespaciadas $c[n]$ y (c) sonido resultante de la convolución $y[n] = c[n] * p[n]$. En la parte derecha y de arriba hacia abajo los gráficos (d),(e) y (f) corresponden al cepstrum de (a), (b) y (c) respectivamente.

eliminar la información de fase la reconstrucción de la señal original a partir de los datos del análisis es prácticamente imposible.

Por otro lado, la escala de frecuencias logarítmica (asociada a la WT) representa mejor las escalas de frecuencias utilizadas en Psicoacústica que la escala lineal (asociada a la TF) puesto que modela mejor la escala de frecuencias del oído humano. En cuanto a la complejidad computacional asociada a cada una de estas escalas, en ambos casos existen algoritmos rápidos de cálculo, pero al utilizar escalas logarítmicas la transformada Wavelet Discreta presenta una gran rigidez a la hora de descomponer las señales en bandas de frecuencias cuyos anchos de banda solo pueden tomar valores relacionados mediante potencias de 2. Para utilizar wavelets con anchos de banda arbitrarios se debe trabajar con Wavelets continuas que computacionalmente son menos eficientes.

Algunas técnicas de análisis como wavelets continuas, wavelet paquet o warped spectrum consiguen buenas aproximaciones de la escala auditiva humana, lo que ha permitido grandes avances en procesado y compresión de audio. Pero si se desea realizar un análisis psicoacústico con precisión, la única opción es implementar los filtros normalizados asociados a las escalas psicuacústicas, para lo que las técnicas más convenientes son las asociadas a las Transformadas de Fourier. Además, los modelos psicoacústicos normalizados se han implementado en base a estos filtros analógicos, por lo que para poder contrastar resultados experimentales con los existentes en la bibliografía se hace necesario utilizar los mismos modelos.

El análisis mediante cepstrum se ha destacado en este capítulo porque permite obtener las características espectrales de los ruidos generados por el motor de un vehículo. Estas características se utilizaran en el capítulo 4 como base para desarrollar un modelo de

generación de señales de motor sintéticas cuyos parámetros psicoacústicos se analizarán en el capítulo 5.

Capítulo 4

Modelado de ruido de motor

4.1. Introducción.

Con el fin de realizar un estudio psicoacústico aplicado al ruido producido por el motor de un vehículo, nos hemos centrado en dos tipos de ruido: el ruido de combustión y el ruido de escape.

Para evaluar el grado de molestia o satisfacción del sonido de un determinado producto es necesario trabajar con un gran número de muestras de estos sonidos; y esto se convierte en una tarea tediosa cuando los sonidos son difíciles de generar. Este es el caso de los ruidos de motor, cuando en un diseño se pretenden estudiar los distintos sonidos que produce, puesto que se hace necesario realizar variaciones en sus parámetros de funcionamiento o incluso en la cámara de combustión de éste [194, 195, 196].

Además, si se desea utilizar estos sonidos en experimentos con parámetros psicoacústicos controlados (Loudness, Roughness, Sharpness o Fluctuation Strength [3]) nos encontramos con un problema adicional: la variación de un parámetro del motor provoca variaciones en todos los parámetros psicoacústicos simultáneamente siendo imposible aislar uno de ellos.

Con el fin de dar una solución a este problema se han desarrollado modelos y programas informáticos capaces de sintetizar este tipo de ruidos [197, 20, 198], obteniendo de esta forma un gran número de señales sin tener que recurrir a la grabación de sonidos reales. Entre los modelos anteriores, destaca el propuesto por Feng [20] puesto que fue específicamente diseñado para analizar los parámetros psicoacústicos (especialmente el Roughness) producido por las señales de motor.

En este capítulo se va a desarrollar un nuevo modelo que permita generar señales sintéticas de combustión y de escape a partir del análisis de señales de motor reales y de su mecanismo de generación.

Según Anderton [199], las características del ruido producido por un motor de combustión interna se pueden explicar a partir de un modelo lineal en el que se deben considerar las fuerzas que actúan como entradas del sistema (combustión y mecánicas) y la respues-

ta del sistema (desde el punto de vista de transmisión de vibraciones y de radiación de ruido).

En el interior del cilindro de un motor se producen variaciones de presión en el rango de frecuencias audibles, con niveles de presión superiores a 200 dB, que son atenuados por el bloque de motor entre 80 y 150 dB. El ruido producido de esta forma se conoce como “ruido de combustión”. Desde el punto de vista mecánico, el bloque del motor se puede modelar como un conjunto de masas, muelles y amortiguadores [199], que al ser combinados entre si forma un sistema que puede ser considerado lineal e invariante en el tiempo (LTI) [200, 201].

En el caso del ruido de escape este se forma mediante la excitación del sistema de escape del motor por la señal procedente de la válvula de escape de los cilindros del motor [202, 203, 197].

En ambos casos los ruidos generados se pueden modelar como la superposición de una señal periódica con ligeras variaciones (que hemos llamado modelo Cuasiperiódico) y un ruido coloreado.

Tras esta introducción, el capítulo comienza con un análisis de los ruidos de combustión y de escape, a continuación se describe el modelo de Feng, posteriormente se propone un nuevo modelo para sintetizar señales de motor y se comparan señales reales con otras sintéticas para probar la validez del modelo. Por último se realiza un análisis frecuencial de los distintos modelos de síntesis de señal que permitirán (en el capítulo siguiente) evaluar los parámetros psicoacústicos de las señales sintéticas.

4.2. Ruido producido por un motor

Esta sección está dedicada al análisis de los sonidos producidos por un motor y su proceso de generación con el fin de sentar las bases de un modelado T-F de este tipo de ruidos.

4.2.1. Presión en el interior de un cilindro.

En el análisis del ruido generado por la combustión, se considera que la fuente de ruido es la presión existente en el interior de los cilindros, y el sistema que transforma ese ruido (que se puede considerar lineal [204, 205, 206]) está compuesto por la estructura metálica que envuelve la cámara de combustión y los elementos físicos existentes en el entorno del motor y del punto de recepción.

El modelo de onda simplificado propuesto por Anderton [207] correspondiente a la presión existente en el interior de un cilindro, consiste en un tramo creciente (formado por un trozo de señal coseno) y otro decreciente (formado por otro trozo de señal coseno de amplitud distinta al anterior) según la expresión:

$$p(t) = \begin{cases} c \frac{P_0}{2} \left(1 + \cos\left(\frac{\pi}{T_0}t\right)\right) & \text{si } -T_0 < t < 0 \\ \frac{P_0}{2} \left(1 + \cos\left(\frac{\pi}{T_0}t\right)\right) & \text{si } 0 \leq t < T_0 \end{cases} \quad (4.1)$$

Donde P_0 es el valor máximo que tiene la presión en el interior del cilindro, T_0 es el tiempo que transcurre desde que la presión tiene valor 0 hasta que alcanza su valor máximo (que coincide con el tiempo que transcurre desde el máximo de presión hasta que vuelve a valer 0); c es el cociente entre la amplitud del coseno que forma la parte izquierda de la señal (el tramo creciente) y la amplitud del coseno de la parte derecha de la señal (el tramo decreciente).

La presión en el interior de un cilindro en un motor diesel es diferente a la que existe en un motor de gasolina, puesto que en el motor diesel se introduce el combustible en el interior del cilindro en un punto próximo al Punto Muerto Superior del ciclo de compresión y esto produce un aumento brusco de la presión en este instante. En los motores de gasolina este incremento no se produce y ello equivale a sustituir el parámetro c por el valor constante 1.

La descomposición frecuencial de la señal de presión en el cilindro, se puede obtener mediante la transformada de Fourier de la señal de presión representada en (4.1):

$$P(f) = \frac{P_0}{2} \frac{\left(2(2\pi f)^2 + (e^{+j2\pi f T_0} - 1) \left(\frac{\pi}{T_0}\right)^2\right) (1 - c) - \left(\frac{\pi}{T_0}\right)^2 2 \cdot j \cdot \text{sen}(2\pi f T_0)}{(j2\pi f) \left((2\pi f)^2 - \left(\frac{\pi}{T_0}\right)^2\right)} \quad (4.2)$$

La figura 4.1 es una representación de la variación de la presión en el interior de un cilindro de un motor que gira a 1000 revoluciones por minuto y de su transformada de Fourier (según el modelo de Anderton [207]). La presión máxima alcanzada en el interior del cilindro en este ejemplo es de 100 bares y las distintas gráficas corresponden a distintos valores del parámetro c . La duración de la señal de presión ($2T_0$) corresponde a la sexta parte de un ciclo de motor (720° o 2 revoluciones de motor); y esto se aproxima bastante bien a la presión en el interior de los cilindros de un motor de cuatro tiempos a carga máxima.

La figura 4.2 representa la variación a lo largo del tiempo de una señal de presión real, capturada en el interior de un cilindro de un motor diesel de cuatro tiempos que gira a 3000 revoluciones por minuto. La señal representada se repite cada 0.04 segundos, lo que corresponde a una frecuencia 25Hz. Este valor de 25 Hz se obtiene dividiendo entre 2 el régimen de giro del motor medido en RPM (puesto que se necesitan 2 revoluciones para completar un ciclo) y dividiendo el resultado entre 60 (para pasar de ciclos por minuto a ciclos por segundo).

Las señales periódicas concentran su potencia en los armónicos de la frecuencia fundamental de la señal, que en el caso de la señal mostrada en la figura 4.2 es de 25 Hz, como se puede apreciar en la figura 4.3 donde se representa su transformada de Fourier.

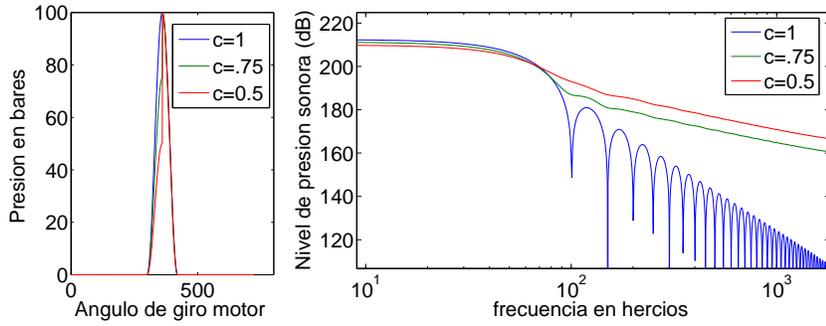


Figura 4.1: Representación del modelo de Anderton [207] de la presión en el interior del cilindro de un motor de cuatro tiempos girando a una velocidad de 1000 R.P.M. para distintos valores de salto de presión. La parte izquierda representa la variación de la presión en función del ángulo de giro del motor mientras que la parte de la derecha representa la densidad espectral de energía del sonido producido por un solo ciclo de la presión en el interior del cilindro.

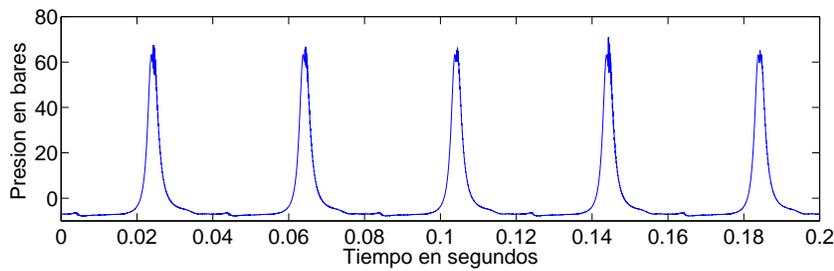


Figura 4.2: Representación de la presión en el interior de uno de los cilindros de un motor diesel de 4 tiempos funcionando a 3000 revoluciones por minuto con una inyección de combustible de $28mm^3$.

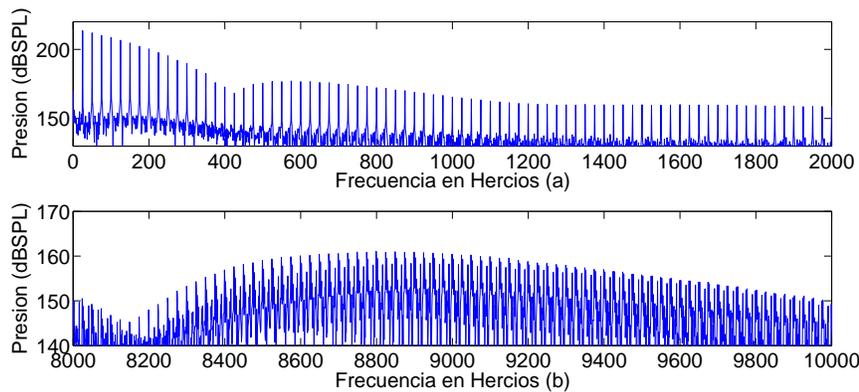


Figura 4.3: Representación de la composición frecuencial de la señal de la figura 4.2 en dos bandas de frecuencias (a) para frecuencias entre 0 y 2000 Hz (b) para frecuencias entre 8000 y 10000 Hz.

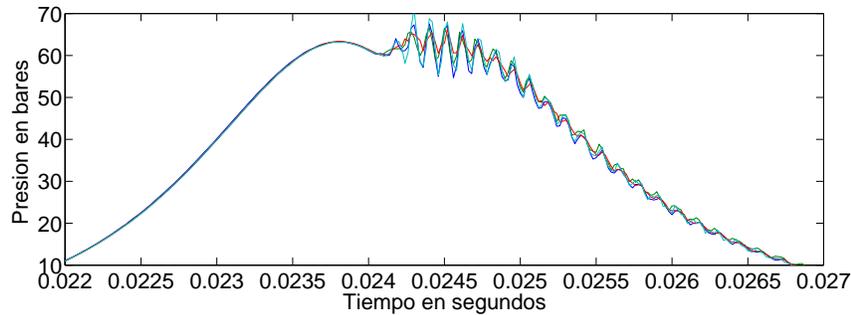


Figura 4.4: Ampliación de la señal mostrada en la figura 4.2 durante los 5 milisegundos alrededor del máximo superponiendo los cuatro primeros periodos.

Si se observa con detalle la gráfica (a) de la figura 4.3 se podrá comprobar que entre dos armónicos consecutivos el espectro no se hace nulo. Esto indica que la señal no es exactamente periódica (a pesar de su apariencia), sino que está mezclada con ruido de un nivel muy reducido.

Por otra parte, al observar la gráfica (b) de figura 4.3, espectro comprendido entre 8000Hz y 10000Hz, se observa, además de un menor nivel de la señal, un aumento del ruido, que supone una distribución que difiere bastante de la estructura de armónicos correspondiente a las señales periódicas, puesto que la potencia no se concentra en los múltiplos de la frecuencia fundamental. Así pues, se puede decir que las frecuencias bajas de la presión en el interior del cilindro se repiten de forma periódica mientras que las frecuencias altas de esta señal de presión no se repiten de forma estrictamente periódica.

En la figura 4.4 se han representado de forma superpuesta los diferentes ciclos de la señal que aparece en figura 4.2 para analizar lo que ocurre en las proximidades de los máximos. Se observa que alrededor del máximo se producen unas oscilaciones debidas a la combustión que son ligeramente diferentes en los distintos ciclos de la señal de presión de motor. Esto se debe a que la combustión en el interior del cilindro no es exactamente igual en todos los ciclos puesto que depende de la distribución de las partículas de combustible en el interior de éste. En la figura 4.4 también se observa que, aunque el instante en el que se inician las oscilaciones es prácticamente idéntico en todos los ciclos, la forma de las oscilaciones difiere entre los distintos ciclos.

La figura 4.5, que representa el contenido espectral de los ciclos mostrados en la figura 4.4, confirma que, aunque todas estas oscilaciones tienen una frecuencia aproximada de 10kHz, tiene un gran ancho de banda y la forma exacta de la onda varía ciclo a ciclo.

La forma de la onda de presión en el interior del cilindro depende de parámetros de funcionamiento del motor como la cantidad de combustible inyectado en la cámara de combustión. La figura 4.6 representa la forma que tiene la onda de presión para distintos valores de inyección de combustible: 0mm^3 , 28mm^3 y 68mm^3 . Cuando la inyección de combustible es de 0mm^3 el motor no gira por sí mismo (puesto que en su interior no se produce la combustión que proporciona la potencia necesaria para moverlo) sino que es movido por una fuerza externa a él. En este caso particular, se observa que no se producen

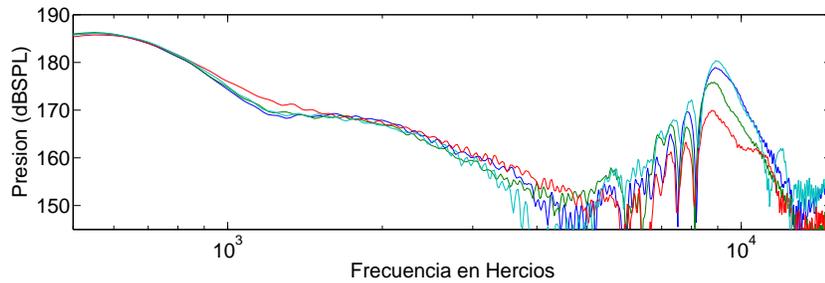


Figura 4.5: Representación espectral de las señales que aparecen en la figura 4.4.

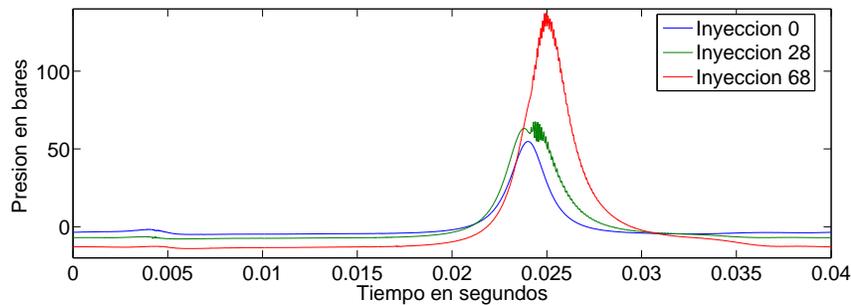


Figura 4.6: Representación de un periodo de la presión en el interior de uno de los cilindros de un motor diesel de 4 tiempos funcionando a 3000 revoluciones por minuto con una inyección de combustible de 0mm^3 , 28mm^3 y 68mm^3 .

estas oscilaciones alrededor del máximo de presión y la comparación de esta señal con el resto se utiliza para comprobar el efecto que produce la combustión. En este caso, la variación de la presión se debe únicamente al movimiento del pistón en el interior del cilindro. Las gráficas de la presión en el interior del cilindro correspondientes a valores de inyección de combustible 28mm^3 y 68mm^3 se pueden descomponer como la superposición de dos señales: una asociada al movimiento del pistón en el interior del cilindro (que varía suavemente y coincide con la gráfica asociada a 0mm^3 de inyección) y otra asociada al incremento de presión que produce la explosión del combustible (contiene las variaciones rápidas de presión).

Para ver más en detalle las diferencias entre estas señales, la figura 4.7 representa el espectro de los tres ciclos mostrados en la figura 4.6 (con distinta inyección de combustible). Aquí se puede apreciar que mientras que el espectro de la señal asociada a una inyección de combustible de 0mm^3 no contiene frecuencias altas, las otras sí que las contienen. Estudios exhaustivos con distintos valores de inyección, de carga y de velocidad asocian el ruido de combustión a valores de frecuencia superiores a 800 Hz [195, 208], ya que para frecuencias inferiores la variación de la presión se debe únicamente al movimiento del pistón.

Esta observación nos permite representar las señales de presión descompuestas en altas frecuencias y bajas frecuencias como se muestra en la figura 4.8. Aunque las altas frecuencias presentan unos niveles muy reducidos en comparación con las bajas, pueden

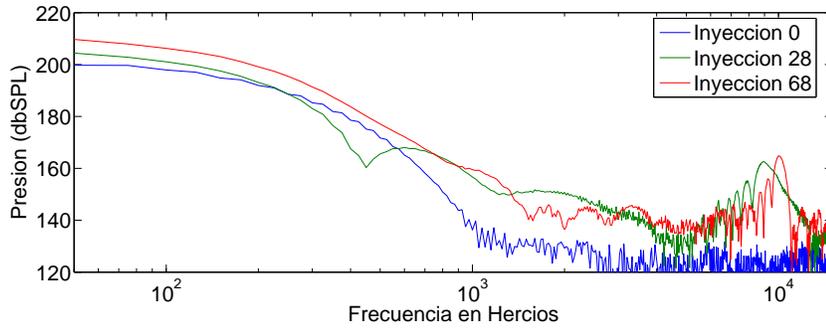


Figura 4.7: Transformada de Fourier de las señales representadas en la figura 4.6.

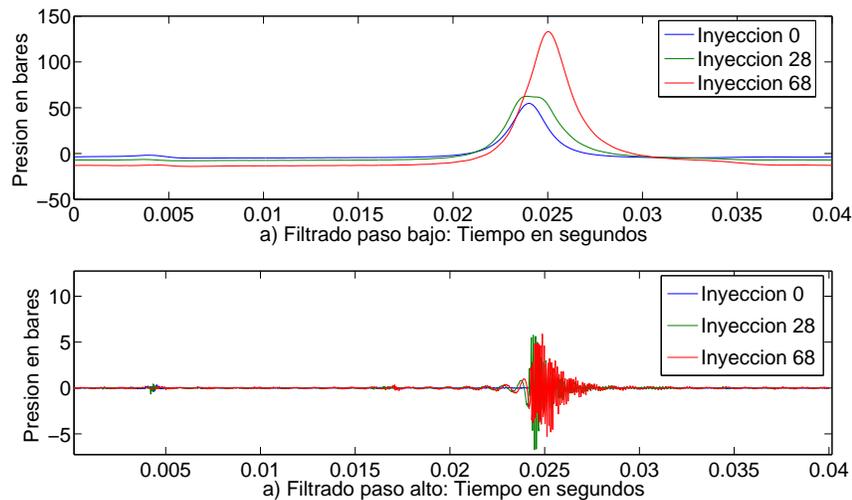


Figura 4.8: Filtrado paso bajo (a) y paso alto (b) con frecuencia de corte de 1000Hz de las señales representadas en las figuras 4.6 y 4.7.

ser de gran ayuda para localizar el instante de tiempo en el que se produce la combustión puesto que están muy localizadas en el dominio temporal.

Una comparación más detallada entre las tres ondas mostradas en las figuras 4.6 y 4.7 permite utilizar técnicas de análisis temporal y frecuencial para separar la señal de presión en el interior de un cilindro en tres señales diferentes [209], asociadas también a procesos físicos diferentes:

Señal pseudo-motriz: debida al movimiento del pistón en el cilindro, con independencia de la existencia o no de combustión, está compuesta únicamente por frecuencias bajas (señal con inyección $0mm^3$).

Señal de combustión: es la responsable del movimiento del motor puesto que le proporcionan casi la totalidad de la energía mecánica que se extrae del motor. También está formada por frecuencias bajas, pero en este caso se deben a la combustión. A diferencia de la Pseudo-motriz, su valor aumenta al aumentar la cantidad de

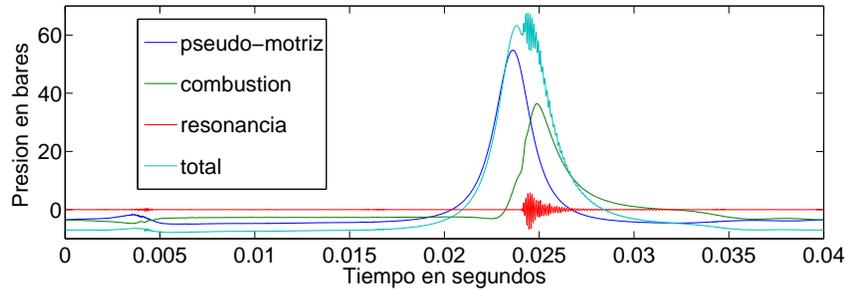


Figura 4.9: Descomposición de la presión en el interior de un cilindro, al que se le inyectan 28mm^3 de combustible, en 3 señales.

combustible en la cámara de combustión.

Señal de resonancia de la cámara de combustión: se produce debido a la distribución irregular de los gases que en el proceso de combustión excitan los modos de resonancia del cilindro y corresponden a las frecuencias altas de la señal de combustión.

La figura 4.9 muestra cómo se puede descomponer la presión en el interior de un cilindro al que se le inyecta en cada ciclo 28mm^3 de combustible en estas tres señales. Si se descompone la presión en el interior del cilindro en tres bandas de frecuencia obtenemos las gráficas que aparecen en la figura 4.10; se observa que, mientras las bajas y medias frecuencias se repiten de forma bastante regular, las altas frecuencias (asociadas a la “resonancia de la cámara de combustión”) se suceden de forma bastante irregular.

Las técnicas de representación Tiempo-Frecuencia como espectrogramas y escalogramas son de gran utilidad en el análisis de estas señales. La utilización la wavelet continua de Shanon que aparece en la figura 4.11, proporciona el escalograma de la figura 4.12.

La figura 4.12 muestra la presión en el interior de un cilindro al que se le inyecta en cada ciclo 28mm^3 de combustible (gráfica (a)) y dos escalogramas de esta señal: un escalograma para valores de escala comprendidos entre 1 y 100 (gráfica (b)); y otro escalograma para valores de escala comprendidos entre 1 y 10 (gráfica (c)). Los colores asociados a cada nivel corresponden a escalas logarítmicas. Hablando en términos de frecuencia, la gráfica (b) correspondería a un análisis entre 50 kHz y 0.5 kHz mientras que la gráfica (c) correspondería a un análisis entre 50 kHz y 5 kHz. En las representaciones anteriores se han situado las escalas más pequeñas en la parte superior de las gráficas para que en términos de frecuencia los valores más altos aparezcan en la parte superior.

Del análisis de estos escalogramas se pueden extraer las siguientes conclusiones:

- El contenido frecuencial de la señal en bajas frecuencias se repite ciclo a ciclo mientras que en las altas frecuencias existen variaciones entre los ciclos, tanto en amplitud como en frecuencia.

- En el momento en que se produce la combustión aparece una señal de aproximadamente 10 kHz ($a = 5$) cuya frecuencia disminuye casi linealmente.

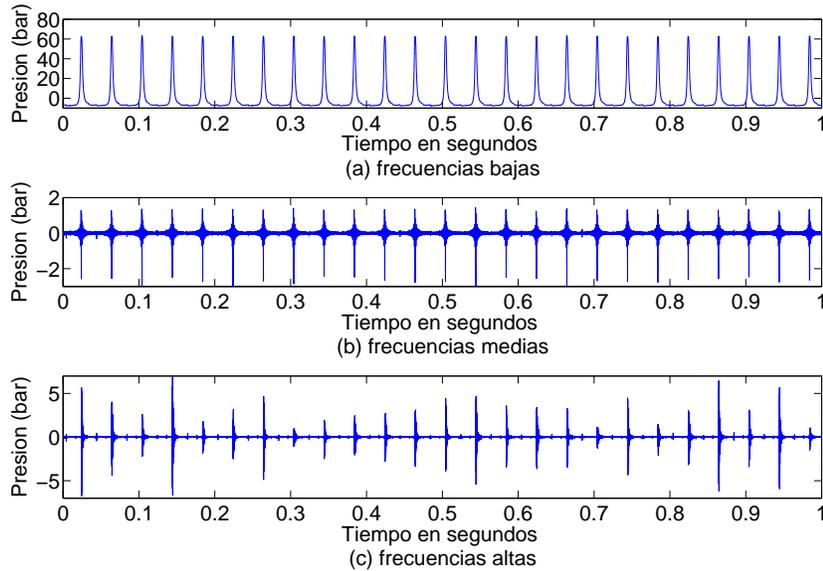


Figura 4.10: Descomposición de la presión en el interior de un cilindro, al que se le inyectan 28mm^3 en tres bandas de frecuencias: se han representado como bajas frecuencias aquéllas inferiores a 400Hz , medias frecuencias las comprendidas entre 400 y 5000Hz y altas a aquéllas que superan los 5000Hz .

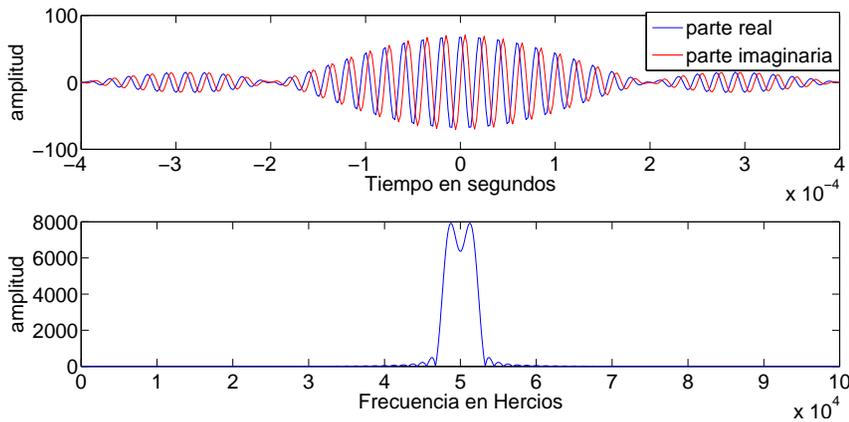


Figura 4.11: Representación temporal y espectral de la wavelet de Shanon ($\Psi(t) = \sqrt{f_b} \text{sinc}(f_b t) e^{j2\pi f_c t}$) con parámetros $F_c = 50000$ $F_b = 5000$.

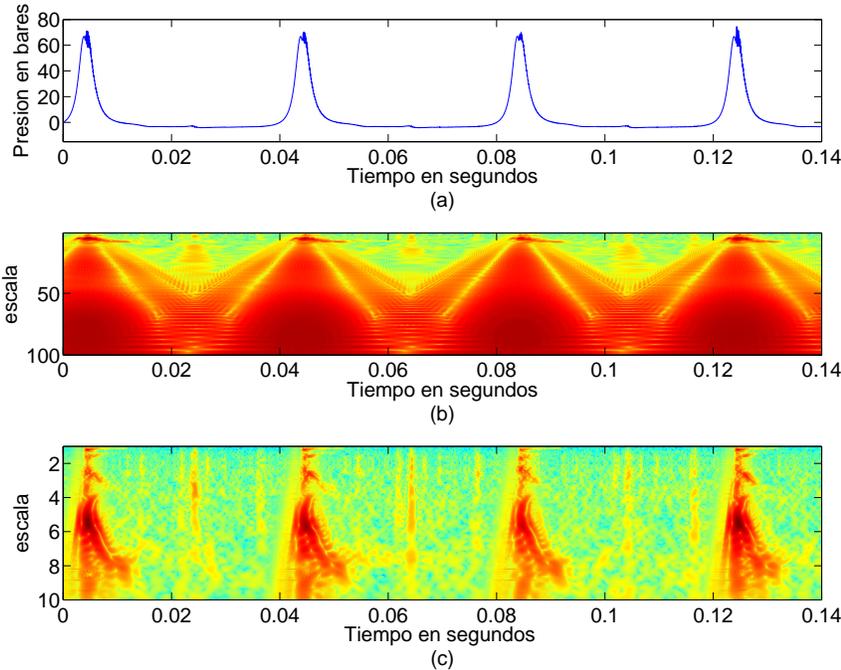


Figura 4.12: Escalograma de la presión en el interior de un cilindro, al que se le inyectan 28mm^3 de combustible.

- Aparecen varias líneas verticales entre dos ciclos consecutivos, que se deben a la apertura y cierre de las válvulas del cilindro.

4.2.2. Ruido de combustión.

El ruido de combustión tiene su origen en el proceso de combustión de gases que se produce dentro de los cilindros de un motor y es una de las fuentes que más ruido produce en el interior del vehículo [205, 210]. El ruido de combustión se puede modelar como una transformación lineal producida sobre la presión existente en el interior del cilindro [199]. La respuesta en frecuencia del sistema asociado al motor [211] tiene un máximo comprendido entre las frecuencias de 1000Hz y 2000Hz, tal y como cabe esperar de una estructura rígida de hierro fundido [199, 212]. Un estudio más detallado del ruido de combustión así como la relación entre el ruido y otros parámetros del motor se puede encontrar en [199, 206, 213].

También hay que tener en cuenta que los motores utilizados en coches y camiones tienen más de un cilindro, por lo que el sonido que se percibe será la combinación del generado por cada uno de ellos.

La figura 4.13, representa la señal de presión medida a una distancia de 1 metro del motor así como su escalograma (obtenido con la wavelet de la figura 4.11 y representado para el margen de escalas [1, 100] y [1, 10]), cuando la presión en el interior de éste es la

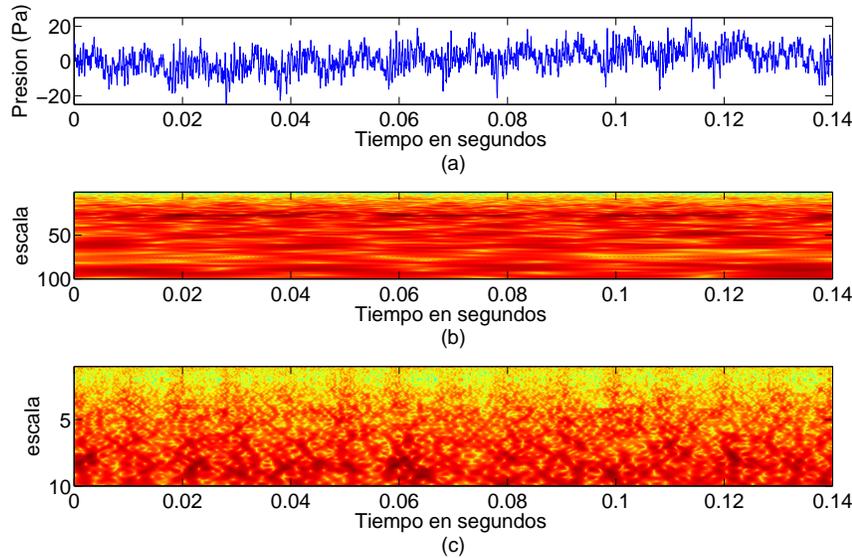


Figura 4.13: Representación del sonido generado por un motor medido a 1 metro de distancia. La gráfica (a) de la figura corresponde a la presión medida en Pascales (utilizando un fondo de escala de $122dB_{SPL}$), mientras que las gráficas (b) y (c) corresponden a escalogramas analizados con distintos rangos de escala.

que aparece en la figura 4.12. En la gráfica (a) de la figura 4.13 se observa que la presión toma valores próximos a 20 pascales (lo que corresponde a un nivel de $120dB_{SPL}$).

A diferencia de lo que ocurría en la sección anterior al analizar la presión en el interior del cilindro motor, en estas figuras no se observan instantes de tiempo concretos en los que aparezcan frecuencias determinadas. Tampoco se observa en la señal que existan frecuencias predominantes sobre otras; ni se aprecia que la señal se repita de forma periódica sino que más bien parece un ruido de banda ancha. Desantes [208] utiliza análisis wavelet para estudiar la señal de presión que se produce en el interior de los cilindros del motor y del ruido de combustión que originan para establecer relaciones entre el ruido y la presión en el interior del cilindro. En este trabajo se observa que mientras que los escalogramas de la presión en el cilindro se repiten de forma bastante regular, los escalogramas asociados al ruido de combustión no muestran esta regularidad.

La figura 4.14 (que contiene la transformada de Fourier de la señal de motor representado en la figura 4.13) permite comprobar que el ruido del motor está formado por la suma de una componente periódica y otra no periódica (y de banda ancha), ya que de forma similar a lo que ocurría cuando se analizaba el espectro de la presión en el cilindro, la potencia de la señal está concentrada en múltiplos de una frecuencia fundamental, que coincide con los 25Hz de la señal de presión en el cilindro. En este caso, el análisis frecuencial nos es de mucho mayor interés que cuando se analiza la señal de presión en un cilindro, pues en el dominio del tiempo es imposible detectar que la señal con la que se está tratando tiene componentes periódicas. La aparición de los armónicos en estas representaciones solamente es posible si se realiza un análisis de la señal durante un intervalo

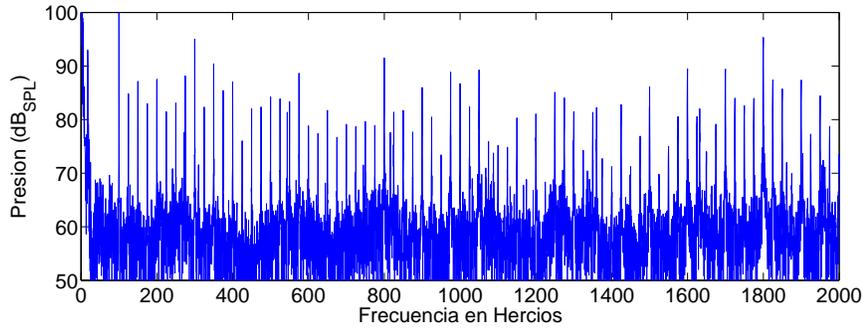


Figura 4.14: Representación frecuencial de la señal mostrada en la figura 4.13.

temporal suficientemente largo.

El ruido de combustión mostrado en la figura 4.14 corresponde a un motor de 4 cilindros cuyas explosiones se producen de forma sincronizada a intervalos regulares de tiempo, por lo que cabría esperar un ruido cuya frecuencia fundamental fuera 4 veces superior a la frecuencia de la señal de presión en los cilindros ($f_0 = 100\text{Hz}$). Debido a que la función de transferencia entre el cilindro y el punto de grabación del sonido es distinta para cada cilindro, la periodicidad del ruido se produce con una frecuencia igual a la de la señal de presión en el cilindro, que coincide con la mitad de la frecuencia de giro del motor. El espectro del ruido de combustión está compuesto, por tanto, por armónicos de la mitad de la frecuencia de giro del motor que reciben el nombre de “armónicos de medio orden”.

La utilidad del análisis frecuencial de esta señal de sonido va más allá de la simple comprobación de la periodicidad de la señal y el cálculo de su periodo, ya que si el espectro mostrado en la figura 4.14 se divide en dos señales, se puede separar la parte periódica de la no periódica. Para esta separación se puede proceder de la siguiente forma: seleccionar un umbral por encima del cual quede el contenido de los picos (150 dB aproximadamente) y dividir el espectro anterior en dos espectros distintos: las frecuencias que superan el umbral (que se ha situado en 70dB_{SPL}) y las que no lo alcanzan (como se muestra en la figura 4.15).

Si ahora representamos el escalograma de la señal resultante tras el proceso de eliminación de ruido descrito en el párrafo anterior obtenemos las gráficas que aparecen en la figura 4.16, donde se puede apreciar en las gráficas (b) y (c), aunque con cierta dificultad, que la representación espectral sigue un patrón que se repite de forma bastante regular cada 0.04 segundos. Además, dentro de estos 40 milisegundos (que corresponden a un ciclo completo de motor) se observan superpuestos varios patrones no exactamente iguales, que se corresponden con cada uno de los cilindros que forman el motor. En definitiva, la eliminación de las componentes no armónicas del espectro de la señal nos ha permitido visualizar un cierto patrón periódico del sonido producido por el motor.

Si nos fijamos en esta parte periódica de la señal, comprobamos que existe una relación entre la presión medida en el exterior del motor en forma de sonido y la registrada en el interior del cilindro. Es habitual considerar [212, 206] que este sonido se puede modelar

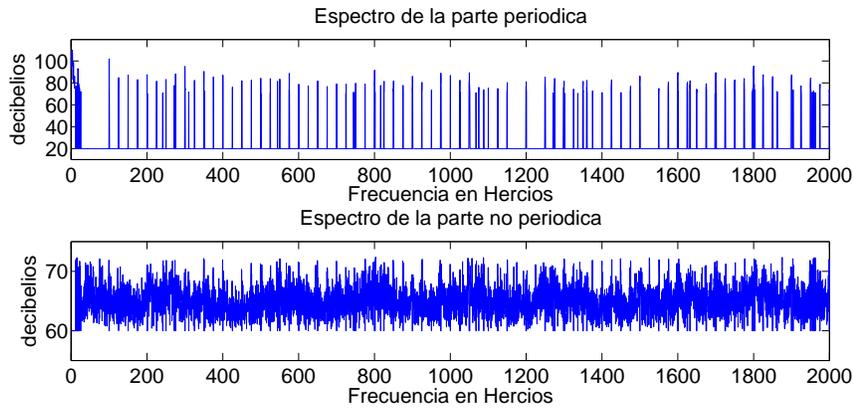


Figura 4.15: Descomposición de la señal de sonido aplicando un umbral de $70dB_{SPL}$ al espectro de la señal de presión.

como la salida de un sistema lineal al que se aplica como señal de entrada la presión existente en el interior del cilindro. Por ejemplo, y utilizando como excitación la presión en el interior de un cilindro al que se le inyectan $28mm^3$ (analizada previamente en la sección 4.2, ver figuras 4.2 a 4.5) y como señal de salida el sonido captado en las proximidades del motor tras la eliminación de ruido (mostrado en la figura 4.16), se puede concluir que el motor se comporta como un sistema cuya función de transferencia es la mostrada en la figura 4.17.

La función de transferencia que aparece en la figura 4.17 se ha obtenido dividiendo la densidad espectral de potencia de la salida del sistema entre la de la entrada obtenida mediante el método de estimación espectral del promediado del periodograma modificado de Welch [214] utilizando ventanas de Hamming de 20ms sin solapamiento. Aunque los resultados mostrados en la figura 4.17 son simplemente una muestra, concuerdan con los resultados obtenidos en estudios exhaustivos de esta función de transferencia [206, 209, 212]. Además, esta función de transferencia con un mínimo de atenuación comprendido entre las frecuencias de 1000 y 2000 Hz es la que cabe esperar de un motor formado por una estructura rígida de hierro.

Al grabar ruidos de combustión, es habitual que los micrófonos recojan también otros ruidos generados en las proximidades del motor, en este entorno son de utilidad las técnicas denominadas Independent Component Analysis (ICA), que permiten mediante el uso de varios micrófonos aislar sonidos procedentes de varias fuentes; y las técnicas de procesamiento tiempo-frecuencia, que aplicadas a cada una de las componentes extraídas permiten identificar la procedencia de cada una de las señales [205, 215, 216, 217].

Para mejorar el diseño de un motor atendiendo al ruido que produce [209, 218, 219], es interesante descomponer el ruido captado por el micrófono en la superposición de tres ruidos distintos asociados a cada una de las señales en que se puede descomponer la presión en el cilindro analizadas en la sección 4.2.1: señal pseudo-motriz, señal de combustión y señal de resonancia de la cámara de combustión.

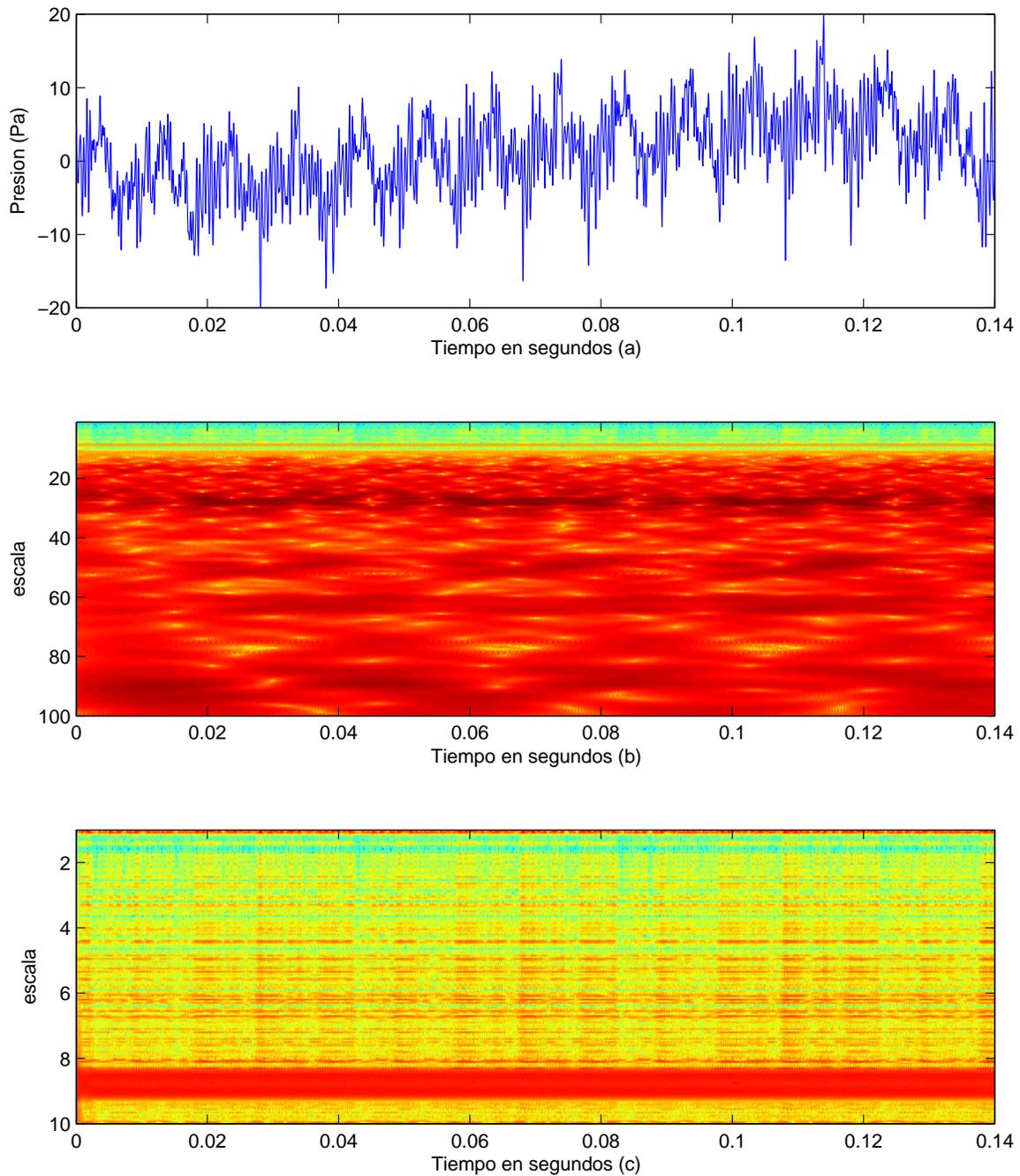


Figura 4.16: Descomposición de la señal de sonido aplicando un umbral de 70dB_{SPL} al espectro de la señal de presión.

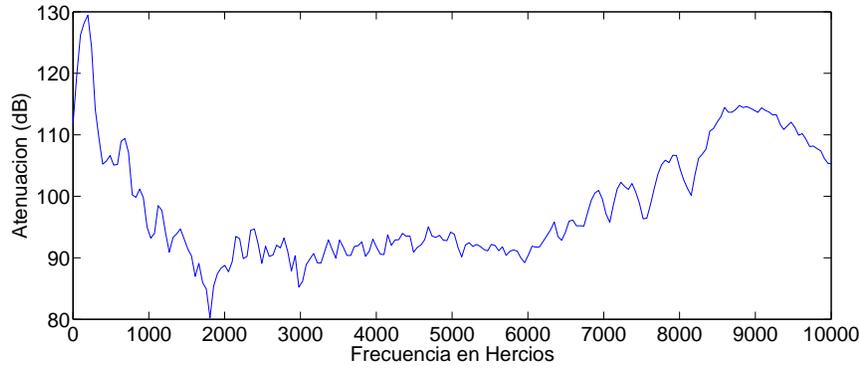


Figura 4.17: Función de transferencia del motor.

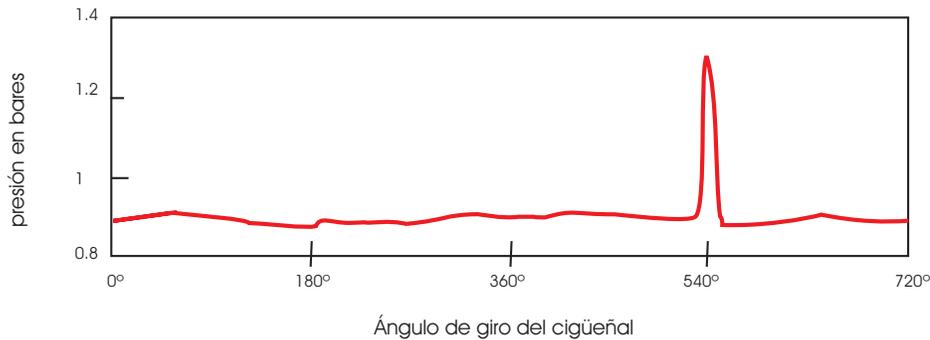


Figura 4.18: Presión en la válvula de escape de un motor.

4.2.3. Ruido de escape.

Otros ruidos también asociados a un motor son los producidos por el sistema de admisión y expulsión de gases. El análisis de estos ruidos se puede realizar mediante el estudio y modelado de las señales que los originan y los sistemas que los transforman. En este caso, se considera como señal excitación la onda de presión que sale de las válvulas de escape de los cilindros de un motor; y el sistema que transforma la onda de presión es el propio sistema de escape conectado al motor [202, 220, 203, 221, 222, 197, 223, 224]

Cada una de las válvulas de escape va conectada al tubo de escape del vehículo mediante un colector que recoge las salidas de gases de varias válvulas y las conduce a un solo tubo. La presión instantánea medida a la salida de la válvula de escape de un motor sin la presencia de silenciador, que se esquematiza en la figura 4.18, presenta un salto en la presión de nivel comparable al valor de la presión atmosférica en condiciones normales [223], y se produce en un periodo muy reducido de tiempo. Esta señal puede ser aproximada mediante una función delta de Dirac [222].

El ruido de escape, al igual que ocurre con el de combustión, aumenta proporcionalmente al logaritmo de la velocidad del vehículo [73], produciéndose un aumento de 10 dB cada vez que se dobla la velocidad. Además, el ruido también aumenta en función de la carga, particularmente en el arranque. En cuanto a la dirección de propagación de estos

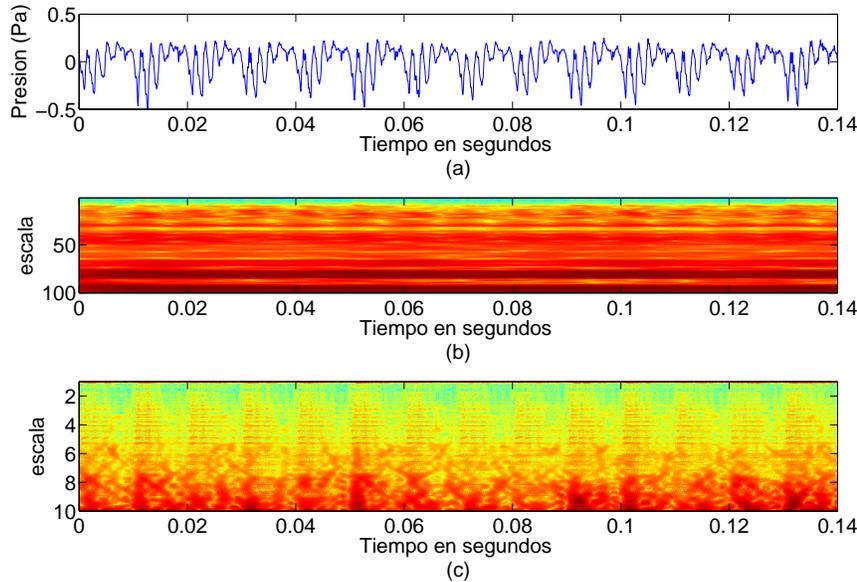


Figura 4.19: Representación del sonido producido en la salida del tubo de escape de un motor. La gráfica (a) de la figura corresponde a la presión medida en Pascales , las gráficas (b) y (c) son escalogramas de la figura (a).

ruidos, es predominante en las direcciones comprendidas entre 20° y 40° respecto a la horizontal, efecto principalmente debido a la reflexión producida en el suelo.

Los camiones pesados, que a menudo tienen el sistema de escape en una posición elevada y el motor a una altura aproximada de 1 metro y paneles laterales, producen un ruido con niveles superiores y un espectro más ancho [73, 225] que el que produce un turismo.

La emisión de ruido de escape está limitada por la directiva europea Directiva 70-157-CEE, y la forma de medirlo está detallada en la norma ISO 5130 [226].

La figura 4.19 muestra la forma de la onda temporal de una señal de ruido de escape así como su escalograma. Tanto del análisis temporal como del análisis del escalograma se desprende que esta onda presenta un patrón de repetición periódico bastante más regular que el que se observaba en las señales producidas por el ruido de combustión del motor.

En la figura 4.19 también se aprecia que el intervalo de repetición de la señal es de 10 ms en lugar de los 40 ms que se observan por ejemplo en la figura 4.16. Esto se debe a que el ruido de escape producido por cada una de las válvulas del motor (4 en este caso) es similar al de las otras mientras que el ruido de combustión que produce una válvula del motor es distinto al que producen las otras.

Al analizar el espectro de la señal de escape 4.20 se puede comprobar que está formado principalmente por los armónicos de una frecuencia fundamental que coincide con el doble de la frecuencia de giro del motor (100Hz para este caso), aunque en algunas bandas frecuenciales aparecen armónicos de medio orden correspondientes a la mitad de

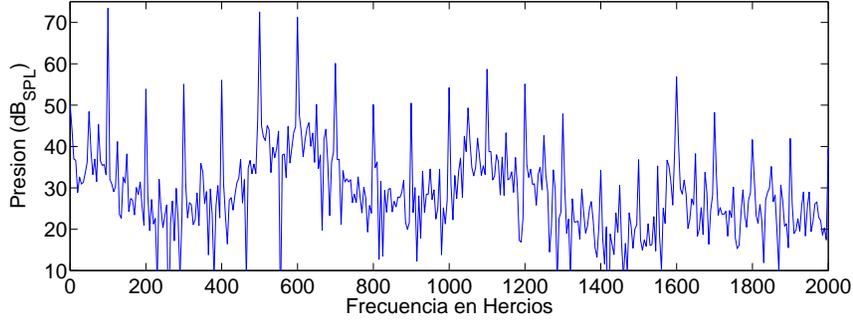


Figura 4.20: Representación del espectro del sonido captado a la salida del tubo de escape de un motor que gira a 3000r.p.m.

la frecuencia de giro del motor.

4.3. Modelo de ruido de motor periódico.

En la literatura relativa al análisis de ruidos de motor se pueden encontrar varios métodos para analizar parámetros psicoacústicos relativos tanto a ruidos reales como sintéticos [194, 21, 20]. El modelo de sintetizador de ruido propuesto por Feng [20], extrae las características espectrales de un ruido real y las utiliza en la generación de nuevos ruidos de motor. El ruido sintético consiste en la combinación de dos señales: una señal determinista, $x(t)$, y otra estocástica, $n(t)$:

La parte determinista del modelo, se forma a partir de una señal periódica consistente en una suma de armónicos cuya frecuencia fundamental es la del ruido del motor y cuyas amplitudes y fases se han obtenido mediante la extracción de los coeficientes de Fourier de una señal de motor real. Esta señal periódica se modula en amplitud con una señal moduladora senoidal según la expresión:

$$x(t) = (1 + m \cdot \cos(2\pi f_m t)) \sum_{k=-20}^{20} c_k e^{jk \frac{2\pi}{T} t} \quad (4.3)$$

donde c_k son los coeficientes de Fourier extraídos de un sonido de motor real, T el periodo de la señal de motor, f_m la frecuencia de modulación y m el índice de modulación con el que se modula la serie de Fourier obtenida con los coeficientes anteriores.

La parte estocástica del modelo, se forma mediante un filtrado autorregresivo (AR) de un ruido blanco. Para extraer los coeficientes del filtro AR se parte de una señal real a la que se le elimina su parte periódica mediante unos filtros en hendidura sintonizados a los armónicos de la frecuencia fundamental del ruido. Una vez eliminada la parte periódica de la señal de motor, los coeficientes AR son estimados

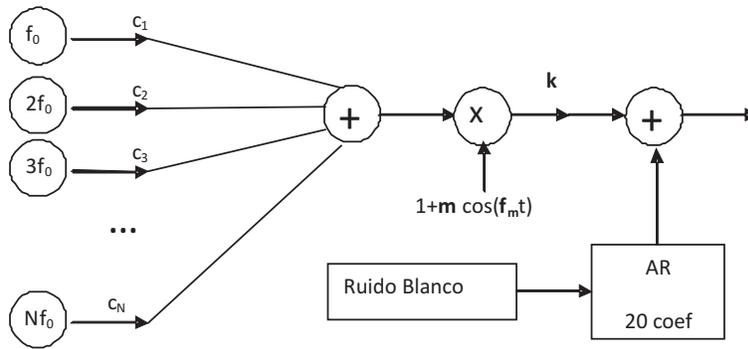


Figura 4.21: Modelo de ruido de motor propuesto por Feng [20].

mediante comparaciones subjetivas. La incorporación de esta parte estocástica a la señal le proporciona un mayor realismo, aunque también modifica los parámetros psicoacústicos de la señal.

En la figura 4.21 aparece el esquema de este sintetizador de ruido: los coeficientes $c_1 \dots c_N$ representan las amplitudes de los N primeros armónicos utilizados para generar la señal periódica (en el modelo original, el número de armónicos utilizados es $N = 20$). Esta señal periódica es modulada en AM con frecuencia de modulación f_m e índice de modulación m , para ser posteriormente mezclada con ruido coloreado (parte estocástica) formado mediante el filtrado de un ruido blanco con un filtro todo polos de orden 20, que es superpuesta a la señal anterior para formar la señal sintética.

4.4. Nuevo modelo de ruido de motor: modelo cuasiperiódico.

El objetivo de esta sección es desarrollar una mejora del modelo de ruido de motor propuesto por Feng para estudiar sus parámetros psicoacústicos y establecer una relación entre éstos y los parámetros mecánicos del motor.

A partir del análisis de los ruidos de combustión y de escape realizado en la sección 4.2 se propone un modelo que permita sintetizar estos ruidos de motor con características próximas a los ruidos reales. El modelo propuesto es una variación del presentado en la sección 4.3 donde el ruido se formará también como la suma de dos componentes: una parte no periódica compuesta por ruido coloreado y una parte “cuasiperiódica”.

4.4.1. Descripción del modelo.

El ruido producido por el motor de un coche funcionando a N_{rpm} revoluciones por minuto se aproxima a una señal periódica, pero no es exactamente periódico debido al hecho de que la presión en el interior de los cilindros del motor sufre pequeñas variaciones

entre ciclos sucesivos [195, 196] tal y como se ha expuesto en la sección 4.2. Así pues, la presión en el interior de un cilindro corresponde a un patrón cuasiperiódico cuya frecuencia fundamental f_0 es la mitad de la frecuencia de giro del motor:

$$f_0 = \frac{N_{rpm}}{120} \text{ Hz.} \quad (4.4)$$

La señal de presión depende del diseño del motor (tiempos asociados a la apertura y cierre de las válvula, relación de compresión, etc.) y de sus parámetros de funcionamiento (tiempo en el que se produce la explosión, temperatura de refrigeración del motor, velocidad del motor, carga del motor, relación aire-combustible, recirculación de los gases de escape, etc.) [195, 196, 194]. El movimiento del motor se produce debido a explosiones en el interior de los cilindros que tienen lugar cuando están llenos de un gas inflamable que explota debido al aumento de la presión en motores diesel, o al efecto de una chispa en motores de gasolina.

Por lo tanto, el ruido percibido es una combinación del ruido de combustión generado dentro de los cilindros y ruido mecánico debido a vibraciones del motor.

Cuando un motor está compuesto por un número de cilindros igual a N_{cyl} , se producen N_{cyl} veces más explosiones por unidad de tiempo que cuando sólo existe un cilindro. Por lo tanto, el tiempo transcurrido entre dos explosiones del motor es igual al tiempo transcurrido entre dos explosiones del mismo cilindro dividido entre N_{cyl} . Es decir, al analizar la señal periódica que genera un motor con varios cilindros, observamos que su periodo de repetición es N_{cyl} veces inferior al de la señal periódica que genera un solo cilindro y, por lo tanto, la frecuencia fundamental de la señal periódica de un motor de varios cilindros es N_{cyl} superior a la frecuencia fundamental de la señal de presión en el cilindro mostrada en (4.4):

$$f_0 = \frac{N_{rpm}}{120} \times N_{cyl} \text{ Hz.} \quad (4.5)$$

Respecto a la expresión anterior hay que añadir que solamente es válida cuando las señales producidas por cada uno de los cuatro cilindros son muy similares, y deja de ser válida si los patrones de señal producidos en cada cilindro son diferentes entre sí. En este caso, aparecen lo que se conoce como “armónicos de medio orden” [198], que corresponden a los múltiplos de la mitad de la frecuencia de giro del motor (4.4).

Esto puede marcar una diferencia entre el análisis del “ruido de combustión” (donde existe bastante diferencia entre las señales producidas por los distintos cilindros y por tanto aparecen armónicos de medio orden con amplitudes considerables) y las “señales de escape” (donde las ondas generadas por cada uno de los cilindros son prácticamente idénticas y los armónicos de medio orden casi desaparecen por completo).

Aunque el modelo que se propone es válido para generar una gran variedad de ruidos asociados al motor (combustión, escape, ...), las gráficas que se presentan en las figuras siguientes corresponden a ruido de escape puesto que en ellas es más fácil de visualizar

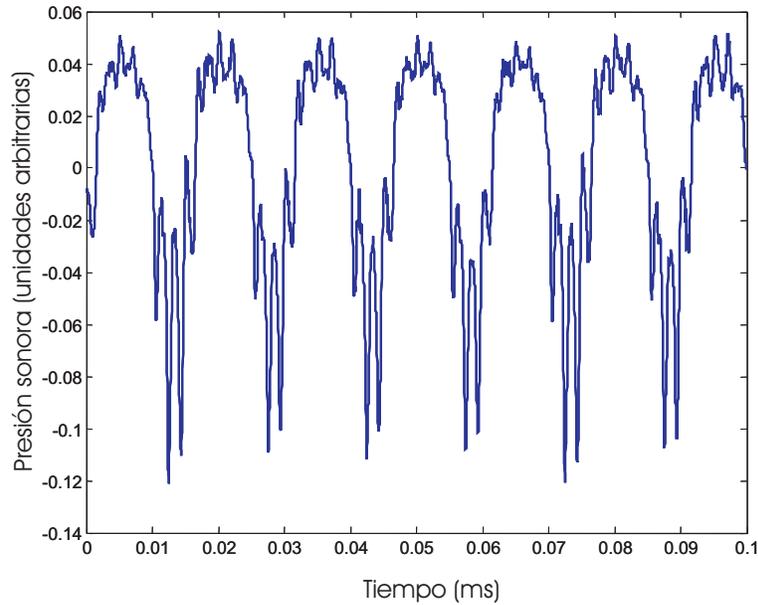


Figura 4.22: Señal de presión correspondiente al ruido de escape generado por un motor de cuatro cilindros trabajando a 2000 r.p.m.

patrones repetitivos que en señales de ruido de combustión sin necesidad de preprocesarlas como se vio en la secciones 4.2.2 y 4.2.3.

Las características “cuasiperiódicas” de este tipo de ruido se pueden apreciar en las figuras 4.22 y 4.23. La figura 4.22 muestra el ruido de escape producido por un motor de cuatro cilindros trabajando a 2000 r.p.m., mientras que la figura 4.23 es una representación de la densidad espectral de potencia [163] de esta señal.

La figura 4.24 esquematiza un modelo simplificado que genera la señal producida como consecuencia de una explosión aislada en el interior del cilindro en el instante de tiempo T_0 . En esta figura, $x(t)$ corresponde a la señal de presión dentro del cilindro cuando se produce una explosión en el instante de tiempo $t = 0$, y $x(t - T_0)$ representa la misma señal de presión cuando la explosión ocurre en $t = T_0$. La señal asociada a la explosión se ha modelado como un impulso (delta de Dirac) que excita a un sistema lineal e invariante en el tiempo cuya respuesta impulsional es $x(t)$, y que según la teoría de sistemas se puede obtener mediante la convolución:

$$x(t - T_0) = \delta(t - T_0) * x(t). \quad (4.6)$$

En este modelo $x(t)$ contiene información de la forma de la señal de presión en el interior del cilindro. Por otro lado, $h(t)$ corresponde a la respuesta impulsional del medio de transmisión, que modela el comportamiento del sistema mecánico entre el punto en el que ocurre la explosión (cilindro) y el punto en el que se escucha o registra el sonido. Este sonido se puede escribir como:

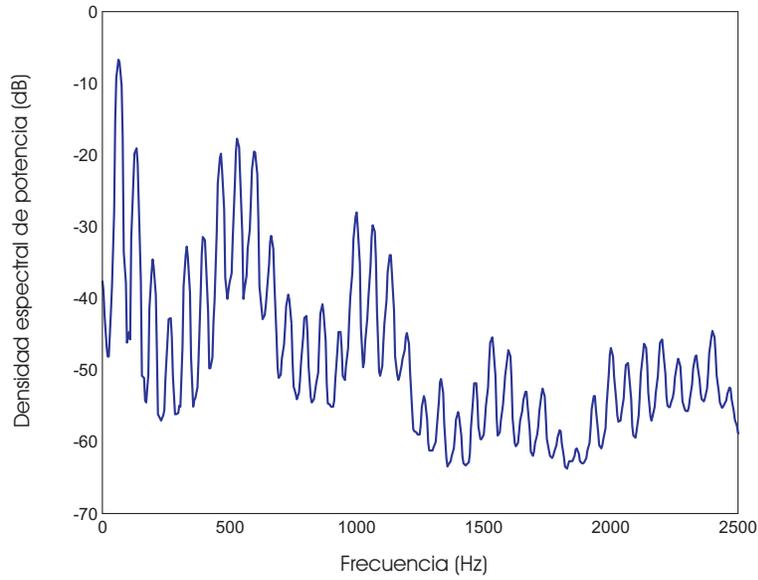


Figura 4.23: Densidad espectral de potencia de la señal representada en la figura 4.22.

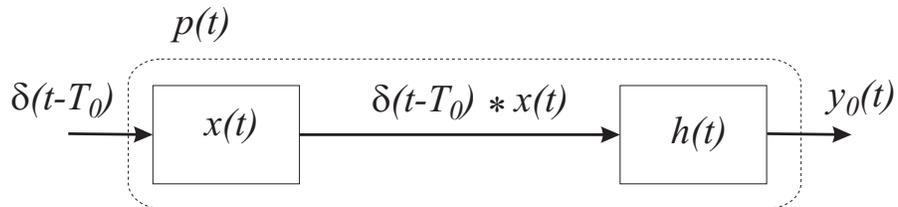


Figura 4.24: Modelo de generación de ruido producido por una explosión aislada en el cilindro del motor.

$$y_0(t) = (\delta(t - T_0) * x(t)) * h(t) \quad (4.7)$$

y utilizando la propiedad asociativa de la convolución:

$$y_0(t) = (\delta(t - T_0) * x(t)) * h(t) = \delta(t - T_0) * (x(t) * h(t)) = \delta(t - T_0) * p(t) \quad (4.8)$$

siendo

$$p(t) = x(t) * h(t). \quad (4.9)$$

Utilizando la teoría de sistemas lineales, los dos sistemas que aparecen en la figura 4.24 se podrían reemplazar por uno solo cuya respuesta impulsional fuera la señal $p(t)$ de la ecuación 4.9, que representa la presión sonora medida en el punto de audición cuando se produce una explosión en el interior del cilindro.

El sonido producido por un motor ideal trabajando a una velocidad angular constante donde todas las explosiones ocurrieran a intervalos regulares de tiempo $T = 1/f_0$, correspondería al modelo presentado en la figura 4.24, pero sustituyendo el impulso aislado por un tren de impulsos periódicos:

$$c(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT). \quad (4.10)$$

Sin embargo, el modelo de excitación de un motor real debería tener en cuenta pequeñas variaciones aleatorias que afectan a las amplitudes y los retrasos existentes entre explosiones sucesivas, por lo que la expresión de la señal de entrada $c(t)$ que aparece en (4.10) se debería de sustituir por:

$$c(t) = \sum_{k=-\infty}^{\infty} a_k \delta(t - T_k), \quad (4.11)$$

siendo a_k la amplitud del k -ésimo ciclo; y T_k el instante de tiempo en que se produce el pulso k -ésimo. La ecuación (4.10) se puede considerar una particularización de la expresión (4.11) donde se ha asumido que los pulsos se producen a intervalos regulares de tiempo y son de la misma amplitud, es decir:

$$T_k = k \cdot T, \text{ y } a_k = 1 \quad (4.12)$$

y el valor de T sería la esperanza matemática del tiempo transcurrido entre dos deltas consecutivas:

$$T = E\{T_k - T_{k-1}\}, \quad (4.13)$$

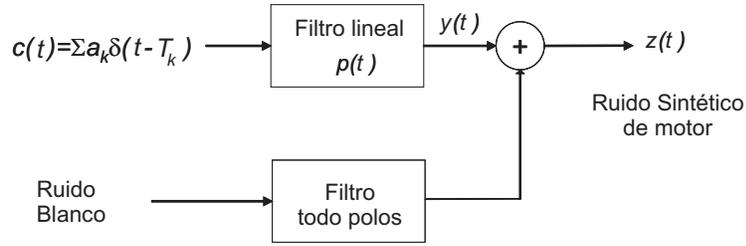


Figura 4.25: Modelo del generador de ruido.

que coincide con el inverso de la frecuencia fundamental de la señal expresada en la ecuación (4.4):

$$T = 1/f_0, \quad (4.14)$$

Asumiendo que la generación de ruido se produce mediante un sistema lineal (de respuesta impulsional $p(t)$) excitado por un tren de pulsos cuasiperiódico $c(t)$ que aparece en (4.11), el ruido de motor percibido por un observador se puede expresar como:

$$y(t) = p(t) * c(t). \quad (4.15)$$

Una de las ventajas que aporta este modelo (que se esquematiza en la figura 4.25), es que permite relacionar de forma sencilla los parámetros psicoacústicos (como Loudness, Sharpness, Roughness o Fluctuation Strength) asociados a estos ruidos con los parámetros de este modelo simplificado y por lo tanto con las variables de funcionamiento del motor.

Mediante este sencillo modelo se ha reducido la señal de motor a unos pocos parámetros que dependen de la estructura física del motor, del régimen de funcionamiento y del medio de transmisión desde el motor hasta el punto en que se escucha el ruido:

La señal $c(t)$: definida en la ecuación (4.11), depende del régimen de giro del motor y de las irregularidades o imperfecciones que puedan existir en este giro debidas al proceso físico/químico asociado a una explosión. Los picos de amplitud (a_k) de la señal $c(t)$ se pueden elegir de forma adecuada para generar señales moduladas en amplitud de la forma que se explica en la sección 5.4.1. Y los valores de T_k se pueden escoger según se indica en la expresión (4.12) o introduciendo variaciones que modelen desajustes en el sincronismo del motor.

La señal $p(t)$: definida en (4.9), depende de la onda de presión en el interior de un cilindro cada vez que se produce una explosión y la transformación que esta onda sufre debida al camino que debe de atravesar hasta el punto de recepción del sonido y que incluye la estructura del motor y todos los elementos que rodean a éste y al punto de recepción.

El ruido: que aparece en la parte inferior de la figura 4.25 y depende de todos los elementos no considerados en los puntos anteriores y que añaden ruido a la señal recibida. Esta señal se debe añadir al conjunto para darle un aspecto más realista [20].

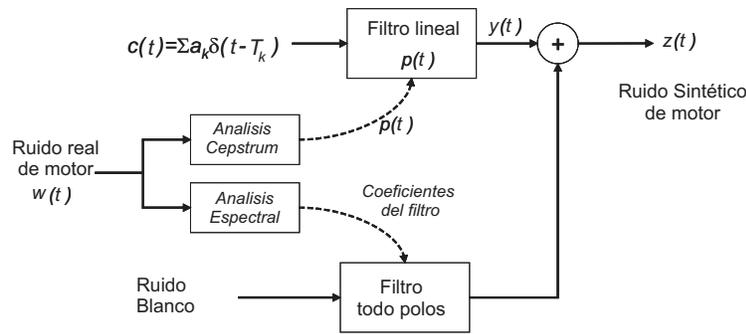


Figura 4.26: Obtención de parámetros del modelo propuesto para síntesis de señales de motor.

4.4.2. Obtención de los parámetros del modelo propuesto.

A partir de un conjunto de grabaciones de señales de escape producidas por motores reales, se han analizado las dos partes que la forman: una parte de la señal podría definirse como cuasiperiódica ($y(t)$) y consiste en la repetición de una señal patrón ($p(t)$) con pequeñas variaciones de amplitud y fase (tal y como aparece en la expresión (4.15) y la otra parte corresponde a un ruido blanco filtrado.

El sintetizador de ruido de motor propuesto en este trabajo, está basado en el análisis mediante cepstrum complejo estudiado en la sección 3.4.3 y puede ser descrito mediante el diagrama de bloques representado en la figura 4.26, que representa el proceso de extracción de parámetros de la señal real y la utilización de estos en la formación de una señal sintética.

Las señales reales de motor $y(t)$ utilizadas en la extracción de parámetros del sintetizador de la figura 4.26, se han obtenido de grabaciones digitales en cámara semi-anecoica utilizando un Head And Torso Simulator (HATS) situado a 3 metros del tubo de escape de un motor. Todas las grabaciones tienen una duración de 10 segundos y se han muestreado a 48000 muestras por segundo.

El motor utilizado en los experimentos es un motor comercial High-Speed Diesel Injection (HSDI) de una marca automovilística Europea, y es representativo del ruido producido por un vehículo de gama alta, tipo sedan. Las grabaciones se han realizado en el exterior del vehículo en un punto cercano al tubo de escape, por lo que el ruido grabado es representativo de lo que perciben los peatones que transitan por la acera en las inmediaciones de este tipo de automóviles. El modelo propuesto en la Figura 4.25 se puede aplicar también a otros tipos de motores y posiciones de grabación ya que el único efecto que esto produciría sería la modificación de la respuesta del filtro que se utiliza para modelar el canal de propagación.

Para analizar la parte de la señal asociada a la repetición del patrón y la extracción de este patrón, se ha procedido de la siguiente forma: primero se calcula el cepstrum complejo de un ruido muestreado $w(t)$ mediante el diagrama de bloques que aparece en la Figura 3.14 de la sección 3.4.3. A continuación, el cepstrum complejo se envientana (*lifter*) para aislar la información relativa al patrón $p(t)$. Finalmente se recupera la señal patrón

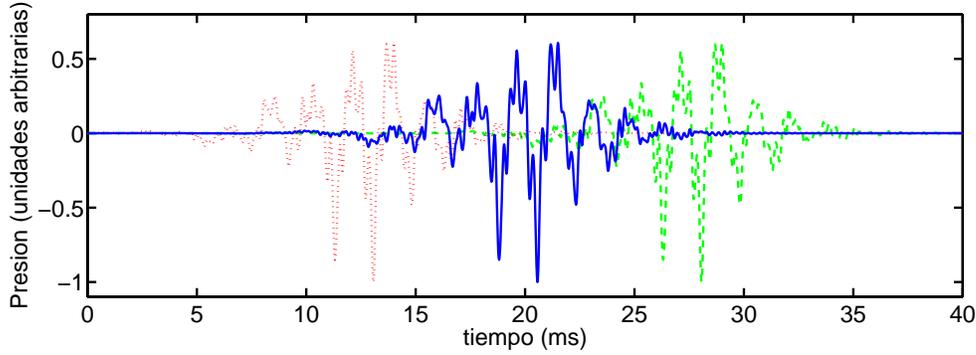


Figura 4.27: Superposición del patrón obtenido de una señal de escape de un motor funcionando a 4000 r.p.m. con dos versiones desplazadas de este: la línea azul corresponde a la forma de onda de la señal patrón mientras que las de color verde y rojo corresponden a repeticiones del patrón retrasadas y adelantadas un periodo respectivamente.

calculando el cepstrum complejo inverso.

Debido a las variaciones temporales que presentan los sonidos con los que se trabaja y para evitar que se produzca aliasing en el dominio log-frecuencia [193, 163], se aplica este procedimiento a ventanas de 3 periodos de la señal obteniendo un patrón para cada ventana y, a continuación, se calcula el patrón global de la señal como la media de los patrones obtenidos en cada ventana. Antes de realizar el promedio se desplazan temporalmente todos los patrones obtenidos con el fin de obtener la máxima correlación entre ellos.

Para ilustrar el funcionamiento del sintetizador, la figura 4.27 muestra la repetición periódica del patrón obtenido de una señal de escape cuando el motor funciona a un régimen de giro de 4000 r.p.m. En este caso, utilizando la relación mostrada en la expresión (4.5), el periodo del sonido producido es de $T = 7,5ms$.

Si en la señal $c(t)$ de la expresión (4.11) se toma $a_k = 1$ y $T_k = k \cdot T = k \cdot 7,5ms$, se generará una señal sintética que corresponde a la suma de las repeticiones periódicas del patrón mostradas en la figura 4.27. Este resultado corresponde a una señal sintética de gran parecido con la original como se puede apreciar en la figura 4.28.

Al escuchar esta señal, que es estrictamente periódica, se tiene la sensación subjetiva de estar escuchando una señal poco realista. Para solucionar este problema el modelo permite realizar variaciones de los parámetros a_k y T_k de forma distinta a la mostrada en el párrafo anterior. Aunque también existe la posibilidad de añadir ruido coloreado mediante la rama inferior de la figura 4.26.

El procedimiento para la extracción de los parámetros del filtro del modelo utilizado para colorear el ruido blanco es similar al utilizado por Feng [20]: la señal real procedente del motor es filtrada con filtros muy estrechos para eliminar los armónicos dominantes. La señal resultante se utiliza para calcular 20 coeficientes de un modelo autorregresivo mediante el método de la autocorrelación [227]. Estos coeficientes se usan en un filtro todo

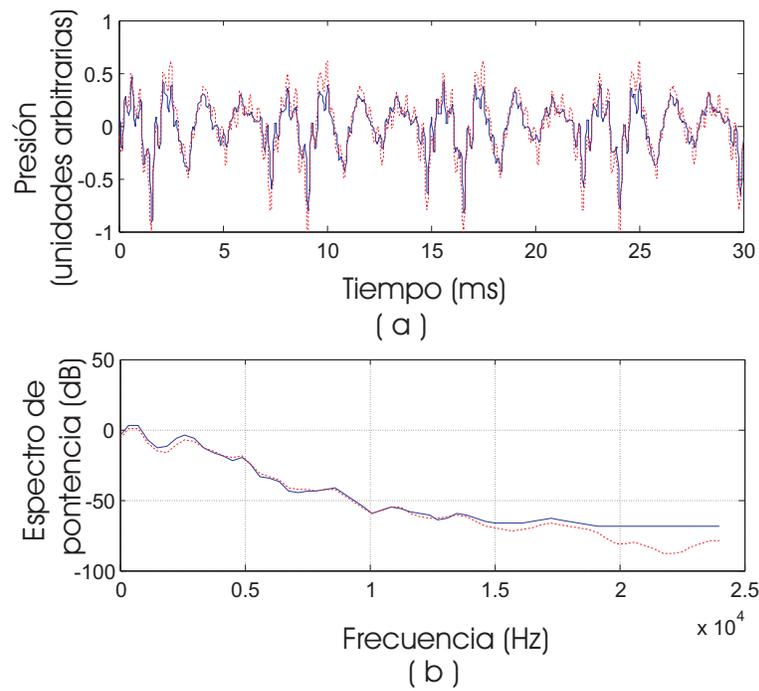


Figura 4.28: Comparación entre el ruido de escape real producido por un motor que gira a 4000 r.p.m (línea continua) y la versión sintética de este sonido (punteado). (a) Forma de onda temporal. (b) Densidad Espectral de Potencia.

polos con el fin de obtener un ruido coloreado mediante el filtrado de un ruido blanco.

El proceso de extracción de parámetros se ha realizado para señales de motor de distintos regímenes de giro obteniendo varios registros que permitirán generar un gran abanico de señales sintéticas [228].

Posteriormente, se ha convolucionado el patrón $p(t)$ con un tren de pulsos formado según la expresión (4.11) para conformar diferentes sonidos sintéticos.

Los valores reales de a_k que aparecen en (4.11) son aleatorios y habitualmente tienen una varianza reducida. Sin embargo, en el modelo propuesto, se han considerado además variaciones de a_k conforme a una modulación en amplitud, para la generación de señales sintéticas según la expresión:

$$a_k = 1 + m \cdot \cos(2\pi f_m t)|_{t=\frac{k}{f_0}} = 1 + m \cdot \cos(2\pi \frac{f_m}{f_0} k), \quad (4.16)$$

Donde m se define como el índice de modulación y toma valores comprendidos en el intervalo $0 < m \leq 1$, f_0 es la frecuencia fundamental del sonido que aparece en (4.4) (medida en Hercios), y relacionada con el régimen de giro del motor según la expresión (4.5), y donde f_m representa la frecuencia de modulación del sonido sintético (también medida en Hercios).

De esta forma, la generación de a_k mediante un modelo de modulación en amplitud y frecuencia (4.16) permite controlar los parámetros que influirán en la percepción subjetiva del ruido.

4.4.3. Validación subjetiva del modelo propuesto: test de realismo

En la figura 4.28 se aprecia una gran similitud entre una señal sintética y la señal original de la que procede. No obstante, y puesto que se va a trabajar con señales sintéticas con índices de modulación distintos a los que tienen las señales reales para estudiar la respuesta del oído, se ha diseñado un “Test de Realismo” para comprobar si señales sintéticas generadas con parámetros de modulación exagerados tienen la apariencia de la señal de motor que están modelando o se perciben como un sonido completamente distinto.

Para la realización de este experimento se han generado señales sintéticas de 1000, 3000 y 5000 r.p.m. a partir de los patrones extraídos de señales reales con la misma velocidad de giro que cada una de las señales a sintetizar. Estos patrones se han convolucionado con un tren de deltas modulado en amplitud con una frecuencia de modulación de 70 Hercios, una amplitud de modulación del 50 % y una relación señal ruido de 30 decibelios. Las señales se han etiquetado de la forma que aparece en la tabla 5.4.

En el experimento intervinieron como jueces, 8 profesores y alumnos de la Escuela Politécnica Superior de Alcoy con edades comprendidas entre los 20 y los 40 años, no

Señal	características
1	señal de escape medida a 1000 r.p.m.
2	señal sintética de 1000 r.p.m.
3	señal de escape medida a 3000 r.p.m.
4	señal sintética de 3000 r.p.m.
5	señal de escape medida a 5000 r.p.m.
6	señal sintética de 5000 r.p.m.

Tabla 4.1: Etiquetas asignadas a las señales utilizadas en el test de realismo.

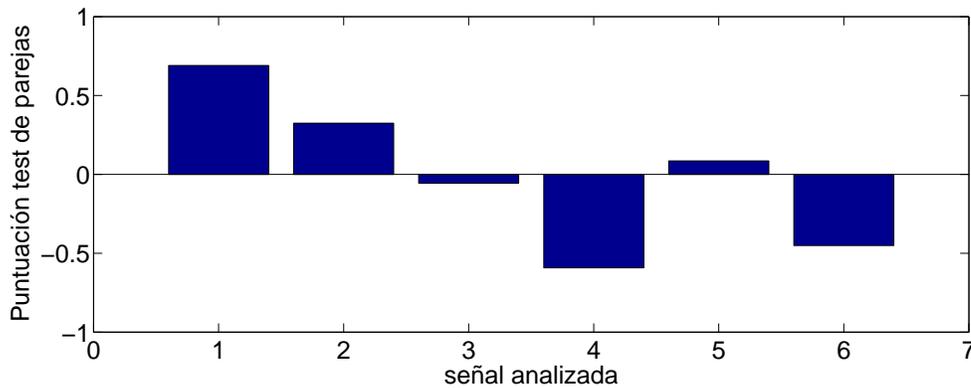


Figura 4.29: Realismo de las señales analizadas.

relacionados con la acústica y el procesado de señal.

La valoración del “realismo” de las señales se obtuvo mediante un test de parejas. Este test de realismo consistió en ofrecer a los jueces distintas parejas de señales formadas por una señal real y otra sintética para que seleccionaran cual de los dos sonidos de la pareja les parecía un sonido de motor más real [229].

Como en las parejas presentadas a los jueces había siempre una señal elegida entre el grupo de las señales reales y otra señal elegida entre el grupo de las señales sintéticas, se obligaba al juez a seleccionar una sola de las dos señales impidiéndole seleccionar que las dos señales le parecían igualmente reales.

Un test de parejas [230, 231, 13] asigna una puntuación a cada una de las señales evaluadas, que es mayor cuanto mayor es la intensidad percibida por los jueces del parámetro analizado (en este caso realismo). Esta puntuación se obtiene a partir de la probabilidad de selección de una señal al ser comparada con otra. La suma de todas las puntuaciones asignadas en el test es cero por lo que algunas señales tendrán puntuación positiva mientras que otras la tendrán negativa. Por otra parte, el rango de puntuaciones obtenidas (diferencia entre la máxima y la mínima puntuación) será mayor cuanto más diferentes sean las señales entre sí en cuanto al parámetro analizado.

En la figura 4.29 se muestran los resultados obtenidos en la evaluación del realismo de las señales mediante el test de parejas.

Señal	1	2	3	4	5	6
1	0	0.75	0	0.875	0	0.75
2	0.25	0	0.875	0	0.75	0
3	0	0.125	0	0.625	0	0.75
4	0.125	0	0.375	0	0.25	0
5	0	0.25	0	0.75	0	0.625
6	0.25	0	0.25	0	0.375	0

Tabla 4.2: Probabilidad de que la señal cuyo número aparece en la fila sea elegida como más real que la señal cuyo número aparece en la columna.

Del análisis de la figura 4.29 se pueden extraer las siguientes conclusiones:

- El margen de valores obtenidos representados en la figura 4.29 es reducido, lo que indica que todas las señales tienen un grado de realismo similar según los jueces.
- Para cualquier valor de revoluciones del motor la señal real presenta un realismo superior a la sintética, como cabía esperar, puesto que la señal sintética se ha generado con valores exagerados de índices de modulación.
- Las señales de bajas revoluciones presentan un valor de realismo superior a las de altas revoluciones. Incluso se da el caso que la señal sintética de 1000 r.p.m. se percibe como más real que señales reales de mayor velocidad de giro.

Para apoyar las conclusiones anteriores, la tabla 4.2 ofrece los datos de distribución de probabilidades a partir de los cuales se ha obtenido la puntuación mostrada en la figura 4.29. En la tabla se indica la probabilidad de que la señal cuyo número aparece en la fila sea elegida por los jueces como más real que la señal cuyo número aparece en la columna. Por ejemplo, el valor 0,75 que aparece en la fila 1 columna 2 significa que al comparar la señal 1 (fila) con la señal 2 (columna) la probabilidad de que los jueces elijan como más real a la señal 1 (que aparece en la fila) es de 0,75. Y por lo tanto la probabilidad de que elijan como más real la señal 2 será 0,25 ($= 1 - 0,75$), que es el valor que aparece en la fila 2 de la columna 1.

Como no tiene sentido comparar una señal consigo misma, los elementos de la diagonal principal contienen 0's. Y como en todas las comparaciones se comparó una señal real con una sintética, los elementos de la matriz asociados a comparaciones entre dos señales reales o dos sintéticas también contienen 0's.

Es interesante resaltar el hecho de que, aparte de los valores anteriores, ningún otro elemento de la matriz contiene un valor 0 ni un valor 1. Esto significa que en cualquier comparación los jueces muestran división de opiniones a la hora de marcar cuál de las dos señales es la real. Por esta razón, se ha considerado que el método propuesto es capaz de generar señales sintéticas con un nivel suficiente de semejanza con una señal de motor real. Al menos, siempre que no se exceda un índice de modulación del 50 %, una frecuencia de modulación de 70 Hz y el régimen de giro no exceda las 5000 r.p.m.

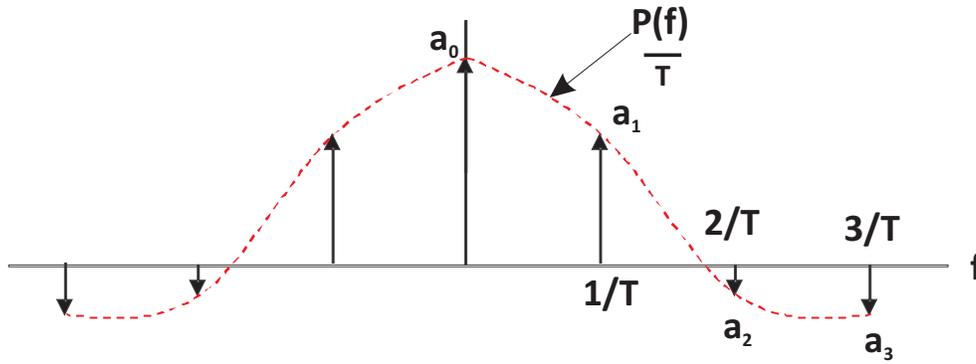


Figura 4.30: Forma general de la representación espectral de una señal que se repite periódicamente a intervalos de tiempo T .

4.5. Análisis frecuencial de los distintos modelos.

Para analizar matemáticamente los parámetros psicoacústicos de los modelos descritos en los párrafos anteriores conviene estudiar la composición frecuencial de las señales proporcionadas por los modelos y su relación con los parámetros que lo definen.

Se analizará el resultado de la excitación de un sistema lineal e invariante mediante distintos trenes de deltas: un tren de deltas periódico, un tren de deltas modulado en A.M. y un tren de deltas con amplitudes aleatorias.

4.5.1. Señal excitación $c(t)$: tren de deltas periódico.

El sonido producido por un motor ideal funcionando en un régimen de giro constante debería ser una señal periódica, al igual que el ruido de escape o el ruido mecánico producido por los elementos móviles de la transmisión.

En este tipo de modelo, la señal está completamente definida por sus coeficiente de Fourier (a_k) y el periodo de repetición de la señal (T), que dependiendo del tipo de ruido con que se trabaje (escape, combustión, mecánico...) coincidirá con la frecuencia de giro del motor o podrá ser un múltiplo de ésta:

$$y(t) = \sum_{k=-\infty}^{\infty} a_k e^{jk \frac{2\pi}{T} t} \quad (4.17)$$

En este caso, la representación frecuencial de la señal $y(t)$ tiene el aspecto mostrado en la Figura 4.30. La potencia de la señal está concentrada en los múltiplos de su frecuencia fundamental, lo que matemáticamente se puede representar por deltas $\delta(f - k/T)$ con un área igual a la amplitud del coeficiente de Fourier del armónico correspondiente (a_k).

Teniendo en cuenta que una señal periódica también se puede escribir como la repetición periódica de una señal patrón $p(t)$ según la expresión:

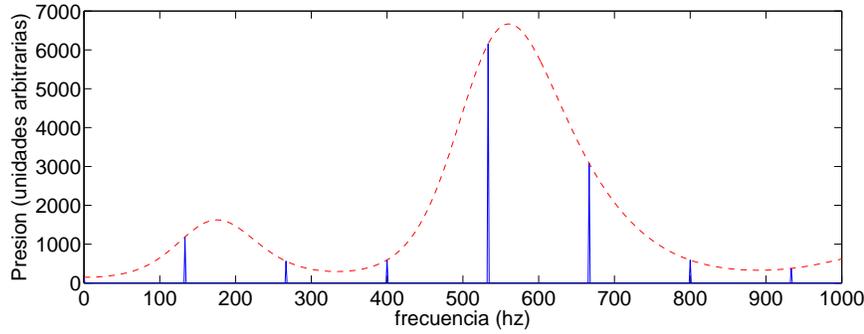


Figura 4.31: Transformada de Fourier de la señal periódica que aparece en la figura 4.28 (a) asociada a un motor que gira a 4000 r.p.m. Con línea discontinua la transformada de Fourier del patrón que se repite y con línea continua la transformada de Fourier de la señal periódica.

$$y(t) = p(t) * c(t) \quad (4.18)$$

donde,

$$c(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT) \quad (4.19)$$

El valor de los (a_k) que aparecen en la figura 4.30 se pueden obtener del análisis de la transformada de Fourier de la señal patrón $p(t)$. Es decir:

$$a_k = \frac{1}{T} P(k/T) \quad (4.20)$$

Este modelo espectral se ajusta bastante bien a la Transformada de Fourier de las señales de escape (como la que aparece en la figura 4.20) si se toma como periodo de repetición el inverso de la frecuencia fundamental f_0 de la ecuación (4.5).

Si se trabaja con ruido de combustión, debido a la elevada amplitud de los armónicos de medio orden, el espectro (ver figura 4.14) se ajusta a un modelo de armónicos donde el periodo de repetición de un motor de 4 cilindros es el doble de la frecuencia de giro del motor (armónicos de medio orden).

La figura 4.31 representa la transformada de Fourier (en el margen de frecuencias desde 0 hasta 1000 Hz) de la señal de escape sintética que aparece en la figura 4.28 (a) obtenida al repetir de forma periódica el patrón representado en la figura 4.27. Se puede observar que el espectro de la señal está concentrado en los múltiplos de la frecuencia de repetición de la señal (133Hz.); y las amplitudes correspondientes a cada frecuencia se ajustan al valor de la transformada de Fourier del patrón.

Si una señal periódica de periodo T como ésta se modula en amplitud con una portadora de frecuencia f_m y un índice de modulación m según el **modelo de Feng** [20] su

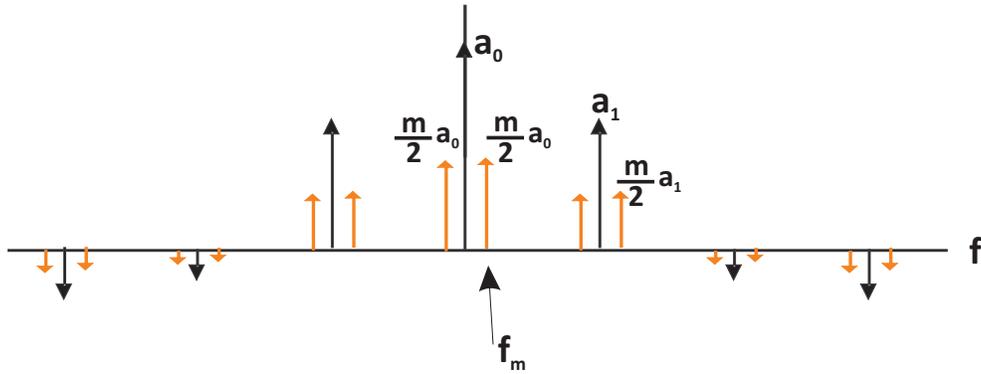


Figura 4.32: Transformada de Fourier de una señal periódica (de periodo T) modulada en amplitud con una frecuencia de modulación f_m y un índice de modulación m .

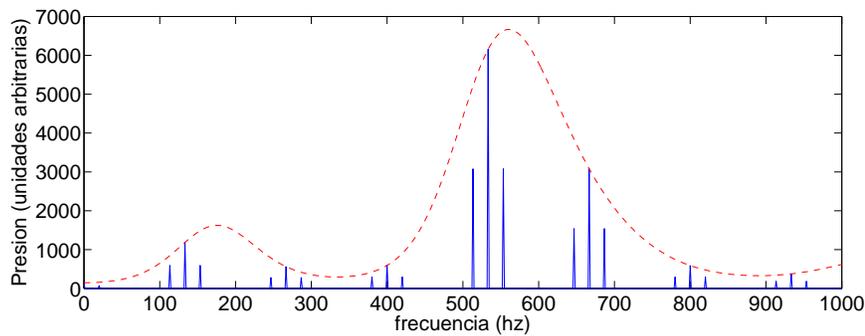


Figura 4.33: Transformada de Fourier de la señal representada en la figura 4.31 modulada en amplitud con una frecuencia de modulación de 20Hz y un índice de modulación $m = 1$ según el modelo de Feng. (Con línea discontinua se representa la transformada de Fourier del patrón utilizado para formar la señal sin modular).

Transformada de Fourier presenta el aspecto que se muestra en la figura 4.32: a la derecha y a la izquierda de cada armónico de la frecuencia fundamental (k/T) aparecen dos deltas (a una distancia f_m de dichos armónicos) cuya amplitud es el resultado de multiplicar la amplitud del armónico original por la mitad del índice de modulación.

La figura 4.32 es un modelo de la transformada de Fourier de este tipo de señales, mientras que la figura 4.33 representa el caso práctico de la Transformada de Fourier de la señal obtenida modulando la señal que aparece en la figura 4.31 según el modelo de Feng, con una frecuencia de modulación de 20Hz y un índice de modulación del 100 % ($m=1$).

Se puede comprobar que la representación que aparece en la figura 4.33 se obtiene superponiendo a ambos lados de los picos de la figura 4.31 dos nuevos picos cuya amplitud es la mitad del pico central (puesto que se ha utilizado un índice de modulación $m = 1$). Teniendo en cuenta que la potencia es proporcional al cuadrado de los valores que aparecen en la gráfica, la aparición de estos nuevos picos supone que la potencia de la señal mostrada en esta figura es superior a la mostrada en la figura 4.31 en un factor de 1.5. Como norma general la potencia de la señal obtenida al modular una señal periódica según el modelo de Feng con un índice de modulación m es superior a la potencia de la señal sin modular

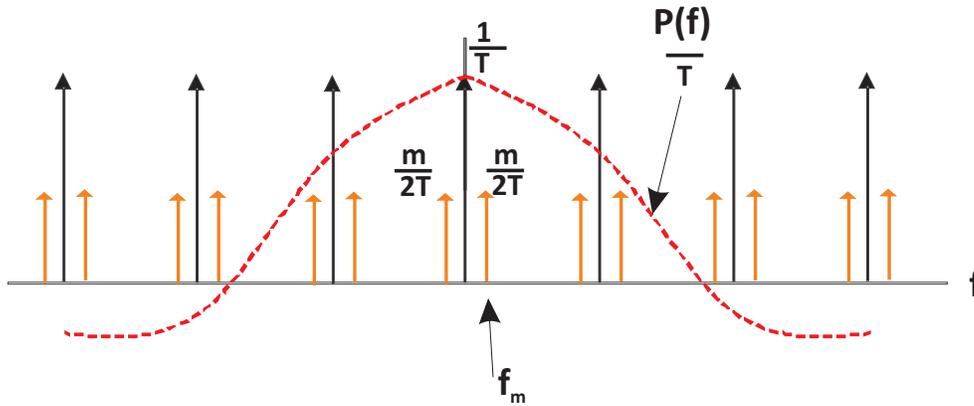


Figura 4.34: Representación de la Transformada de Fourier de un tren de deltas periódicas de periodo T moduladas en AM con frecuencia de modulación f_m e índice de modulación m . La línea discontinua representa el espectro de la señal patrón que se convoluciona en el dominio del tiempo con las deltas cuyo espectro aparece en las líneas continuas.

en un factor $1 + m^2/2$, por lo que para índices de modulación pequeños, la potencia de la señal modulada será similar a la potencia de la señal sin modular.

4.5.2. Señal excitación $c(t)$: tren de deltas modulado en AM.

Este modelo de señal se forma utilizando en la expresión (4.19) la señal excitación que aparece en (4.11). Para representar la Transformada de Fourier de este modelo, resulta imprescindible conocer la forma en la que se seleccionan los valores de a_k y T_k en la expresión (4.11).

Se puede obtener una forma de onda similar a la mostrada en la sección anterior si se seleccionan los valores T_k retrasados uniformemente ($T_k = kT$) y los valores de a_k modulados en AM según 4.16, obteniéndose:

$$c(t) = \sum_{k=-\infty}^{\infty} (1 + m \cdot \cos(2\pi f_m t)) \delta(t - kT), \quad (4.21)$$

En este caso, la transformada de Fourier de la señal $c(t)$ tiene la forma que se indica mediante las líneas continuas de la figura 4.34: deltas de amplitud $1/T$ que se repiten en las frecuencias armónicas de ($f_0 = 1/T$) que tienen a ambos lados (separados una frecuencia f_m) dos deltas de área $m/2T$.

Si la señal de trabajo $y(t)$ se forma convolucionando la señal $c(t)$ con una señal patrón $p(t)$ tal y como se indica en (4.15), la representación frecuencial de la señal resultante corresponderá al producto entre la transformada de Fourier de la señal $c(t)$ y la de $p(t)$ (que aparece representada mediante la línea discontinua de la figura 4.34).

El espectro resultante de la convolución de estas dos señales se muestra en la figura

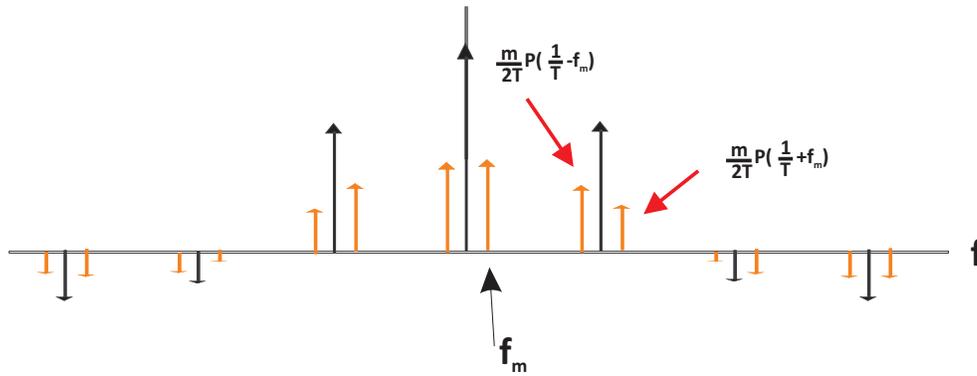


Figura 4.35: Transformada de Fourier de una señal sintética generada mediante la repetición de un patrón a intervalos regulares de tiempo con amplitudes determinadas por una modulación AM con índice de modulación m .

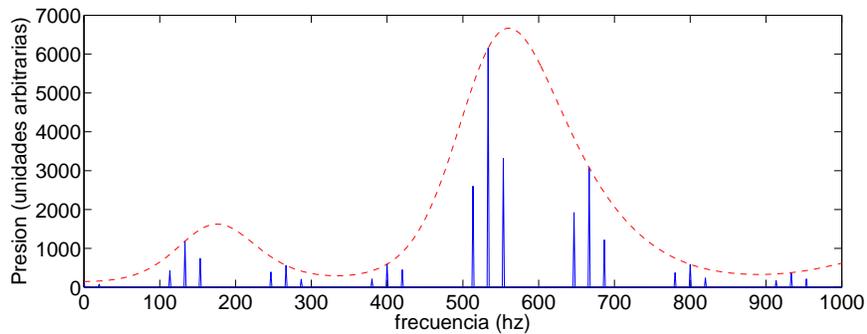


Figura 4.36: Transformada de Fourier de la señal obtenida al convolucionar una señal patrón con un tren de deltas modulado en amplitud con una frecuencia de modulación de 20Hz y un índice de modulación $m = 1$. La línea discontinua corresponde a la transformada de Fourier del patrón utilizado.

4.35. Al comparar la figura 4.32 con la figura 4.35 se puede observar que aunque las deltas que aparecen en los armónicos de la frecuencia fundamental tienen la misma amplitud en ambas representaciones, las deltas que aparecen a los lados de estos armónicos son diferentes en ambas representaciones: mientras que en la figura 4.32 las dos deltas en $f_k \pm f_m$ tienen la misma amplitud $\frac{m}{2}a_k$, en la figura 4.35 sus amplitudes son diferentes. Si bien este hecho no es demasiado relevante a la hora de comparar parámetros psicoacústicos como el Loudness o el Sharpness, puede suponer diferencias importantes a la hora de evaluar el Roughness o el Fluctuation Strength de los dos tipos de modelos de señal.

La figura 4.36 corresponde a un ejemplo práctico de este modelo de señal: representa la Transformada de Fourier de una señal sintética formada a partir del mismo patrón empleado en las figuras 4.31 y 4.33 convolucionado con un tren de deltas modulado en amplitud con una frecuencia de modulación de 20Hz y un índice de modulación $m = 1$. En este caso, no se puede establecer una relación directa entre la potencia de la señal y el índice de modulación, pero, al igual que en el modelo de señal descrita en el apartado 4.5.1, se puede afirmar que si el índice de modulación es reducido la potencia de la señal

obtenida será similar a la de la señal sin modular.

4.5.3. Señal excitación $c(t)$: tren de deltas con amplitud aleatoria.

Otra forma diferente de generar la señal $c(t)$ de la expresión (4.15) consiste en seleccionar los valores de a_k de forma aleatoria con una media 1 y desviación típica σ ; y los valores de T_k equiespaciados a intervalos T . Para la obtención de la representación espectral de esta señal $c(t)$ resulta conveniente reescribir los valores de a_k como la suma de la constante 1 y una variable aleatoria b_k de media 0 y desviación típica σ :

$$c(t) = \sum_{k=-\infty}^{\infty} (1 + b_k)\delta(t - kT) = \sum_{k=-\infty}^{\infty} \delta(t - kT) + \sum_{k=-\infty}^{\infty} b_k\delta(t - kT), \quad (4.22)$$

En este caso, la transformada de Fourier de la señal $c(t)$ (representada mediante las líneas continuas de la figura 4.37) se puede descomponer en dos partes, asociadas a cada uno de los sumandos de la expresión (4.22):

- Un conjunto de deltas con la misma amplitud situadas en los armónicos de la frecuencia fundamental $f_0 = 1/T$, asociadas al primer sumando de la ecuación (4.22)
- Y un ruido blanco cuya amplitud depende de la desviación típica de los valores de a_k , asociadas al primer sumando de la ecuación (4.22). Se puede observar en la figura 4.37 que este ruido blanco presenta un patrón de repetición periódico (como cabe esperar puesto que la transformada de Fourier de una señal compuesta por un tren de deltas debe ser periódica).

En la figura 4.37 también se ha representado mediante un trazo discontinuo la transformada de Fourier de una señal patrón que se convolucionará con la señal $c(t)$ según la expresión (4.15) para formar la señal sintética $x(t)$. La transformada de Fourier de la señal $x(t)$ corresponderá por lo tanto a la multiplicación de la función representada en la línea discontinua y la representada por la línea continua.

La figura 4.38 representa la transformada de Fourier de la señal sintética resultante al repetir una señal patrón de forma aleatoria conforme a lo descrito en los párrafos anteriores; y en ella se observa que tanto las deltas como el ruido siguen la forma del espectro del patrón.

La figura 4.39 es un ejemplo de este tipo de señales obtenido a partir de la convolución del mismo patrón empleado en los ejemplos de las figuras 4.31, 4.33 y 4.36, con un tren de deltas aleatorias de media 1 y varianza 1. De nuevo se puede apreciar que la mayor parte de la potencia de la señal se encuentra concentrada en las frecuencias múltiplo de la frecuencia del tren de deltas. Una vez más, la potencia de la señal generada será similar a la de la señal periódica si la varianza de las deltas es reducida.

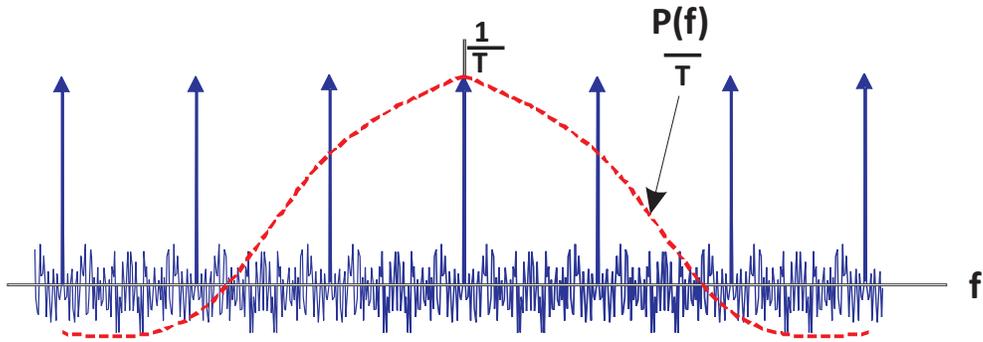


Figura 4.37: Las líneas continuas representan la transformada de Fourier de un tren de deltas periódicas de periodo T cuya amplitud ha sido modulada por una señal aleatoria de media 1 y varianza σ . La línea discontinua representa el espectro de la señal patrón que se utilizará en (4.15) para formar la señal sintética $x(t)$.

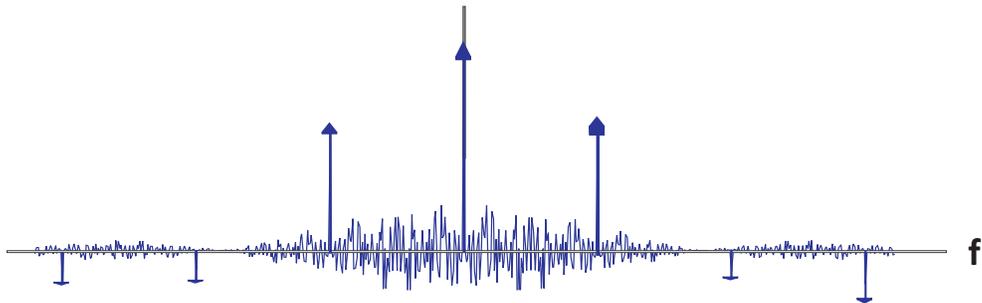


Figura 4.38: Transformada de Fourier de una señal sintética generada mediante la repetición de un patrón a intervalos regulares de tiempo con amplitudes seleccionadas de forma aleatoria según una distribución estadística de media 1 y varianza σ .

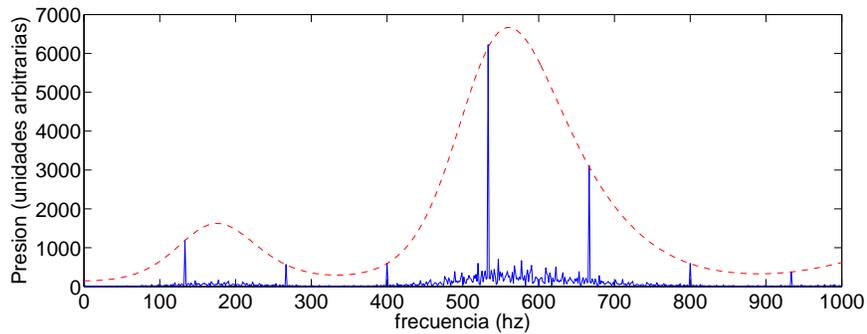


Figura 4.39: Transformada de Fourier de la señal obtenida al convolucionar una señal patrón con un tren de deltas aleatorias de media 1 y varianza 1. La línea discontinua corresponde a la transformada de Fourier del patrón utilizado.

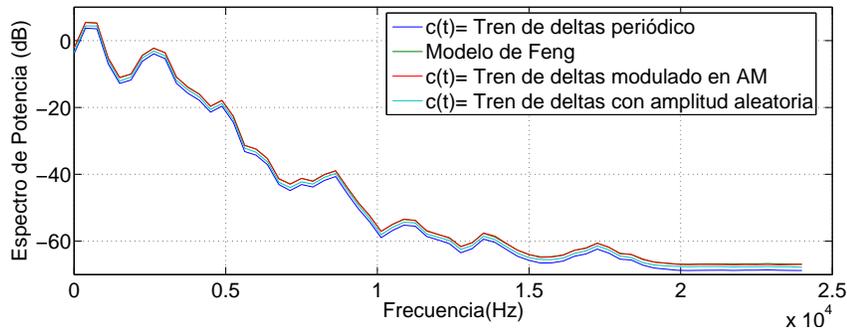


Figura 4.40: Densidad Espectral de Potencia de las señales sintéticas generadas a partir del patrón de la figura 4.27 y cuya Transformada de Fourier aparece en las figuras 4.31, 4.33, 4.36 y 4.39

4.5.4. Comparación entre modelos.

Las representaciones frecuenciales utilizadas en los ejemplos de esta sección son Transformadas de Fourier en el rango de frecuencias entre 0 y 1000 Hz y todas están formadas por varios picos puesto que la potencia de la señal se acumula en frecuencias concretas. Este tipo de representación es adecuado para mostrar el detalle de la distribución espectral, pero si se desea obtener una visión global del espectro de la señal en un rango de frecuencias más amplio es preferible utilizar la Densidad Espectral de Potencia (DEP) que se puede obtener utilizando el método de Welch [214] de estimación espectral mediante promediado de periodograma.

La figura 4.40 muestra la DEP hasta 25kHz de las señales sintéticas obtenidas con los distintos métodos presentados en las secciones 4.5.1, 4.5.2 y 4.5.3 a partir de un mismo patrón (el de la figura 4.27). Como se puede apreciar, todos los métodos generan señales con una DEP muy similar que, además, coincide con la representación de la figura 4.28 (b). Se puede observar también que existe una ligera diferencia entre las amplitudes de las distintas representaciones: la señal estrictamente periódica presenta menor potencia que el resto. Esto se debe a que en todas ellas se ha utilizado un índice de modulación muy elevado, pero si el índice de modulación es reducido la diferencia de potencias es despreciable.

La figura 4.41 representa la DEP de 4 señales sintéticas periódicas generadas con el patrón de la figura 4.27 y distintas velocidades de giro de motor. De nuevo se observa una gran similitud entre ellas, puesto que tan solo se aprecia una diferencia de nivel entre ellas cuya interpretación temporal es obvia: al aumentar el número de veces que se repite un patrón por unidad de tiempo aumenta la potencia de la señal generada.

En definitiva, al utilizar el modelo propuesto para generar señales sintéticas, la forma de la DEP permanece constante al variar el periodo de repetición de la señal o sus parámetros de modulación, puesto que la forma de la DEP depende únicamente del patrón de repetición.

Este hecho nos permitirá generar varias señales sintéticas con una misma DEP y

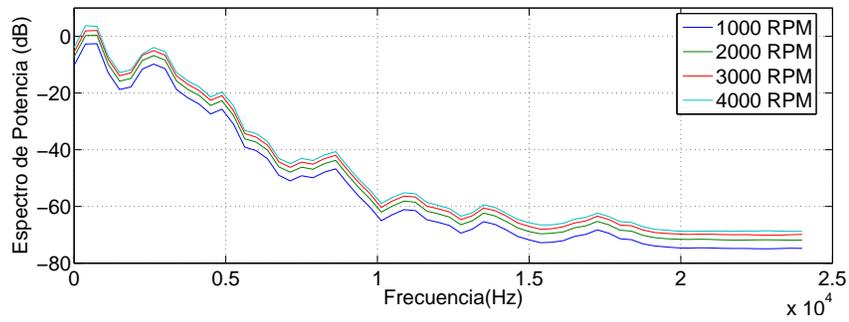


Figura 4.41: Densidad Espectral de Potencia de las señales sintéticas generadas a partir del patrón de la figura 4.27 con distintas velocidades de giro de motor.

distintos tipos de modulación para estudiar la respuesta del oído ante la modulación, eliminando la distorsión que pudieran suponer en los experimentos trabajar grupos de señales donde se modifiquen simultáneamente varios parámetros psicoacústicos.

Capítulo 5

Parámetros Psicoacústicos en el modelo de señales Cuasiperiódicas.

5.1. Introducción.

El objetivo de este capítulo es el estudio de los parámetros psicoacústicos de las señales sintéticas generadas mediante el modelo cuasiperiódico descrito en la sección 4.4.

En este modelo, las señales se generan mediante la convolución de un pulso con un tren de deltas moduladas según la ecuación 4.15. En la sección 4.5 se comprobó cómo se puede fijar la división de la potencia de las señales sintéticas en el espectro de frecuencia y, por lo tanto, entre los distintos filtros de banda crítica seleccionando la forma de la señal patrón $p(t)$. La forma de la densidad espectral de potencia dentro de un filtro de banda crítica (repartida en toda la banda o concentrada en una o varias frecuencias) depende exclusivamente de la señal $c(t)$ (distancia entre las deltas, tipo de modulación, amplitud y frecuencia de modulación ...).

Esto hace que el Loudness de las señales sintéticas se pueda variar libremente modificando la amplitud de $p(t)$, mientras que el Sharpness se puede variar modificando la distribución espectral de $p(t)$. Al realzar las frecuencias altas de la señal patrón se realzarán también las frecuencias altas de la Densidad Espectral de Potencia del sonido producido y por tanto aumentará el Sharpness (y viceversa).

En cuanto a los parámetros Fluctuation Strength y Roughness, su valor se podrá variar modificando la señal $c(t)$. Al trabajar con ruidos de motor modelados como señales cuasiperiódicas, el Roughness es uno de los parámetros más influyentes en la valoración de la molestia producida [19, 228, 20, 21, 22, 23, 24].

Existe un gran número de estudios dedicados a analizar el Roughness producido por señales simples como dos tonos de distintas frecuencias, o unos pocos armónicos de una frecuencia fundamental con distintas relaciones de amplitudes y fases entre sí [232, 233, 60, 3, 234, 235]. Y también existen distintos algoritmos que proporcionan medidas del Roughness muy próximas a las valoraciones obtenidas en experimentos con jurados para

este tipo de señales simples [3, 8, 5]. Pero los distintos ruidos producidos por un motor no se ajustan a estos modelos tan simples. Aunque algunos de estos algoritmos se han aplicado con éxito a otro tipo de señales más complejas, no presentan buenos resultados cuando se aplican a ruidos de motor ni a señales cuasi periódicas en general.

En este capítulo se propondrá una relación entre los parámetros de las señales cuasi-periódicas y el Roughness percibido y evaluado mediante tests de jurado. Para esto, en la sección 5.2 se analiza matemáticamente la envolvente de las señales presentes en cada uno de los filtros de banda crítica asociados al modelo de señal propuesto; así como el Roughness y Fluctuation Strength derivados de su análisis. Como consecuencia de este análisis, en la sección 5.3 nos planteamos la adecuación de los métodos tradicionales de cálculo de Roughness para este tipo de señales y en la sección 5.4 analizamos el Roughness y Fluctuation Strength de señales sintéticas de escape mediante varios experimentos. Los resultados de los experimentos indican una relación entre el Roughness y los parámetros de la señal sintética, pero con el propósito de generalizar los resultados, la sección 5.5 propone la utilización de un nuevo tipo de patrones sintéticos mientras que la sección 5.6 analiza el Roughness de las señales generadas con esos nuevos patrones. La sección 5.7 resume las conclusiones de todos estos experimentos.

5.2. Envolvente en los canales de banda crítica.

Si se analizan los parámetros psicoacústicos de una señal periódica de frecuencia f_0 , la primera parte del análisis consistirá hacer pasar la señal por un banco de filtros que modele el comportamiento del oído (filtros de banda crítica), como muestra la figura 5.1. La salida obtenida en cada uno de los filtros que componen el banco de filtros consistirá en un conjunto de armónicos de la frecuencia fundamental f_0 , es decir:

$$\dots, nf_0, (n + 1)f_0, (n + 2)f_0, \dots \quad (5.1)$$

Por ejemplo, si se trabaja con la señal de escape de un motor de 4 cilindros funcionando a 3000 r.p.m. la frecuencia fundamental de la señal periódica obtenida será de $f_0 = 100$ Hz. Puesto que el ancho de banda crítica para frecuencias inferiores a 500 Hz es de aproximadamente 100 Hz, al descomponer esta señal periódica en bandas críticas las bandas críticas bajas contendrán tan sólo un armónico. Para frecuencias superiores a los 500 Hz, el ancho de banda crítica es aproximadamente el 20 % de la frecuencia central de la banda, por lo que el ancho de banda crítica aumenta a medida que aumenta la frecuencia de análisis y para 2000 Hz es de aproximadamente 400 Hz. Esto significa que a la salida del filtro de banda crítica de 2000 Hz existirán un total de 4 armónicos. Como la mayor parte de la potencia (desde el punto de vista acústico) de este tipo de señales se encuentra por debajo de los 2000 Hz, en dicho rango de frecuencias los filtros de banda crítica dejan pasar entre 1 y 4 armónicos de la señal.

El cálculo del Roughness de un sonido se realiza a partir del análisis de la envolvente de las señales obtenidas a la salida cada uno de los filtros anteriores por lo que resul-

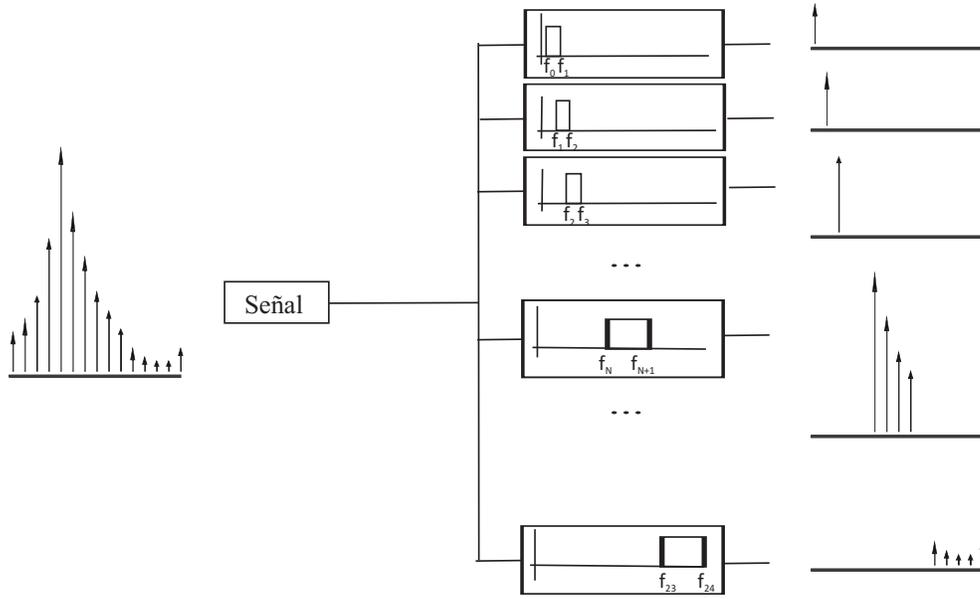


Figura 5.1: Descomposición de una señal periódica por el sistema auditivo.

ta interesante disponer de una expresión analítica que nos permita trabajar con estas envolventes.

La envolvente de estas señales se puede obtener calculando el modulo de la señal analítica [89, 236] que resulta de la utilización de la transformada de Hilbert según la expresión (3.7) y que en el caso de trabajar con señales discretas se puede obtener de forma muy eficiente mediante el uso de la FFT [237].

Cuando la señal a analizar está formada por una combinación de señales coseno según la expresión:

$$x(t) = \sum_{i=1}^N A_i \cos(\omega_i t + \phi_i) \quad (5.2)$$

dicha envolvente se puede calcular [233] mediante la expresión:

$$E^2(t) = \sum_{i=1}^N A_i^2 + 2 \sum_{\substack{j>i \\ i=1}}^N A_i A_j \cos((\omega_i - \omega_j)t + (\phi_i - \phi_j)) \quad (5.3)$$

Si se trabaja con una señal formada por la suma de 4 señales coseno, la envolvente se puede obtener como [23]:

$$E(t) = \sqrt{DC^2 + AC_1^2 + AC_2^2 + AC_3^2} \quad (5.4)$$

donde

$$DC^2 = A_1^2 + A_2^2 + A_3^2 + A_4^2 \quad (5.5)$$

$$\begin{aligned} AC_1^2 = & 2A_1A_2 \cos((\omega_2 - \omega_1)t + (\phi_2 - \phi_1)) \\ & + 2A_2A_3 \cos((\omega_3 - \omega_2)t + (\phi_3 - \phi_2)) \\ & + 2A_3A_4 \cos((\omega_4 - \omega_3)t + (\phi_4 - \phi_3)) \end{aligned} \quad (5.6)$$

$$\begin{aligned} AC_2^2 = & 2A_1A_3 \cos((\omega_3 - \omega_1)t + (\phi_3 - \phi_1)) \\ & + 2A_2A_4 \cos((\omega_4 - \omega_2)t + (\phi_4 - \phi_2)) \end{aligned} \quad (5.7)$$

$$AC_3^2 = 2A_1A_4 \cos((\omega_4 - \omega_1)t + (\phi_4 - \phi_1)) \quad (5.8)$$

Pero en el caso que nos ocupa, las frecuencias que componen la señal son armónicos de una frecuencia fundamental por lo que se cumplen la relaciones:

$$\omega_2 - \omega_1 = \omega_3 - \omega_2 = \omega_4 - \omega_3 = 2\pi f_o \quad (5.9)$$

$$\omega_3 - \omega_1 = \omega_4 - \omega_2 = 2\pi 2f_o \quad (5.10)$$

$$\omega_3 - \omega_1 = \omega_4 - \omega_2 = 2\pi 3f_o \quad (5.11)$$

Si se emplea notación compleja para representar mediante un solo valor complejo el módulo y la fase de cada uno de los armónicos que forman la señal que aparece en (5.2):

$$Z_1 = A_1 e^{j\phi_1}; Z_2 = A_2 e^{j\phi_2}; Z_3 = A_3 e^{j\phi_3}; Z_4 = A_4 e^{j\phi_4} \quad (5.12)$$

Las expresión (5.4) se puede reescribir como:

$$E(t) = \sqrt{\text{Real} \{R_0 + 2R_1 e^{j2\pi f_o t} + 2R_2 e^{j2\pi 2f_o t} + 2R_3 e^{j2\pi 3f_o t}\}} \quad (5.13)$$

Siendo

$$R_0 = Z_1 Z_1^* + Z_2 Z_2^* + Z_3 Z_3^* + Z_4 Z_4^* \quad (5.14)$$

$$R_1 = Z_2 Z_1^* + Z_3 Z_2^* + Z_4 Z_3^* \quad (5.15)$$

$$R_2 = Z_3 Z_1^* + Z_4 Z_2^* \quad (5.16)$$

$$R_3 = Z_4 Z_1^* \quad (5.17)$$

Donde se puede comprobar que R_i corresponde a la autocorrelación para un desplazamiento de i muestras del vector formado por los coeficientes $[Z_1, Z_2, Z_3, Z_4]$, que se define mediante la expresión:

$$R_i = \sum_{n=1}^{N-i} Z_{i+n} Z_n^* \quad (5.18)$$

donde N es la longitud del vector (que coincide con el número de armónicos).

Teniendo en cuenta esta definición podemos generalizar las expresiones (5.4) y (5.13) para el caso de tener N armónicos:

$$E(t) = \sqrt{\text{Real} \left\{ R_0 + 2 \sum_{i=1}^{N-1} R_i e^{j2\pi i f_0 t} \right\}} \quad (5.19)$$

Como síntesis de la expresión anterior cabe decir que la envolvente de una señal formada por N armónicos es una señal periódica, con frecuencia de repetición igual a la frecuencia fundamental asociada a los N armónicos. La expresión analítica de dicha envolvente consiste en la raíz cuadrada de la suma de N armónicos cuyas amplitudes y fases se pueden obtener mediante la operaciones algebraicas simples realizadas con los coeficientes de Fourier de la señal original.

5.2.1. Señal excitación $c(t)$: tren de deltas periódico.

Una señal de escape ideal sin ningún tipo de modulación producida por un motor de 4 cilindros trabajando a 3000 r.p.m. tiene una forma periódica de frecuencia fundamental $f_0 = 100$ Hz. Como se vio en la sección 4.5.1 la señal se podrá expresar matemáticamente mediante la ecuación (4.17) y su espectro tendrá la forma que aparece en la figura 4.30, donde la separación entre las deltas será de 100 Hz y la amplitud de cada una de ellas dependerá de la forma de onda de la señal patrón.

Si se estudia el Roughness de este tipo de señales según el modelo de Daniel y Weber (ver apartado 2.6.6), habrá que hacer pasar la señal a través de un banco de filtros de ancho de banda igual a un bark. Esto dará como resultado en frecuencias inferiores a 2000 Hz canales que contendrán entre 1 y 4 armónicos del total de armónicos que aparecen en la ecuación (4.17):

$$x_l(t) = \text{Re} \left\{ \sum_{k=k_0}^{k_0+i_l-1} 2a_k e^{jk \frac{2\pi}{T} t} \right\} \quad (5.20)$$

Donde $x_l(t)$ es la salida del filtro de la banda crítica l , k_0 es el armónico de frecuencia más baja contenido en la banda crítica e i_l el número de armónicos contenidos en la banda crítica l . Para estudiar los parámetros Fluctuation Strength y Roughness de cada canal

es necesario obtener la envolvente de la señal para lo que resulta útil escribir la ecuación (5.20) como:

$$x_l(t) = \text{Re} \{ E(t) e^{j\theta(t)} \} \quad (5.21)$$

Pudiéndose escribir la señal $x(t)$ como la suma de las señales en cada banda crítica:

$$x(t) = \sum x_l(t) \quad (5.22)$$

Dependiendo del número de armónicos en cada canal de banda crítica (valor de i_l de cada canal en la ecuación (5.20)) podremos distinguir los siguientes casos:

Canales con 1 armónico: en este caso $E(t)$ toma el valor constante:

$$E(t) = 2 |a_k| \quad (5.23)$$

y la salida del filtro es una señal coseno cuya envolvente es constante y por lo tanto no contribuye a incrementar el Roughness ni el Fluctuation Strength de la señal.

Canales con 2 armónicos: aquellos canales que contengan 2 armónicos presentarán siempre valores elevados de Roughness y Fluctuation Strength puesto que el batido de dos cosenos produce siempre una señal con envolvente no constante, ya que la expresión (5.19) se convierte en:

$$E(t) = \sqrt{a^2 + b^2 + 2 \cdot a \cdot b \cdot \cos(2\pi f_0 t + \phi)} \quad (5.24)$$

donde a y b son las amplitudes de los respectivos armónicos y ϕ la diferencia de fase entre los dos armónicos en el instante de tiempo $t = 0$.

En esta expresión se puede observar que la envolvente es periódica con periodo $T = 1/f_0$. En cada periodo solamente se produce un máximo y un mínimo que valen:

$$\min \{ E(t) \} = \sqrt{a^2 + b^2 - 2 \cdot a \cdot b} \quad (5.25)$$

$$\max \{ E(t) \} = \sqrt{a^2 + b^2 + 2 \cdot a \cdot b} \quad (5.26)$$

La figura 5.2 corresponde a la representación de la señal obtenida mediante la suma del armónico 10 y 11 de una frecuencia fundamental de 10Hz con amplitudes de 1 y 0.5 respectivamente. En esta gráfica se puede observar que la envolvente es aproximadamente un coseno de frecuencia igual a la frecuencia fundamental de la señal aunque no es exactamente un coseno (como también se desprende de la ecuación (5.24)).

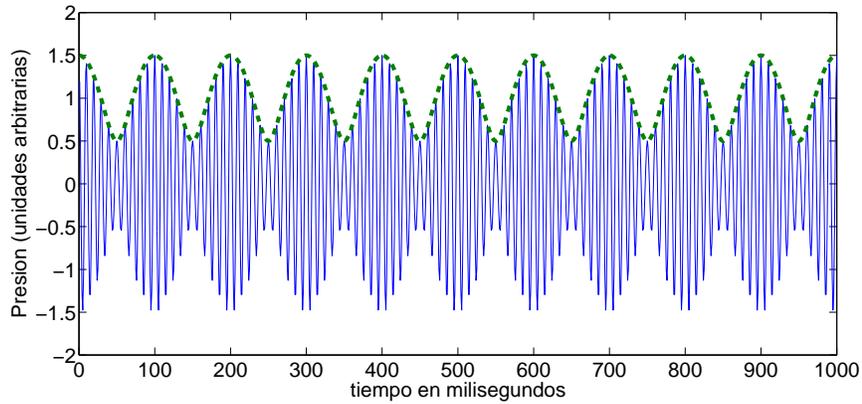


Figura 5.2: Batido producido por la suma de un coseno de frecuencia 100Hz y amplitud 1 y otro seno de frecuencia 110Hz y amplitud 0.5. La línea discontinua corresponde a la envolvente de la señal.

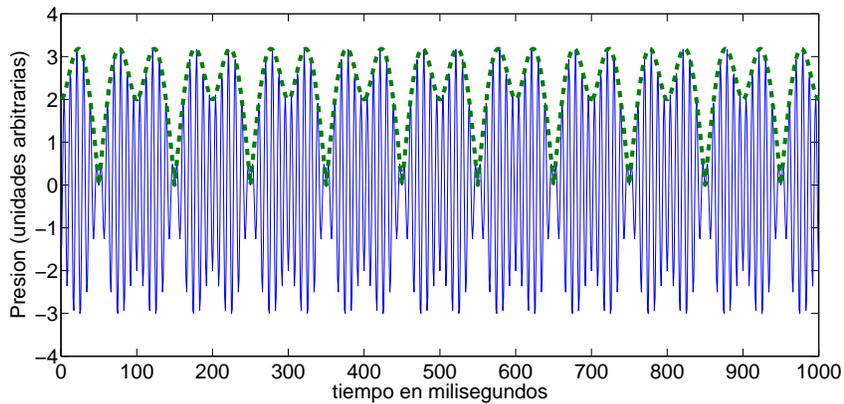


Figura 5.3: Forma de onda producida al sumar tres cosenos de frecuencias 100Hz 110Hz y 120Hz; amplitudes 1, 1 y 2; y fases 0, π y π respectivamente. La línea discontinua corresponde a la envolvente de la señal.

El índice de modulación de la envolvente será mayor cuanto menor sea la diferencia de amplitudes de los dos armónicos que forman la señal y es independiente del desfase que haya entre ellos.

Canales con 3 armónicos: En canales con 3 armónicos se produce una envolvente de mayor frecuencia que cuando solamente existen dos armónicos puesto que al desarrollar la ecuación (5.19) aparece dentro de la raíz un coseno de frecuencia doble a la frecuencia fundamental de la señal periódica. Este mayor contenido en frecuencias altas de la envolvente se puede apreciar en la figura 5.3 que representa la señal producida al superponer un coseno de frecuencia 100Hz y amplitud 1 y fase 0, otro coseno de frecuencia 110Hz amplitud 1 y fase π y un tercer coseno de frecuencia 120Hz, amplitud 2 y fase π .

En el caso más común las amplitudes y fases de los tres cosenos son independientes

entre sí y se producen envolventes con una gran amplitud y anchura espectral. Aunque se pueden dar casos particulares cuando los armónicos de frecuencia mayor y menor tienen la misma amplitud (modulaciones AM y cuasi-FM [236] según estén en fase o en oposición de fase con el armónico de frecuencia central). En el caso de la modulación AM, la envolvente de la señal consiste en un coseno con frecuencia igual a la frecuencia fundamental de los armónicos. En el caso de modulaciones cuasi-FM, la envolvente es prácticamente plana.

Canales con 4 o más armónicos: Al incrementar el número de armónicos que componen la señal, la envolvente de la señal resultante aumenta en ancho de banda. La forma exacta de la envolvente depende de la relación existente entre amplitudes y fases de los armónicos que la forman. Si estas amplitudes y fases tienen forma aleatoria, la envolvente presenta un índice de modulación cercana al 100 %.

Cuando se analiza el Roughness de una señal de escape de las características indicadas mediante el modelo de Daniel y Weber (analizado en la sección 2.6.6), encontramos que únicamente los filtros asociados a frecuencias bajas contienen envolventes continuas y por lo tanto no producen Roughness. Pero todos aquellos canales asociados a frecuencias medias y altas contienen señales cuyas envolventes presentan variaciones de amplitudes considerables (puesto que las amplitudes y fases de los distintos armónicos del canal observados en la práctica no están correlados entre sí). A pesar de las grandes amplitudes observadas en las envolventes de cada uno de los filtros de banda crítica (profundidad de modulación generalizada), el grado de correlación entre ellas es pequeño y como resultado de esto el Roughness específico en cada banda suele presentar valores muy reducidos.

Como el Roughness total se calcula como la suma de Roughness específicos de cada una de las bandas (que siempre toma valores positivos) para algunas configuraciones especiales de armónicos se podrían obtener valores de Roughness elevados, aún tratándose de una señal periódica sin ningún tipo de modulación [238].

5.2.2. Modelo de Feng.

Si partimos de una señal periódica escrita mediante su descomposición en bandas críticas según aparece en la expresión (5.22) y se modula en amplitud con una portadora senoidal se obtiene:

$$(1 + m \cdot \cos(2\pi f_m t)) \sum x_i(t) = \sum (1 + m \cdot \cos(2\pi f_m t)) x_i(t) \quad (5.27)$$

Es decir, la modulación de la señal periódica equivale a modular cada una de sus bandas críticas con la misma señal moduladora. Pero al modular una señal en AM, aumenta su ancho de banda en un valor igual a la frecuencia de modulación por lo que solamente si la frecuencia de modulación es pequeña se podrá afirmar que el contenido de cada banda crítica es:

$$x_l(t) = (1 + m \cdot \cos(2\pi f_m t)) \operatorname{Re} \left\{ \sum_{k=k_0}^{k_0+i_l-1} 2a_k e^{jk \frac{2\pi}{T} t} \right\} \quad (5.28)$$

Cuando una señal periódica como la mostrada en la figura 4.30 se modula en AM con una portadora senoidal, se obtiene una señal cuyo espectro queda modelado en la figura 4.32. Las deltas que aparecen a ambos lados de cada armónico están separadas de éste una frecuencia igual a la frecuencia de modulación f_m , tienen una amplitud igual al producto de la amplitud del armónico por el índice de modulación m dividido entre dos.

Si el valor de f_m es reducido, las deltas que aparecerán al lado de cada armónico (en la figura 5.1) estarán próximas a éste y, por lo tanto, al descomponer la señal en bandas críticas, estarán en la misma banda que el armónico correspondiente. Analizando por separado lo que ocurre en los filtros de banda crítica según el número de armónicos que estos contengan al igual que se hizo al estudiar las señales periódicas:

Canales con 1 armónico (bandas de frecuencia baja): La salida de cada uno de estos canales contendrá 1 armónico con sus dos deltas asociadas a la señal moduladora. Esto producirá oscilaciones de la envolvente con un índice de modulación igual a m . Además, las oscilaciones de todos los canales con un solo armónico estarán sincronizadas entre sí por lo que la correlación entre canales será elevada, dando origen a valores elevados en el Roughness específico de estas bandas.

Canales con varios armónicos (bandas de frecuencia altas): La forma de la envolvente en estas bandas depende de dos factores: por un lado, la existencia de varios armónicos produce una envolvente periódica cuya energía se concentra en los múltiplos de la frecuencia fundamental del sonido analizado.

Por otro lado, la modulación AM produce una envolvente con una amplitud proporcional al índice de modulación y una frecuencia igual a la frecuencia de modulación. Por lo tanto, la influencia de la modulación en la envolvente se limita a la región espectral de la envolvente (coincidente con la frecuencia de modulación f_m) y la relación con el índice de modulación no es lineal, pues también depende de la estructura espectral de los armónicos.

- En el caso excepcional de que los armónicos sin modular tuvieran una **envolvente continua** o con oscilaciones reducidas, la aparición de la señal moduladora provocaría un aumento en el Roughness específico de la banda proporcional al índice de modulación. Además, como la señal moduladora es la misma en todas las bandas, las envolventes estarán correladas, por lo que el Roughness específico de estas bandas aumentará al aumentar el índice de modulación.
- Si, como resulta habitual en las señales de motor, la envolvente de la señal existente a la salida del filtro de banda crítica presenta **variaciones importantes**, la presencia de una señal moduladora no necesariamente aumentará el valor de la profundidad de modulación generalizada sino que dependiendo de la amplitud y fase de la señal original y la moduladora ésta se podría incrementar

o decrementar. En este caso, la envolvente de la señal en cada banda crítica no es debida a la modulación sino a su composición armónica, por lo que la correlación entre las bandas críticas adyacentes será reducida. Como consecuencia, el Roughness específico de estas bandas suele tener valores reducidos [2.6.6](#).

Se puede deducir de los párrafos anteriores que si la frecuencia de modulación es baja, el Roughness total (calculado mediante el método de Daniel y Weber) se debe principalmente a las bandas bajas del espectro y este crece proporcionalmente al índice de modulación m . Pero si tenemos en cuenta que las bandas bajas tienen una anchura espectral aproximada de 100Hz, para poder hablar de frecuencias de modulación reducidas, deben ser inferiores a los 50 Hz. Pero para oscilaciones con frecuencias muy por debajo de los 50 Hz la sensación percibida por el oído es Fluctuación en lugar de Roughness, por lo que las conclusiones anteriores solamente tienen validez para frecuencias de modulación ligeramente inferiores a los 50 Hz.

Para frecuencias de modulación altas (superiores a 50 Hz), deja de ser cierta la igualdad mostrada en la ecuación [\(5.28\)](#) en las bandas de frecuencia baja, pero continua siendo cierta para las frecuencias más altas. La contribución al Roughness de las distintas bandas se puede resumir del siguiente modo:

Bandas de frecuencia baja: Contribuyen en menor medida al Roughness total, porque a pesar de los elevados valores de las profundidades de modulación generalizadas en cada canal, la correlación entre ellos es reducida. Esto se debe a que las deltas que aparecen en el espectro a cada lado del armónico debido a la modulación, se encuentran tan separadas de este que se desplazan al canal $n-1$ y al canal $n+1$. De este modo, las deltas más próximas a cada armónico provienen del canal superior e inferior y la distancia entre estas y el armónico central ya no es la frecuencia de modulación (f_m) sino la diferencia entre esta y la frecuencia fundamental de la señal ($f_m - f_0$). En estas circunstancias la envolvente de cada canal ya no está en fase con los contiguos y por lo tanto la correlación se reduce. Además la frecuencia de modulación de la señal en la banda no coincide con la frecuencia de la señal moduladora.

Bandas de frecuencia altas: En las bandas críticas más altas, la anchura espectral del canal toma también valores elevados. En este caso, coexistirán en el canal varios armónicos y todos ellos serán modulados por la misma moduladora (excepto aquellos más alejados del centro del canal que sufrirán la interferencia de los canales adyacentes). Esto dará como resultado que la envolvente del canal sea aproximadamente igual al producto entre la envolvente que tendría el canal sin modular multiplicado por la señal moduladora. El índice de modulación de la señal moduladora influirá en la profundidad de modulación generalizada pero únicamente en la zona de frecuencias próximas a la frecuencia de modulación, por lo tanto puede aumentar la correlación entre distintas bandas, pero depende de la configuración de los armónicos.

A la vista de la discusión anterior y en contra de lo que cabría esperar, el Roughness de una señal periódica modulada no crece al aumentar el índice de modulación de la señal moduladora sino que la relación entre estos dos parámetros es bastante más compleja. En los experimentos realizados por Feng [20]) con este tipo de señales, se confirmó que el aumento del índice de modulación no producía necesariamente un aumento del Roughness.

5.2.3. Señal excitación $c(t)$: tren de deltas modulado en AM.

Aunque este tipo de señales se asemeja a las obtenidas mediante la modulación AM analizada en la sección anterior, la figura 4.35 muestra como las deltas que aparecen a ambos lados de los armónicos tienen distinta amplitud (a diferencia de lo que se aprecia en la gráfica 4.32); y este hecho además de afectar a la amplitud de la envolvente [233], disminuirá la correlación entre las distintas bandas críticas.

Aunque el procedimiento de cálculo de Roughness propuesto por Daniel y Weber 2.6.6 es capaz de explicar el hecho de que no aumente el Roughness al aumentar el índice de modulación, ofrece unos resultados que no coinciden con la valoración obtenida en un test de jurado cuando se utilizan señales sintéticas de motor obtenidas mediante la convolución de una señal patrón con un tren de deltas moduladas en AM con portadora senoidal [229]. En estos experimentos se observó que tanto el algoritmo de Zwicker [3] como el de Daniel y Weber [8], subestiman el Roughness para bajas revoluciones de motor y lo sobreestiman para altas revoluciones.

5.2.4. Señal excitación $c(t)$: tren de deltas con amplitud aleatoria.

Este tipo de señales se generan utilizando a partir de la expresión (4.38) donde los valores de a_k presentan una distribución de ruido gaussiano blanco. La distribución espectral de potencia asociada a estas señales es la que aparece en la figura 4.38, donde se puede apreciar la superposición de dos distribuciones: una distribución concentrada formada por conjunto de deltas situadas en los armónicos de la frecuencia del ruido, y una distribución dispersa repartida por todo el espectro de frecuencias. La potencia asociada a la distribución concentrada aumenta cuando crece la media de a_k , mientras que la potencia asociada a la distribución dispersa aumenta al incrementar la varianza de a_k . Un incremento en la varianza de a_k no supone necesariamente un aumento en las envolventes de las salidas de los filtros de banda crítica que modelan el comportamiento del oído, ni tampoco un aumento de la correlación entre distintas bandas críticas. Por lo tanto, un incremento en la varianza de a_k no supone un aumento del Roughness ni del Fluctuation Strength cuando se aplican los métodos de cálculo de estos parámetros analizados en el capítulo 2.

Como resumen de esta sección podemos decir que los modelos tradicionales de cálculo de Roughness (capítulo 2) proporcionan resultados correctos cuando se aplican a señales simples, pero al trabajar con señales complejas como las descritas en el capítulo 4 los

resultados podrían no ajustarse a la sensaciones percibidas. En los siguientes experimentos se estudiará la dependencia del Roughness y el Fluctuation Strength con los parámetros de las señales sintéticas y se propondrá una mejora del modelo de cálculo de Roughness.

5.3. Evaluación de la calidad de los estimadores de Roughness.

El análisis teórico realizado en la sección 5.2 parece indicar una discordancia entre la sensación de Roughness y los resultados proporcionados por los métodos tradicionales de cálculo de este parámetro al aplicarlos a señales cuasiperiódicas. Puesto que muchas señales reales (como las producidas en el tubo de escape de un motor) se pueden modelar como señales cuasiperiódicas, estos métodos de cálculo podrían proporcionar resultados inexactos al trabajar con ellas. En esta sección se evalúa el grado de exactitud de estos métodos a la hora de calcular el Roughness producido por señales cuasiperiódicas tomando varias señales de escape como representativas de esta clase.

5.3.1. Los jueces.

En los experimentos han intervenido 14 jueces con edades comprendidas entre los 20 y los 40 años. Los jueces fueron reclutados de entre los alumnos y profesores de la titulación de Ingeniero de Telecomunicación (impartida en la Universidad Politécnica de Valencia). Ninguno de ellos indicó padecer de problemas de audición cuando se les preguntó; y todos ellos trabajaban en temas relacionados con el procesado de señal.

5.3.2. Estímulos utilizados.

Se diseñó un test donde se utilizaron ocho señales reales de escape grabadas con distintos esquemas de funcionamiento del motor: cuatro asociadas a un motor funcionando a un régimen de giro de 1000 r.p.m. y otras cuatro asociadas a un motor girando a 2000 r.p.m..

5.3.3. Igualación del Loudness de las señales.

Es bien conocida la influencia del Loudness en la percepción de sensaciones como la molestia causada por un sonido. Valores de Loudness altos pueden hacer que los jueces califiquen los sonidos como muy molestos y atribuyan, de forma errónea, esta molestia al Roughness dando valoraciones de Roughness altas a sonidos con valores altos de Loudness. Para evitar este efecto en los experimentos se han ajustado los niveles de todas las señales utilizadas a un nivel de 8 sones (equivalente a un tono de 1 kHz con un nivel de presión sonora de 70 dB).

5.3.4. Instrumentación utilizada en el experimento.

Las pruebas se han realizado utilizando un ordenador portátil con procesador Pentium Centrino y un software no comercial desarrollado por el Instituto de Telecomunicaciones y Aplicaciones Multimedia (iTEAM) [239] que presenta las señales a los jueces y registra las respuestas.

Los sonidos almacenados en el ordenador portátil se reprodujeron mediante una tarjeta de sonido de 24 bits (M-Audio FireWire 410) a la que se conectaron unos auriculares Sennheiser HD 600.

5.3.5. Procedimiento.

En cada uno de los test se le pidió al jurado que seleccionara la señal que tuviera el Roughness mayor (no permitiendo la opción de seleccionar “valores iguales”) y se analizaron los resultados según la metodología de test de comparación de parejas [230, 231, 240].

Al principio de cada sesión de test se entrenaba a los jueces para que identificaran la sensación de Roughness mediante un ejemplo que contenía 6 sonidos diferentes en los que se variaba el Roughness de forma creciente. El grupo de seis sonidos de Roughness creciente estaba formado por un tono de 1000 Hz modulado en amplitud con una frecuencia de 70 Hz y con un índice de modulación que valía of 0 %, 20 %, 30 %, 70 %, 90 % and 100 % en cada una de las seis señales. Estos sonidos se ordenaron en orden creciente de índice de modulación de forma que el aumento en la sensación de Roughness percibido fuera notable. Se indicó a los miembros del jurado que la vibración percibida al modular los sonidos anteriores recibe el nombre de Roughness, que cada uno de los sonidos escuchados tiene un Roughness superior al sonido anterior y que a continuación escucharían parejas de sonidos y deberían de identificar el sonido de la pareja que tuviera mayor Roughness. El propósito de estos seis sonidos fue únicamente el entrenamiento del jurado en la sensación de Roughness, y no se utilizaron como referencia en ninguno de los test posteriores.

5.3.6. Resultados del test.

Los resultados del test se muestran en la figura 5.4. Los valores negativos de Roughness que aparecen en la figura se deben a la utilización de test de parejas, que proporciona puntuaciones con media 0. Los resultados del test muestran una conclusión clara: que un jurado perciba un sonido con mayor Roughness que otro no significa que los métodos tradicionales de cálculo de Roughness le asignen un valor calculado mayor. Por esta razón debemos concluir diciendo que: “los métodos tradicionales de cálculo de Roughness no proporcionan buenas estimaciones de este parámetro cuando se aplican a señales complejas como las señales de motor”.

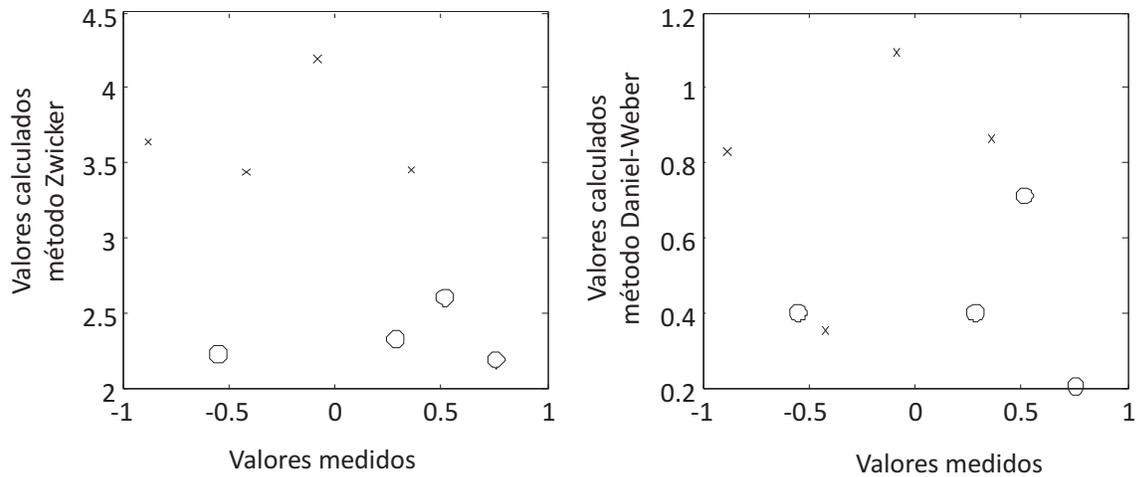


Figura 5.4: Las representaciones de la figura muestran en el eje de abscisas los valores de Roughness medidos en el test y en el eje de ordenadas los valores de Roughness calculados mediante el método de Zwicker (izquierda) y Daniel-Weber (derecha). Los círculos están asociados a señales de 1000 r.p.m y las aspas a señales de 2000 r.p.m.

5.4. Evaluación de parámetros psicoacústicos de las señales sintéticas de escape

Para estudiar las causas de la falta de exactitud de los métodos tradicionales de cálculo de Roughness puesta de manifiesto en la sección 5.3.6, así como en otros experimentos realizados con señales similares [20, 19, 22, 238], en esta sección se analizará el Roughness de señales de escape sintéticas. Esto nos permitirá además experimentar acerca de la capacidad del modelo propuesto en el capítulo 4 (y representado en la figura 4.25), en la generación de sonidos con parámetros psicoacústicos controlados.

A continuación se describe como se han desarrollado cuatro test de jurado para valorar las sensaciones subjetivas que producen sonidos sintéticos de 1000, 2000, 3000 y 4000 r.p.m.

5.4.1. Estímulos utilizados.

Se diseñaron cuatro test de jurado, uno para cada uno de las velocidades ensayadas: 1000, 2000, 3000 and 4000 r.p.m. En cada test se utilizaron 5 señales sintéticas diferentes generadas con el sintetizador mostrado en la figura 4.25.

Las variaciones en los parámetros psicoacústicos Loudness o Sharpness son fáciles de generar modificando la amplitud de la señal (para producir variaciones en el Loudness) o variando el contenido espectral del patrón (para producir variaciones en el Sharpness). Los experimentos se centran, por lo tanto, en la capacidad del modelo para generar señales

Señales	Frecuencia de Modulación (Hz)	Índice de Modulación (%)
1	0	0
2	50	20
3	50	40
4	70	20
5	70	40

Tabla 5.1: Índice de modulación y frecuencia de modulación utilizados en la ecuación (4.16) para generar las amplitudes a_k de las cinco señales involucradas en los test de 1000 r.p.m., para obtener ruidos sintéticos que produzcan diferentes sensaciones psicoacústicas.

con Roughness y Fluctuation Strength variables.

Para construir una señal sintética se ha utilizado un patrón extraído de una señal real con el mismo régimen de giro que la señal a sintetizar. Y se ha utilizado el cepstrum complejo para obtener el patrón ($p(t)$) de la señal.

Para generar la señal sintética, se combina este patrón $p(t)$ con una señal tren de deltas $c(t)$ según la expresión (4.15). Esta señal $c(t)$ tiene la forma mostrada en (4.11) y para construirla se utilizan valores de a_k según la expresión (4.16), tomando distintos índices y frecuencias de modulación, como se indica en la tabla 5.1.

La tabla 5.1 presenta, a modo de ejemplo, los valores del índice de modulación (m) y la frecuencia de modulación (f_m) utilizados en la generación de señales sintéticas evaluadas en el test asociado al régimen de giro del motor de 1000 r.p.m. Los índices de modulación utilizados para generar las señales sintéticas de 2000, 3000 y 4000 r.p.m. fueron los mismos que para 1000 r.p.m. (0 %, 20 % y 40 %), pero las frecuencias de modulación fueron distintas en cada caso. Esto se debe a que la máxima frecuencia de modulación utilizable depende del régimen de giro de la señal [228].

Según Zwicker [3], el Roughness de una señal modulada en AM de banda estrecha aumenta con el índice de modulación, alcanzando su máximo para valores de frecuencia de modulación aproximados a 70 Hz. Sin embargo, para una modulación de banda ancha, la dependencia del Roughness con la frecuencia y el índice de modulación no es tan sencilla.

El análisis mediante cepstrum nos permite controlar el Roughness con un solo parámetro a_k , dependiendo del índice y la frecuencia de modulación según la ecuación (4.16). En este sentido, el objetivo principal de los test asociados a este experimento ha sido la validación del método para describir el Roughness que percibe un grupo de jueces que representen a un consumidor del producto.

5.4.2. Procedimiento.

En estos experimentos han intervenido ocho jueces con edades comprendidas entre los 20 y los 40 años. Los jueces fueron reclutados de entre los alumnos y profesores de la titulación de Ingeniero Técnico de Telecomunicación, especialidad Telemática que se imparte en la Escuela Politécnica Superior de Alcoy. Ninguno de ellos indicó padecer de problemas de audición cuando se les preguntó; y tampoco ninguno de ellos trabajaba en temas relacionados con los motores ni la acústica, por lo que desde el punto de vista técnico se les puede considerar como un jurado de usuarios potenciales del producto.

Para evitar que diferencias en el Loudness de las señales afectara a la valoración del Roughness por parte de los jueces, se modificó la amplitud de todos los sonidos para igualar su Loudness a 8 sonos mediante como se indicó en la sección 5.3.3. La instrumentación utilizada en los experimentos coincide con la descrita en la sección 5.3.4.

En cada uno de los test, se le pidió al jurado que seleccionara la señal que tuviera el Roughness mayor (no permitiendo la opción de seleccionar “valores iguales”) y se analizaron los resultados según la metodología de test de comparación de parejas [230, 231, 240].

Al principio de cada sesión de test se entrenaba a los jueces para que identificaran la sensación de Roughness mediante el procedimiento descrito en la sección 5.3.5.

A continuación se realizaron cuatro experimentos (uno para cada régimen de giro del motor) con el fin de evaluar el Roughness de 5 señales sintéticas. En cada experimento se presentaron a los jueces todas las posibles parejas que se pueden formar con las señales a analizar (10 comparaciones para 5 señales distintas), junto con otras parejas redundantes de control (con el fin de evaluar la calidad de las respuestas de cada uno de los jueces). Para evaluar la repetitividad de un juez (responder de la misma forma que la primera vez cuando se le repite una pregunta [230, 241]), en cada test se añadieron a las 10 comparaciones anteriormente citadas 2 comparaciones más, correspondientes a la repetición de dos parejas ya evaluadas cambiadas de orden. Con el fin de evitar sesgo en la evaluación de las señales, el orden en el que se presentaban al jurado era aleatorio. Como ejemplo del contenido de un experimento, la Tabla 5.2 contiene las comparaciones que se presentaban a los jueces en el test asociado a la frecuencia de giro del motor de 1000 r.p.m.

Otro parámetro utilizado para medir la calidad de las respuestas proporcionadas por los jueces fue la consistencia [230, 241]. Si al comparar las señales A y B, un juez decide que B tiene un valor de Roughness superior y al comparar las señales B y C decide que es la señal C la que presenta un valor superior de Roughness, para ese juez las señales ordenadas de menor a mayor Roughness serían primero A a continuación B y por último C. Por lo tanto, si se pide que compare las señales A con C debería seleccionar la señal C como señal con más Roughness, si no lo hiciera así se diría que el juez no es consistente.

Tras la realización de los test se analizaron la repetitividad y la consistencia jueces y se desecharon aquellos test en los que los jueces tuvieran puntuaciones inferiores al 80% en alguno de estos dos parámetros. Como consecuencia, al analizar los resultados, no se tuvieron en cuenta las respuestas proporcionadas por uno de los jueces en el test 1, y se

Comparación	Primera señal	Segunda señal
1	1	2
2	3	4
3	4	5
4	3	5
5	2	5
6	2	3
7*	2	1
8*	4	3
9	1	4
10	1	5
11	1	3
12	2	4

Tabla 5.2: Ordenación de las señales a analizar en el test asociado al régimen de giro del motor de 1000 r.p.m. Las parejas 7 y 8 se han marcado porque son las mismas parejas etiquetadas como 1 y 2 ordenadas de forma inversa. La comparación de las evaluaciones de los jueces para estas 4 parejas proporcionan un indicador de la calidad del juez.

descartaron los resultados de dos jueces (distintos al anterior) en el test 4.

5.4.3. Resultados del test de jurado para ruido sintético de 1000 r.p.m.

En un test de parejas se asigna una puntuación final a las señales analizadas de forma creciente en relación al parámetro analizado, es decir que para este caso en particular cuanto mayor sea la puntuación mayor será la percepción de Roughness el jurado tiene respecto a la señal. La figura 5.5 representa las puntuaciones obtenidas por las distintas señales analizadas en el test relativo a 1000 r.p.m. (donde los parámetros de las señales utilizados se representan en la Tabla 5.1). Como la suma de las puntuaciones proporcionadas por el método del test de parejas es siempre cero, algunas puntuaciones son positivas mientras que otras toman valores negativos. Del análisis de los resultados mostrados en la figura 5.5 se desprenden las siguientes conclusiones:

- a) Comparación de las señales 1, 2 y 3: estas señales corresponden a frecuencias de modulación de 50 Hz con índices de modulación de 0% (para la señal 1 donde no existe modulación y se usa como referencia) 20% (para la señal 2) y 40% (para la señal 3). Al comparar las puntuaciones de la señales 2 con la señal 3 se observa que la señal 2 es percibida con mayor Roughness que la señal 3, en contra de lo que cabría esperar puesto que la señal 2 tiene un menor índice de modulación.
- b) Comparación de las señales 1, 4 y 5: estas señales corresponden a señales con fre-

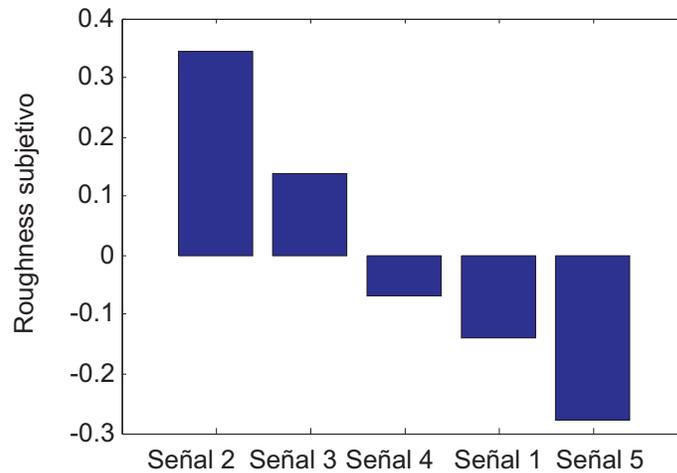


Figura 5.5: Puntuaciones obtenidas en el test de jurado realizado con las señales sintéticas de 1000 descritas en la tabla 5.1.

cuencia de modulación 70 Hz con índices de modulación 0 % (la misma señal de referencia comentada en el punto anterior), 20 % and 40 % respectivamente. En este caso volvemos a encontrar resultados opuestos a los esperados puesto que la señal 5 (la de mayor índice de modulación) es percibida como la señal de menor Roughness del grupo.

- c) Comparación entre las señales 2 y 4 y entre las señales 3 y 5. Las señales 2 y 4 tienen el mismo índice de modulación (20 %) pero difieren en la frecuencia de modulación. Según la teoría una frecuencia de modulación de 70 Hz debería producir mayor sensación de Roughness que una frecuencia de modulación de 50 Hz por lo que la puntuación obtenida para la Señal 4 debería de ser mayor que la obtenida para la Señal 2. Y con el mismo argumento, la puntuación para la señal 5 debería ser mayor que obtenida para la señal 3, pero los valores obtenidos en el experimento indican precisamente lo contrario.
- d) El rango de puntuaciones obtenido para las cinco señales es muy reducido (varía tan sólo entre el intervalo comprendido entre $-0,3$ y $0,4$). En un test de parejas un intervalo de puntuaciones reducido indica que las señales utilizadas son muy similares (respecto a la característica analizada). De hecho, algunos miembros del jurado indicaron que les resultaba complicado identificar la sensación de Roughness en las señales analizadas. Aunque los valores empleados como índices y frecuencias de modulación son lo bastante diferentes como para producir un variado margen de valores de Roughness.

Una explicación a los resultados contradictorios obtenidos en este test la podemos encontrar si realizamos un análisis más profundo del modelo propuesto en la figura 4.25. El tren de deltas $c(t)$ utilizado en la formación de la señal tiene una frecuencia fundamental que viene dado por (4.5), y para un régimen de giro del motor de 1000 r.p.m. vale:

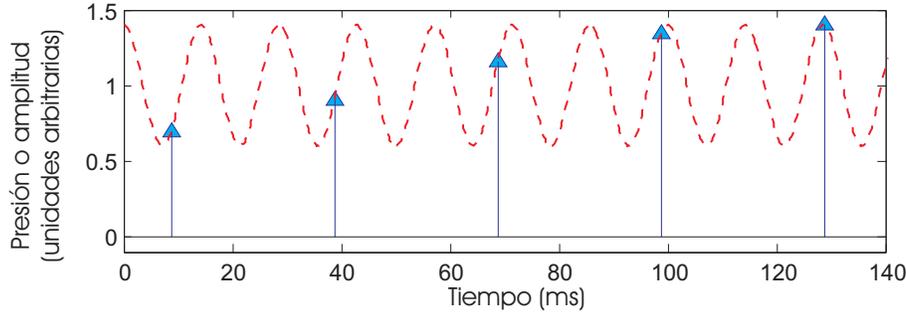


Figura 5.6: Tren de deltas $c(t)$ utilizado en la generación de la señal sintética de 1000 r.p.m. modulada en amplitud con frecuencia de modulación de 70 Hz e índice de modulación 40 % (Señal 5 de la Tabla 5.1).

$$f_0 = \frac{1000}{120} \times 4 = 33,33\text{Hz}, \quad (5.29)$$

cuya inversa indica que el periodo fundamental de este tren de deltas es de 30 milisegundos (o lo que es lo mismo $N = 1440$ muestras para una frecuencia de muestreo de 48 kHz). Es decir, se genera un impulso cada 30 ms cuya amplitud viene dada por la ecuación (4.16).

Esta expresión muestra que los valores de a_k provienen del muestreo, con frecuencia de muestreo f_0 , de una señal moduladora de frecuencia f_m . Pero este valor de f_0 no cumple el teorema del muestreo de Nyquist para valores de f_m utilizados para sintetizar las señales de este experimento, ya que con una frecuencia de muestreo f_0 sólo se pueden muestrear correctamente señales que tengan una frecuencia inferior a $f_0/2$ [136]. El submuestreo de una señal hace que la señal muestreada se comporte como el “alias” de ésta cuya frecuencia esté comprendida en el intervalo que va desde 0 hasta $f_0/2$, entendiéndose como “alias” de la frecuencia f_m a las frecuencias:

$$f_m^{\text{alias}} = \pm f_m + l \cdot f_0, \quad \text{con } l = 0, \pm 1, \pm 2, \dots \quad (5.30)$$

Volviendo al caso particular de las señales sintéticas de nuestro experimento asociado a un régimen de giro de 1000 r.p.m., las amplitudes a_k de los impulsos se han generado muestreando un coseno de frecuencia $f_m = 70$ Hz con una frecuencia de muestreo de $f_0 = 33,33$ Hz mediante la ecuación (4.16). Debido al efecto del submuestreo la modulación efectiva es mucho menor de la esperada, ya que, según (5.30), la frecuencia moduladora de 70 Hz al ser muestreada mediante una frecuencia de $f_0 = 33,33$ Hz (inferior a la de Nyquist), se comporta como su “alias” de frecuencia $f_m^{\text{alias}} = f_m - 2f_0 = 3,33$ Hz. La Figura 5.6 ilustra este efecto de submuestreo mostrando el tren de deltas generado mediante (4.16) con $f_m = 70$ Hz y $f_0 = 33,33$ Hz: las variaciones de amplitud son tan lentas que coinciden con las que produciría un coseno de 3,33 Hz muestreado con la misma frecuencia de muestreo f_0 .

De forma similar, las señales utilizadas en el experimento en las que se empleó una frecuencia moduladora de $f_m = 50$ Hz, generan trenes de impulsos que se ajustan al “alias” de frecuencia $f_m^{\text{alias}} = f_m - f_0 = 16,66$ Hz.

A la vista de este análisis podemos concluir que los resultados del experimento, mostrados en la Figura 5.5, sí se ajustan a la teoría en cuanto que las Señales 2 y 3 se perciben como señales moduladas con una frecuencia de 16,66 Hz (no de 50 Hz) mientras que las Señales 4 y 5 se perciben como señales moduladas con una frecuencia de 3,33 Hz (no de 70 Hz) por lo que es lógico que las primeras tengan una puntuación mayor que las últimas al evaluar su Roughness.

También se puede explicar, a partir de este análisis, el hecho de que todas las señales analizadas tengan valores muy similares (Figura 5.5) y que los jueces mostraran dificultad para identificar la sensación de Roughness puesto que con frecuencias moduladoras de 3,33 Hz y 16,66 Hz se perciben sensaciones de Fluctuación más que de Roughness.

Además, cuando se trabaja con un motor funcionando a 1000 r.p.m. (que genera señales cuasiperiódicas con frecuencia fundamental de $f_0 = 33,33$) es imposible generar frecuencias de modulación superiores a 18 Hz. Por lo tanto no se pueden generar señales con diferencias significativas en su Roughness; y por esta razón los resultados del test asociados a 1000 r.p.m se han descartado en las conclusiones globales del experimento.

5.4.4. Resultados del test de jurado para ruidos sintéticos de 2000 r.p.m., 3000 r.p.m. y 4000 r.p.m.

Como ya se indicó en la sección 5.4.3, el máximo valor de frecuencia moduladora que se puede utilizar en la síntesis de ruidos de motor es la mitad de la frecuencia asociada al régimen de giro del motor. Es decir, que para un régimen de giro de 2000 r.p.m. la máxima frecuencia de modulación que se puede conseguir es de 33,3 Hz, para 3000 r.p.m. de 50 Hz y para 4000 r.p.m. de 66,6 Hz.

Debido a estas restricciones, las señales sintéticas utilizadas para los test se han generado utilizando dos frecuencias moduladoras para cada régimen de giro del motor:

- La máxima frecuencia de modulación permitida para el régimen de giro.
- El 70 % de la frecuencia máxima anterior.

La Tabla 5.3 resume los parámetros utilizados en la generación de las señales sintéticas utilizadas en los experimentos asociados a 2000, 3000 y 4000 r.p.m.

Antes de entrar a analizar los resultados proporcionados por el jurado en los test de parejas, hay que indicar que a pesar de haber obtenido 15 puntuaciones distintas (una para cada señal), estas puntuaciones sólo tienen sentido en el contexto del test en el que fueron obtenidas. Es decir que solamente se pueden comparar entre sí las puntuaciones obtenidas en el test de 2000 r.p.m. o las puntuaciones obtenidas en el test de 3000 r.p.m., pero no tiene sentido comparar la valoración de una señal de 2000 r.p.m. con de la valoración de

		2000 r.p.m.	3000 r.p.m.	4000 r.p.m.
Señales	Índice Mod. (%)	Frecuencia Mod. (Hz)	Frecuencia Mod. (Hz)	Frecuencia Mod. (Hz)
1	0	0	0	0
2	20	23,33	35	46,66
3	40	23,33	35	46,66
4	20	33,33	50	66,66
5	40	33,33	50	66,66

Tabla 5.3: Índices y frecuencias de modulación utilizados en la ecuación (4.16) para generar las amplitudes a_k de las cinco señales utilizadas en los test de 2000, 3000 and 4000 r.p.m.

otra señal de 3000 r.p.m puesto que dichas valoraciones fueron obtenidas en diferentes contextos (diferentes test).

La Figura 5.7 muestra (de forma superpuesta) los resultados de los tres test asociados a 2000, 3000 y 4000 r.p.m. La Figura 5.7 (a) ilustra la dependencia del Roughness respecto al índice de modulación para frecuencias de modulación correspondientes al 70 % del máximo permitido para cada velocidad de giro del motor. Los resultados indican que un incremento del índice de modulación produce un incremento del Roughness percibido y la dependencia de estos dos parámetros es casi lineal. El amplio rango de valores en el que varían las puntuaciones obtenidas para las señales (desde -4 hasta 3) indican una unanimidad casi total entre los miembros del jurado a la hora de calificar el Roughness de las señales analizadas, así como una clara distinción entre las señales con mayor y menor Roughness.

De forma similar en el apartado (b) de la Figura 5.7 se representa la dependencia del Roughness con respecto al índice de modulación, pero ahora se analizan señales moduladas con una frecuencia de modulación del 100 % de la máxima frecuencia de modulación posible para el régimen de giro. Los resultados son muy similares (prácticamente idénticos) a los comentados en el párrafo anterior. Por esto se puede afirmar que, independientemente de la frecuencia de modulación y del régimen de giro del motor, el aumento del índice de modulación produce un aumento en el Roughness existiendo una dependencia lineal entre ambos parámetros.

Para analizar la dependencia del Roughness con la frecuencia de modulación, la Figura 5.8 representa la valoración que los jueces han asignado al Roughness de señales con el mismo índice de modulación pero diferentes frecuencias de modulación (ver la tabla 5.3 para conocer los valores reales de estas frecuencias). En la gráfica (a) de la Figura 5.8 se representa la variación del Roughness respecto a la frecuencia de modulación cuando se utiliza un índice de modulación del 20 %. Aquí se puede observar que esta dependencia es mayor para 2000 r.p.m. que para velocidades mayores. Sin embargo, en todo el test se observa que a pesar de que existe un incremento del Roughness a medida que aumenta la frecuencia de modulación, este aumento es muy débil. Los bajos valores obtenidos para las señales que se muestran en la gráfica (a) respecto a los que aparecen en la gráfica (b)

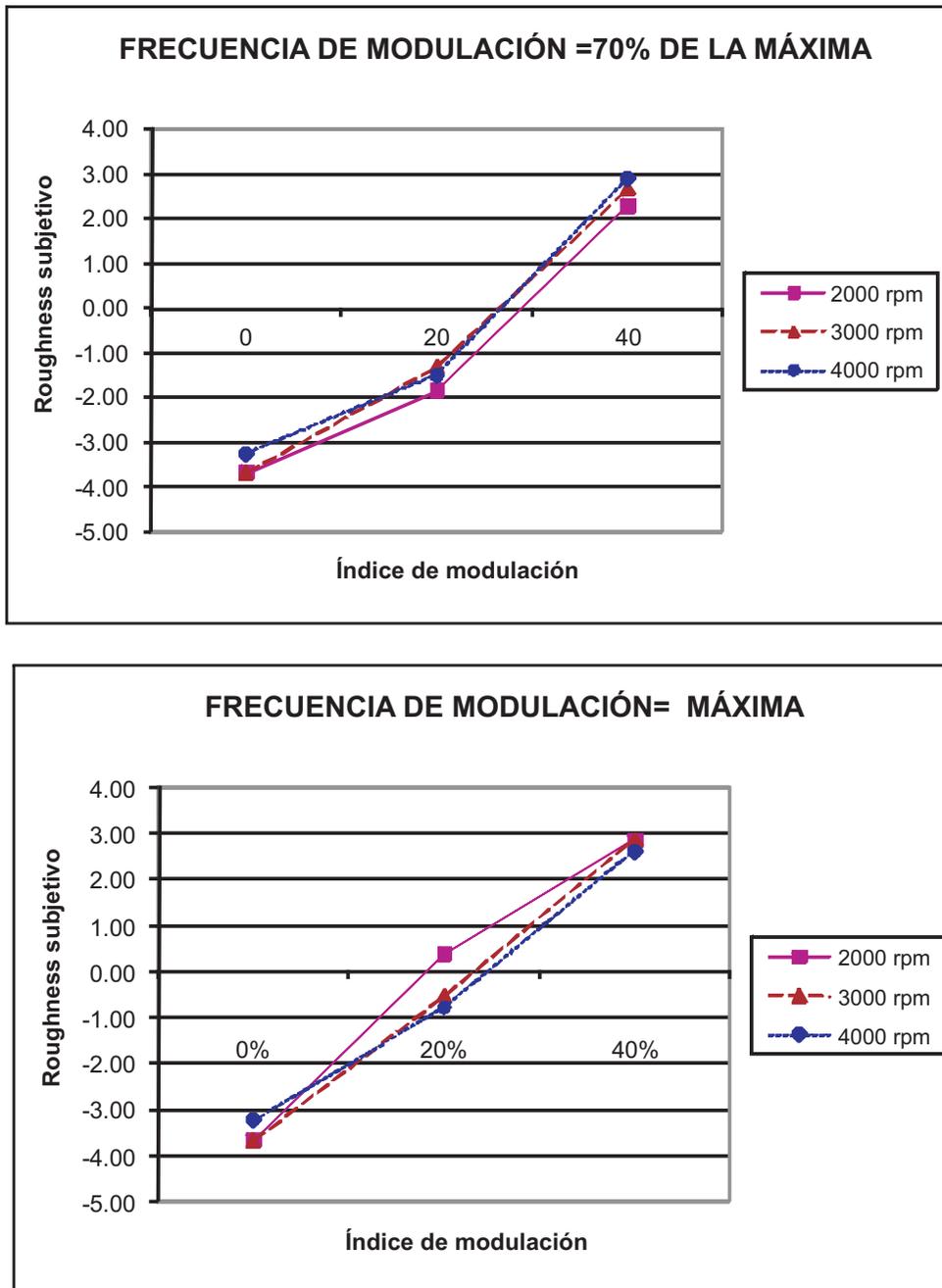


Figura 5.7: Roughness subjetivo en función del índice de modulación (0 %, 20 % y 40 %) para los test de 2000, 3000 y 4000 r.p.m. En el gráfico (a) se ha utilizado la máxima frecuencia de modulación posible para cada régimen de giro mientras que en (b) se ha utilizado una frecuencia de modulación igual al 70 % de dicho máximo.

de la Figura 5.8 se pueden explicar por el pequeño índice de modulación utilizado (20 %).

Por último, si analizamos la gráfica (b) de la Figura 5.8 (que representa la relación entre Roughness y frecuencia de modulación cuando el índice de modulación utilizado es de 40 %), comprobamos que las gráficas presentan tendencias similares a las anteriores (donde se utilizó un índice de modulación del 20 %) excepto para el ruido de 4000 r.p.m. En este caso particular, un incremento de la frecuencia de modulación produce una reducción del Roughness percibido.

En las dos gráficas que aparecen en la Figura 5.8 se observa que, aunque existe variación del Roughness al variar la frecuencia de modulación, la dependencia entre estos dos parámetros no es demasiado fuerte.

A pesar de que los test de jurado llevados a cabo con estas señales sintéticas no son exhaustivos en cuanto al número de velocidades de motor, frecuencias de modulación e índices de modulación utilizados, sí que son representativas del comportamiento del modelo propuesto en la Figura 4.25. Los resultados de la Figura 5.7 confirman que el modelo basado en el análisis mediante cepstrum es capaz de generar ruidos sintéticos con el Roughness controlado. Además, el modelo propuesto identifica el índice de modulación de la ecuación (4.16) como el parámetro que más influye en el Roughness percibido.

Como resumen, se pueden extraer dos conclusiones importantes de las Figuras 5.7 y 5.8:

- El modelo propuesto ha sido validado como un buen método a la hora de producir ruidos sintéticos de motor con un valor de Roughness controlado. Además, estos resultados permiten considerar parámetros del cepstrum complejo de las señales analizadas para mejorar los métodos de cálculo de Roughness cuando se trabaja con señales de banda ancha, en lugar de utilizar únicamente parámetros extraídos de modelos de banda estrecha.
- El Roughness de los sonidos provenientes de un motor tienen una dependencia mayor con el índice de modulación que con la frecuencia de modulación (por comparación de los distintos rangos que tienen las puntuaciones obtenidas en las Figuras 5.7 y 5.8).

5.5. Nuevos patrones sintéticos.

El modelo de señal cuasiperiódica descrito en la sección 4.4 consiste en la convolución de dos señales según se muestra en la ecuación (4.15): un patrón $p(t)$ y un tren de deltas modulado $c(t)$. Las secciones anteriores se han ocupado de analizar la influencia de $c(t)$ en las señales, en las referencias [229, 238, 228] se estudia la influencia de los valores a_k y T_k en el Roughness de señales sintéticas formadas a partir de trenes de deltas $c(t)$, generadas mediante la ecuación (4.11).

En esta sección y en la sección 5.6 se explora la influencia del patrón $p(t)$ de la ecuación (4.15) en el modelo. En la sección 2.6.6 se expuso como depende el Roughness

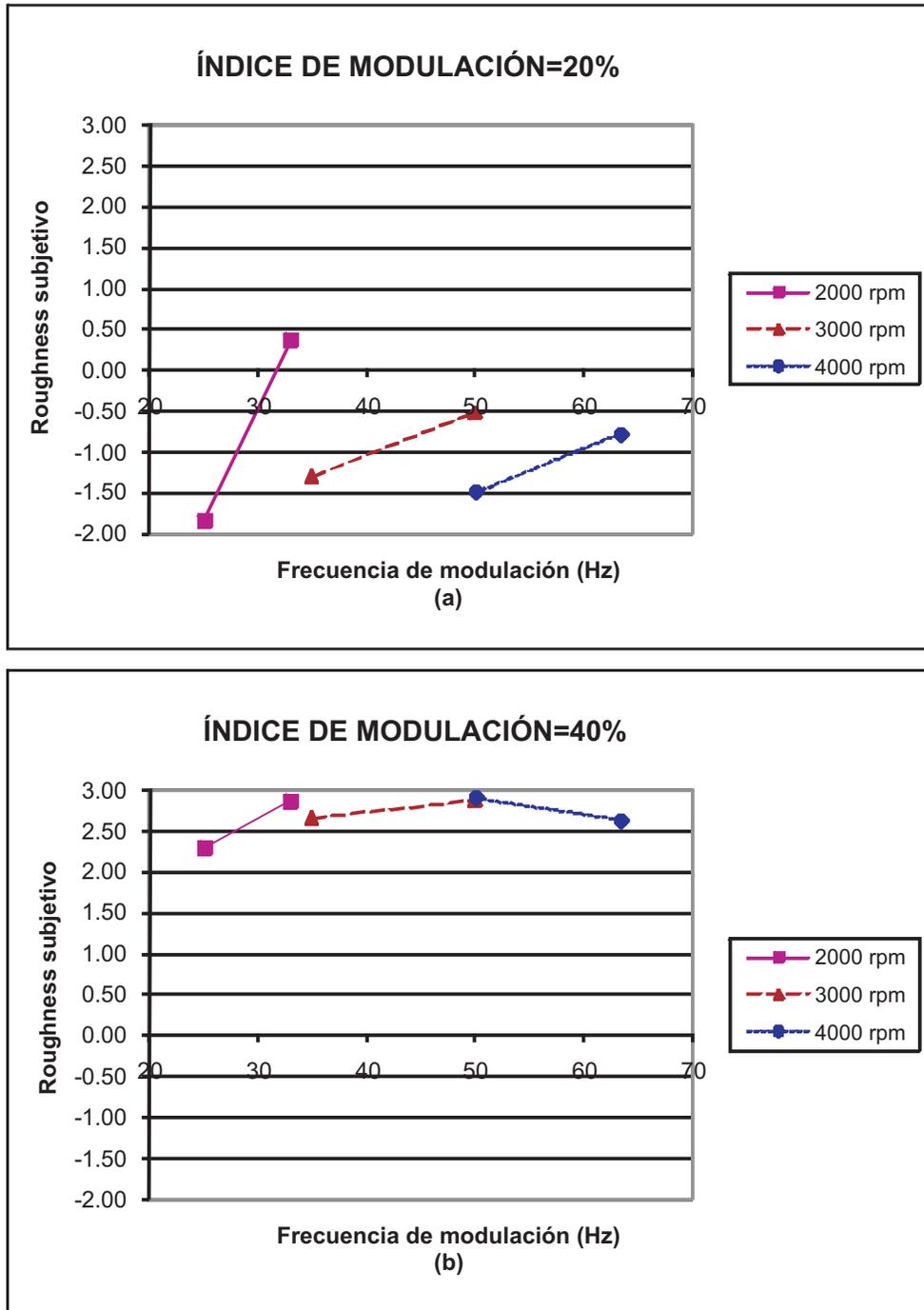


Figura 5.8: Roughness subjetivo en función de la frecuencia de modulación (en la Tabla 5.3 aparecen los valores exactos). En la parte superior (a) se ha utilizado un índice de modulación del 20%, y en la parte inferior (b) se ha utilizado un índice de modulación del 40%.

Total del Roughness Específico y la correlación entre las bandas adyacentes. Para estudiar la importancia de la correlación entre bandas, en esta sección vamos a generar distintas señales con el mismo Roughness Específico en todas las bandas críticas pero con diferentes correlaciones entre bandas.

Para controlar el número de bandas críticas excitadas podríamos utilizar como señal patrón ruido blanco filtrado, pero esta opción no permite controlar el Roughness Específico en cada banda puesto que sería aleatorio. Para resolver este problema se podrían utilizar las respuestas impulsionales de filtros paso banda como señal patrón, pero esta opción no permite controlar la duración de la señal patrón o la correlación entre bandas adyacentes. Por esta razón el patrón $p(t)$ seleccionado ha sido la señal chirp. Esta señal, diseñada de forma adecuada, permite el control del ancho de banda de excitación y de la correlación entre bandas. Además, como la señal chirp puede excitar diferentes bandas durante instantes de tiempo muy cortos, permite generar valores máximos de profundidad de modulación en cada banda. En esta sección, indicaremos primero cómo se ha diseñado una señal chirp básica (que hemos llamado la chirp de barrido uniforme en la escala Bark) y después indicaremos cómo es posible combinar varias chirp para generar nuevas señales patrón que permitan modificar la correlación entre bandas.

La experimentación con patrones distintos a los extraídos de señales asociadas a los vehículos a motor permitirá extrapolar las conclusiones obtenidas a una gama más amplia de señales cuasiperiódicas, no sólo a las de motor.

5.5.1. La señal chirp de barrido uniforme en la escala Bark.

Como los procedimientos habitualmente empleados en el cálculo del Roughness descomponen las señales en fragmentos con la misma anchura espectral en la escala bark, hemos diseñado una chirp cuya frecuencia instantánea varíe uniformemente en la escala bark. Por lo tanto, la frecuencia instantánea en la escala bark de este tipo de señales chirp debe de ser:

$$z(t) = kt, \quad (5.31)$$

Siendo k la velocidad de barrido. Para diseñar esta chirp partimos de la relación entre la escala bark y la escala de frecuencia en hercios propuesta por H. Traunmüller [68]:

$$z(Bark) = \frac{26,81}{1 + \frac{1960}{f(Hz)}} - 0,53 \quad (5.32)$$

La ecuación (5.32) nos permite escribir la relación inversa:

$$f(Hz) = \frac{m \cdot z(Bark) + n}{a - z(Bark)} \quad (5.33)$$

Siendo $m = 1960$, $n = 1038,8$, $a = 26,28$.

Para construir una señal cuya frecuencia instantánea varíe según la ecuación 5.31, la frecuencia instantánea en la escala herciana $f(t) = f(z(t))$ debe corresponder a la sustitución de la ecuación (5.31) en la ecuación (5.33). Por lo tanto la fase instantánea corresponderá a la integral:

$$\theta(t) = 2\pi \int f(z(t)) dt = 2\pi \int f(z(t)) \frac{dt}{dz} dz = 2\pi \int f(z(t)) \frac{1}{dz/dt} dz \quad (5.34)$$

De la ecuación (5.31) obtenemos:

$$\frac{dz}{dt} = k \quad (5.35)$$

y,

$$\theta(t) = \frac{2\pi}{k} \{-mz - (ma + n) \cdot \log(a - z)\} + \theta_0 \quad (5.36)$$

Por lo que la chirp de barrido uniforme en la escala bark tiene la forma:

$$x(t) = \sin \left(2\pi \left\{ -mt - \frac{ma + n}{k} \log(a - kt) \right\} + \theta_0 \right) \quad (5.37)$$

La ecuación (5.32) se puede considerar exacta en el intervalo $2 \leq z \leq 20,1$, pero fuera de estos valores se obtiene mayor precisión [68] modificando los parámetros de la ecuación (5.32) de la siguiente forma:

$$f = \frac{m \cdot z + n}{a - z} \begin{cases} m = 1960, n = 294.98, a = 22.038 & \text{si } z_{ap} < 2 \\ m = 1960, n = 1038.8, a = 26.28 & \text{si } 2 \leq z_{ap} \leq 20.1 \\ m = 1960, n = 9934.456, a = 27.6396 & \text{si } z_{ap} > 20.1 \end{cases} \quad (5.38)$$

Los parámetros necesarios para diseñar una chirp de estas características son: duración (D), frecuencia inicial (f_0) y frecuencia final (f_1).

Utilizando estos parámetros y inventanando la ecuación (5.32) podemos generar una chirp con frecuencia inicial (f_0) y frecuencia final (f_1). En estos experimentos se han utilizado ventanas de Tukey [242] con tiempos de inicio y fin del 5%.

La figura 5.9 representa una chirp con $f_0 = 20Hz$, $f_1 = 2000Hz$ y duración $D = 30ms$, inventanada con una ventana de Tukey [242] con tiempos de inicio y fin del 5%. Como la diferencia entre la frecuencia inicial f_0 y la frecuencia final f_1 es de 12.9 barks y su duración D son 30ms, la velocidad de barrido en (Eq. (5.31)) es $k = 0,43barks/ms$.

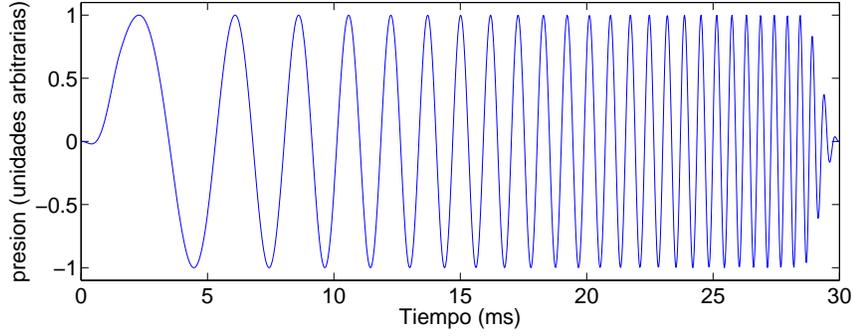


Figura 5.9: Chirp con $f_0 = 20\text{Hz}$, $f_1 = 2000\text{Hz}$ y duración $D = 30\text{ms}$.

5.5.2. Combinación de varias chirp.

En cada instante de tiempo, la señal patrón diseñada únicamente produce excitación en una frecuencia del espectro y, por lo tanto, en un solo filtro de banda crítica. Para producir la excitación simultánea de una zona más amplia del espectro utilizaremos como señal patrón la suma de dos señales chirp retrasadas un tiempo l :

$$p(t) = x(t) + x(t - l) \quad (5.39)$$

Los parámetros necesarios para diseñar un patrón de estas características son: duración de las chirps (D), retardo entre las dos chirps (l), frecuencia inicial (f_0) y frecuencia final (f_1).

La figura 5.10 representa la señal generada utilizando dos copias de la chirp representadas en la figura 5.9 retrasadas $l = 10\text{ms}$ mientras que la figura 5.11 es el espectrograma de la señal representada en la figura 5.10. Si seleccionamos un rango de frecuencias en la figura 5.9 y analizamos su evolución temporal (horizontal) se puede apreciar este retardo de 10ms . Si analizamos la composición frecuencial de la señal en un instante de tiempo (vertical) podemos observar dos componentes frecuenciales. La separación entre estas dos componentes (desplazamiento bark) es:

$$\Delta z = k \cdot l \quad (5.40)$$

En este caso el desplazamiento bark entre las chirps es de 4,3 barks.

En este ejemplo se ha utilizado un valor elevado del parámetro l con el fin de generar representaciones claras: se puede apreciar visualmente la separación entre dos frecuencias simultáneas. En los experimentos realizados en la sección 5.6 se han utilizado valores inferiores del parámetro l , que, en lugar de producir representaciones con dos líneas como las mostradas en la figura 5.11, producen una sola línea cuyo grosor está controlado por el parámetro l .

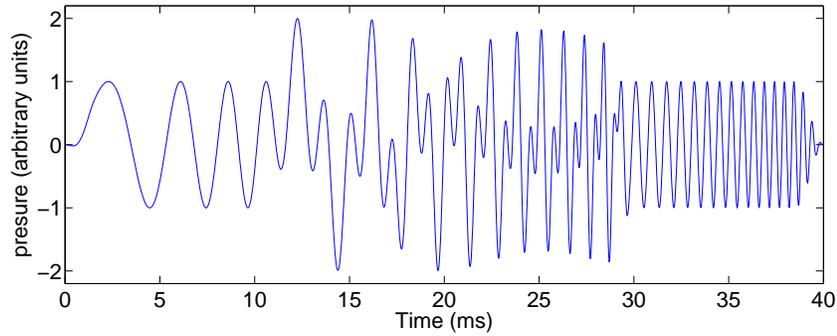


Figura 5.10: Suma de dos chirp básicas ($f_0 = 20\text{Hz}$, $f_1 = 2000\text{Hz}$ y $D = 30\text{ms}$) retrasadas $l = 10\text{ms}$.

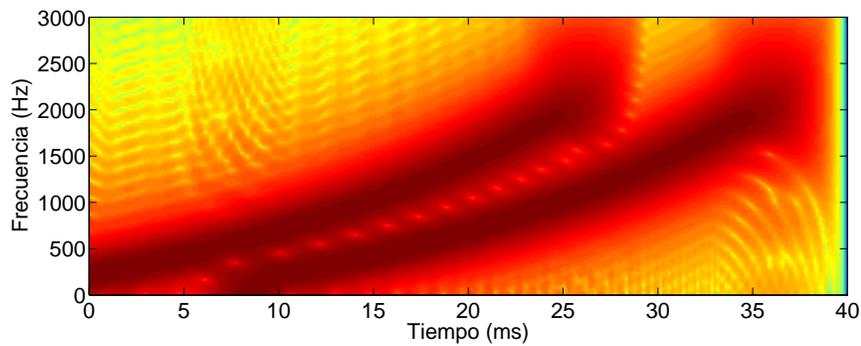


Figura 5.11: Espectrograma de la señal representada en la Figura 5.10

5.6. Evaluación de parámetros psicoacústicos de señales sintéticas con patrón chirp.

Utilizando los patrones descritos en la sección 5.5, hemos construido varias señales sintéticas que representan el ruido de combustión asociado a motores que giran a un con un régimen de giro de 3000 r.p.m. Esta velocidad está asociada a un periodo de repetición $T_k = 10ms$ en la ecuación (4.11).

Cada experimento consiste en la estimación del Roughness de varias señales utilizando el método del test de comparación de parejas [231, 230, 240]. En cada uno de los test se le pidió al jurado que seleccionara la señal que tuviera el Roughness mayor (no permitiendo la opción de seleccionar “valores iguales”). Los jurados estuvieron compuestos por un número de jueces comprendido entre 14 y 16 con capacidad auditiva normal. Tras la realización de los test se analizaron la repetitividad y la consistencia jueces y se desecharon aquellos test en los que los jueces tuvieran puntuaciones inferiores al 80 % en alguno de estos dos parámetros.

Al principio de cada sesión de test se entrenaba a los jueces para que identificaran la sensación de Roughness mediante el procedimiento descrito en la sección 5.3.5. Para evitar que diferencias en el Loudness de las señales afectara a la valoración del Roughness por parte de los jueces, se modificó la amplitud de todos los sonidos para igualar su Loudness a 8 sonos mediante como se indicó en la sección 5.3.3. La instrumentación utilizada en los experimentos coincide con la descrita en la sección 5.3.4.

5.6.1. Dependencia de la forma del patrón (una chirp).

El objetivo de este experimento es validar las conclusiones obtenidas en la sección 4.4.3 con otro patrón diferente al obtenido a partir de una señal real de motor. El patrón elegido ha sido la señal chirp.

Para formar las señales sintéticas de 3000 r.p.m. se ha partido de un tren de deltas con periodo de repetición de 10 ms y se ha utilizando un patrón de 20 ms, por lo que las repeticiones tendrán un solapamiento temporal de un 50 %. Para formar la señal patrón se han utilizado chirps simples con frecuencia inicial $f_0 = 20 Hz$, $f_1 = 3000 Hz$ con el fin de generar ruidos con un margen de frecuencias similares a los producidos por un motor.

El tren de deltas periódicas ha sido modulado en amplitud con las frecuencias de modulación $f_{mod} = 0 Hz$, $35 Hz$ y $50 Hz$ (lo que supone 0 % 70 % y 100 % de la máxima frecuencia de modulación que admiten las señales de 3000 r.p.m. según se indicó en las secciones 5.4.4 y 5.4.3) en cuanto al índice de modulación se han utilizado los valores 0 %, 20 % y 40 %.

Aunque en principio se trata de 9 señales, todas aquellas con frecuencia de modulación 0 ó índice de modulación 0 corresponden a la misma señal (señal sin modular) y esto supone un total de 5 señales tal y como se muestra en la tabla 5.4. Para comparar todas las señales entre si se necesitan 10 comparaciones ($5 \times 4 / 2$); y añadiendo 2 para control de

frecuencia de modulación	Índice de modulación		
	0 %	20 %	40 %
0 %	X		
35 %		X	X
50 %		X	X

Tabla 5.4: Combinaciones utilizadas de frecuencia de modulación e índice de modulación.

repetitividad se quedan 12 señales.

Los resultados del test de parejas son los que se muestran en la figura 5.12, y de su análisis se pueden extraer las siguientes conclusiones:

- El Roughness aumenta al aumentar el índice de modulación: tanto los resultados obtenidos para las señales sintéticas utilizando una frecuencia de modulación de 35Hz como para las asociadas a la frecuencia de modulación de 50Hz indican un Roughness creciente a medida que aumenta el índice de modulación. Estos resultados coinciden con los mostrados en la figura 5.7 casi de forma exacta, no sólo en la tendencia creciente sino también en los valores obtenidos: la línea continua de la figura 5.12 coincide con la gráfica superior de la figura 5.7 (ambas asociadas a frecuencia de modulación de 35 Hz), mientras que la línea discontinua de la figura 5.12 coincide con la gráfica inferior de la figura 5.7 (ambas asociadas a frecuencia de modulación de 50 Hz).
- El Roughness aumenta al aumentar la frecuencia de modulación: el Roughness aumenta ligeramente al aumentar el índice de modulación cuando se trabaja con un índice de modulación del 40 % mientras que el aumento es mayor cuando se trabaja con un índice de modulación del 20 %. Y este resultado también coincide con el observado en 5.7
- Por tanto, se puede afirmar que la frecuencia de modulación usada en $c(t)$ es un parámetro determinante en la percepción del Roughness, independientemente de la forma de $p(t)$.

5.6.2. Dependencia de la duración del patrón (dos chirps superpuestas).

Para formar las señales utilizadas en este experimento, las amplitudes del tren de delta correspondientes a los valores de a_k en la expresión (4.11) se han generado a partir de una variable aleatoria gaussiana de media 1 y varianza 0.2, y las chirp utilizadas en la generación de los patrones tienen frecuencia inicial $f_0 = 20Hz$ y frecuencia final $f_1 = 3000Hz$.

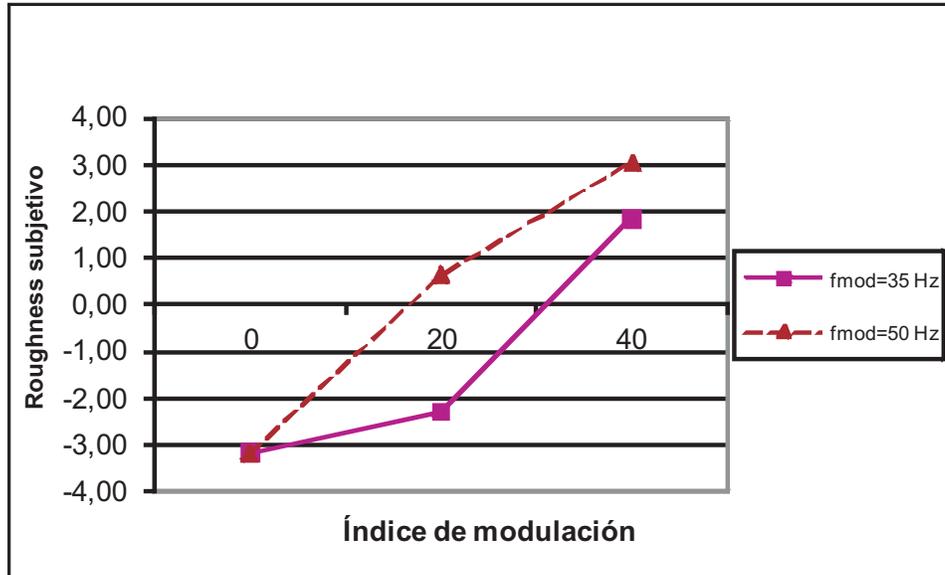


Figura 5.12: Roughness subjetivo en función del índice de modulación (0 %, 20 % y 40 %) para frecuencias de modulación de 35Hz y 50Hz.

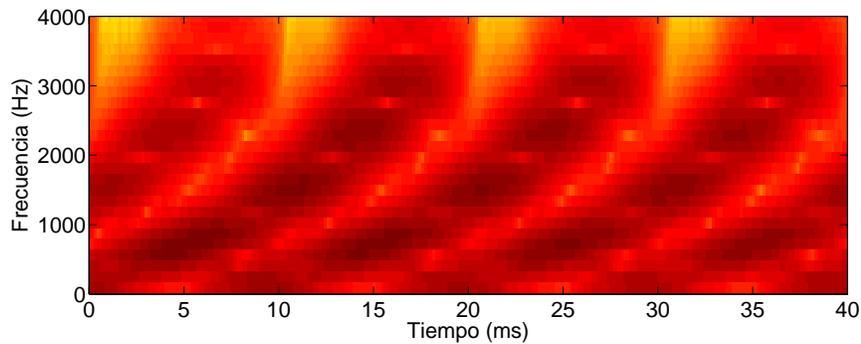


Figura 5.13: Espectrograma de una señal generada con $D = 20$.

En este experimento se han generado cuatro señales utilizando chirps de distinta duración ($D = 1, 10, 20$ y $30ms$). La señal patrón se ha generado sumando dos chirps retrasadas l . El parámetro l se ha ajustado para cada señal con el objetivo de conseguir una separación de frecuencias de 1 bark entre las dos chirp. La figura 5.13 representa el espectrograma de la señal generada con $D = 20$. Se puede observar que a causa del reducido valor del parámetro l , el patrón produce una única línea en la representación. Este patrón es repetido (con ligeras variaciones en la amplitud) cada $T_k = 10ms$.

El modelo de cálculo de Roughness expuesto en la sección 2.6.6 comienza dividiendo la señal en bandas. Al hacer pasar la señal representada en la figura 5.13 a través de un banco de filtros paso banda de ancho un bark (filtros de banda crítica) y calcular la envolvente de estas señales paso banda, se obtienen las representaciones mostradas en la figura 5.14. Esta gráfica (correspondiente a la salida de los filtros de banda crítica 10, 11 y 12) muestra tres señales muy similares con pequeños retardos entre sí.

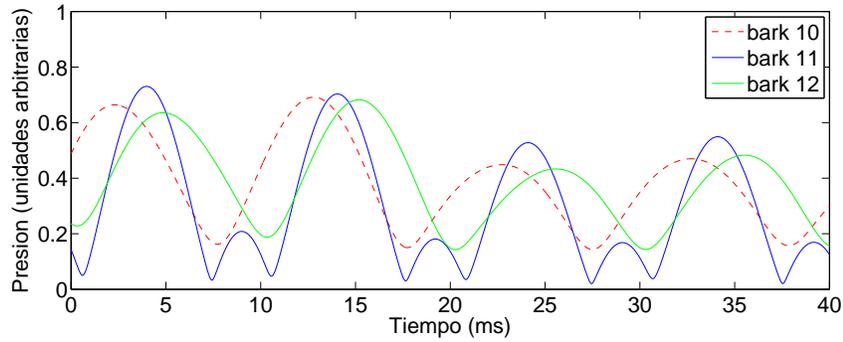


Figura 5.14: Envoltura de una tras un filtrado de banda crítica para las bandas 10, 11 y 12

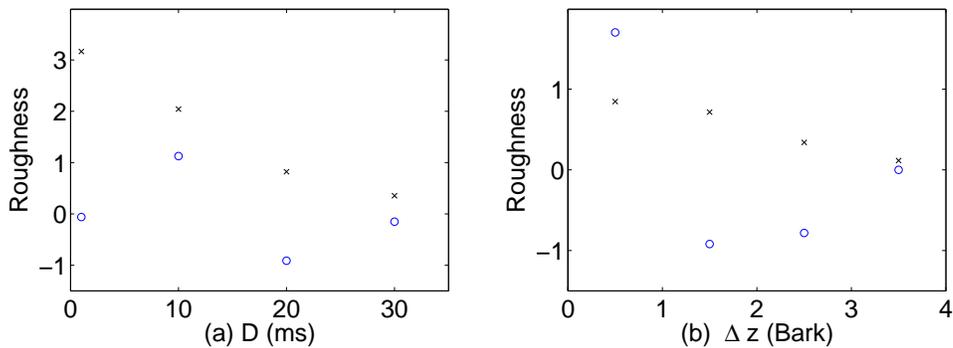


Figura 5.15: Roughness calculado (x) y medido (o) en el test. Izquierda (a) para duraciones diferentes D , y derecha (b) para retardos l diferentes

Las distintas señales utilizadas en este experimento producen espectrogramas como los de la figura 5.13, pero con líneas de distinta pendiente (cuanto mayor es D menor es la pendiente). Pequeñas pendientes producen grandes retrasos entre las líneas mostradas en la figura 5.14 y esta situación produce menor correlación entre las diferentes bandas críticas. Como resultado, al incrementar la duración del patrón (D), el Roughness debería de decrecer. Esta tendencia se puede observar en las marcas (x) de la parte izquierda de la figura 5.15 que representa el Roughness calculado (conforme al método de Daniel-Weber método) para las señales del experimento. Sin embargo, los resultados del experimento, círculos (o) en la parte izquierda de la figura 5.15, no muestran esta tendencia.

Los resultados del experimento muestran un máximo de Roughness próximo a $D = 10ms$. Este hecho podría deberse al retardo de grupo de la membrana basilar, que produce retardos mayores en las frecuencias bajas que en las altas. Para duraciones próximas a $D = 10ms$ este retraso podría sincronizar todas las componentes espectrales de la chirp y producir una excitación simultánea a lo largo de toda la membrana basilar [243]. Aunque algunas audiciones informales realizadas invirtiendo el patrón $p(t)$ suenan prácticamente igual que los sonidos del experimento y esto contradice la teoría de la sincronización de señales en la membrana basilar.

5.6.3. Dependencia de la excitación instantánea del patrón (dos chirps superpuestas).

Para formar las señales utilizadas en este experimento, las amplitudes del tren de delta correspondientes a los valores de a_k en la expresión (4.11) se han generado a partir de una variable aleatoria gaussiana de media 1 y varianza 0.2, y las chirp utilizadas en la generación de los patrones tienen frecuencia inicial $f_0 = 20Hz$ y $f_1 = 3000Hz$. El patrón se ha generado sumando dos chirps de duración $D = 20ms$, y retardo (l) variable con el fin de que cada patrón excite simultáneamente puntos separados 0,5, 1,5, 2,5 y 3,5barks. En este experimento todos los patrones tienen la pendiente que se muestra en la figura 5.13. El efecto asociado al incremento de los retardos entre las chirp (l) es el aumento del ancho de las líneas del espectrograma, incluso si el retardo es demasiado elevado llegan a aparecer dos líneas separadas.

Si observamos la salida de distintos filtros de banda crítica que se muestran en la figura 5.14, el efecto del incremento del retraso entre chirps (l) que se aprecia en estas representaciones temporales es un aumento de la duración de los máximos y una disminución de la anchura de los mínimos. Estos hechos provocan una disminución en la profundidad de modulación y un aumento en la correlación entre distintos canales. El resultado global de estos dos efectos es una pequeña reducción del Roughness calculado mediante el método de Daniel a medida que aumenta el parámetro (l), como se puede observar en el gráfico de la derecha de la figura 5.15.

En gráfico de la derecha de la figura 5.15 se representa con el símbolo (x) el Roughness calculado conforme al método de Daniel y Weber de cada señal, mientras que los resultados obtenidos por el test de parejas se representan con el símbolo (o). A pesar de que el Roughness calculado decrece a medida que aumenta el retraso entre chirps (l), el Roughness medido es alto cuando la separación entre chirps es reducida, pero disminuye cuando esta separación es mayor de 1 bark.

Estos resultados indican que la ponderación del Roughness Específico con la correlación entre bandas adyacentes para calcular el Roughness Total, se podría mejorar añadiendo algún otro factor de ponderación. También se confirma la falta de precisión ofrecida por los métodos tradicionales en el cálculo del Roughness cuando se analizan señales complejas como las de un motor, y la poca influencia del patrón utilizado en el Roughness de la señales.

5.7. Conclusiones.

Se ha realizado un análisis matemático del Roughness que predice el mal funcionamiento de los métodos clásicos de cálculo de Roughness cuando se aplican a ruidos asociados al automóvil. Además, este hecho se ha confirmado mediante los experimentos realizados.

Se ha comprobado que el Roughness tiene una gran dependencia del índice de modulación y una dependencia muy débil de la frecuencia de modulación cuando las amplitudes

de las repeticiones del patrón se calculan mediante un modelo de modulación de amplitud (AM).

Se ha comprobado que, en las señales modeladas como la convolución entre un tren de deltas y un patrón, el Roughness percibido depende en gran medida de la variación de la amplitud de las deltas, mientras que la forma del patrón es poco influyente

Se ha desarrollado una nueva señal (la chirp de barrido uniforme en la escala Bark) para analizar el Roughness de ruido de motor sintético. Esta nueva señal, debido a su característica de excitación uniforme en la escala sensorial, proporciona una nueva herramienta en psicoacústica.

Por lo tanto, los métodos de cálculo del Roughness se podrían mejorar considerando otros parámetros, además de la ponderación del Roughness específico con la correlación entre canales adyacentes.

Capítulo 6

Conclusiones y líneas futuras

Conclusiones.

Para poder estimar la molestia producida por los ruidos generados en un motor de combustión interna es necesario trabajar con parámetros psicoacústicos. Primero se deben estimar los parámetros psicoacústicos de estos sonidos y a continuación combinarlos de forma adecuada para obtener este parámetro de molestia.

El Roughness es un parámetro psicoacústico de gran relevancia en el cálculo de la molestia producida por los ruidos asociados a un motor, aunque su cálculo no ha sido estandarizado. A pesar de que existen procedimientos para estimar el Roughness de sonidos simples, no son generalizables a otros sonidos más complejos como los tratados en esta tesis.

Para estudiar los parámetros psicoacústicos que los ruidos asociados a un motor producen en las personas, es necesario realizar experimentos con jurado. Cuando se le pide a un jurado que compare el Roughness de varios sonidos, sus respuestas se ven influenciadas por el resto de parámetros psicoacústicos de los sonidos que se comparan. Por este motivo, en un test de jurado dedicado a estudiar el Roughness es imprescindible que todas las señales posean el resto de parámetros psicoacústicos iguales. Puesto que cualquier combinación de señales de motor real presenta variaciones en todos sus parámetros psicoacústicos, en estos experimentos se hace necesario recurrir a señales sintéticas.

En esta tesis se ha desarrollado un nuevo modelo que permite generar señales sintéticas de combustión y de escape a partir del análisis de señales de motor reales y de su mecanismo de generación. Estas señales se han modelado como la superposición de una señal periódica con ligeras variaciones (que hemos llamado modelo Cuasiperiódico) y un ruido coloreado. Tanto los ruidos de combustión como los de escape se han modelado matemáticamente como la convolución entre dos señales: una señal patrón y un tren de deltas modulado.

Aunque se podría haber desarrollado un modelo más complejo, se ha buscado un compromiso entre simplicidad y exactitud para poder investigar el modelo de Roughness de este tipo de señales. Además, el modelo de señal propuesto permite manipular de forma

individual cada uno de los parámetros psicoacústicos.

Para la síntesis de señales de motor se ha partido del análisis tiempo-frecuencia de señales de motor reales, aunque en esta tesis también se ha desarrollado otro tipo de señal, la chirp de barrido uniforme en la escala Bark, que ha permitido profundizar en el conocimiento de la percepción del Roughness y supone el desarrollo de una nueva herramienta en psicoacústica.

Como resultado de la experimentación con estas señales sintéticas se ha comprobado que el Roughness tiene una gran dependencia de los parámetros asociados al tren de deltas del modelo (sobre todo del índice de modulación y en menor medida de la frecuencia de modulación), mientras que la forma del patrón utilizado no influye demasiado en este parámetro.

Líneas futuras.

Una primera línea de investigación a desarrollar se centra en relacionar los parámetros de las señales sintéticas más influyentes en el Roughness con los parámetros físicos de funcionamiento del motor.

Una segunda línea de investigación consiste en ampliar el modelo de síntesis de sonido incluyendo nuevos parámetros relacionados con el motor y estudiar su influencia en el Roughness.

Otra línea de investigación consiste en profundizar en el estudio sobre la influencia que el Roughness de las señales de motor tiene sobre la valoración global de la molestia que producen estos sonidos.

Otra línea de investigación interesante se centraría en aplicar el modelo cuasiperiódico estudiado en esta tesis a otros productos distintos de los motores de vehículos, como puede ser maquinaria rotativa o motores eléctricos.

Una última línea de investigación consistiría en profundizar en el estudio de la percepción de las cualidades de sonidos a partir de la señal chirp de barrido uniforme en la escala Bark.

Apéndice 1: Wavelets y filtros de reconstrucción perfecta

Bancos de filtros de reconstrucción perfecta.

La finalidad de un banco de filtros es separar las señales en distintas bandas de frecuencia. Pero los filtros reales tienen bandas de transición y “fugas”; y esto hace que al diseñar bancos de filtros no se pueda evitar el solapamiento de frecuencias entre varios filtros (una misma frecuencia “pasa” a través de distintos filtros del banco de filtros). Por lo tanto, la reconstrucción de las señales a partir de las salidas ofrecidas por cada uno de los filtros que componen un banco de filtros no es trivial: no es suficiente con sumar las salidas de cada uno de los filtros, ya que si se hiciera esto se produciría distorsión de amplitud y de fase. Es necesario desarrollar una teoría de diseño de filtros que permita la “reconstrucción perfecta” de las señales.

La figura 6.1 representa un banco de reconstrucción perfecta con dos canales. La parte izquierda está formada por dos filtros H_0 y H_1 (el primero paso bajo y el segundo paso alto), conocidos como filtros de análisis, que descomponen la señal $x[n]$ en dos señales, seguidos cada uno de ellos por un diezmador ($\downarrow 2$) que elimina las muestras impares de las señales existentes a la salida de los filtros. Los puntos suspensivos del centro, indican que cada una de las dos señales diezmadas pueden ser tratadas de distintas formas: pueden ser codificadas para ser almacenadas o transmitidas, pueden ser procesadas con distintos propósitos, etc. De momento y con el fin de demostrar el funcionamiento del sistema se supondrá que no se realiza ninguna operación sobre las muestras que forman las dos señales. Por último, la parte derecha de la figura está formada por sistemas intercaladores de ceros ($\uparrow 2$) (que intercalan un cero entre cada dos muestras) y de filtros (filtros de

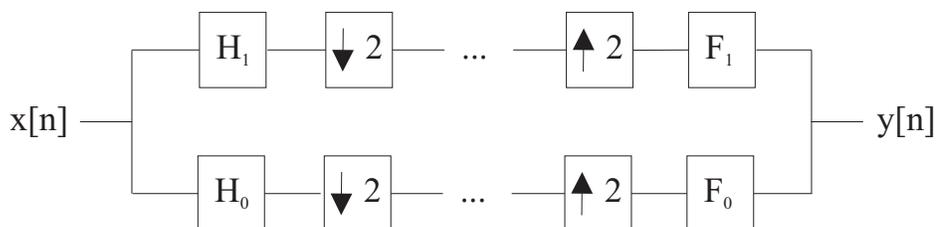


Figura 6.1: Banco de filtros de reconstrucción perfecta con dos canales.

reconstrucción): F_0 será un filtro paso bajo y F_1 un filtro paso alto.

El uso de los diezmadores ($\downarrow 2$) e intercaladores de ceros ($\uparrow 2$) de la figura 6.1 tiene una finalidad práctica: si no se utilizaran los diezmadores resultaría que al descomponer una señal $x[n]$ en N bandas distintas (2 bandas para el caso de la figura) para poder almacenarlas o transmitir las, se tendrían que almacenar o transmitir N señales de la misma longitud que la original (lo que resulta claramente ineficiente por contener redundancia). El uso de los diezmadores permite, por lo tanto, descomponer una señal en N nuevas señales cuyo tamaño será N veces inferior al de la señal original.

El uso de los intercaladores de ceros ($\uparrow 2$) se hace necesario en el proceso de reconstrucción de señales por el hecho de haber utilizado diezmadores. Pero el efecto que produce en cada uno de los ramales del banco de filtros el uso combinado de un diezmador y un intercalador de ceros es el aliasing, o repetición de espectros (como se verá más adelante).

Diseñar un banco de filtros de “reconstrucción perfecta” consiste en seleccionar de forma adecuada los filtros H_0 , F_0 , H_1 y F_1 de la figura 6.1 para evitar los efectos de “distorsión de amplitud y de fase” y “aliasing”.

Volviendo a la figura 6.1, y centrando nuestra atención en el efecto producido por un diezmador (que elimina las muestras impares) seguido de un intercalador de 0's (que pone 0's en el lugar donde estaban las muestras que eliminó el intercalador), se puede comprobar que cuando estos dos elementos se colocan en cascada (como ocurre en la figura 6.1) su efecto es el de sustituir por 0's las muestras impares, lo que equivale a multiplicar una señal por la secuencia:

$$s[n] = \frac{1}{2} (1 + (-1)^n) \quad (6.1)$$

y al multiplicar una señal $a[n]$ por esta secuencia se obtiene la señal:

$$b[n] = \frac{1}{2} (a[n]s[n]) = \frac{1}{2} (a[n] + a[n](-1)^n) \quad (6.2)$$

cuya transformada z vale:

$$B(z) = \frac{1}{2} (A(z) + A(-z)) \quad (6.3)$$

Teniendo esto en cuenta, la señal obtenida a la salida del filtro F_0 (ramal superior) de la figura 6.1 tendrá la transformada z :

$$\frac{1}{2}X(z)H_0(z)F_0(z) + \frac{1}{2}X(-z)H_0(-z)F_0(z) \quad (6.4)$$

mientras que la señal obtenida a la salida del filtro F_1 (ramal inferior) de la figura 6.1 tendrá como transformada z :

$$\frac{1}{2}X(z)H_1(z)F_1(z) + \frac{1}{2}X(-z)H_1(-z)F_1(z) \quad (6.5)$$

Tanto en la expresión (6.4) como en la (6.5) los términos de la izquierda suponen versiones de la señal original $X(z)$ filtradas paso bajo ($X(z)H_0(z)F_0(z)$) o paso alto ($X(z)H_1(z)F_1(z)$) puesto que $H_0(z)$ y $F_0(z)$ representan filtros paso bajo y $H_1(z)$ y

$F_1(z)$ son filtros paso alto. Son señales muy útiles pues corresponden a las frecuencias bajas y altas de la señal original, y al sumarlas deberían proporcionar la señal original perfectamente reconstruidas.

Sin embargo, el término de la derecha de la expresión (6.4) corresponde a una versión modulada de la señal original ($X(-z)$) filtrada por un filtro paso alto ($H_0(-z)$) y posteriormente por un filtro paso bajo ($F_0(z)$). El resultado de estas operaciones proporciona una señal atenuada (ha sido filtrada paso alto y paso bajo) y de ninguna utilidad práctica (puesto que no contiene información de la señal $x[n]$, sino de una versión modulada de esta). Lo mismo ocurre con el término de la derecha de la expresión (6.5): corresponde a una señal residual atenuada de ninguna utilidad práctica y que se debería eliminar.

La señal obtenida a la salida de la figura 6.1 corresponde a la suma de las señales representadas en las expresiones (6.4) y (6.5). Si el sistema constituye un banco de filtros de reconstrucción perfecta, esta suma debería de coincidir con la señal original $y[n] = x[n]$. Para ello, la suma de las versiones filtradas paso bajo y paso alto deberán coincidir exactamente con la señal original y la suma de las señales residuales deberán de anularse:

$$\frac{1}{2}X(z)H_0(z)F_0(z) + \frac{1}{2}X(z)H_1(z)F_1(z) = X(z) \quad (6.6)$$

$$\frac{1}{2}X(-z)H_0(-z)F_0(z) + \frac{1}{2}X(-z)H_1(-z)F_1(z) = 0 \quad (6.7)$$

Para esto deberán cumplirse las condiciones:

$$H_0(z)F_0(z) + H_1(z)F_1(z) = 2 \quad (6.8)$$

$$H_0(-z)F_0(z) + H_1(-z)F_1(z) = 0 \quad (6.9)$$

La condición expresada en la ecuación (6.8) se conoce como condición de no distorsión mientras que la ecuación (6.9) se conoce como condición de cancelación de alias. Para que se cumpla la ecuación (6.8) es necesario utilizar filtros no causales: esto no representa ningún inconveniente desde el punto de vista teórico, pero a veces sí que representa un problema práctico por lo que en ocasiones la ecuación (6.8) se sustituye por:

$$H_0(z)F_0(z) + H_1(z)F_1(z) = 2z^{-l} \quad (6.10)$$

Que es una versión desplazada de (6.8) que sí que se podrá implementar utilizando filtros causales.

Aunque de forma menos intuitiva, también se puede llegar a las conclusiones expresadas en los párrafos anteriores utilizando notación matricial. Ya que la salida proporcionada por el sistema de la figura 6.1 se puede escribir como:

$$Y(z) = \frac{1}{2} \begin{bmatrix} F_0(z) & F_1(z) \end{bmatrix} \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} \begin{bmatrix} X(z) \\ X(-z) \end{bmatrix} \quad (6.11)$$

Y para que la salida sea una reconstrucción perfecta de la entrada: $Y(z) = X(z)$, se deberá de cumplir:

$$\begin{bmatrix} F_0(z) \\ F_1(z) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 0 \end{bmatrix} \quad (6.12)$$

definiendo $\mathbf{H}_m(z)$ como la matriz:

$$\mathbf{H}_m(z) = \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} \quad (6.13)$$

y calculando su inversa y el producto de matrices en (6.12) se obtiene:

$$\begin{bmatrix} F_0(z) \\ F_1(z) \end{bmatrix} = \frac{2}{\det(\mathbf{H}_m(z))} \begin{bmatrix} H_1(-z) \\ -H_0(-z) \end{bmatrix} \quad (6.14)$$

La relación (6.14) indica como se deben elegir los filtros de reconstrucción una vez seleccionado los filtros de análisis.

Un parámetro interesante del banco de filtros es lo que se conoce como filtro producto, que se define como:

$$P_0(z) = \frac{2H_0(z)H_1(-z)}{\det(\mathbf{H}_m(z))} \quad (6.15)$$

que recibe el nombre de filtro producto porque a la vista de (6.14), se puede comprobar que:

$$P_0(z) = H_0(z)F_0(z) \quad (6.16)$$

Es decir el filtro producto corresponde al producto de las funciones de transferencia de los filtros paso bajos utilizados en el banco de filtros.

También se define $P_1(z)$ como el producto de las funciones de transferencia de los filtros paso alto:

$$P_1(z) = H_1(z)F_1(z) \quad (6.17)$$

A la vista de la relación existente entre los filtros representada por la ecuación (6.14) y teniendo en cuenta que $\det(\mathbf{H}_m(-z)) = -\det(\mathbf{H}_m(z))$, se puede comprobar que:

$$P_1(z) = -P_0(-z) \quad (6.18)$$

Si ahora se desarrolla la ecuación (6.11) y se seleccionan los filtros de reconstrucción (F_0 y F_1) como se indica en la expresión (6.14), se llega a la conclusión de que la condición necesaria y suficiente para una reconstrucción perfecta ($Y(z) = X(z)$) es:

$$P_0(z) - P_0(-z) = 2 \quad (6.19)$$

La expresión anterior, también se puede obtener al reescribir la condición de no distorsión (6.8) y la condición de cancelación de alias (6.9) haciendo uso del filtro producto $P_0(z)$.

Históricamente, el diseño de banco de filtros se abordaba seleccionando primero el filtro paso bajo $H_0(z)$ y construyendo el filtro $H_1(z)$ a partir del anterior. En el año 1975 Esteban y Galand [244] seleccionaron el filtro $H_1(z) = H_0(-z)$, lo que en el dominio del tiempo supone alternar los signos de la respuesta impulsional del filtro y a esto le llamaron filtro espejo en cuadratura (porque el filtro paso bajo es simétrico al filtro paso alto). Con esta opción cabía la posibilidad de utilizar filtros IIR (necesarios si se quería conseguir la reconstrucción perfecta). más tarde en 1984 Smith y Barnwell [245] y en 1985 Mintzer [130] mejoraron el diseño seleccionando el filtro $H_1(z) = -z^{-n}H_0(-z^{-1})$. Y esto, en el dominio del tiempo supone además de alternar los signos de su respuesta impulsional reflejar la señal respecto al eje horizontal y al vertical y desplazarla en el tiempo.

Actualmente, la tarea de diseño de un banco de filtros de reconstrucción perfecta deja poca libertad, y se puede resumir en tres pasos:

1. Diseñar un filtro paso bajo $P_0(z)$ que satisfaga la ecuación (6.19).
2. Obtener los filtros paso bajo del banco de filtro mediante la factorización de $P_0(z)$ en $F_0(z)H_0(z)$.
3. Obtener los filtros paso alto del banco mediante el uso de las relaciones expresadas en la igualdad (6.14).

Análisis Wavelet usando filtros.

La expresión (3.12) se utiliza para definir la STFT de una señal continua: conceptualmente esta transformación equivale a realizar una Transformada de Fourier a la señal $x(t')$ multiplicada por una ventana centrada en un instante de tiempo t . Resulta bastante sencillo trasladar estos conceptos (inventanar y realizar una transformada de Fourier) al mundo discreto y obtener la definición de STFT proporcionada por la ecuación (3.40) o su versión muestreada en frecuencia (3.51). Sin embargo esto no es posible para la transformada Wavelet. La ecuación (3.21) corresponde a la de definición de transformada Wavelet de una señal continua, conceptualmente consiste en una correlación de la señal $x(t')$ con una versión escalada (o estirada) de una función de análisis. Puesto que la operación de escalado (o estiramiento) no se puede realizar sobre señales discretas, no existe ninguna expresión similar a las anteriores que defina la transformada Wavelet de una señal discreta.

Al trabajar de forma práctica con transformadas wavelet de señales continuas, es habitual realizar un muestreo del plano tiempo-frecuencia (o tiempo-escala) como se indicó en la sección 3.2.1. Y para realizar un muestreo eficiente en el plano tiempo frecuencia es típico utilizar un valor de $\alpha = 2$ (computación octava por octava) como se indicó en la sección 3.2.1. Una computación octava por octava realizada sobre una señal continua significa que si se utiliza una wavelet que realice un análisis entre las frecuencias de 1Hz y 2Hz, la siguiente wavelet realizará un análisis de la señal entre las frecuencias 0.5Hz y 1Hz, y la siguiente entre 0.25Hz y 0.5Hz y así sucesivamente.

En la sección 6 se indicó como, mediante el uso de filtros, se puede descomponer una señal discreta en otras dos señales (con la mitad de longitud) de forma que sea posible la reconstrucción de la señal original partir de ellas. Si los filtros de análisis utilizados son buenas aproximaciones de filtros paso bajo y paso alto ideales, al utilizar esta descomposición se divide una señal en dos bandas de frecuencia: una de ellas contiene las frecuencias bajas de la señal $[0, \frac{1}{4}]$ (si se trabaja con señales reales) y la otra contiene las frecuencias altas de la señal $[\frac{1}{4}, \frac{1}{2}]$. Si se vuelve a realizar esta descomposición sobre la señal que contiene las frecuencias bajas de la señal original ($[0, \frac{1}{4}]$), se obtendrán dos nuevas señales: una de ellas contendrá las frecuencias $[0, \frac{1}{8}]$ de la señal original y la otra contendrá las frecuencias $[\frac{1}{8}, \frac{1}{4}]$ de la señal original. Las señales obtenidas a las salidas de cada uno de los filtros paso bajo reciben el nombre de *aproximaciones* mientras que las que resultan de los filtrados paso alto reciben el nombre de *detalles*.

Si se continúa realizando descomposiciones únicamente de la banda baja obtenida en la descomposición de forma iterativa, se obtiene el esquema representado en la figura 6.2, donde se puede observar que la rama superior de cada descomposición se obtiene un filtrado paso banda de la señal original (conteniendo información de las frecuencias $[\frac{1}{4}, \frac{1}{2}]$, $[\frac{1}{8}, \frac{1}{4}]$, $[\frac{1}{16}, \frac{1}{8}]$, ...) mientras que en la rama inferior de cada descomposición se obtiene un filtrado paso bajo de la señal original (conteniendo información de las frecuencias $[0, \frac{1}{4}]$, $[0, \frac{1}{8}]$, $[0, \frac{1}{16}]$, ...). Observando la figura 6.2 se puede comprobar que la rama superior de cada una de las descomposiciones realiza sobre una señal discreta original ($x[n]$) el mismo tipo de análisis que la transformada wavelet realiza sobre una señal continua.

Así pues, la transformada wavelet de una señal discreta se calcula, de forma práctica, utilizando la descomposición que aparece en la figura 6.2. La gran diversidad de transformadas wavelet empleadas en la práctica se debe a las diferentes posibilidades existentes para seleccionar los filtros $H_0(z)$ y $H_1(z)$, dependiendo del propósito de la transformada.

Relación wavelet continua-wavelet discreta.

La figura 6.3 representa la sucesión de filtrados y diezmados correspondientes al ramal inferior de la figura 6.2.

La figura 6.4 muestra una equivalencia entre dos sistemas: un diezmador ($\downarrow 2$) seguido de un sistema LTI con función de transferencia $H_1(z)$, es equivalente a un sistema LTI con función de transferencia $H_1(z^2)$ seguido por ese mismo diezmador ($\downarrow 2$). Teniendo en

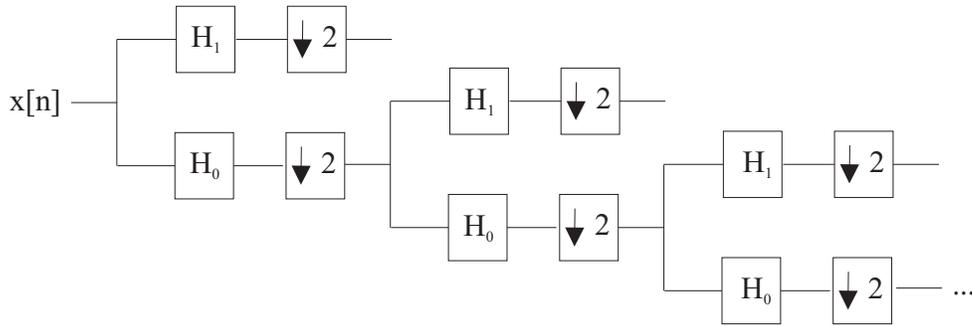


Figura 6.2: Cascada de filtros para obtener un análisis mediante wavelets.

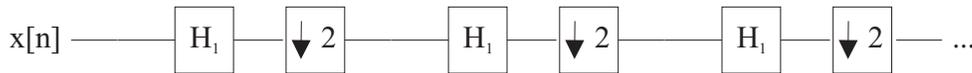


Figura 6.3: Rama inferior de la iteración representada en la figura 6.2.

cuenta esta relación, la salida del sistema representado en la figura 6.3 (después de pasar por i filtros), sería equivalente a utilizar un sistema con función de transferencia:

$$H_0^i(z) = H_0(z) \cdot H_0(z^2) \cdot H_0(z^4) \dots = \prod_{l=0}^{i-1} H_0(z^{2^l}) \quad (6.20)$$

seguido por un diezmador ($\downarrow 2^i$).

Pero la salida del ramal inferior de la figura 6.2 no representa un análisis wavelet puesto que esta salida se corresponde con un filtrado paso bajo mientras que un análisis wavelet se corresponde con un filtrado paso banda (como ya se indicó en la sección 6). El análisis wavelet se corresponde con la salida del ramal superior de cada iteración; y de la misma forma que la ecuación (6.20) proporciona la función de transferencia del filtro que permite obtener la salida del ramal inferior a partir de la señal original, la función de transferencia del filtro que permite obtener la salida de los ramales superiores, se puede escribir como:

$$H_1^i(z) = H_0(z) \cdot H_0(z^2) \dots H_0(z^{2^{i-2}}) \cdot H_1(z^{2^{i-1}}) = \left(\prod_{l=0}^{i-2} H_0(z^{2^l}) \right) \cdot H_1(z^{2^{i-1}}) \quad (6.21)$$

A medida que las iteraciones representadas en la figura 6.2) crecen, también lo hacen las longitudes de las respuestas impulsionales de los filtros representados por las expresiones

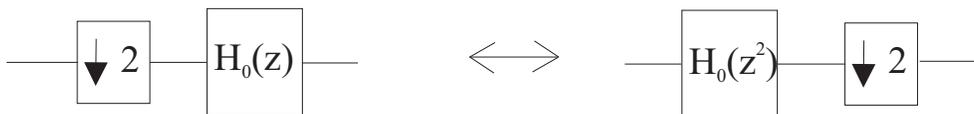


Figura 6.4: Equivalencia de un sistema formado por un filtro y un diezmador.

(6.20) y (6.21) ($h_0^i[n]$ y $h_1^i[n]$). En ocasiones, las respuestas impulsionales de estos filtros crecen de forma “regular” y reciben el nombre de *filtros regulares*.

Para ver cómo se comportan estos filtros a medida que crece el valor de i , se define la función continua $h_0^i(t)$ a partir de la función discreta $h_0^i[n]$ de la siguiente manera:

$$h_0^i(t) = h_0^i(n) \text{ en el intervalo } \left[\frac{n}{2^i}, \frac{n+1}{2^i} \right] \quad (6.22)$$

En la figura 6.5 se puede observar la convergencia de uno de estos filtros. En concreto, se ha utilizado un filtro de Daubechies de 4 coeficientes ($h_0[n] = [0,4830 \ 0,8365 \ 0,2241 \ -0,1294]$). Esta función continua que tiene forma de filtro paso bajo se la conoce como *función de escalado* y se representa por la expresión $\Phi(t)$.

Si ahora se estudia en el dominio del tiempo el comportamiento del sistema representado por la expresión (6.21) y se define $h_1^i[n]$ como:

$$h_1^i(t) = h_1^i(n) \text{ en el intervalo } \left[\frac{n}{2^i}, \frac{n+1}{2^i} \right] \quad (6.23)$$

Se puede estudiar del mismo modo la convergencia de la función $h_1^i(t)$. En la figura 6.6 se puede observar al convergencia de la función $h_1^i(t)$ cuando se utiliza el filtro $h_0[n]$ empleado en la figura 6.5 junto con el filtro $h_1[n] = [-0,4830 \ 0,8365 \ -0,2241 \ -0,1294]$ diseñado por Daubechies como complemento del primero. La función a la que converge $h_1^i(t)$ se conoce como *función wavelet* y se representa mediante $\Psi(t)$. Las funciones wavelet continuas así obtenida se pueden emplear para el análisis de funciones continuas utilizándola como función de análisis en la expresión (3.21).

Aunque se entiende por wavelet continua a la transformación que aparece en la ecuación (3.21) y que se realiza sobre una función continua, de forma práctica, esta transformación se realiza utilizando una versión discretizada de esta expresión y trabajando con señales y funciones de análisis muestreadas (y por lo tanto discretas). Por esta razón, es habitual hablar de la transformadas wavelets continua de una señal discreta para referirse a la aplicación de la ecuación (3.21) discretizada sobre una señal continua muestreada.

En resumen, cuando se realiza un análisis wavelet de una secuencia discreta, no se emplea la señal wavelet sino una sucesión de filtrados para obtener una descomposición de la señal en bandas de octava. Por lo que de forma práctica, no suele tener demasiado interés conocer la forma de la wavelet con la que se está trabajando puesto que se trabaja con coeficientes de filtros discretos. Si por cualquier razón se desea conocer la forma de esta función wavelet basta con realizar las iteraciones indicadas en esta sección.

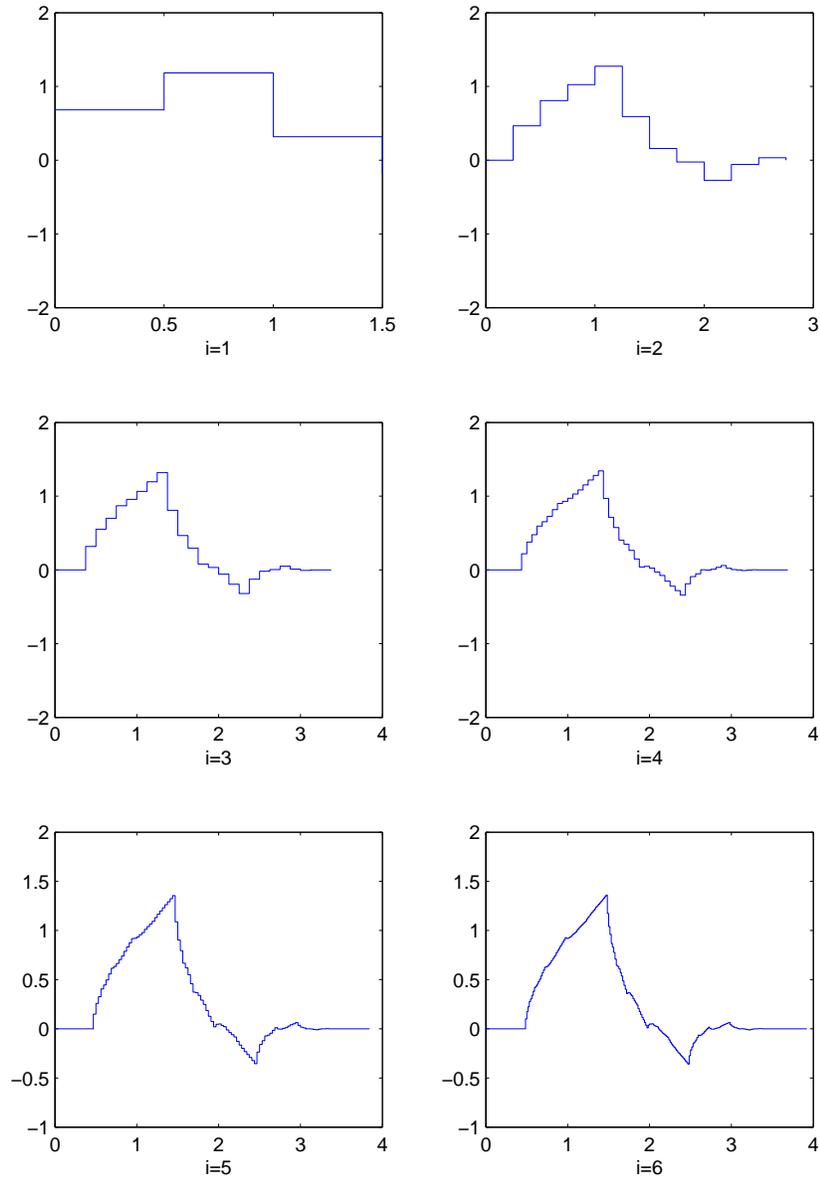


Figura 6.5: Convergencia del filtro paso bajo iterado (función de escalado) para el filtro de Daubechies de 4 coeficientes.

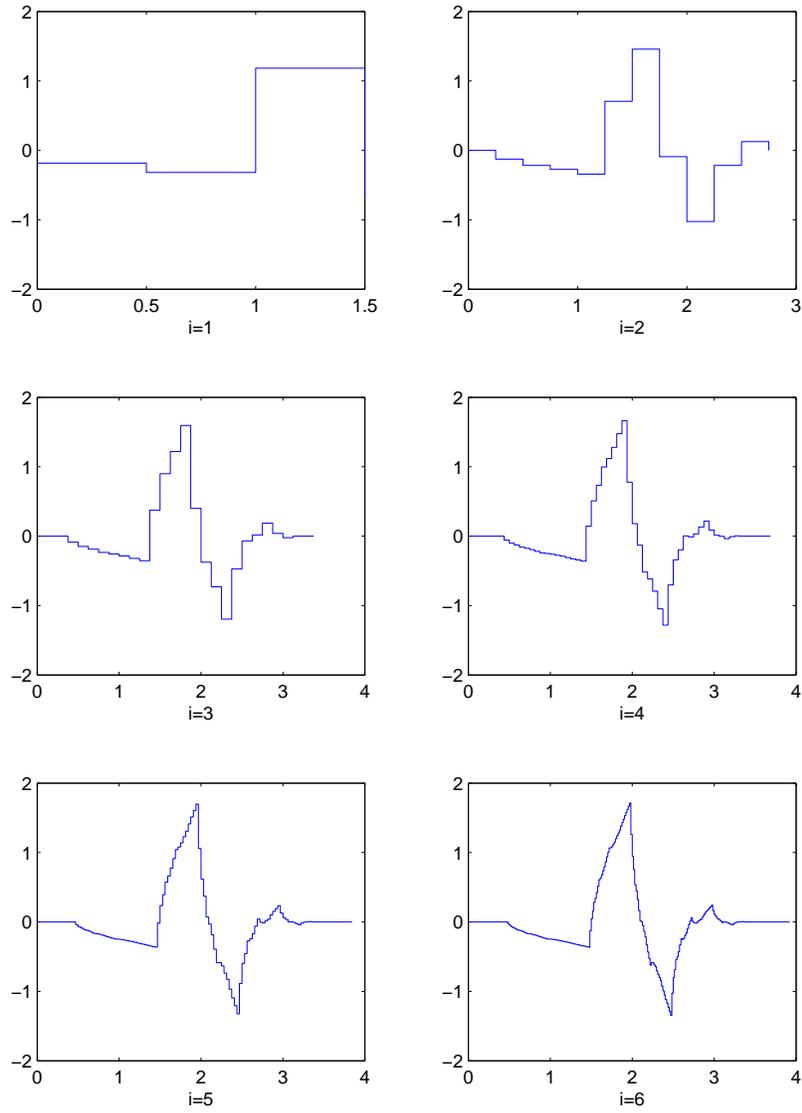


Figura 6.6: Convergencia del filtro paso banda iterado (función wavelet) para el filtro de Daubechies de 4 coeficientes.

Apéndice 2: Publicaciones Relacionadas.

Revista

[1] A. Camacho, G. Piñero, M. de Diego, and A. Gonzalez. Exploring roughness perception in car engine noises through complex cepstrum analysis. *Acta acustica*, 94(1):130–140, 2008.

Congreso internacional

[2] A. Camacho, G. Piñero, and R Llinares. Evaluation of roughness for a pattern repeated with random amplitude. In *17th International congress on sound and vibration proceedings*, pages 2997–3004, El Cairo, Egipto, Julio 2010. International Institute of Acoustics and Vibration IIAV.

[3] G. Piñero, M. de Diego, A. Gonzalez, Fuster. L., and A. Camacho. Application of different time-frequency analysis to psychoacoustic description of car engine noise. In *AES Actas del congreso*, Barcelona, 2005.

[4] A. Camacho, G. Piñero, M. de Diego, and M. Ferrer. On the use of complex cepstrum in psychoacoustic evaluations of engine noise. In *11th International congress on sound and vibration proceedings*, pages 2997–3004, San Petersburgo, Rusia, Julio 2004. International Institute of Acoustics and Vibration IIAV.

[5] A. Camacho, G. Piñero, L. Fuster, and M. de Diego. Evaluation of human perception of roughness for diesel car noise. In *10th International congress on sound and vibration proceedings*, Estocolmo, Suecia, Julio 2003. International Institute of Acoustics and Vibration IIAV.

[6] G. Piñero, A. Camacho, L. Vergara, A. González, and M. de Diego. Relation between time-frequency parameters and psychoacoustics features in noise quality evaluation. In *17th International Congress on Acoustics (ICA) Actas del Symposium*, 2001.

Congreso Nacional

[7] A. Camacho, G. Piñero, G. Igual, and R. Llinares. Uso del cepstrum complejo en la síntesis de ruidos de motor. In *URSI 2004 Actas del Congreso*, 2004.

[8] A. Camacho, G. Piñero, and R. Llinares. Síntesis de sonidos periódicos usando técnicas tiempo-frecuencia. In Compobell S.L., editor, *La investigación ante la sociedad del conocimiento*, pages 93 – 96, Alcoy, Noviembre 2004. Escuela Politécnica Superior de Alcoy, Escuela Politécnica Superior de Alcoy.

[9] A. Camacho, G. Igual, G. Piñero, and D. Villodre. Generador de señales de audio. In *URSI 2002 actas del congreso*, 2002.

Referencias

- [1] ISO. 532 method for calculating loudness level. Technical report, ISO, 1975. [1](#), [2.6](#), [2.6.3](#), [2.6.3](#), [2.6.3](#), [6](#), [2.6.3](#), [3.4](#)
- [2] DIN 45692 Measurement technique for the simulation of the auditory sensation of sharpness, 2009. [1](#), [2.6](#)
- [3] E. Zwicker and H. Fastl. *Psychoacoustics: facts and models*. Springer, Berlin ; New York, 1999. [1](#), [2.4.1](#), [2.4.2](#), [2.4.4](#), [2.6](#), [2.6.2](#), [2.6.2](#), [2.6.3](#), [2.6.3](#), [2.6.4](#), [2.6.5](#), [2.6.6](#), [2.6.6](#), [2.7](#), [3.4](#), [4.1](#), [5.1](#), [5.2.3](#), [5.4.1](#)
- [4] H. Fastl. The psychoacoustics of sound-quality evaluation. *Acustica*, 83(5):754–764, 1997. [1](#), [2.6](#), [2.7](#), [2.8](#)
- [5] W. Aures. Ein berechnungsverfahren der rauigkeit. *Acustica*, 58(5):268–281, 1985. [1](#), [2.6.6](#), [5.1](#)
- [6] B.C.J. Moore. *An introduction to the psychology of hearing*. San Diego (CA) [etc.], Academic Press, 2003. [1](#), [2.3](#), [2.4.1](#), [2.4.2](#), [2.4.4](#), [2.6](#), [2.6.1](#), [2.6.2](#)
- [7] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. Mit Press, 1997. [1](#)
- [8] P. Daniel and R. Weber. Psychoacoustical roughness: Implementation of an optimized model. *Acta acustica united with acustica*, 83(1):113–123, 1997. [1](#), [2.6.6](#), [2.6.6](#), [5.1](#), [5.2.3](#)
- [9] Head Acoustics. Artemis Reference Manual. Technical report, 2010. [1](#), [3.4](#)
- [10] Brüel & Kjær. Pulse-Sound Quality Reference Manual. Technical report, 1996. [1](#), [2.6.6](#), [3.4](#)
- [11] S.H. Shin. Comparative study of the commercial software for sound quality analysis. *Acoustical Science and Technology*, 29(3):221–228, 2008. [1](#), [2.6](#)
- [12] M. Bodden. Instrumentation for sound quality evaluation. *Acustica*, 83(5):775–783, 1997. [1](#), [2.7](#)

- [13] W. Ellermeier, M. Mader, and P. Daniel. Scaling the unpleasantness of sounds according to the btl model: Ratio-scale representation and psychoacoustical analysis. *Acta acustica united with acustica*, 90(1):101–107, 2004. [1](#), [2.6.6](#), [2.7](#), [4.4.3](#)
- [14] H. Fastl. Psychoacoustics and sound quality. *Fortschritte der Akustik*, 28:765–766, 2002. [1](#), [2.7](#)
- [15] H. Fastl. Psychoacoustic basis of sound quality evaluation and sound engineering. In *Proc. 13th Intern. Congress on Sound and Vibration, Vienna*, 2006. [1](#), [2.6](#), [2.7](#)
- [16] R. Jurc, O. Jiricek, and M. Brothánek. Methods for the assessment of pleasantness in sound quality. *Noise Control Engineering Journal*, 58(1):62, 2010. [1](#), [2.7](#)
- [17] C.V. Beidl and W. Stucklschwaiger. Application of the avl-annoyance index for engine noise quality development. *Acta acustica united with acustica*, 83(5):789–795, 1997. [1](#), [2.7](#)
- [18] M.S. Khan. Development of an annoyance index for heavy-duty diesel engine noise using multivariate analysis. *Noise Control Engineering Journal*, 45(4):157–167, 1997. [1](#), [2.5.5](#), [2.7](#)
- [19] B. Feng, GH Wakefield, MD Good, NC Otto, and R. Rebandt. On the perceived roughness of automotive powertrain sounds. *SAE transactions*, 104(6):2309–2317, 1995. [1](#), [5.1](#), [5.4](#)
- [20] BJ Feng and NC Otto. Synthesis of powertrain sounds for investigations in roughness. pages 571–578, 1993. [1](#), [4.1](#), [4.3](#), [4.21](#), [4.4.1](#), [4.4.2](#), [4.5.1](#), [5.1](#), [5.2.2](#), [5.4](#)
- [21] R. Hoeldrich and M. Pflueger. A generalized psychoacoustical model of modulation parameters (roughness) for objective vehicle noise quality evaluation. *SAE transactions*, 108(6; PART 2):2999–3002, 2000. [1](#), [3.4](#), [4.3](#), [5.1](#)
- [22] M. Pflueger, R. Hoeldrich, F.K. Brandl, and W. Biermayer. Subjective assessment of roughness as a basis for objective vehicle interior noise quality evaluation, sae paper 1999-01-1850. 1999. [1](#), [2.6.6](#), [5.1](#), [5.4](#)
- [23] L.P.R. de Oliveira, K. Janssens, P. Gajdatsy, H. Van der Auweraer, P.S. Varoto, P. Sas, and W. Desmet. Active sound quality control of engine induced cavity noise. *Mechanical Systems and Signal Processing*, 23(2):476–488, 2009. [1](#), [2.6.6](#), [5.1](#), [5.2](#)
- [24] D. Berckmans, P. Kindt, P. Sas, and W. Desmet. Evaluation of substitution monopole models for tire noise sound synthesis. *Mechanical Systems and Signal Processing*, 24(1):240–255, 2010. [1](#), [5.1](#)
- [25] J.A. Millán, R. Millán. *Diccionario de la lengua española*. [Madrid], Espasa-Calpe, 1995. [2.2](#)

-
- [26] J. Catalá, J. Aguilar, and F. Senent. *Física general*. Valencia, Tip. Puertes, 1977. [2.2.1](#), [2.2.1](#), [2.2.1](#)
- [27] S. Burbano and E. Burbano. *Física general*. Zaragoza, Librería General, 1977. [2.2.1](#)
- [28] D. Howard and J. Angus. *Acoustics & psychoacoustics*. Focal Press, 2009. [2.2.1](#), [2.6.1](#)
- [29] M. Recuero and C. Gil. *Acústica arquitectónica*. Alcorcón (Madrid), Izquierdo-Electricidad, 1991. [2.2.3](#), [2.2.4](#)
- [30] Brüel & Kjør. *Sound Intensity*. 1986. [2.2.3](#), [2.2.3](#), [2.2.4](#)
- [31] J.E. Hind. Physiological acoustics - introduction. In *Handbook of acoustics, Engineering-Computing-and-Technology*, pages 1085–1091. New York: John Wiley & Sons, 1998. [2.3](#)
- [32] W.F. Ganong. *Fisiología médica*. México, El Manual Moderno, 1990. [2.3](#)
- [33] J. Blauert. *Communication acoustics*. Springer, 2005. [2.3](#)
- [34] D. Havelock, S. Kuwano, and M. Vorlander. *Handbook of signal processing in acoustics*. Springer Verlag, 2008. [2.3](#)
- [35] E.A.G. Shaw. Acoustical characteristics of the outer ear. In *Handbook of acoustics, Engineering-Computing-and-Technology*, pages 1093–1103. New York: John Wiley & Sons, 1998. [2.3.1](#), [2.3.1](#)
- [36] B. Gardner and K. Martin. Hrtf measurements of a kemar dummy-head microphone. Technical Report 280, 1994. [2.1](#), [2.3.1](#), [2.3](#)
- [37] M.R. Bai and K.Y. Ou. Head-related transfer function (HRTF) synthesis based on a three-dimensional array model and singular value decomposition. *Journal of Sound and Vibration*, 281(3-5):1093, 2005. [2.3.1](#)
- [38] C Porschmann, J Sahrhage, J Blauert, and T. Djelani. An interactive virtual-environment generator for psychoacoustic research II: Collection of head-related impulse responses and evaluation of auditory localization. *Acustica*, 86(6):1046, 2000. [2.3.1](#)
- [39] J. Blauert, H. Lehnert, J. Sahrhage, and H. Strauss. An interactive virtual-environment generator for psychoacoustic research. I: Architecture and implementation. *Acustica*, 86(1):94, 2000. [2.3.1](#)
- [40] V.R. Algazi. Elevation localization and head-related transfer function analysis at low frequencies. *The Journal of the Acoustical Society of America*, 109(3):1110–1122, 2001. [2.3.1](#)

- [41] B.D. Vanveen, K.E. Hecox, and J.S. Chen. External ear transfer-function modeling a beamforming approach. *The Journal of the Acoustical Society of America*, 92(4):1933, 1992. [2.3.1](#)
- [42] J. Chen, B.D. Veen, and K.E. Hecox. A spatial feature-extraction and regularization model for the head-related transfer-function. *The Journal of the Acoustical Society of America*, 97(1):439, 1995. [2.3.1](#)
- [43] S. Mehrgardt and V. Mellert. Transformation characteristics of external human ear. *The Journal of the Acoustical Society of America*, 61(6):1567, 1977. [2.3.1](#)
- [44] S.E. Voss and Allen J.B. Measurement of acoustic-impedance and reflectance in the human ear canal. *The Journal of the Acoustical Society of America*, 95(1):372, 1994. [2.3.1](#)
- [45] M. Vorlander. Acoustic load on the ear caused by headphones. *The Journal of the Acoustical Society of America*, 107(4):2082, 2000. [2.3.1](#)
- [46] G. Keidser. Relative loudness perception of low and high frequency sounds in the open and occluded ear. *The Journal of the Acoustical Society of America*, 107(6):3351–3357, 2000. [2.3.1](#)
- [47] W.T. Peake. Acoustic properties of the middle ear. In *Handbook of acoustics, Engineering-Computing-and-Technology*, pages 1105–1114. New York: John Wiley & Sons, 1998. [2.3.2](#)
- [48] A. Hudde and H. Engel. Measuring and modeling basic properties of the human middle ear and ear canal. part I: Model structure and measuring techniques. *Acustica*, 84(4):720, 1998. [2.3.2](#)
- [49] A. Hudde and H. Engel. Measuring and modeling basic properties of the human middle ear and ear canal. part II: Ear canal, middle ear cavities, eardrum, and ossicles. *Acustica*, 84(4):720, 1998. [2.3.2](#)
- [50] A. Hudde and H. Engel. Measuring and modeling basic properties of the human middle ear and ear canal. part III: Eardrum impedances, transfer functions and model calculations. *Acustica*, 84(6):1091, 1998. [2.3.2](#)
- [51] WD. Ward. Handbook of acoustics. Engineering-Computing-and-Technology, pages 1125–1137. New York: John Wiley & Sons, 1998. [2.3.3](#)
- [52] G. von Békésy, EG Wever, and W.T. Peake. Experiments in hearing. *The Journal of the Acoustical Society of America*, 88:2905, 1990. [2.3.3](#)
- [53] M. Cooke. *Modeling auditory processing and organisation*. Cambridge University Press, New York, 1993. [2.3.3](#), [2.6.1](#)
- [54] J. Blauert. Sound-quality evaluation - a multi-layered problem. *Acustica*, 83(5):747–753, 1997. [2.3.3](#)

-
- [55] P.C. Loizou. Mimicking the human ear. *IEEE signal processing magazine*, 15(5):101, 1998. [2.3.3](#)
- [56] J.L. Eriksson and A. Robert. The representation of pure tones and noise in a model of cochlear nucleus neurons. *The Journal of the Acoustical Society of America*, 106(4):1865, 1999. [2.3.3](#)
- [57] N.B. Slepecky. Anatomy of the cochlea and auditory nerve. In *Handbook of acoustics, Engineering-Computing-and-Technology*, pages 1115–1123. New York: John Wiley & Sons, 1998. [2.3.3](#)
- [58] R.D. Patterson. Time-domain modeling of peripheral auditory processing - a modular architecture and a software platform. *The Journal of the Acoustical Society of America*, 98(4):1890–1894, 1995. [2.3.3](#)
- [59] C.J. Plack, A.J. Oxenham, and V. Drga. Linear and nonlinear processes in temporal masking. *Acta acustica united with acustica*, 88(3):348, 2002. [2.3.3](#)
- [60] Y.I. Fishman, D.H. Reser, J.C. Arezzo, and M. Steinschneider. Complex tone processing in primary auditory cortex of the awake monkey. i. neural ensemble correlates of roughness. *The Journal of the Acoustical Society of America*, 108(1):235, 2000. [2.3.3](#), [5.1](#)
- [61] M.R. Leek, M.L. Dent, and R.J. Dooling. Masking by harmonic complexes in budgerigars (*melopsittacus undulatus*). *The Journal of the Acoustical Society of America*, 107(3):1737, 2000. [2.3.3](#)
- [62] ISO226:1987 / UNE 74003:1992. Curvas normalizadas de igual sonoridad. Technical report, AENOR, 1992. [2.5](#)
- [63] S.S. Stevens. On the psychophysical law. *Psychological Review*, 64(3):153–181, 1957. [2.4.1](#)
- [64] S. Buus. Auditory masking. In *Handbook of acoustics, Engineering-Computing-and-Technology*, pages 1147–1165. New York: John Wiley & Sons, 1998. [2.4.2](#)
- [65] P. Noll. Mpeg digital audio coding - setting the standard for high-quality audio compression. *IEEE signal processing magazine*, 14(5):59–81, 1997. [2.4.3](#)
- [66] D. Campbell, E. Jones, and M. Glavin. Audio quality assessment techniques—A review, and recent developments. *Signal Processing*, 89(8):1489–1500, 2009. [2.4.3](#)
- [67] G. Grimm, V. Hohmann, and Verhey J.L. Loudness of fluctuating sounds. *Acta acustica united with acustica*, 88(3):359, 2002. [2.4.4](#), [2.6.3](#)
- [68] H. Traunmuller. Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, 88:97, 1990. [2.4.4](#), [5.5.1](#), [5.5.1](#)
- [69] UNE-EN606051 CEI:651:1979. Sonómetros. Technical report, AENOR, 1990. [2.5](#)

- [70] UNE-EN60804 CEI:60804:2000. Sonómetros integradores-promediadores. Technical report, AENOR, 2002. [2.5](#)
- [71] Brüel & Kjær. Basic frequency analysis of sound. Technical report, Brüel & Kjær, 1998. [2.5.1](#)
- [72] UNE-EN61260 CEI:1260:1995. Filtros de banda de octava y de una fracción de banda de octava, 1997. [2.5.3](#), [2.2](#), [2.3](#)
- [73] D. Ouis. Annoyance from road traffic noise: a review. *Journal of Environmental Psychology*, 21(1):101–120, 2001. [2.5.4](#), [2.6](#), [2.7](#), [4.2.3](#)
- [74] A. Papoulis, S.U. Pillai, and S. Unnikrishna. *Probability, random variables, and stochastic processes*, volume 73660116. McGraw-Hill New York, 2002. [2.5.4](#), [2.5.4](#)
- [75] DIN. 45631 procedure for calculating loudness level and loudness. Technical report, DIN, 1991. [2.6](#), [2.6.3](#)
- [76] ANSI S3.4-2007 Procedure for the Computation of Loudness of Steady Sound. [2.6](#)
- [77] Klaus Genuit. Need for standardization of psychoacoustics. *The Journal of the Acoustical Society of America*, 127(3):1880–1880, 2010. [2.6](#)
- [78] G. Lemaitre, P. Susini, S. Winsberg, S. McAdams, and B. Letinturier. The sound quality of car horns: Designing new representative sounds. *Acta acustica united with acustica*, 95(2):356–372, 2009. [2.6](#)
- [79] G. Lemaitre, P. Susini, S. Winsberg, S. McAdams, and B. Letinturier. The sound quality of car horns: a psychoacoustical study of timbre. *Acta acustica united with acustica*, 93(3):457–468, 2007. [2.6](#), [2.6.6](#)
- [80] C. Spence and M. Zampini. Auditory contributions to multisensory product perception. *Acta acustica united with acustica*, 92(6):1009–1025, 2006. [2.6](#), [2.7](#)
- [81] J.Y. Jeon, J. You, and H.Y. Chang. Sound radiation and sound quality characteristics of refrigerator noise in real living environments. *Applied Acoustics*, 68(10):1118–1134, 2007. [2.6](#), [2.7](#)
- [82] R.P. Leite, S. Paul, and S.N.Y. Gerges. A sound quality-based investigation of the HVAC system noise of an automobile model. *Applied Acoustics*, 70(4):636–645, 2009. [2.6](#), [2.7](#)
- [83] A. Nykanen and A. Sirkka. Specification of component sound quality applied to automobile power windows. *Applied Acoustics*, 70(6):813–820, 2009. [2.6](#), [2.7](#)
- [84] R.B. Raggam, M. Cik, R.R. Holdrich, K. Fallast, E. Gallasch, M. Fend, A. Lackner, and E. Marth. Personal noise ranking of road traffic: Subjective estimation versus physiological parameters under laboratory conditions. *International journal of hygiene and environmental health*, 210(2):97–105, 2007. [2.6](#), [2.7](#)

-
- [85] S. Sandrock, B. Griefahn, T. Kaczmarek, H. Hafke, A. Preis, and T. Gjestland. Experimental studies on annoyance caused by noises from trams and buses. *Journal of Sound and Vibration*, 313(3-5):908–919, 2008. [2.6](#), [2.7](#)
- [86] M. Klemenz. Sound synthesis of starting electric railbound vehicles and the influence of consonance on sound quality. *Acta acustica united with acustica*, 91(4):779–788, 2005. [2.6](#), [2.7](#)
- [87] B.C.J. Moore. Frequency analysis and pitch perception. In *Handbook of acoustics, Engineering-Computing-and-Technology*, pages 1167–1180. New York: John Wiley & Sons, 1998. [2.6.1](#), [2.6.2](#), [2.6.3](#)
- [88] B.C.J. Moore and B.R. Glasberg. A revision of Zwicker’s loudness model. *Acta acustica*, 82(335-345):17, 1996. [2.6.1](#), [2.6.1](#), [2.6.3](#)
- [89] T. Dau, B. Kollmeier, and A. Kohlrausch. Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, 102:2892, 1997. [2.6.1](#), [5.2](#)
- [90] V. Hohmann. Frequency analysis and synthesis using a gammatone filterbank. *Acta acustica united with acustica*, 88(3):433, 2002. [2.6.1](#)
- [91] S. Strahl and A. Mertins. Analysis and design of gammatone signal models. *The Journal of the Acoustical Society of America*, 126:2379, 2009. [2.6.1](#), [2.6.1](#)
- [92] T. Irino and R.D. Patterson. A compressive gammachirp auditory filter for both physiological and psychophysical data. *The Journal of the Acoustical Society of America*, 109:2008, 2001. [2.6.1](#)
- [93] B. Scharf. Loudness. In *Handbook of acoustics, Engineering-Computing-and-Technology*, pages 1181–1195. New York: John Wiley & Sons, 1998. [2.6.3](#)
- [94] M. Ferri, JA Martínez, J. Alba, J. Ramis, and C. Linares. A fuzzy set based approach for the estimation of loudness of environmental noise. *Noise Control Engineering Journal*, 57(5):551, 2009. [2.6.3](#)
- [95] J. Chalupper and H. Fastl. Dynamic loudness model (dln) for normal and hearing-impaired listeners. *Acta acustica united with acustica*, 88(3):378, 2002. [2.6.3](#)
- [96] Jin Yong Jeon and Shin-ichi Sato. Annoyance caused by heavyweight floor impact sounds in relation to the autocorrelation function and sound quality metrics. *Journal of Sound and Vibration*, 311(3-5):767–785, APR 8 2008. [2.6.5](#)
- [97] N. Grimault, S.P. Bacon, and C. Micheyl. Auditory stream segregation on the basis of amplitude-modulation rate. *The Journal of the Acoustical Society of America*, 111:1340, 2002. [2.6.6](#)
- [98] H. von Helmholtz. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. F. Vieweg und sohn, 1870. [2.6.6](#)
-

- [99] J.H. McDermott, A.J. Lehr, and A.J. Oxenham. Individual Differences Reveal the Basis of Consonance. *Current Biology*, 2010. [2.6.6](#)
- [100] E. Terhardt. Ein psychoakustisch begründetes konzept der musikalischen konsonanz. *Acustica*, 36:121–136, 1976. [2.6.6](#)
- [101] P. Daniel. Psychoacoustical roughness. In David Havelock, Sonoko Kuwano, and Michael Vorländer, editors, *Handbook of Signal Processing in Acoustics*, pages 263–274. Springer New York, 2009. [2.6.6](#)
- [102] A. Moreno, C. Simón, F. and delaColina, and A. Marchioni. Factores determinantes en el calculo de la aspereza de señales acústicas por el método de aures. Avila, 1999. [2.6.6](#)
- [103] J. Fels and M. Klemenz. Annoyance Perception of Spatially Distributed Sound Sources. *Acta acustica united with acustica*, 89(3):547–555, 2003. [2.6.6](#)
- [104] M.M.J. Houben, A. Kohlrausch, and D.J. Hermes. Perception of the size and speed of rolling balls by sound. *Speech communication*, 43(4):331–345, 2004. [2.6.6](#)
- [105] G Canévet, S Meunier, A Marchioni, X. Regal, J. L. Carles, and I. López Barrio. Nuevos estudios de validacion subjetiva de los indices de calidad sonora. In *XXX Jornadas Nacionales de Acúsica, Tecniacústica*, volume 99, 1999. [2.6.6](#)
- [106] K. Zimmer, W. Ellermeier, and C. Schmid. Using probabilistic choice models to investigate auditory unpleasantness. *Acta acustica united with acustica*, 90(6):1019–1028, 2004. [2.6.6](#)
- [107] M. Leman. Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model (sim). In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), December*, pages 7–9, 2000. [2.6.6](#)
- [108] Y. Wang. A Study on Sound Roughness Evaluation Based on an Auditory Synchronization Index Model. In *Image and Signal Processing, 2009. CISP'09. 2nd International Congress on*, pages 1–5. IEEE, 2009. [2.6.6](#)
- [109] W. De Baene, A. Vandierendonck, M. Leman, A. Widmann, and M. Tervaniemi. Roughness perception in sounds: behavioral and erp evidence. *Biological Psychology*, 67(3):319–330, 2004. [2.6.6](#)
- [110] E. Zwicker and H. Fastl. *Psychoacoustics* . Springer, Berlin, 1990. [2.6.6](#)
- [111] E. Terhardt. On the perception of periodic sound fluctuations (roughness). *Acustica*, 30(4):201–213, 1974. [2.6.6](#)
- [112] M. Guirao and JM Garavilla. Perceived roughness of amplitude-modulated tones and noise. *The Journal of the Acoustical Society of America*, 60:1335, 1976. [2.6.6](#)

-
- [113] RC Mathes and RL Miller. Phase effects in monaural perception. *The Journal of the Acoustical Society of America*, 19:735, 1947. [2.6.6](#)
- [114] A. Vogel. Roughness and its relation to the time-pattern of psychoacoustical excitation. In *Facts and models in hearing: proceedings of the symposium on psychophysical models and physiological facts in hearing, held at Tutzing, Oberbayern, Federal Republic of Germany*, page 241. Springer, 1974. [2.6.6](#)
- [115] H. Fastl. Roughness and temporal masking patterns of sinusoidally amplitude modulated broadband noise. *Psychophysics and Physiology of Hearing, EF Evans and JP Wilson, eds., Academic: London*, pages 403–414, 1977. [2.6.6](#)
- [116] S. Kemp. Roughness of frequency-modulated tones. *Acustica*, 50(3):126–133, 1982. [2.6.6](#)
- [117] E. Terhardt. Calculating virtual pitch. *Hearing research*, 1(2):155–182, 1979. [2.6.6](#)
- [118] M. Takada, S. Arase, T. Keiichiro, and S. Iwamiya. Economic valuation of the sound quality of noise emitted from vacuum cleaners and hairdryers by conjoint analysis. *Noise Control Engineering Journal*, 57(3):263, 2009. [2.7](#)
- [119] C. Marquis-Favre, E. Premat, and D. Aubree. Noise and its Effects A Review on Qualitative Aspects of Sound. Part II: Noise and Annoyance. *Acta acustica united with acustica*, 91(4):626–642, 2005. [2.7](#)
- [120] S.H. Shin, J.G. Ih, T. Hashimoto, and S. Hatano. Sound quality evaluation of the booming sensation for passenger cars. *Applied Acoustics*, 70(2):309–320, 2009. [2.7](#)
- [121] S.K. Lee, H.W. Kim, and E.W. Na. Improvement of impact noise in a passenger car utilizing sound metric based on wavelet transform. *Journal of Sound and Vibration*, 2010. [2.7](#)
- [122] J.B.J. Fourier. *The analytical theory of heat*. University Press, Cambridge [England], 1878. [3.1](#)
- [123] J.B. Allen. Short-term spectral analysis, synthesis, and modification by discrete fourier-transform. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 25(3):235–238, 1977. [3.1](#)
- [124] J.B. Allen. Corrections to "short term spectral analysis, synthesis, and modification by discrete fourier transform". *IEEE Transactions on Acoustics, Speech and Signal Processing*, 25(6):589–589, Dec 1977. [3.1](#)
- [125] A. Haar. *On the Theory of Orthogonal Functional Systems*. PhD thesis, University of Szeged, 1909. [3.1](#)
- [126] A. Grossmann and Morlet J. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis*, 15(4):723–736, 1984. [3.1](#)

- [127] O. Rioul and M. Vetterli. Wavelets and signal processing, 1991. [3.1](#), [3.2.1](#), [3.2.1](#)
- [128] J. Youngberg and S. Boll. Constant-q signal analysis and synthesis. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'78.*, volume 3, pages 375–378. IEEE, 1978. [3.1](#)
- [129] J. Speiser. Wide-band ambiguity functions. *IEEE Transactions on Information Theory*, 13(1):122–123, 1967. [3.1](#)
- [130] F. Mintzer. Filters for distortion-free two-band multirate filter banks. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33(3):626–630, 1985. [3.1](#), [6](#)
- [131] M. Smith and T. Barnwell III. Exact reconstruction techniques for tree-structured subband coders. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(3):434–441, 1986. [3.1](#)
- [132] P. Burt and E. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on communications*, 31(4):532–540, 1983. [3.1](#)
- [133] PJ Loughlin and Cohen L. The uncertainty principle: Global, local, or both? *IEEE Transactions on Signal Processing*, 52(5):1218, 2004. [3.1](#), [3.2.1](#)
- [134] E. Sejdić, I. Djurović, and J. Jiang. Time-frequency feature representation using energy concentration: An overview of recent advances. *Digital Signal Processing*, 2008. [3.1](#)
- [135] F. Hlawatsch and G. F. Boudreaux-Bartels. Linear and quadratic time-frequency signal representations, 1992. [3.1](#), [3.2.1](#), [3.2.2](#)
- [136] A.V. Oppenheim, A.S. Willsky, and I.T. Young. *Signals and systems*. Englewood Cliffs, N.J., Prentice-Hall, 1983. [3.1.2](#), [3.1.3](#), [5.4.3](#)
- [137] P. Maragos, J.F. Kaiser, and T.F. Quatieri. Energy separation in signal modulations with application to speech analysis. *IEEE Transactions on Signal Processing*, 41(10):3024, 1993. [3.1.3](#)
- [138] YD Wang and R. Kumaresan. Real time decomposition of speech into modulated components. *The Journal of the Acoustical Society of America*, 119(6):EL68, 2006. [3.1.3](#)
- [139] R. Kumaresan and A. Rao. Model-based approach to envelope and positive instantaneous frequency estimation of signals with speech applications. *The Journal of the Acoustical Society of America*, 105(3):1912, 1999. [3.1.3](#)
- [140] D. Qian and Chen S. Joint time-frequency analysis. *IEEE signal processing magazine*, 16(2):52, 1999. [3.2.1](#)
- [141] S. Mallat. *A Wavelet Tour of Signal Processing*. Elsevier Academic Press, 1999. [3.2.1](#)

-
- [142] M.J Bastiaans. A sampling theorem for the complex spectrogram, and gabor expansion of a signal in gaussian elementary signals. *Optical engineering*, 20(4):594–598, 1981. [3.2.1](#)
- [143] I Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005, 1990. [3.2.1](#), [3.2.1](#)
- [144] I Daubechies. A simple wilson orthonormal basis with exponential decay. *SIAM journal on mathematical analysis*, 22(2):554–572, 1991. [3.2.1](#), [3.2.1](#)
- [145] G Battle. Heisenberg proof of the balian low theorem. *Letters in mathematical physics*, 15(2):175–177, 1988. [3.2.1](#)
- [146] O. Rioul. A discrete-time multiresolution theory. *IEEE Transactions on Signal Processing*, 41(8):2591–2606, 1993. [3.2.1](#)
- [147] S.G. Mallat. A theory for multiresolution signal decomposition - the wavelet representation. *IEEE Transactions on pattern analysis and machine intelligence*, 11(7):674–693, 1989. [3.2.1](#)
- [148] S.G. Mallat. Multiresolution approximations and wavelet orthonormal bases of $l_2(\mathbb{R})$. *Transactions of the American Mathematical Society*, 315(1):69–87, 1989. [3.2.1](#)
- [149] Y. Meyer. *Ondelettes et operateurs*. Paris, Hermann, 1990. [3.2.1](#)
- [150] I. Daubechies. 2-scale difference-equations .1. existence and global regularity of solutions. *SIAM journal on mathematical analysis*, 22:1388–1410, 1991. [3.2.1](#)
- [151] I. Daubechies. 2-scale difference-equations .2. local regularity, infinite products of matrices and fractals. *SIAM journal on mathematical analysis*, 23(4):1031–1079, 1992. [3.2.1](#)
- [152] F. Plante, G Meyer, and WA Ainsworth. Improvement of speech spectrogram accuracy by the method of reassignment. *IEEE Transactions on Speech and Audio Processing*, 6(3):282, 1998. [3.2.2](#)
- [153] CG Soares and Z. Cherneva. Spectrogram analysis of the time-frequency characteristics of ocean wind waves. *Ocean engineering*, 32(14-15):1643, 2005. [3.2.2](#)
- [154] D. Sussillo, A Kundaje, and D. Anastassiou. Spectrogram analysis of genomes. *EURASIP journal on applied signal processing*, 2004(1):29, 2004. [3.2.2](#)
- [155] HB Li, WZ Zhao, H Cao, FC Yao, and LY Shao. Measures of scale based on the wavelet scalogram with applications to seismic attenuation. *Geophysics*, 71(5):V111, 2006. [3.2.2](#)
- [156] I. Yesilyurt. The application of the conditional moments analysis to gearbox fault detection - a comparative study using the spectrogram and scalogram. *NDT E international*, 37(4):309, 2004. [3.2.2](#)
-

- [157] ZK Peng, FL Chu, and PW Tse. Detection of the rubbing-caused impacts for rotor-stator fault diagnosis using reassigned scalogram. *Mechanical Systems and Signal Processing*, 19(2):391, 2005. [3.2.2](#)
- [158] I. Yesilyurt. End mill breakage detection using mean frequency analysis of scalogram. *International journal of machine tools manufacture*, 46(3-4):450, 2006. [3.2.2](#)
- [159] O. Rioul and Pflandrin. Time-scale energy-distributions - a general-class extending wavelet transforms. *IEEE Transactions on Signal Processing*, 40(7):1746–1757, 1992. [3.2.2](#)
- [160] J.J. O'Donovan and D.J. Furlong. Perceptually motivated time-frequency analysis. *The Journal of the Acoustical Society of America*, 117:250, 2005. [3.2.2](#)
- [161] G. Matz and F. Hlawatsch. Wigner distributions (nearly) everywhere: time-frequency analysis of signals, systems, random processes, signal spaces, and frames. *Signal processing*, 83(7):1355, 2003. [3.2.2](#)
- [162] N Yen. Time and frequency representation of acoustic-signals by means of the wigner distribution function - implementation and interpretation. *The Journal of the Acoustical Society of America*, 81(6):1841–1850, 1987. [3.2.2](#)
- [163] A.V. Oppenheim, R.W. Schaffer, and R.B. John. *Discrete-time signal processing*. Upper Saddle River, N.J., Prentice Hall, 1999. [3.3.1](#), [3.3.1](#), [3.3.1](#), [3.4](#), [3.4.3](#), [4.4.1](#), [4.4.2](#)
- [164] G. Strang and Nguyen T. *Wavelets and filter banks*. Wellesley, Wellesley-Cambridge Press, 1997. [3.3.2](#)
- [165] G. Strang. Wavelets. *American Scientist*, 82(3):250–255, 1994. [3.3.2](#)
- [166] A. Akansu and R. Haddad. *Multiresolution signal decomposition*. London [etc.], Academic Press, 1992. [3.3.2](#)
- [167] I. Daubechies. *Ten lectures on wavelets*. Society for Industrial Mathematics, 1992. [3.3.2](#)
- [168] P. Abry. *Ondelettes et turbulences : multirésolutions, algorithmes de décomposition, invariance d'échelle et signaux de pression*. Paris [etc.], : Diderot Editeur, Arts & Sciences, 1997. [3.3.2](#)
- [169] A. Teolis. *Computational signal processing with wavelets*. Boston, : Birkhäuser, 1997. [3.3.2](#)
- [170] D. Cabrera. Psysound: A computer program for psychoacoustical analysis. In *Proceedings of the Australian Acoustical Society Conference*, volume 24, pages 47–54, 1999. [3.4](#)

-
- [171] G. Evangelista and S. Cavaliere. Frequency-warped filter banks and wavelet transforms: A discrete-time approach via Laguerre expansion. *IEEE Transactions on Signal Processing*, 46(10):2638–2650, 1998. [3.4](#)
- [172] G. Evangelista and S. Cavaliere. Discrete frequency warped wavelets: theory and applications. *IEEE Transactions on Signal Processing*, 46(4):874–885, 1998. [3.4](#)
- [173] YS Wang. Sound quality estimation for nonstationary vehicle noises based on discrete wavelet transform. *Journal of Sound and Vibration*, 324(3-5):1124–1140, 2009. [3.4](#)
- [174] Y.S. Wang, C.-M. Lee, D.-G. Kim, and Y. Xu. Sound-quality prediction for nonstationary vehicle interior noise based on wavelet pre-processing neural network model. *Journal of Sound and Vibration*, 299(4-5):933 – 947, 2007. [3.4](#)
- [175] B. Carnero and A. Drygajlo. Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms. *IEEE Transactions on Signal Processing*, 47(6):1622–1635, 1999. [3.4](#), [3.4.1](#), [3.4.1](#)
- [176] T. Painter and A. Spanias. Perceptual coding of digital audio. *Proceedings of the IEEE*, 88(4):451–515, 2000. [3.4](#)
- [177] H. Xing, S. Michael, et al. Psychoacoustic music analysis based on the discrete wavelet packet transform. *Journal of Electrical and Computer Engineering*, 2008, 2008. [3.4](#)
- [178] K. Abid and K. Ouni. An improved psycho-acoustic model for MPEG 1 using a Morlet Cambridge wavelet. In *Signals, Circuits and Systems (SCS), 2009 3rd International Conference on*, pages 1–4. IEEE, 2010. [3.4](#)
- [179] He Xing and M.S. Scordilis. An enhanced psychoacoustic model based on the discrete wavelet packet transform. *Journal of the Franklin Institute*, 343(7):738–755, 2006. [3.4](#)
- [180] A. Karmakar, A. Kumar, and RK Patney. A multiresolution model of auditory excitation pattern and its application to objective evaluation of perceived speech quality. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(6):1912–1923, 2006. [3.4](#)
- [181] P. Philippe, F.M. de Saint-Martin, and M. Lever. Wavelet packet filterbanks for low time delay audio coding. *IEEE Transactions on Speech and Audio Processing*, 7(3):310–322, 1999. [3.4](#)
- [182] Y. Shao and C.H. Chang. A Generalized Time–Frequency Subtraction Method for Robust Speech Enhancement Based on Wavelet Filter Banks Modeling of Human Auditory System. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 37(4):877–889, 2007. [3.4](#)
-

- [183] A. Karmakar, A. Kumar, and RK Patney. Design of optimal wavelet packet trees based on auditory perception criterion. *Signal Processing Letters, IEEE*, 14(4):240–243, 2007. [3.4](#)
- [184] K. Abid and K. Ouni. Audio Compression Using a Munich and Cambridge Morlet Wavelet. In *Advances in Multimedia, 2009. MMEDIA'09. First International Conference on*, pages 1–5. IEEE, 2009. [3.4](#)
- [185] A.V. Oppenheim, D. Johnson, and K. Steiglitz. Computation of spectra with unequal resolution using the fast fourier transform. *Proceedings of the IEEE*, 59(2):299–301, feb. 1971. [3.4](#)
- [186] J.O. Smith and J.S. Abel. Bark and ERB bilinear transforms. *IEEE Transactions on Speech and Audio Processing*, 7(6):697–708, 1999. [3.4](#), [3.4.2](#)
- [187] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U.K. Laine, and J. Huopaniemi. Frequency-warped signal processing for audio applications. *J. Audio Eng. Soc.*, 48(11):1011–1031, 2000. [3.4](#)
- [188] M. Parfieniuk and A. Petrovsky. Warped DFT as the basis for psychoacoustic model. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 4. IEEE, 2004. [3.4](#)
- [189] M. Parfieniuk and A. Petrovsky. Approximating the critical bands using warped filter banks based on multiplierless allpass chains. In *International Conference on Signals and Electronic Systems, 2008. ICSES'08.*, pages 351–354. IEEE, 2008. [3.4](#)
- [190] B.P. Bogert, MJR Healy, and J.W. Tukey. The quefrequency alalysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In *Proceedings of the Symposium on Time Series Analysis*, pages 209–243. New York, 1963. [3.4](#)
- [191] E. Teresi, G. De Poli, and F. Ferrero. Mel Cepstrum Parameterization and Classification of Opera Singers' Voices. *Acta acustica united with acustica*, 86(3):568–577, 1997. [3.4](#)
- [192] DG Childers, DP Skinner, and RC Kemerait. The cepstrum: A guide to processing. *Proceedings of the IEEE*, 65(10):1428–1443, 1977. [3.4](#)
- [193] A.V. Oppenheim and R.W. Schafer. From frequency to quefrequency: a history of the cepstrum. *Signal Processing Magazine, IEEE*, 21(5):95–106, 2004. [3.4](#), [4.4.2](#)
- [194] T.N. Patro. Combustion induced powertrain nvh: a time-frequency analysis. *SAE transactions*, 106(6):2545–2555, 1997. [4.1](#), [4.3](#), [4.4.1](#)
- [195] K. Schmillen and J. Wolschendorf. Cycle-to-cycle variations of combustion noise in diesel engines. In Society of Automotive Engineers, editor, *Technical Paper 890129*. SAE, Warrendale, PA, Feb 1989. [4.1](#), [4.2.1](#), [4.4.1](#), [4.4.1](#)

-
- [196] S. Cerdá, J. Romero, J. Navasquillo, and G. Zurita. Análisis tiempo-frecuencia de las resonancias en la combustión. *Revista de acustica*, 32(1-2):17, 2001. [4.1](#), [4.4.1](#), [4.4.1](#)
- [197] AD Jones. Modelling the exhaust noise radiated from reciprocating internal combustion engines: A literature review. *Noise Control Engineering Journal*, 23(1):12–, 1984. [4.1](#), [4.2.3](#)
- [198] SA Amman and M. Das. An efficient technique for modeling and synthesis of automotive engine sounds. *IEEE Transactions on Industrial Electronics*, 48(1):225–234, 2001. [4.1](#), [4.4.1](#)
- [199] D. Anderton. *Engine Noise and Vibration Control. Course Notes*, chapter Basic Origins of Automotive Engine Noise, pages 1–21. University of Southampton, 2003. [4.1](#), [4.2.2](#)
- [200] S.S. Rao. *Mechanical vibrations*. Addison-Wesley, Reading [etc.], 1995. [4.1](#)
- [201] R.C. Hibbeler. *Principles of statics and dynamics*. Pearson Educación, Upper Saddle River, 2006. [4.1](#)
- [202] P. Davies. Ic engine intake and exhaust noise assessment. *Journal of Sound and Vibration*, 223(3):425–, 1999. [4.1](#), [4.2.3](#)
- [203] P. Davies. Piston engine intake and exhaust system design. *Journal of Sound and Vibration*, 190(4):677–, 1996. [4.1](#), [4.2.3](#)
- [204] F. Payri. New methodology for in-cylinder pressure analysis in direct injection diesel engines-application to combustion noise. *Measurement science and technology*, 16(2):540–, 2005. [4.2.1](#)
- [205] W. Li, F. Gu, AD Ball, AYT Leung, and CE Phipps. A study of the noise from diesel engines using the independent component analysis. *Mechanical Systems and Signal Processing*, 15(6):1165–1184, 2001. [4.2.1](#), [4.2.2](#), [4.2.2](#)
- [206] D. Anderton. Relation between combustion system and engine noise. *SAE Paper 790270*, February 1979. [4.2.1](#), [4.2.2](#), [4.2.2](#), [4.2.2](#)
- [207] D. Anderton. *Engine Noise and Vibration Control. Course Notes*, chapter Frequency Domain Studies of Engine Combustion, pages 1–21. University of Southampton, 2003. [4.2.1](#), [4.2.1](#), [4.1](#)
- [208] J.M. Desantes, Torregrosa A.J., and Broatch A. Wavelet transform applied to combustion noise analysis in high-speed di diesel engines. *SAE Trans. J. Engines*, 110:690–700, 2001. [4.2.1](#), [4.2.2](#)
- [209] A.J. Torregrosa, A. Broatch, J. Martin, and L. Monelletta. Combustion noise level assessment in direct injection diesel engines by means of in-cylinder pressure components. *Measurement Science and Technology*, 18(7):2131–2142, 2007. [4.2.1](#), [4.2.2](#)

- [210] R. Hickling, D.A. Feldmaier, F.H.K. Chen, and J.S. Morel. Cavity resonances in engine combustion-chambers and some applications. *The Journal of the Acoustical Society of America*, 73(4):1170–1178, 1983. [4.2.2](#)
- [211] O. Schmitz. Measurement and auralisation of sound radiation from a combustion engine. *Acta acustica united with acustica*, 86(1):183–, 2000. [4.2.2](#)
- [212] M. Lee, J.S. Bolton, and S. Suh. Estimation of the combustion-related noise transfer matrix of a multi-cylinder diesel engine. *Measurement Science and Technology*, 20(1):015106, 2009. [4.2.2](#), [4.2.2](#), [4.2.2](#)
- [213] T. E. Reinhart. *Handbook of Noise and Vibration Control*, chapter 84: Internal Combustion Engine Noise Prediction and Control—Diesel and Gasoline Engines, pages 1024–1033. Wiley, 2007. [4.2.2](#)
- [214] P. Welch. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967. [4.2.2](#), [4.5.4](#)
- [215] X. Liu and RB Randall. Blind source separation of internal combustion engine piston slap from other measured vibration signals. *Mechanical Systems and Signal Processing*, 19(6):1196–1208, 2005. [4.2.2](#)
- [216] Z. Hao, Y. Jin, and C. Yang. Study of engine noise based on independent component analysis. *Journal of Zhejiang University-Science A*, 8(5):772–777, 2007. [4.2.2](#)
- [217] C. Servi re, J.L. Lacoume, and M. El Badaoui. Separation of combustion noise and piston-slap in diesel engine—part ii: Separation of combustion noise and piston-slap using blind source separation methods. *Mechanical Systems and Signal Processing*, 19(6):1218–1229, 2005. [4.2.2](#)
- [218] F. Payri, A. Broatch, X. Margot, and L. Monelletta. Sound quality assessment of Diesel combustion noise using in-cylinder pressure components. *Measurement Science and Technology*, 20:015107, 2009. [4.2.2](#)
- [219] M.E. Badaoui, J. Dani re, F. Guillet, and C. Servi re. Separation of combustion noise and piston-slap in diesel engine—part i: Separation of combustion noise and piston-slap in diesel engine by cyclic wiener filtering. *Mechanical Systems and Signal Processing*, 19(6):1209–1217, 2005. [4.2.2](#)
- [220] P. Davies. Predictive acoustic modelling applied to the control of intake/exhaust noise of internal combustion engines. *Journal of Sound and Vibration*, 202(2):249–, 1997. [4.2.3](#)
- [221] N.S. Dickey. Predictive acoustic modeling applied to the control of intake/exhaust noise of internal combustion engines. *The Journal of the Acoustical Society of America*, 99(4):2460–, 1996. [4.2.3](#)

-
- [222] E. Dokumaci. Prediction of source characteristics of engine exhaust manifolds. *Journal of Sound and Vibration*, 280(3-5):925–, 2005. [4.2.3](#)
- [223] F. Payri. Application of maccormack schemes to ic engine exhaust noise prediction. *Journal of Sound and Vibration*, 195(5):757–, 1996. [4.2.3](#)
- [224] F. Payri, A.J. Torregrosa, and R. Payri. Evaluation through pressure and mass velocity distributions of the linear acoustical description of ic engine exhaust systems. *Applied Acoustics*, 60(4):489–, 2000. [4.2.3](#)
- [225] M. McCallum-Clark, R. Hardy, and M. Hunt. *Transportation and Noise: Land Use Planning Options for a Quieter New Zealand*. Land Transport New Zealand, 2006. [4.2.3](#)
- [226] ISO5130:2007. Acoustics – measurements of sound pressure level emitted by stationary road vehicles. (iso 5130:2007). [4.2.3](#)
- [227] L.B. Jackson. *Digital filters and signal processing 2nd ed.* Kluwer Academic Publishers Boston, 1989. [4.4.2](#)
- [228] A. Camacho, G. Piñero, M. de Diego, and A. Gonzalez. Exploring roughness perception in car engine noises through complex cepstrum analysis. *Acta acustica*, 94(1):130–140, 2008. [4.4.2](#), [5.1](#), [5.4.1](#), [5.5](#)
- [229] A. Camacho, G. Piñero, M. de Diego, and M. Ferrer. On the use of complex cepstrum in psychoacoustic evaluations of engine noise. In *11th International congress on sound and vibration proceedings*, pages 2997–3004, San Petersburgo, Rusia, Julio 2004. International Institute of Acoustics and Vibration IIAV. [4.4.3](#), [5.2.3](#), [5.5](#)
- [230] H.A. David. The method of paired comparison. *London: Griffin*, 1988. [4.4.3](#), [5.3.5](#), [5.4.2](#), [5.4.2](#), [5.6](#)
- [231] RA Bradley and ME Terry. The rank analysis of incomplete block diagrams. i. the method of paired comparisons. *Biometrika*, 39:324–345, 1952. [4.4.3](#), [5.3.5](#), [5.4.2](#), [5.6](#)
- [232] A. Miskiewicz, A. Rakowski, and T. Rosciszewska. Perceived roughness of two simultaneous pure tones. *Acta acustica united with acustica*, 92(2):331–336, 2006. [5.1](#)
- [233] D. Pressnitzer and S. McAdams. Two phase effects in roughness perception. *The Journal of the Acoustical Society of America*, 105:2773, 1999. [5.1](#), [5.2](#), [5.2.3](#)
- [234] D. Pressnitzer and S. Mcadams. An effect of the coherence between envelopes across frequency regions on the perception of roughness. *Psychophysics, Physiology and Models of Hearing. Singapore: World Scientific*, pages 105–108, 1999. [5.1](#)
- [235] G. Krantz, B. Merker, and G. Madison. Subjective reactions to musical intervals assessed by rating scales. In *Proceedings of The Eighth International Conference on Music Perception and Cognition*, 2004. [5.1](#)
-

- [236] W.M. Hartmann. *Signals, sound, and sensation*. Amer Inst of Physics, 1997. [5.2](#), [5.2.1](#)
- [237] SL Marple. Computing the discrete-time analytic signal via FFT. *IEEE Transactions on Signal Processing*, 47(9):2600–2603, 1999. [5.2](#)
- [238] A. Camacho, G. Piñero, L. Fuster, and M. de Diego. Evaluation of human perception of roughness for diesel car noise. In *10th International congress on sound and vibration proceedings*, Estocolmo, Suecia, Julio 2003. International Institute of Acoustics and Vibration IIAV. [5.2.1](#), [5.4](#), [5.5](#)
- [239] M. Ferrer, M. de Diego, G. Piñero, and González A. Evaluación subjetiva de la calidad del sonido mediante test de parejas. In *Proc. of Tecniacústica in CD-ROM*, pages 1–7, Valladolid, 1999. [5.3.4](#)
- [240] N. Otto, S. Amman, C. Eaton, and S. Lake. Guidelines for jury evaluations of automotive sounds. *Sound & vibration*, 35(4):24–47, 2001. [5.3.5](#), [5.4.2](#), [5.6](#)
- [241] N.C.. Otto and G.H. wakefield. A subjective evaluation and analysis of automotive starter sounds. *Noise Control Engineering Journal*, 41(3):377–382, 1993. [5.4.2](#), [5.4.2](#)
- [242] F.J. Harris et al. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978. [5.5.1](#)
- [243] O. Fobel and T. Dau. Searching for the optimal stimulus eliciting auditory brainstem responses in humans. *The Journal of the Acoustical Society of America*, 116:2213, 2004. [5.6.2](#)
- [244] D. Esteban and C. Galand. Application of quadrature mirror filters to split band voice coding schemes. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP 77.*, 2:191–195, May 1977. [6](#)
- [245] M. Smith and T. Barnwell. A procedure for designing exact reconstruction filter banks for tree-structured subband coders. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP 84.*, 9:421–424, 1984. [6](#)