

Tesis Doctoral

Estudio de la percepción tímbrica en sonidos con modulación mediante experimentos psicofísicos y modelado de la periferia auditiva

Riera, Pablo Ernesto

2015-03-30

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Riera, Pablo Ernesto. (2015-03-30). Estudio de la percepción tímbrica en sonidos con modulación mediante experimentos psicofísicos y modelado de la periferia auditiva. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Riera, Pablo Ernesto. "Estudio de la percepción tímbrica en sonidos con modulación mediante experimentos psicofísicos y modelado de la periferia auditiva". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2015-03-30.

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales
Departamento de Física

Estudio de la percepción tímbrica en sonidos con modulación mediante experimentos psicofísicos y modelado de la periferia auditiva

Tesis presentada para optar al título de
Doctor de la Universidad de Buenos Aires en el área Ciencias Físicas

Por
Lic. Pablo Ernesto Riera

Director de tesis: Dr. Manuel Camilo Eguía

Consejero de estudios: Dr. Mariano Sigman

Lugar de trabajo: Laboratorio de Acústica y Percepción Sonora,
Departamento de Ciencias Sociales, Universidad Nacional de Quilmes.

Buenos Aires, 30 de Marzo de 2015

Estudio de la percepción tímbrica en sonidos con
modulación mediante experimentos psicofísicos y
modelado de la periferia auditiva

Pablo Ernesto Riera

Buenos Aires, 2015

Resumen

La percepción tímbrica es uno de los fenómenos más complejos que realiza el sistema auditivo. El timbre, originalmente establecido como la cualidad que permite discriminar instrumentos musicales entre sí, en la actualidad se encuentra más vinculado con el concepto ecológico de identificación de las fuentes sonoras en un entorno natural. Su inherente carácter multidimensional, hace que la relación entre la forma de onda de un sonido y el timbre percibido sea una relación compleja, en la cual distintas partes del sistema auditivo están involucradas. A pesar de estas limitaciones, es sabido que hay ciertos atributos acústicos temporales y espectrales que suelen destacarse en la percepción tímbrica. Los más relevantes son, por un lado, el tiempo de ataque de la envolvente como característica temporal y por otro el brillo o centroide espectral como característica espectral.

En este trabajo se utilizaron dos metodologías para estudiar el timbre. Por un lado, se realizaron dos experimentos psicofísicos de comparación de pares de sonidos y, por otro, los timbres fueron estudiados mediante técnicas de procesamiento de señales y el modelado de la periferia auditiva. Uno de los experimentos fue realizado utilizando sonidos de amplitud modulada sinusoidalmente como estímulos y llevado a cabo masivamente a través de Internet, mientras que para el otro experimento, se utilizaron multifónicos del saxo alto en un ambiente controlado. Estas dos familias de sonidos poseen modulaciones como atributo temporal característico a diferencia de los que son utilizados en muchos de los estudios de percepción tímbrica donde la principal característica temporal es el tiempo de ataque.

Los experimentos demostraron que la modulación temporal fue un atributo saliente junto con el centroide espectral para ambos conjuntos de sonidos. Además, se observó que sonidos que se encontraban agrupados en un espacio de parámetros de control se encontraban agrupados también, en la mayor parte de los casos, en los correspondientes espacios perceptuales. En cambio, las distancias entre los grupos en estos espacios se vieron modificadas.

Por otra parte, los sonidos fueron analizados en base al procesamiento de sus señales y mediante simulaciones de la periferia auditiva. Para esto se desarrolló un modelo de la periferia auditiva con énfasis en la mecánica coclear, donde se optó por un modelo dinámico no lineal y con alimentación lateral. Este modelo fue desarrollado teniendo en consideración la capacidad de reproducir los factores dominantes de la transducción coclear como la saturación en intensidad, el mecanismo activo de amplificación y las curvas de enmascaramiento entre otros. Los parámetros fueron ajustados de manera satisfactoria para reproducir respuestas clásicas de experimentos psicoacústicos en humanos y datos de experimentos fisiológicos *in vivo* en mamíferos.

Palabras Clave: psicoacústica, percepción tímbrica, sistema auditivo, mecánica coclear.

Studies of timbre perception of modulated sounds through psychophysics experiments and auditory periphery models

Abstract

Timbre perception is one of the most complex tasks the auditory system performs. Timbre, originally defined as the quality that allows the discrimination of musical instruments, it is nowadays more related to the ecological concept of the identification of sound sources in natural environments. Because of its inherently multidimensional character, the relationship between the sound wave and the perceived timbre is a complex matter, where different parts of the auditory systems are involved. Besides this limitations, it is known that there are certain temporal and spectral acoustic features that are fundamental to the perception of timbre. The most relevant are the attack time of the envelope as a temporal feature and the brightness or spectral centroid as a spectral feature.

In this work we use two methodologies for studying timbre. Firstly, two psychophysical experiments of sound pairs comparison were done, and secondly the timbres were analyzed by signal processing techniques and through a model of the auditory periphery. One of the experiments was carried out using sinusoidally amplitude modulated sounds as stimuli, massively through internet and the other experiment was done using alto saxophone multiphonics but in a controlled environment. These two families of sounds have modulations as a characteristic temporal feature which is not present on many timbre studies, where the attack time is more common.

The experiments showed that the temporal modulation was a salient feature, along with the spectral centroid for both sound families. In addition, it was observed that sounds grouped in a control parameter space were also grouped, in most cases, in the corresponding perceptual spaces. In contrast, it were the distances between the groups within these spaces that changed.

Also the sounds were analysed by signal processing techniques and through auditory periphery simulations. To this end, we developed an auditory periphery model with emphasis on the cochlear mechanics, using a non linear dynamical model with feedforward and feedbackward properties. This model was developed with the aim of reproducing dominant factors of cochlear transduction, for example, the saturation with intensity, the active mechanism of amplification and the masking curves among other capabilities. The parameters where successfully adjusted to reproduce typical human psychoacoustical experiments and data from physiological *in vivo* mammals experiments.

Keywords: psychoacoustics, timbre perception, auditory system, cochlear mechanics.

Índice general

1. Introducción	1
1.1. Definiciones del timbre	2
1.1.1. Espacios tímbricos	6
1.1.2. Atributos acústicos	8
1.2. Experimentos de percepción tímbrica	9
1.2.1. Identificación de timbres	9
1.2.2. Comparación de timbres y espacios tímbricos	11
1.3. Procesamiento del sonido en la periferia auditiva	16
1.3.1. El oído interno y la descomposición en frecuencias	16
1.3.2. Filtros auditivos	17
1.4. Percepción de tonos con modulación	22
1.4.1. Tonos SAM	25
1.4.2. Rugosidad	27
2. Multifónicos del saxofón	33
2.1. Introducción	33
2.2. Análisis acústico de los multifónicos	41
2.2.1. Procedimientos	41
2.2.2. Resultados	43
2.2.3. Resumen	48
2.3. Experimento de comparación de multifónicos	48
2.3.1. Desarrollo experimental	49
2.3.2. Resultados	50
2.3.3. Resumen	53
3. Tonos de amplitud modulada sinusoidalmente	55
3.1. Introducción	55
3.2. Método experimental	57
3.3. Resultados	60
3.4. Discusión	68
4. Modelado de la periferia auditiva	71
4.1. Introducción	72
4.1.1. Oído externo y medio	72

4.1.2. Oído interno y la cóclea	73
4.1.3. Mecánica coclear	77
4.2. Modelo de la cóclea	83
4.2.1. Sistema pasivo: osciladores con acoplamiento hidrodinámico	84
4.2.2. Sistema activo: alimentación lateral	89
4.2.3. Oído medio	92
4.2.4. Células ciliadas internas y sinapsis	92
4.3. Resultados numéricos	93
4.3.1. Cóclea	93
4.3.2. Células ciliadas internas y sinapsis	99
4.4. Resultados con tonos modulados	100
4.4.1. Modelo temporal de la rugosidad	102
4.4.2. Espacios tímbricos de modulación	108
4.5. Resumen y discusión	113
5. Conclusiones generales	115
A. Apéndices	119
A.1. Escalamiento multidimensional	119
A.2. Cocleograma de banco de filtros <i>gammatones</i>	121
A.3. Modelo espectral de rugosidad	122
A.4. Hidrodinámica de la cóclea	123
A.5. Relación de dispersión del sistema activo	126
Bibliografía	131
Agradecimientos	145

Capítulo 1

Introducción

I think the definition of timbre by the American Standards Association should be this: "We do not know how to define timbre, but it is not loudness and it is not pitch." Albert S. Bregman.

Esta tesis esta dedicada al estudio del timbre en sonidos con modulación en amplitud. El abordaje de este tema se realiza tanto desde el análisis de los parámetros acústicos, como de la realización de experimentos psicofísicos y el modelado de la respuesta de la periferia auditiva. Teniendo en cuenta la variedad de enfoques a adoptar, este capítulo esta dedicado a introducir las herramientas que se utilizaran en cada uno de estos enfoques y a revisar los antecedentes en la literatura en los campos de estudio respectivos. La introducción está organizada de la siguiente manera: primero se establece el concepto de timbre y las problemáticas que subyacen al tratar de definirlo, luego, se exponen las diferentes clases de experimentos psicofísicos que existen para estudiar el timbre y en particular los métodos usados en este trabajo y finalmente se detallan las principales características del procesamiento del sonido en el sistema auditivo que influyen la percepción tímbrica.

1.1. Definiciones del timbre

El uso de la palabra timbre intenta identificar de una manera un tanto sencilla y vaga a un conjunto complejo de fenómenos psicoacústicos y musicales. Una manera práctica de definir el timbre es diciendo lo que no es. En ese sentido, el timbre abarcaría los parámetros de la percepción que no están dados por la altura musical, la intensidad sonora, la posición espacial, la duración e incluso las características del espacio como la reverberación y otros fenómenos. Esto se ve reflejado en la definición dada por el *American National Standards Institute* [56]: “El timbre es aquella característica del sentido auditivo por el cual un sujeto puede evaluar dos sonidos con la misma nota y misma intensidad como distintos”.

Por otro lado, es difícil tratar de definir el timbre por lo que sí es, dado que tiene un carácter multidimensional. No es posible medir el timbre utilizando una sola dimensión a diferencia de lo que ocurre con otros atributos sonoros, como por ejemplo la altura musical que permite ordenar los sonidos de mas graves a mas agudos en una escala.

Aunque históricamente se utilizó la palabra timbre para describir aquello que permite distinguir los instrumentos musicales entre sí, actualmente entendemos que el timbre tiene varias acepciones diferentes. Podemos empezar mencionando dos enfoques básicamente distintos de definir el timbre [77]

1. El timbre hace referencia a un gran conjunto de atributos perceptuales, algunos que pueden ser variados continuamente (por ejemplo el brillo, velocidad de ataque, armonicidad, etc) y otros que son más bien discretos o categóricos (p. ej. las consonantes en el habla)
2. El timbre es la principal herramienta para el reconocimiento, identificación y seguimiento en el tiempo de las fuentes sonoras (por ejemplo lo que permite distinguir al viento, un perro ladrando, un clarinete, etc.).

La primera definición relaciona al timbre con un proceso de construcción en el cual el sonido es analizado en base a su información acústica a través de los atributos perceptuales que son extraídos por el sistema nervioso. Podemos pensar de manera idealizada que esta definición es la que rige en los ambientes controlados de la música y de los laboratorios,

en los cuales el sonido se puede modificar de manera precisa para controlar el timbre. Las técnicas de síntesis permiten cambiar el timbre del sonido de manera continua.

La segunda definición, en cambio, concibe al timbre como aquella cualidad del sonido que permite establecer las identidades de las fuentes sonoras en lo cotidiano y en los ambientes naturales. Esto se asocia con la idea de que en la naturaleza, cada fuente sonora posee restricciones acústicas que limitan la posibilidad de emitir ciertos sonidos. En contraposición a la primera definición, en este caso se plantea una experiencia sensorial más bien discreta que permite la categorización *directa* de los sonidos y vincularlos con referencias externas a él [35].

Ambas definiciones permiten generar categorías de sonidos. Tomando como base la primera, los sonidos pueden ser agrupados en categorías a partir de los atributos que poseen. Puede realizarse mediante un mecanismo de emparejamiento de los atributos perceptuales de dos sonidos en busca de coincidencias y utilizando un nivel de umbral para determinar si dos sonidos pertenecen o no a la misma categoría [111]. En la segunda definición las categorías están dadas por el hecho de que la información sonora recibida implica que existe una causa que generó el sonido. Este mecanismo se llama inferencia por causalidad, y sugiere que hay un fuerte vínculo entre la representación mental del evento sonoro y el objeto que lo produjo.

Estas definiciones pueden relacionarse con dos tipos de escuchas. Schaeffer [103] acuñó el término de *escucha reducida* para referirse a una escucha no referencial, librada de los indicios de la fuente sonora. Luego, en la misma línea de trabajo, Chion and Schaeffer [15] establecieron como *escucha causal* a la acción de dar un significado a lo que se escucha.

Las dos definiciones corresponden a dos usos distintos del timbre. No obstante, ambas pueden ser aplicadas para estudiar una misma situación sonora, resaltando diferentes aspectos.

Tomemos el caso de un sintetizador electrónico analógico. Este instrumento por un lado permite ejercer un control preciso y continuo de varias magnitudes, lo cual se asocia al primer enfoque del timbre, mientras que al mismo tiempo posee un sonido característico asociado a su electrónica y al sistemas de reproducción, lo cual lo individualiza como fuente sonora y está mas asociado al segundo enfoque. Otro ejemplo posible sería el de un

pequeño ensamble de instrumentos orquestales, donde las múltiples fuentes pueden sonar de forma simultánea y permiten generar un continuo de sonoridades complejas y obtener cierta precisión para el control del timbre, obteniendo un sonido resultante, que no permite identificar cada fuente por separado. Por lo que en este caso, es útil el primer enfoque del timbre ya que el reconocimiento de las fuentes queda ofuscado. Estos ejemplos ilustran que dependiendo de cada situación sonora una definición podrá ser más útil o no que la otra.

Un punto de contacto entre las dos posturas puede establecerse en el marco de la evolución de los sistemas sensoriales. Uno de los objetivos de los sistemas sensoriales es establecer representaciones coherentes de los objetos del mundo externo. Estos objetos tienen propiedades físicas inherentes y el sistema auditivo en particular se desarrolló bajo la presión evolutiva de ser capaz de extraer estas propiedades. Para ilustrar este punto podemos tomar como ejemplo a los sistemas acústicos fuente - filtro [48]. Numerosas fuentes sonoras quedan bien descritas por estos sistemas, como es el caso de la voz humana y la gran mayoría de los instrumentos musicales. En estos sistemas podemos identificar a la fuente con la parte del instrumento que transforma la energía suministrada en vibraciones y al filtro como una característica morfológica del objeto que filtra a la fuente. El sistema auditivo bien pudo desarrollarse con la ventaja evolutiva de que diferenciar la fuente del filtro facilita en la mayor parte de los casos la identificación de la fuente sonora. Esta información debe ser extraída mediante algún mecanismo en base a distintos atributos acústicos para la fuente y el filtro. Esta visión es compatible con las dos definiciones de timbre, es decir, la fuente sonora es identificable en base a una conexión entre el evento sonoro y los atributos acústicos ligados a la naturaleza de aquel evento.

En el contexto del estudio de los instrumentos musicales también suele hablarse de los sistemas fuente-filtro, donde a la fuente se la refiere como el excitador y al filtro como el resonador. Esto permite generar las clásicas familias de instrumentos en base a cómo estos son ejecutados y cuáles son los elementos que resuenan. Es posible generar lo que se llama un agrupamiento jerárquico [50], donde hay grupos generales y subgrupos ¹. Esta herramienta de agrupamiento jerárquico ha sido utilizada para el estudio de la relación entre el timbre y la causalidad del evento sonoro [24].

¹un ejemplo de esto es la taxonomía Hornbostel-Sachs[116] de los instrumentos musicales

En cierta manera podemos pensar en un continuo de definiciones del timbre. Hablar del timbre en base a los atributos acústicos es similar a una descripción de proximidad o de cercanía, en cambio hablar de la fuente sonora corresponde a una descripción distante o lejana [48]. Si observamos un sonido con lupa podremos desarmarlo en todas las variables que lo componen, en cambio si lo vemos a la distancia podremos distinguir su identidad sin poner atención en los detalles.

Esta diferenciación entre observar en detalle y globalmente ocurre con otras modalidades sensoriales. Por ejemplo, se podría comparar al término timbre con el término *apariencia*, el cual puede tener muchas acepciones según el tipo de observación y recae en la misma problemática que el timbre. Podemos trazar un paralelismo entre el sentido de la audición y el de la visión, tomando la regla de que el color se corresponde a la nota musical y la forma al timbre. Un piano y un violín tocando la misma nota podría compararse con un cuadrado y un triángulo del mismo color. O en cambio, dos sonidos provenientes de un mismo instrumento con diferentes notas pero timbres muy parecidos podrían compararse con dos círculos (casi idénticos) que tienen distinto color. [48]

Es evidente que hay ciertas cuestiones que pueden ser un poco arbitrarias a la hora de marcar un paralelismo entre dos modalidades sensoriales. Dependiendo del conjunto de objetos con los que se cuente, se podrá trazar uno u otros paralelismos más o menos adecuados en cada caso. Un caso interesante es el de ciertas palabras que se suelen usar para describir atributos tímbricos pero que no necesariamente se corresponden con su origen. Por ejemplo el brillo de un sonido suele referirse en la mayoría de los casos a la presencia de frecuencias altas, cuando desde el punto de vista de la luz, el brillo, está más relacionado a la intensidad. Una posible explicación para esta correspondencia puede darse si consideramos que muchos instrumentos musicales suelen tener más brillo cuando son tocados intensamente.

Existen muchas palabras que se aplican en las descripciones tímbricas, llamados atributos verbales [112]. En muchos casos se basa en utilizar escalas semánticas, brillante-opaco, ordenado-caótico, natural-artificial, suave-rugoso [42], y en otros casos palabras aisladas, metálico, profundo, filoso, etc [31].

A partir del siguiente ejemplo podemos ilustrar algunas de las diferentes apreciaciones que se mencionaron sobre el timbre. En la figura 1.1 se muestra un conjunto de sonidos

organizados de manera jerárquica en grupos y subgrupos. El conjunto de sonidos fue tomado para ejemplificar ciertas relaciones, pero la manera de agruparlos no es única. Dado que hay que imaginarse cómo suenan estos sonidos, suponemos que los instrumentos musicales son ejecutados de manera canónica ² y para los sonidos que no son de instrumentos suponemos que son ejemplos claros y reconocibles de lo que su nombre indica.

El primer nivel distingue sonidos armónicos que poseen altura musical de aquellos que son impulsivos o ruidosos y no poseen una altura definida. El siguiente nivel distingue sonidos por las familias de instrumentos orquestales típicas a las que pertenecen, las cuerdas, los metales y los de percusión. El siguiente nivel agrupa sonidos en base a sus cualidades espectrales promedio, donde por ejemplo dos violines aunque distintos objetos físicos pueden ser considerados similares con respecto a la organización de los armónicos y el ruido de la lluvia se puede considerar que tiene un perfil espectral similar al del ruido rosa. Luego en el siguiente nivel se considera a la fuente sonora como el delimitador, por lo que se puede distinguir entre el timbre del violín A y del B.

Este diagrama de ejemplo busca mostrar de forma resumida que es posible generar muchos distintos tipos de agrupaciones de sonidos mediante las diferentes definiciones del timbre. Otra forma de generar estructuras de relaciones entre timbres es mediante los denominados *espacios tímbricos*.

1.1.1. Espacios tímbricos

Los espacios tímbricos permiten representar un conjunto de sonidos mediante posiciones, especificadas como coordenadas en dichos espacios. Pueden construirse asociando una dimensión para cada atributo acústico o perceptual u otros tipos de mapeo entre sonidos y sistemas de coordenadas.

Por otra parte es posible generar espacios tímbricos mediante experimentos psicofísicos. En este caso las dimensiones se obtienen de procesos cognitivos que pueden corresponder o no con atributos acústicos o con combinaciones de ellos.

La construcción de estos espacios puede servir para el estudio del timbre desde el punto

²Para las cuerdas frotadas supondremos que son tocadas con arco y que todos los sonidos son ejecutados con una dinámica intermedia

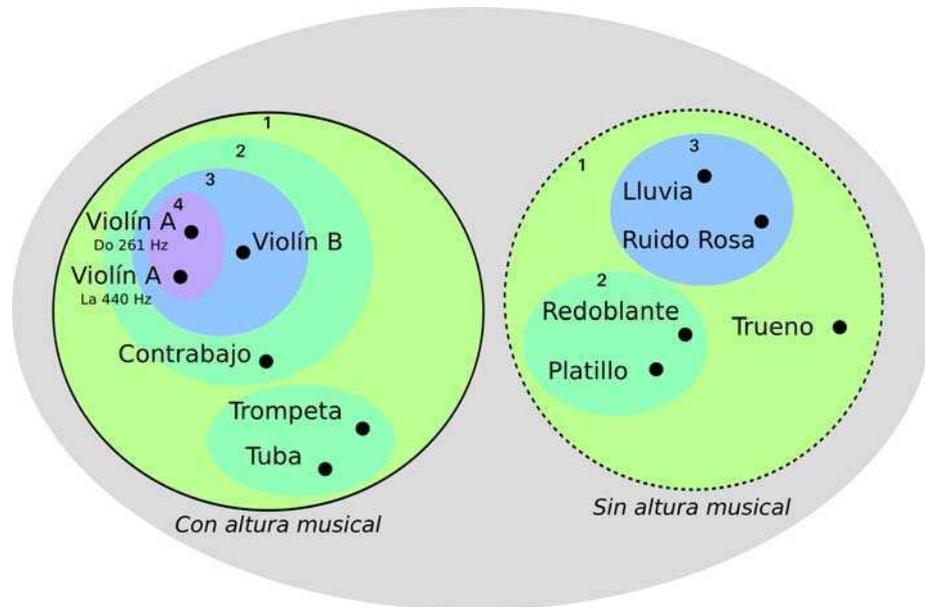


Figura 1.1: Conjunto de sonidos organizados forma jerárquica. Nivel 1: distingue sonidos con o sin altura musical, izquierda y derecha respectivamente. Nivel 2: sonidos que pertenecen a una misma familia de instrumentos musicales. Nivel 3: sonidos con cualidades espectrales similares. Nivel 4: sonidos provenientes de la misma fuente sonora.

de vista acústico solamente [53], para el estudio desde el punto de vista perceptual [78] o para el uso de estos espacios como control musical [120].

Un desafío que proponen los espacios tímbricos es suponer que existen dimensiones continuas en las cuales el timbre puede representarse. Los atributos acústicos toman por lo general valores continuos y es matemáticamente posible pensar en una transición entre un timbre y otro. Al hablar de dimensiones perceptuales, el espacio de representación interno puede no ser suave y la percepción puede ser categórica, a pesar de que exista una representación física continua [106]. Al mismo tiempo, en los espacios definidos por atributos acústicos, es posible definir una distancia entre dos puntos, en este caso dos sonidos. En cambio en un espacio perceptual la noción de distancia puede no ser tan clara e incluso muy dependiente del sujeto y el conjunto de sonidos.

1.1.2. Atributos acústicos

Existen numerosos atributos acústicos que pueden extraerse de un sonido [123], muchos de los cuales son relevantes para la percepción tímbrica. Podemos agrupar estos atributos relevantes en tres grupos, los puramente temporales, los puramente espectrales y los que son espectrotemporales.

Los atributos temporales son aquellos que están relacionados exclusivamente con la envolvente temporal de la amplitud del sonido. Esta suele dividirse en cuatro segmentos, el ataque, la transición entre ataque y la parte estacionaria, la parte estacionaria y el decaimiento final [46]. De estos el que más influencia al timbre es el tiempo de duración del ataque, aunque ciertamente la forma exacta de la envolvente posee es parte del timbre.

Los atributos espectrales están relacionados a las estadísticas de la magnitud del espectro. El más prominente es el valor medio del espectro (centroide espectral) dado que permite una descripción básica en la escala grave-agudo. El valor del centroide espectral correlaciona con la sensación de un sonido filoso o punzante (del inglés *perceptual sharpness*), que se calcula como el valor medio sobre el espectro pero ponderado dando mayor peso a las frecuencias altas [30]. Luego existen varias otras medidas que han sido relevantes en diferentes estudios como la suavidad de la envolvente espectral, también llamada irregularidad espectral, que serviría para cuantificar pozos en el espectro y permite distinguir sonidos con todos los armónicos de aquellos con solo los impares como el caso del clarinete [79].

Los atributos espectrotemporales están dados por aquellas magnitudes del espectro que varían en el tiempo. Las magnitudes mencionadas en el párrafo anterior bien pueden variar en el tiempo y generar características tímbricas particulares. Una de las más relevantes es la del flujo espectral, que indica el flujo de energía de una parte del espectro a otra en el tiempo. En el caso de un sonido armónico, es posible medir las envolventes de amplitud de cada armónico y sus frecuencias en función del tiempo y medir las variaciones en las trayectorias. En muchos sonidos de instrumentos hay marcas tímbricas particulares dadas por estas variaciones. Por ejemplo en los instrumentos de doble caña los armónicos individuales comienzan de manera sincrónica mientras que en los de la familia de los metales se ven

asincronías, los armónicos graves comienzan primero y luego los más agudos, lo que genera un flujo espectral. Finalmente los sonidos impulsivos como el de un piano o una guitarra suelen estar caracterizados por un ataque muy corto donde la mayoría de los armónicos comienzan juntos pero luego la energía se traslada hacia los más graves generando un flujo en la otra dirección.

1.2. Experimentos de percepción tímbrica

Existen numerosos experimentos sobre los distintos aspectos de la percepción tímbrica [77, 25, 47]. Muchos de estos experimentos están orientados a identificar los atributos acústicos que le otorgan a un sonido su timbre característico. Sin embargo la mayor parte de ellos se han centrado de forma exclusiva en la tímbrica de los instrumentos musicales occidentales.

En esta sección nos referiremos a dos grandes clases de experimentos. La primera de ellas está orientada a la identificación de sonidos modificados. En estos experimentos se editan ciertos atributos de forma selectiva y se evalúa si los sujetos mantienen un nivel similar de identificación comparado con los sonidos originales. En la segunda clase se encuentran los experimentos en los que los sujetos juzgan la similitud de pares de sonidos en base a sus características tímbricas para luego construir espacios tímbricos a partir de técnicas de escalamiento multidimensional.

1.2.1. Identificación de timbres

Una de las primeras preguntas planteadas acerca del timbre de instrumentos musicales estuvo orientada a dilucidar si era el ataque de un sonido o su parte estacionaria lo que más influía en su identificación. El trabajo de Hajda [45] fue realizado con sonidos de doce instrumentos impulsivos o continuos (en *stacatto* y sostenidos) en su versión original, editados para incluir sólo el ataque o la parte estacionaria, e invertidos temporalmente. Los participantes, luego de escuchar todos los fragmentos, realizaban una elección forzada entre distintas alternativas para identificar el instrumento correspondiente a cada uno de los fragmentos.

Los sonidos en su versión original fueron identificados correctamente el 93 % de las veces. En el caso de los sonidos continuos sostenidos, para las condiciones en las que se removió el ataque se observó un nivel de identificación mayor que para aquellas en las cuales se removió la parte sostenida y dicho nivel fue similar que el observado para los sonidos originales. En cambio, los sonidos continuos en *staccato* fueron mejor identificados para la condición de solo ataque (fijo en 80 ms) que para la condición de solo estacionario. Los sonidos impulsivos no fueron identificados en reversa mientras que si lo fueron los sonidos continuos. Este trabajo concluyó que la discusión sobre el ataque o la parte sostenida no puede aplicarse de manera clara a los sonidos impulsivos ya que no pueden ser particionados de una manera precisa por no tener un parte estacionaria. Finalmente para los sonidos continuos sostenidos la parte estacionaria fue suficiente para la identificación y los indicios estarían en las relaciones de las trayectorias del centroide espectral y la envolvente de amplitud.

En los trabajos de Charbonneau [11], Grey and Moorer [40] y McAdams et al. [79] en lugar de estudiar la identificación de la fuente de los sonidos, se estudió la discriminación entre sonidos originales y sonidos con características espectrotemporales simplificadas mediante métodos de síntesis y resíntesis. En particular en el trabajo de McAdams et al. [79] se encontró que solo ciertas simplificaciones del espectro daban como resultado una mayor discriminación entre el sonido original y el modificado. Si bien estas modificaciones variaban para cada instrumento aquellas que produjeron mayor discriminación fueron tanto las que simplificaban la envolvente del espectro (haciéndolo más regular), como las que simplificaban la envolvente de amplitud de los armónicos, reduciendo el flujo espectral.

Con una perspectiva similar, pero utilizando sonidos no musicales, en el trabajo de Geffen et al. [33], se estudiaron qué parámetros espectrotemporales eran los que permiten identificar el sonido del agua en movimiento. Luego de analizar y modificar diversos parámetros espectrotemporales los autores concluyeron que la característica fundamental del sonido del agua fluyendo (en términos de su identificación) la constituía la invariancia de escala temporal. Esto es, aquellas características que no se modifican al reproducir el sonido a diferentes velocidades y permiten que siga siendo identificado. Mientras que si esa invariancia temporal era perturbada los sonidos eran percibidos como poco realistas.

1.2.2. Comparación de timbres y espacios tímbricos

La otra clase de experimentos que describiremos es la de comparaciones de pares y, posee un interés espacial, ya que será la utilizada para los dos experimentos de la presente tesis.

Este tipo de experimentos se basa en las relaciones entre timbres, a diferencia de los anteriores que eran sobre la identificación. Para esto, los sujetos deben juzgar las disimilitudes entre dos sonidos a partir de una escala que suele tomar valores entre timbres muy similares y timbres muy disímiles. Estas disimilitudes se pueden pensar como las distancias entre sonidos en el espacio perceptual donde se representan los timbres, que a priori puede ser un espacio de muchas dimensiones. Para poder simplificar el análisis se utiliza la técnica de escalamiento multidimensional (MDS de sus siglas en inglés, ver apéndice A.1), que permite generar un espacio de baja dimensionalidad que puede ponerse en correspondencia con el espacio tímbrico perceptual. De esta manera es posible representar el grupo de sonidos mediante una configuración geométrica sencilla.

En la figura 1.2 se representa el proceso de este tipo de experimentos. Primero se constituye un conjunto de sonidos con el cual se quieren estudiar las relaciones tímbricas de sus elementos. Usualmente el usuario escucha al comienzo todos los sonidos del conjunto para poder *calibrar* cuáles son las máximas o mínimas distancias entre los sonidos. Luego todos los pares posibles de sonidos son presentados y se asigna un valor de similitud. Estos valores suelen ser tomados de una escala que va desde un extremo que dice *diferentes* a otro que dice *iguales* (u otras palabra afines). Luego esta información se transforma mediante el escalamiento multidimensional en una configuración geométrica de 1, 2 o 3 dimensiones donde se pueden estudiar las relaciones y distancias entre sonidos de manera visual. Esta técnica por sí sola no identifica la naturaleza física de estas dimensiones y es trabajo del que realiza el experimento interpretar la relación entre las dimensiones perceptuales y los atributos acústicos de los sonidos.

Mediante esta técnica se han realizado trabajos con sonidos sintetizados [81, 91, 102, 8], resintetizados o simulaciones de instrumentos [40, 66, 78, 120], grabaciones de instrumentos [58, 71], grabaciones de díadas de instrumentos [65, 109] y también sonidos ambientales [44].

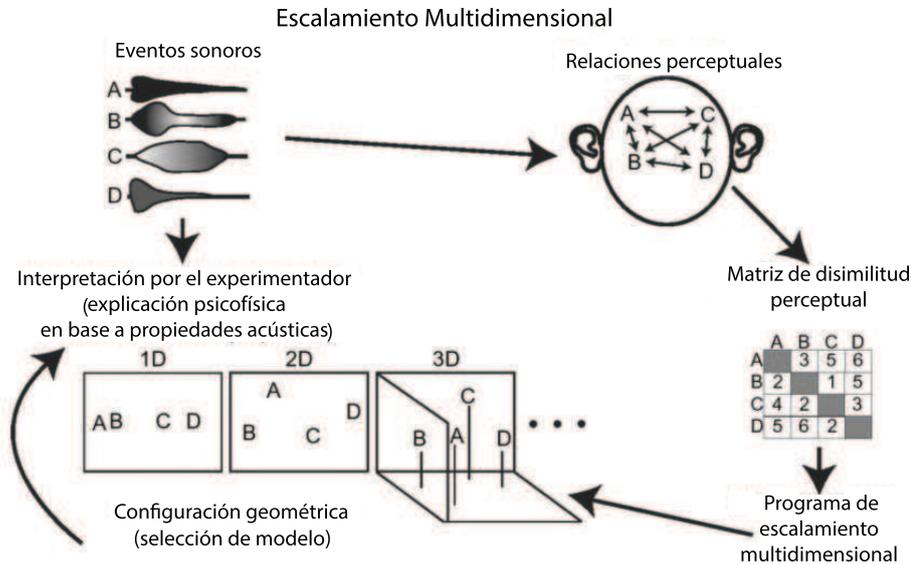


Figura 1.2: Diagrama de procedimiento de los experimentos de comparaciones de pares con análisis de escalamiento multidimensional. Adaptado de [77]

Para ejemplificar el método podemos ver en la figura 1.3 uno de los resultados del trabajo de Samson et al. [102]. Se utilizaron como estímulos nueve sonidos sintetizados a partir de un atributo espectral y temporal. En la dimensión espectral se modificó la cantidad de armónicos y en la temporal, el tiempo de ataque. Los resultados MDS muestran un organización casi perfecta entre las dimensiones espectrales y temporales. En este caso, como los sonidos fueron sintetizados en base a dos parámetros, resulta razonable pensar que basten dos dimensiones para generar el espacio tímbrico perceptual. Las comparaciones fueron realizadas con pares de sonidos y fragmentos melódicos, tanto con intensidades normalizadas como aleatorias, en ambos casos obtuvieron configuraciones MDS similares.

Otro experimento de comparación de pares pero realizado con sonidos sintetizados de instrumentos musicales es el de McAdams et al. [78]. El principal resultado de este trabajo se muestra en la figura 1.4. En este caso se utilizaron sonidos sintetizados ya que además se introdujeron sonoridades híbridas como combinación de dos instrumentos. En este caso fueron necesarias tres dimensiones para generar un espacio tímbrico que representase las respuestas por los participantes. Cada dimensión de este espacio corresponde a un atributo tímbrico que tiene un correlato acústico claro. La dimensión 1 corresponde al tiempo de

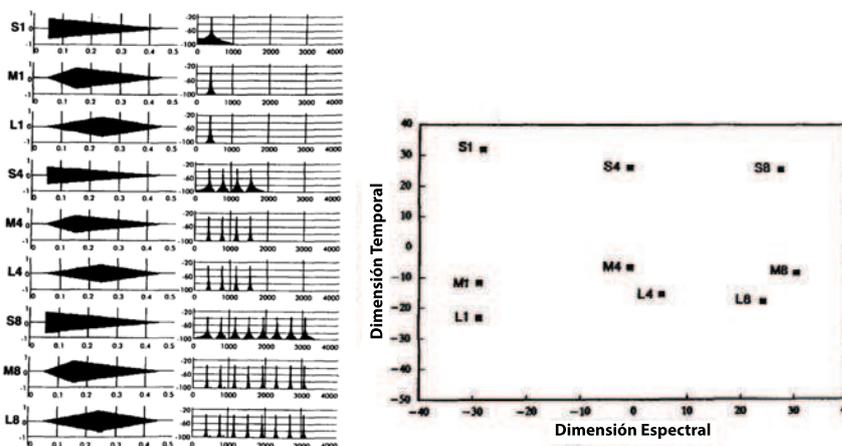


Figura 1.3: Experimento de Samson et al. [102]. (a) Conjunto de sonidos utilizados en el experimento. Las letras S, M y L corresponden a tiempos de ataque cortos, medianos y largos respectivamente y los números indican la cantidad de armónicos. (b) Configuración MDS

ataque, la dimensión 2 al centroide espectral y la dimensión 3 al flujo espectral.

Hasta este punto, los atributos que hemos mencionado y que cumplen un rol relevante en la percepción tímbrica han sido estudiados en su gran mayoría en el contexto de las sonoridades de los instrumentos musicales. Sin embargo, existen otros trabajos que se adentran en el estudio de otras sonoridades. En el trabajo de Gygi et al. [44] se realizaron comparaciones de pares sobre un conjunto de 100 sonidos ambientales, incluyendo sonidos generados por humanos, de animales, de máquinas y texturales. En los resultados del MDS de este experimento, exhibidos en la figura 1.5 puede observarse que los sonidos fueron agrupados según distintas cualidades, como por ejemplo, sonidos impulsivos o continuos, sonidos armónicos o no y sonidos vocalizados o no. Aunque las dimensiones encontradas por el MDS no parecen en principio corresponder a atributos acústicos únicos que muestren una variación continua, estas dimensiones pueden explicarse luego de un análisis posterior mediante regresores de combinaciones de atributos. Mencionaremos solamente el resultado que fue encontrado para la dimensión 1. Este fue dado por una combinación de tres atributos, uno relacionado a la saliencia de la altura tonal, otro relacionado negativamente a la profundidad de modulación y el tercero relacionado positivamente a la asimetría (*skewness*) del espectro. De esta forma, un desplazamiento a lo largo de la dimensión 1 correspondería

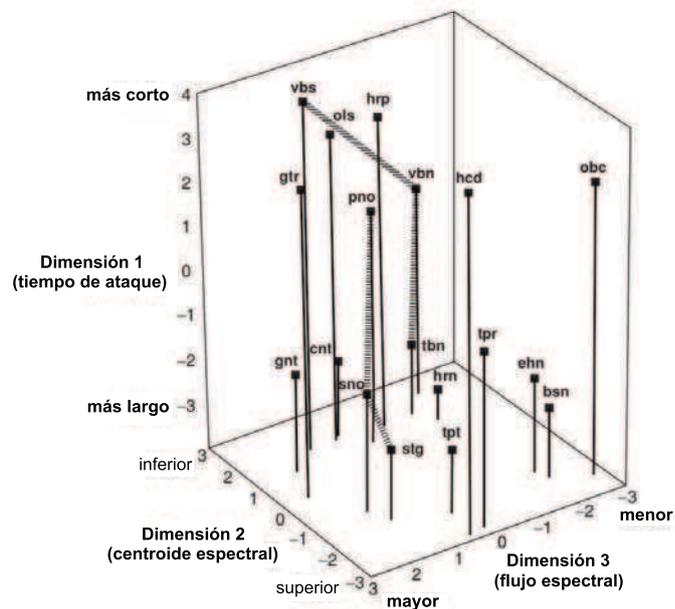


Figura 1.4: Configuración MDS 3D para el experimento de McAdams et al. [78]. *hrn:* corno francés. *tpt:* trompeta. *tbn:* trombón. *hrp:* harp. *tpr:* trumpar (trompeta/guitarra). *ols:* obolesta (oboe/celesta). *vbs:* vibráfono. *sno:* striano (cuerda frotada/piano). *hcd:* clavicordio. *ehn:* corno inglés. *bsn:* fagót. *cnt:* clarinete. *vbn:* vibrone (vibráfono/trombone). *obo:* obochord (oboe/clavicordio). *gtr:* guitarra. *stg:* cuerda frotada. *pno:* piano. *gnt:* guitarret (guitarra/clarinete)

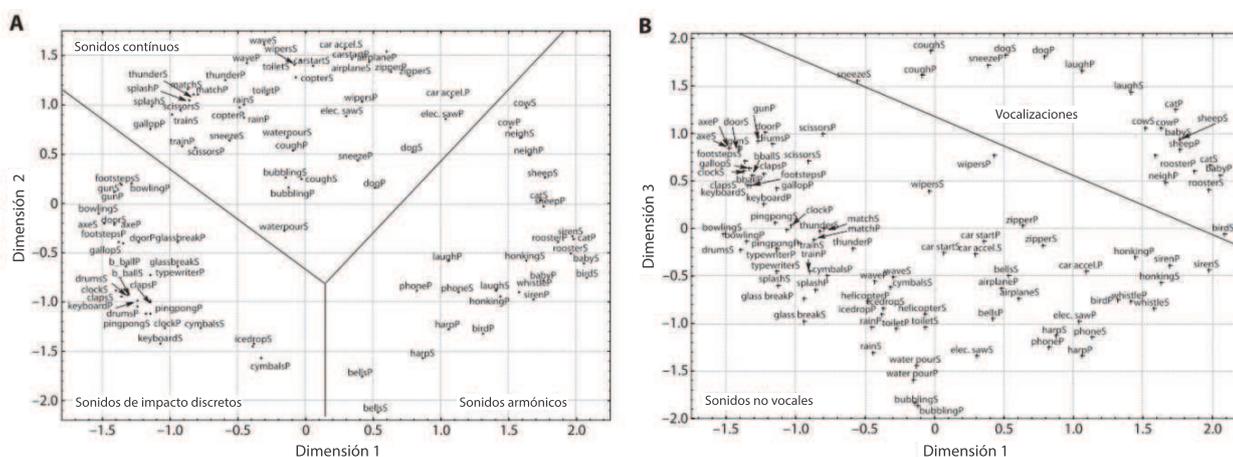


Figura 1.5: Configuración MDS 3D del experimento de Gygi et al. [44]. Los sonidos resultaron agrupados en grande categorías. En la figura de la izquierda (dimensiones 1 y 2): sonidos armónicos, sonidos de impacto discretos y sonidos continuos. En el panel de la derecha (dimensiones 1 y 3): vocalizaciones y no vocalizaciones

a una transición entre sonidos con modulación o repetitivos, sin un tono, hasta sonidos tonales, continuos, con mayor energía en los graves.

Existe una gran variedad de sonoridades que todavía no han sido estudiadas lo suficiente con experimentos psicofísicos de percepción tímbrica, incluso sonoridades relevantes en la música como las de instrumentos electrónicos y las dadas por las técnicas extendidas de los instrumentos. Entre los pocos trabajos podemos destacar el estudio de sonoridades texturales de la música electroacústica y contemporánea [41] y el estudio de nuestro grupo sobre los multifónicos del saxofón [97].

En la presente tesis se estudiaron las características tímbricas de dos grupos de sonidos. Por un lado los multifónicos del saxofón con sonoridades muy complejas y por otro lado sonidos menos complejos como los tonos modulados en amplitud sinusoidalmente. Estos dos grupos de sonidos tienen ciertos atributos perceptuales en común que se originan en las modulaciones de amplitud.

Para profundizar en la percepción de la modulación, debemos introducir ciertas funciones del procesamiento que ocurren en la periferia auditiva.

1.3. Procesamiento del sonido en la periferia auditiva

En esta sección describiremos, de forma resumida, el procesamiento básico de las señales que ingresan a la periferia auditiva ³. Este procesamiento es la primera etapa dentro de la ruta auditiva que termina, en etapas posteriores, en la percepción del sonido. La periferia ha sido estudiada con mucho detalle ya que en esta instancia se ponen de manifiesto muchos fenómenos que tienen correlatos perceptuales, como el enmascaramiento o los tonos diferenciales.

1.3.1. El oído interno y la descomposición en frecuencias

El sonido que se propaga por el aire ingresa al sistema auditivo periférico a través del canal auditivo donde es transformado en vibraciones por la membrana timpánica. Estas vibraciones son propagadas por la cadena de huesecillos del oído medio para ser suministradas al oído interno formado principalmente por el órgano de la cóclea. Dentro de este órgano se encuentra la membrana basilar que se extiende sobre toda la cóclea separándola en dos cámaras llenas de fluido. Esta membrana mide aproximadamente 35 mm en humanos y sus propiedades mecánicas varían gradualmente a lo largo de su extensión, lo que permite que las vibraciones que ingresan en la cóclea se propaguen sobre ella formando una onda viajera. Las frecuencias bajas producen vibraciones máximas cerca del final o ápex de la cóclea y las frecuencias altas producen vibraciones máximas cerca del comienzo o la base. Esto genera una de las características más importantes del procesamiento que ocurre en la periferia auditiva, la descomposición en frecuencias de la señal. Se puede pensar a la membrana basilar como un analizador espectral o un banco de filtros, separando sonidos complejos en sus componentes en frecuencias. El hecho de que las frecuencias se segregan espacialmente se conoce como tonotopía [90].

Sobre la membrana basilar, a lo largo de toda su extensión, se encuentra el órgano de Corti. Este órgano contiene a las células ciliadas externas e internas, que se encargan de amplificar las vibraciones de la membrana y traducirlas a eventos sinápticos respectivamente.

³En la sección 4.1.2 se expone con más detalle los mecanismos fisiológicos que intervienen en el procesamiento de sonido

Poseemos aproximadamente 3000 células ciliadas internas y cada una está conectada con una decena de fibras auditivas que llevan la señal en forma de disparos neuronales hacia el núcleo coclear, el primer relevo en la vía auditiva dentro del sistema nervioso.

Desde el punto de vista del procesamiento de las señales podemos mencionar los siguientes fenómenos que ocurren cuando el sonido ingresa al oído. Al comienzo la señal es filtrada por el canal auditivo y el oído medio con una respuesta pasa banda ancha que realza las señales en las frecuencias centrales (1 a 3 kHz) y atenúa los sonidos por debajo de 20 Hz y por encima de 20 kHz. Una vez dentro de la cóclea, la señal es amplificada y comprimida por un proceso activo en las células ciliadas externas. La amplificación puede alcanzar ganancias de 60 dB para sonidos tenues, y es nula para sonidos intensos, por lo tanto genera una compresión y transforma un rango de estímulos de 120 dB en aproximadamente 30 dB de rango de respuestas. Esto permite percibir 7 órdenes de magnitud en variaciones de presión, desde 2×10^{-5} pascuales en el límite de la audición a 1 kHz (0 dB SPL), hasta más de 100 pascuales (134 dB SP) cruzando el umbral del dolor. Luego las señales son rectificadas y filtradas (pasa bajos con corte en aproximadamente 3 kHz) por el proceso de transducción de las células ciliadas internas. El destino final de la señal dentro de la célula es generar eventos sinápticos para estimular la fibra auditiva.

El conjunto de disparos neuronales que salen de la cóclea posee toda la información necesaria para la escucha, pero codificada. Por un lado, la organización tonotópica, es preservada en las fibras auditivas y en gran parte de la ruta auditiva, incluyendo la corteza [90]. Por otro lado, las fibras auditivas tienen la capacidad de disparar de forma *enganchada con la fase* de las vibraciones de la membrana (del inglés *phase-locking*), lo que permite que se codifique la estructura fina de la forma de onda. Por lo tanto los principales mecanismos de codificación utilizados por la fibra son la codificación temporal y la tonotópica [84].

1.3.2. Filtros auditivos

El accionar de la cóclea y de la membrana basilar se puede modelar por un banco de filtros pasa bandas, llamados filtros auditivos. Estos filtros deberían conectarse en serie o cascada [73], para mantener una interpretación acorde con los datos de la fisiología, pero también es habitual conectarlos en paralelo con el estímulo. Muchos fenómenos perceptuales

están relacionados con la forma, solapamiento y ordenamiento de los filtros auditivos, como por ejemplo el enmascaramiento y la posibilidad de resolver o separar las componentes sinusoidales en un sonido complejo. Para ejemplificar esto observemos el diagrama de la figura 1.6. Allí se muestra como un sonido armónico formado de muchas componentes atraviesa el banco de filtros. A partir de la salida de cada filtro, computando el valor cuadrático medio, se puede construir el denominado *patrón de excitación*. Este patrón permite observar cómo las diferentes componentes de un sonido pueden estar resueltas o no y por ende la posibilidad de ser percibidas individualmente. Al mismo tiempo las vibraciones de cada sección de la membrana basilar representan la salida de los filtros. Es necesario aclarar que nos referiremos a la salida de los filtros como canales auditivos, en el sentido de canales de comunicación, y no debe ser confundido con el canal auditivo del oído externo.

El ancho de banda efectivo de los filtros auditivos se denomina *banda crítica* y permite establecer qué tan separados tienen que estar dos tonos sinusoidales para que no se enmascaren o interactúen. Es posible medir experimentalmente la banda crítica y caracterizar los filtros auditivos. Por mucho tiempo se utilizó una fórmula para la banda crítica en base a la definición de Zwicker [32, 30], pero actualmente se suele utilizar el denominado ancho de banda equivalente rectangular (ERB del inglés *equivalent rectangular bandwidth*) de Moore and Glasberg [83]. La siguiente expresión relaciona el ancho de banda ERB y la frecuencia central f en Hz:

$$ERB(f) = 24,7 + 0,108f \quad (1.1)$$

Esta definición de los anchos de banda se suele usar en conjunto con la escala ERB_N (número de ERBs) en la cual un ERB_N representa aproximadamente 0.85 mm de la membrana basilar [83] y cuya expresión está dada por:

$$ERBn(f) = 21,4 \log_{10}(0,00437f + 1) \quad (1.2)$$

Estas expresiones fueron construidas a partir del ajuste de experimentos psicofísicos de enmascaramiento con un tipo de filtros en particular, propuestos para representar los filtros

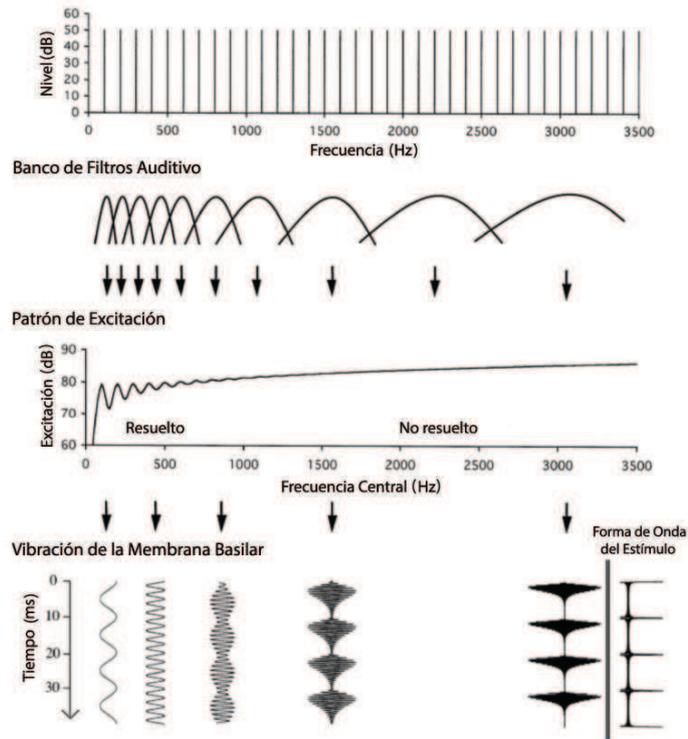


Figura 1.6: Diagrama del banco de filtros auditivo. Un sonido con una serie de armónicos de igual amplitud pasa por el banco de filtros para generar un patrón de excitación. La señal de salida de cada filtro simboliza el movimiento de la membrana basilar

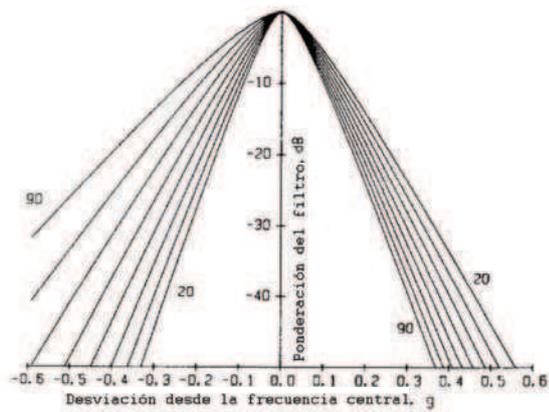


Figura 1.7: Forma de los filtros auditivos ROEX y su variación con la intensidad. Adaptado de Moore and Glasberg [83]

auditivos (figura 1.7). Estos filtros se basan en una función exponencial redondeada en la punta, (ROEX del inglés *rounded exponential*)[89]. Como puede observarse en la figura, los filtros auditivos cambian su forma con la intensidad debido al procesamiento no lineal de la cóclea. Perceptualmente esto significa que a medida que se incrementa el volumen se modifican las regiones del espectro enmascaradas, sobre todo la parte de altas frecuencias.

Mediante un banco de filtros ROEX se pueden generar los patrones de excitación como el de la figura 1.6 [83]. Estos patrones son sólo una representación estática del sonido. Dado que los filtros ROEX no poseen una respuesta impulso bien definida [57], no pueden usarse para generar una respuesta dinámica de la cóclea. Para suplir este problema, es necesario utilizar otros filtros como pueden ser los *gammatones* [88], que tienen una respuesta impulso sencilla dada por un tono sinusoidal con una función *gamma* como envolvente. Sin embargo, estos filtros son lineales y no pueden usarse para ajustar los cambios con la intensidad.

Los filtros *gammatones* suelen usarse en conjunto con la ERB para construir una representación de tiempo-frecuencia llamada *cocleograma* (ver el Apéndice A.2 para una descripción completa). En la figura 1.8 se muestra un ejemplo de cocleograma compuesto de 500 filtros para una señal armónica. El eje vertical está en la escala ERB_N . Aunque el banco de filtros es alimentado en paralelo, se puede apreciar que los primeros filtros activados son los de las frecuencias altas, simulando una onda viajera que se propaga desde los agudos hacia los graves. En el panel de la derecha se puede ver el valor cuadrático medio (RMS) de cada filtro, que equivaldría al patrón de excitación.

El uso de los filtros *gammatone* para realizar cocleogramas es muy útil y se ha extendido mucho [7], sin embargo como se mencionó no incluye las no linealidades presentes en el procesamiento que tiene lugar en la cóclea. Para suplir esta falta se han desarrollado otros esquemas no lineales de procesamiento, entre otros el *gammachirp* de Irino and Patterson [57], el *dual resonance nonlinear* (DRNL) de Meddis et al. [80] y el *multiple band-pass non linearity* (MBPNL) de Goldstein [38].

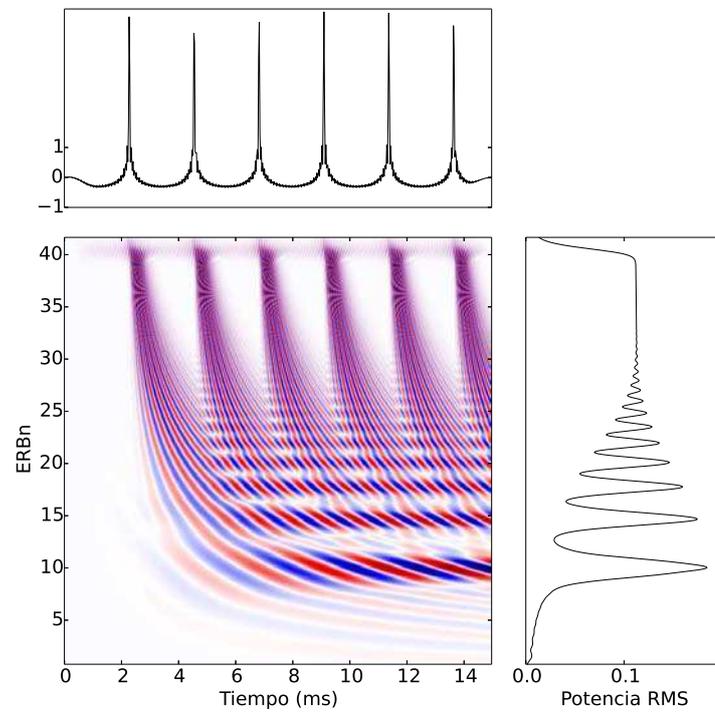


Figura 1.8: *Cocleograma gammatone de un tono armónico. Panel superior: forma de onda del tono. Panel central: cocleograma gammatone. Panel derecho: patrón de excitación*

1.4. Percepción de tonos con modulación

En acústica, cuando se habla de modulación, generalmente se refiere a la modificación o variación en el tiempo de algún parámetro de una señal base o portadora, con el requerimiento que la frecuencia característica de la señal que modula sea menor que la frecuencia característica de la señal portadora. Los tipos más comunes de modulación son: la modulación en amplitud (AM), la modulación en frecuencia (FM) y la modulación en fase (PM).

Esta clase de estímulos pueden considerarse los tonos mas sencillos que evocan una sensación de timbre mas compleja que la de un tono puro. Por otro lado presentan la ventaja de poder controlar sus características mediante la variación de un numero muy pequeño de parámetros (la frecuencia portadora, la frecuencia de modulación y la profundidad de modulación) y finalmente se acercan a los sonidos reales por el hecho de poseer mas de una escala temporal asociada (la que corresponde a su portadora o la estructura fina y la que corresponde a la envolvente o a la magnitud modulada).

Esto se debe, en principio a lo que se menciona anteriormente: esta clase de estímulos presenta características en común con sonidos relevantes ecológicamente, es decir poseer mas de una escala temporal. En el caso especifico de una senoide modulada en amplitud, si la escala temporal de la envolvente esta en el orden de las décimas de segundo esta clase de estimulo se acerca a una version simplificada del habla humana donde la frecuencia portadora esta en el rango audible y la de modulación en el rango temporal en el que ocurren los fonemas [14].

Otro factor que contribuye a la relevancia perceptual de los sonidos con modulación es el hecho de que la segregación de sonidos concurrentes se facilita si los mismos poseen modulación ya que el sistema auditivo es capaz de utilizar la información de coherencia de modulación de las distintas componentes provenientes de una fuente [76].

Finalmente, para el caso de sonidos AM hay evidencia de que el sistema auditivo procesa la envolvente temporal de los sonidos de forma selectiva en las frecuencias de modulación [61]. Este procesamiento tiene lugar en la ruta auditiva a partir de que la información ingresa de la periferia auditiva (vía las fibras del nervio auditivo) a los primeros núcleos neuronales

del tronco cerebral (en particular núcleo coclear y colículo inferior [60]). En particular existe evidencia de que existen poblaciones de neuronas sintonizadas de forma selectiva a diferentes frecuencias de modulación de la envolvente y que se encuentran organizadas de forma complementaria a la organización tonotópica (vinculada a la estructura fina), en lo que se conoce como la organización periodotópica (en el colículo inferior [72])

Con respecto a la periferia, ya sea para tonos AM como FM con modulación sinusoidal, la percepción de modulación da lugar a los siguientes tres fenómenos perceptuales: fluctuaciones, rugosidad, separación de componentes. Cuando la frecuencia de modulación es muy baja el sistema nervioso es capaz de seguir las modulaciones y de percibir las como cambios a lo largo del tiempo como fluctuaciones. En el caso de AM se percibe una fluctuación en la intensidad del sonido y en el caso de FM se perciben las variaciones en la frecuencia. A medida que la frecuencia de modulación se incrementa el sistema nervioso no puede seguir con tanta facilidad las modulaciones y se empieza a percibir un valor promediado de la magnitud. En el caso de FM se empieza a escuchar primero una frecuencia promedio y luego también un percepto de rugosidad. Para el caso de AM ya no se perciben fluctuaciones en la intensidad y esta toma un valor promedio y también se agrega el percepto de rugosidad. A medida que la frecuencia de modulación se sigue incrementando el percepto de rugosidad disminuye y se empieza a escuchar un sonido estático compuesto de varias componentes sinusoidales separadas.

El origen de estos fenómenos está relacionado con la banda crítica. Para ejemplificar esto consideremos un sonido con dos tonos sinusoidales:

$$\sin(2\pi f_1 t) + \sin(2\pi f_2 t)$$

Si las frecuencias f_1 y f_2 son lo suficientemente parecidas entonces los dos tonos interferirán en un mismo filtro auditivo provocando batidos. Es posible expresar la suma de dos funciones seno como el producto de una función seno y una función coseno:

$$2 \cos\left(2\pi \frac{f_2 - f_1}{2} t\right) \sin\left(2\pi \frac{f_1 + f_2}{2} t\right)$$

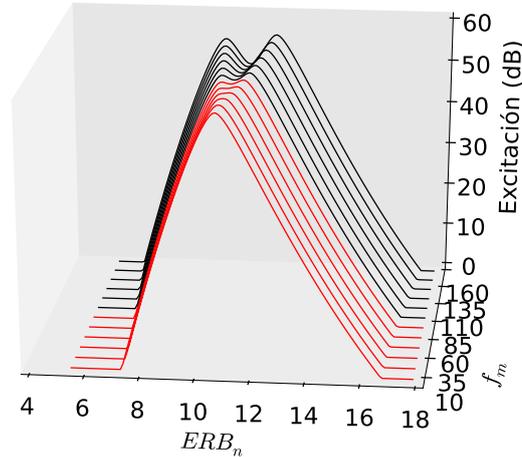


Figura 1.9: Patrón de excitación de un tono con batidos en función de la separación en frecuencias f_m . En rojo se indican los patrones hasta que la separación es igual a la banda crítica en ERB

Para poder expresar esta fórmula como el resultado de una modulación en amplitud hay que hacer la siguiente simplificación:

$$2 |\cos(2\pi(f_2 - f_1)t)| \sin(2\pi \frac{f_1 + f_2}{2} t) \quad (1.3)$$

Esto muestra que la frecuencia de modulación a la cual se perciben los batidos es $f_m = f_2 - f_1$ ⁴. A medida que se incrementa la distancia entre las dos frecuencias f_1 y f_2 , la frecuencia de modulación se hace más grande y el sistema nervioso no puede seguir los cambios en la amplitud y se empieza a producir el concepto de rugosidad, una vez que las frecuencias están separadas más allá de la banda crítica, los tonos pueden ser resueltos individualmente y no interfieren entre sí. En la figura 1.9 se muestran los patrones de excitación a medida que las dos componentes se separan. En rojo se muestran los patrones hasta que la frecuencia de modulación f_m es igual a la ERB, que para una frecuencia central de 500 Hz es $f_m \approx 80$ Hz.

⁴Esta simplificación es perceptualmente equivalente a la original solo para frecuencias de modulación bajas. Cuando la frecuencia de modulación empieza a crecer se pueden oír discrepancias.

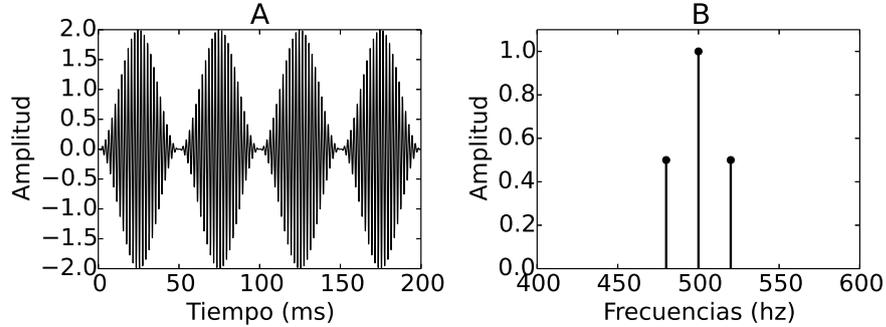


Figura 1.10: (A) Tono SAM para un índice de modulación de 1, una frecuencia portadora de 500 Hz y una moduladora de 20 Hz. (B) Espectro del tono SAM

1.4.1. Tonos SAM

A continuación vamos a estudiar con más detalle los tonos de amplitud modulada sinusoidalmente, llamados tonos SAM (del inglés *sinusoidally amplitud modulated*). Estos tonos a pesar de tener una formulación muy sencilla, permiten evocar una variedad de perceptos modificando sólo dos parámetros. Estos tonos serán utilizados en uno de los experimentos de percepción tímbrica del presente trabajo (capítulo 3). Su expresión matemática está dada por:

$$SAM = (1 + m \sin(2\pi f_m t)) \sin(2\pi f_c t)$$

Donde m es el índice de modulación ($0 \leq m \leq 1$), f_c es la frecuencia portadora (c del inglés *carrier*) y f_m la frecuencia de modulación ($f_m \ll f_c$). La ecuación puede expandirse y descomponerse en la suma de tres sinusoides:

$$SAM = \frac{m}{2} \sin(2\pi(f_c - f_m)t) + \sin(2\pi f_c t) + \frac{m}{2} \sin(2\pi(f_c + f_m)t)$$

por lo que el espectro de un tono SAM está dado por una componente central con frecuencia f_c y dos componentes laterales con amplitud $m/2$ en las frecuencias $f_c \pm f_m$ (figura 1.10)

Los tonos SAM pueden generar sensaciones similares a las comentadas para el caso de tonos modulados en general [60]. En la figura 1.11 se pueden ver las diferentes regiones y

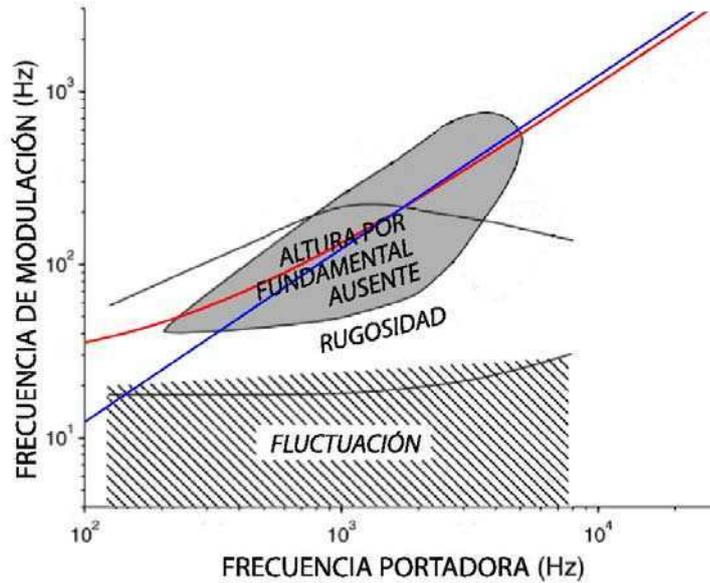


Figura 1.11: *Espacio de perceptos generados por los tonos SAM adaptado de Joris et al. [60]. La curva roja representa la banda crítica en ERBs, y la curva azul un intervalo musical de un tono.*

las sensaciones que son evocadas para los parámetros de f_c y f_m . Nuevamente encontramos el concepto de fluctuación, de rugosidad, y de separación de componentes y también una región dada por la percepción de altura tonal a partir del fenómeno de la fundamental ausente o altura virtual [84].

En el extremo de frecuencias de modulación bajas, el concepto de fluctuaciones se puede explicar mediante un proceso temporal prácticamente independiente de la frecuencia portadora. En el otro extremo el concepto de separación de componentes está más relacionada a la banda crítica y la tonotopía. Luego, el concepto de rugosidad, actúa como una transición de un fenómeno temporal hacia uno espectral. En la figura 1.12, se muestran esta transición con los patrones de excitación a medida que la frecuencia de modulación se incrementa. El cambio de color indica aproximadamente el punto de corte (para $f_c = 500$ Hz, $f_m \approx 80$ Hz) donde las componentes laterales del tono empiezan a separarse. Nótese que estas componentes son resueltas de manera asimétrica, la componente más grave es resuelta primero.

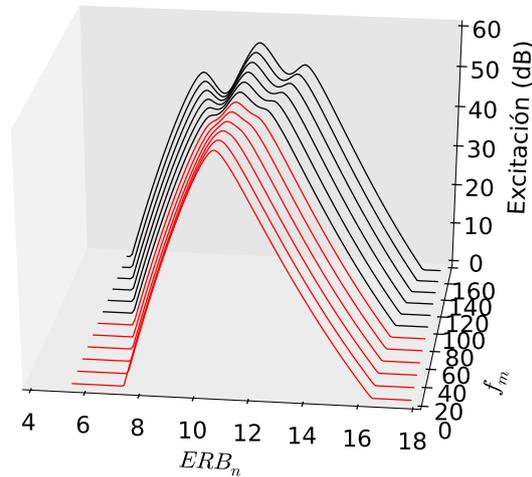


Figura 1.12: Patrones de excitación para tonos SAM en función de la frecuencia de modulación F_m . En rojo se indican los patrones hasta que la separación es igual a la banda crítica en ERB

1.4.2. Rugosidad

El percepto de rugosidad, definido originalmente por Helmholtz [51] a partir de experimentos con dos sinusoides, hoy en día se ha estudiado en relación a la disonancia musical [55, 104] y la tensión en la música occidental [94].

Existen numerosos experimentos que han medido la rugosidad para diferentes tipos de sonidos. Para el caso de tonos sinusoidales, uno de los resultados más conocidos es el dado por el trabajo de Plomp and Levelt [92]. En la figura 1.13 se muestra una curva que indica que la máxima disonancia o rugosidad ocurre cuando los dos tonos están separados en un 25% de la banda crítica.

A pesar de que este resultado es utilizado como una referencia, se pueden encontrar trabajos que presentan una información diferente. En la figura 1.14 se pueden ver una recopilación de varios experimentos adaptado de Pressnitzer [93]. En azul se muestran los resultados de experimentos que han medido la separación mínima necesaria para que dos tonos concurrentes no evoquen el percepto de rugosidad y en negro la separación que evoca el mayor nivel de rugosidad.. Los ajustes indican para el caso de mínima separación y el

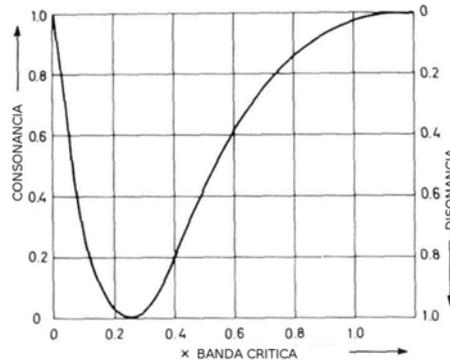


Figura 1.13: Nivel de consonancia al escuchar dos tonos en función de la separación en unidades de ancho de banda crítica

caso de máxima rugosidad un 110 % y 45 % del ancho de la banda crítica medido en ERB respectivamente.

Como podemos notar hay una gran diferencia entre establecer que el máximo de rugosidad ocurre para un 25 % o para un 45 % de la banda crítica. Esta diferencias tan grande puede entenderse parcialmente si consideramos que el trabajo de Plomp and Levelt [92], mide la banda crítica con la fórmula de Zwicker and Terhardt [127] y mientras que en el de Pressnitzer [93] se utiliza la fórmula de ERB. Al mismo tiempo la información utilizada para los ajustes en ambos casos proviene de experimentos muy antiguos. Actualmente se han realizado pocos experimentos para revisar este fenómeno, pero podemos destacar el trabajo de Miśkiewicz et al. [82] en el cual observan que no hay una relación constante entre la fracción de banda crítica y la frecuencia central. Por ejemplo, se establece para frecuencias bajas como 125 Hz, un porcentaje cercano al 60 % y para frecuencias de 4000 Hz un porcentaje cercano al 25 %.

Con respecto a los tonos SAM, también se ha medido experimentalmente como varía la rugosidad en función de la frecuencia de modulación. En la figura 1.15 se muestran los resultados para tonos SAM con $m = 1$ y varias frecuencias portadoras [30]. Es interesante notar como el máximo se desplaza en función de la frecuencia portadora hasta que esta llega a 1 kHz, y luego permanece en un valor fijo en la posición de 70 Hz.

Existen varios tipos de modelos para calcular la rugosidad. Por un lado están los modelos espectrales (a partir de la idea original de Helmholtz) [104, 55, 113], que utilizan una

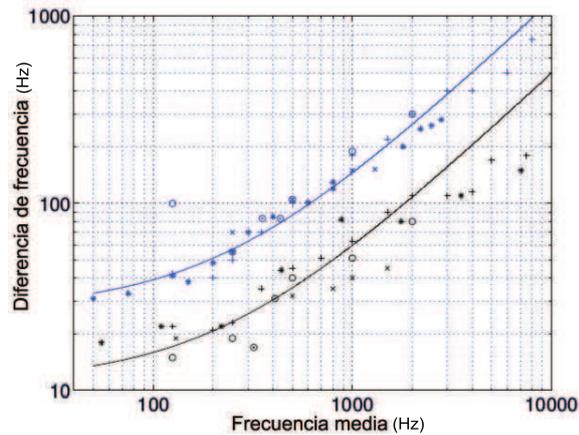


Figura 1.14: En azul, la separación más pequeña sin rugosidad en función de la frecuencia promedio: Plomp et Levelt (1965) + : Plomp et Steeneken (1968) : Mayer (1894) : Cross et Goodwin (1893) : Guthrie et Morrill (1928) et Kaestner (1909). En negro, la separación con máxima rugosidad. : Plomp et Levelt (1965) + : Plomp et Steeneken (1968) : Kameoka et Kuryiagawa (1969a) : Cross et Goodwin (1893) : Guthrie et Morrill (1928) et Kaestner (1909). Los ajustes corresponden para el caso azul a un 110 % y para el negro a un 45 % de la banda crítica medida en ERB. Figura tomada de Pressnitzer [93]

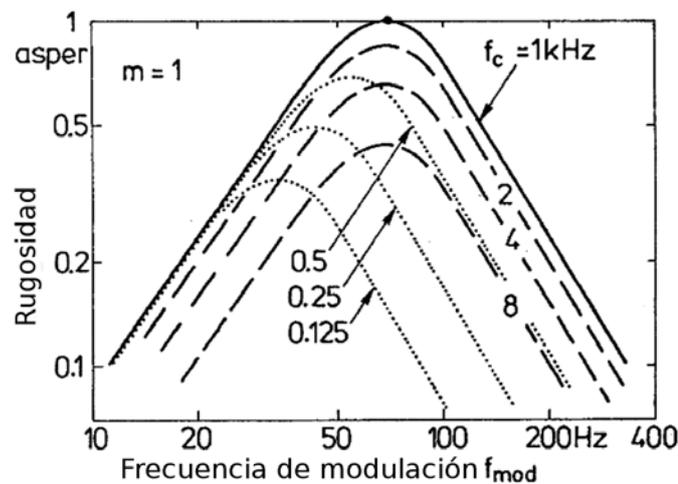


Figura 1.15: Valores de rugosidad para tonos SAM en función de la frecuencia moduladora para varias frecuencias portadoras. Los valores de rugosidad están en unidades de asper que valen 1 para la máxima rugosidad de un tono de 1 kHz. Tomado de Fastl and Zwicker [30]

relación como la que se muestra en la figura 1.13 para computar la rugosidad de cada par de componentes de un espectro. Este tipo de modelo es detallado en el apéndice A.3.

Por otro lado existen modelos temporales para calcular la rugosidad [17, 3]. Estos utilizan un banco de filtros auditivos, para luego medir la rugosidad por cada canal. Esto es realizado analizando la frecuencia de modulación de la envolvente para cada canal auditivo.

Se ha mostrado que la medida de rugosidad es aditiva [110], por lo que la rugosidad total se puede computar sumando la rugosidad presente a lo largo del eje tonotópico. En sonidos con más de dos componentes es necesario computar las rugosidades parciales entre todos los pares de componentes. Para el caso de sonidos armónicos, si se superponen dos notas formando un intervalo musical, el nivel de rugosidad equivalente al nivel de disonancia, y se puede asociar una sensación de relajación para intervalos consonantes o de tensión para los disonantes [92].

Para el caso de los tonos SAM, si se elige una frecuencia de modulación que genera rugosidad, es posible establecer una relación sencilla entre el valor de la rugosidad y la profundidad de modulación:

$$R = const \times m^p \quad (1.4)$$

Donde R es el valor de la rugosidad y p toma valores entre 1 y 2 dependiendo de que trabajos se tomen como referencia [17]. Para una señal arbitraria es posible medir la profundidad de modulación efectiva a partir del coeficiente de variación (CV) de su envolvente:

$$m_{ef} = \sqrt{2} CV = \sqrt{2} \frac{\sigma}{\mu} \quad (1.5)$$

Con σ y μ la desviación estándar y el valor medio de la envolvente respectivamente. Es posible expresar esta relación en términos del valor cuadrático medio (amplitud RMS) de la envolvente si consideramos que $A_{RMS}^2 = \sigma^2 + \mu^2$, por lo que la ecuación queda:

$$m_{ef} = \sqrt{2} \sqrt{\left(\frac{A_{RMS}}{\mu}\right)^2 - 1} \quad (1.6)$$

Mediante estas ecuaciones es posible establecer la profundidad de modulación efectiva del tono con batidos de la ecuación 1.3 y por lo tanto su rugosidad. El valor que se obtiene para la profundidad de modulación es $m_{ef} = 0,68$, lo cual está en perfecto acuerdo con los resultados experimentales que establecen que un tono SAM con valores de m entre 0.65 y 0.7 tiene la misma rugosidad que un tono con batidos[110].

Sin pretender agotar todos los aspectos que abarca el fenómeno del timbre de los sonidos con modulación, en este capítulo se introdujeron los conceptos básicos, el vocabulario y gran parte de las herramientas que serán empleadas en esta tesis para el estudio de esta clase de sonidos. En lo que sigue la exposición del trabajo desarrollado específicamente en esta tesis parte de la percepción tímbrica a alto nivel de sonidos complejos en el ámbito musical, para luego pasar al ámbito mucho más controlable de los sonidos SAM y finalmente realizar una modelización del procesamiento que sufren esta última clase de sonidos en la periferia auditiva.

Capítulo 2

Multifónicos del saxofón

*...il qual ordine ricerca, prima, che le
percosse fatte dentro all'istesso tempo
siano commensurabili di numero, acciò
che la cartilagine del timpano non
abbia a star in un perpetuo tormento...*
Galileo Galilei.

Este capítulo está dedicado al estudio del timbre en los sonidos multifónicos del saxofón. El origen de este estudio se debe al trabajo en colaboración con el saxofonista y compositor Martín Proscia, un especialista en la ejecución y composición con multifónicos en la música contemporánea. Esta colaboración permitió expandir el tipo de sonoridades típicamente utilizados por los estudios de percepción tímbrica, generando nuevas perspectivas para el estudio del timbre. En este capítulo se presentan, primero, los resultados del análisis de los multifónicos en base a sus atributos acústicos y, segundo, en base a un experimento de comparación de pares de sonidos.

2.1. Introducción

Los estudios psicofísicos del timbre en la música se han centrado mayormente en estudiar los sonidos de instrumentos musicales de las orquestas occidentales. Los resultados de estos

experimentos lograron identificar que las propiedades espectrotemporales de estos sonidos pueden ser descritas por un mismo espacio tímbrico de pocas dimensiones. Esto hace referencia a que estos instrumentos tienen muchas propiedades en común, como por ejemplo, pueden ser representados por un sistema fuente-filtro.

El uso de nuevas técnicas de ejecución en la música del siglo XX, las denominadas técnicas extendidas, ha expandido el universo de sonoridades que los instrumentos occidentales pueden alcanzar. Estos avances todavía no han sido reflejados en los estudios psicofísicos de percepción tímbrica, quizás debido a su carácter de novedad, pero también debido a la dificultad de sistematización de conjuntos de sonidos muy heterogéneos. En este contexto, nos propusimos abordar el estudio de la percepción tímbrica de los multifónicos del saxofón.

Suelen conocerse por multifónicos aquellos sonidos que poseen más de una nota reconocible y que son producidos por un instrumento típicamente monofónico. Para su producción se requieren técnicas distintas a las usadas convencionalmente, por lo que los multifónicos caen dentro de la categoría de sonidos generados mediante técnicas extendidas. En el caso de los instrumentos de viento de la familia de las maderas las técnicas incluyen digitaciones, embocaduras y disposiciones del tracto vocal especiales. Los multifónicos son en su gran mayoría sonidos inarmónicos, y por lo tanto la afinación de cada altura no respeta estrictamente las de una escala temperada. Las alturas en los sonidos multifónicos surgen al formarse nuevas resonancias dentro del tubo del instrumento. Esto ocurre por el cambio de la longitud efectiva del tubo, que se produce al utilizar una digitación especial que combina llaves abiertas y cerradas.

Para comprender más en detalle cómo surgen las resonancias a partir de una digitación es necesario observar la impedancia acústica del instrumento. En la figura 2.1 se pueden observar dos espectros de impedancias de un saxo soprano en Bb, medidas a la entrada de la boquilla sin tener la caña puesta [12]. En la figura 2.1a se puede ver el espectro para la nota B4 del saxo (A4 440 Hz), dada por el primer pico de resonancia junto con la digitación necesaria para lograr esa nota, mientras que en la figura 2.1b se puede observar el espectro de la impedancia para una digitación que produce un multifónico. En el multifónico se introduce un nuevo pico de resonancia cerca de la nota del saxo D#5 (C# 554 Hz), lo que permite generar un intervalo de tercera mayor. Para lograr un sonido estable es

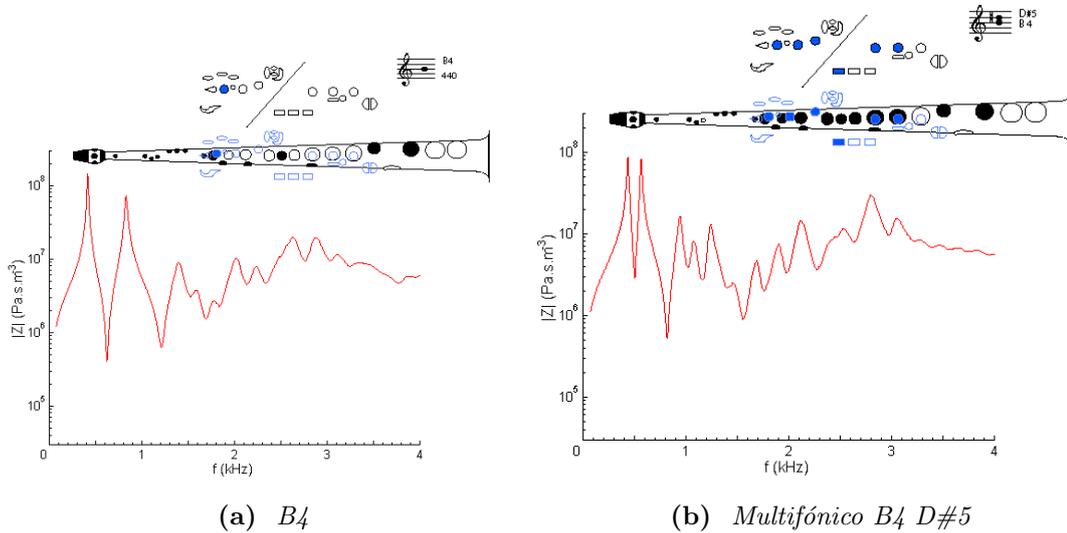


Figura 2.1: Impedancia acústica de un saxo soprano en Bb para dos digitaciones: (a) una nota sola, (b) un multifónico. Las notas están transpuestas. Adaptado de Chen et al. [12]

necesario sintonizar el tracto vocal y la embocadura de manera tal que los dos picos tengan una impedancia resultante similar. En la figura 2.2 se pueden observar mediciones de la impedancia del tracto vocal mientras se ejecuta un multifónico en un saxo tenor. En el primer espectro se puede ver que el tracto vocal se ha sintonizado con los primeros picos de resonancia mientras que en el segundo espectro lo hace con un pico mucho más elevado en frecuencia. La habilidad de manejar las resonancias del tracto vocal en un registro de frecuencias amplio es un indicador de la experiencia del ejecutante según el estudio de Chen et al. [13]. Finalmente el ejecutante también puede modificar la embocadura y la presión sobre la caña para generar cambios finos en la afinación de las alturas y otras características tímbricas.

Debido a todos estos factores que afectan su producción existe una gran cantidad de posibles multifónicos con diferentes sonoridades. Sin embargo, hay ciertas propiedades físicas que son comunes a todos estos sonidos. Las frecuencias de las alturas del multifónico están sujetas a distorsión por intermodulación debido a la saturación en las oscilaciones de la caña [5]. A medida que la presión de soplido es más fuerte, además de subir la intensidad, también aumenta el nivel de distorsión y se generan nuevas componentes en frecuencias

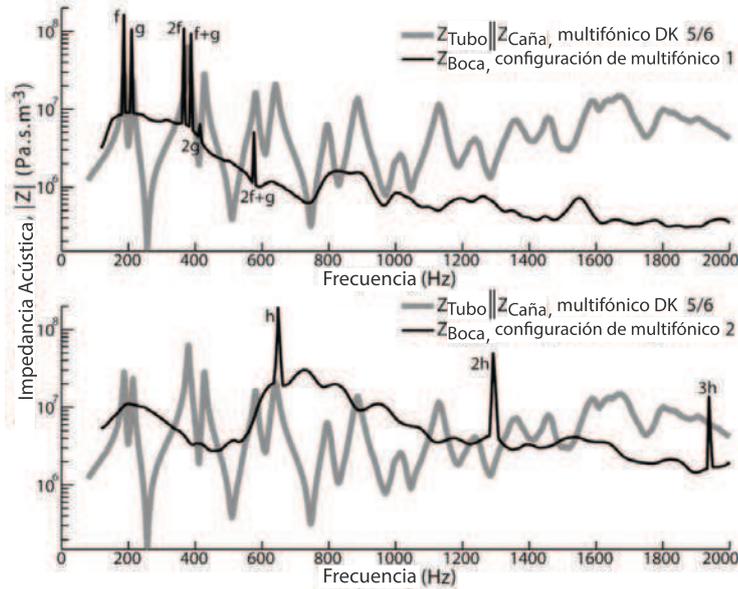


Figura 2.2: Impedancia acústica de un saxo Tenor en Bb para una digitación de un multifónico, superpuesta con la impedancia del tracto vocal en dos configuraciones. Los picos sobre la impedancia del tracto vocal indican las frecuencias más prominentes en el sonido resultante. Adaptado de Chen et al. [12].

(productos de distorsión) que modifican sustancialmente el timbre del sonido mediante batidos y rugosidad¹. Esta relación entre la intensidad y el aumento de contenido armónico es muy común en la ejecución de casi todos los instrumentos musicales. Sin embargo, en el caso de los multifónicos, el cambio tímbrico es mucho más notorio por la disonancia que genera la distorsión y la rugosidad. Otra característica física de los multifónicos es que en muchos casos el sistema acústico en cuestión puede comportarse de forma caótica, lo cual puede agregar efectos similares al ruido [63].

Los fenómenos mencionados anteriormente muestran que tan solo agregando nuevas resonancias al sistema pueden generarse cambios muy drásticos en el sonido resultante. Más aún, debido a que el control de la embocadura, presión y tracto vocal puede modificarse de manera gradual, se genera un abanico continuo de sonidos multifónicos. De todas formas, existe una cantidad limitada de multifónicos desde el punto de vista de las digitaciones, ya que no son infinitas las combinaciones de llaves abiertas y cerrada y además no toda digitación genera un espectro de impedancias con picos lo suficientemente altos para producir

¹Si la separación en frecuencias de los productos de distorsión es menor que la banda crítica.

un sonido estable. Existen varios libros y catálogos de multifónicos con las digitaciones y las notas que producen [68, 117], sin embargo en esas referencias no se profundiza en las cuestiones tímbricas de los sonidos.

En los últimos años ha habido un incremento en el estudio de los sonidos multifónicos, y en particular los producidos por el saxofón [4] [95]. Estos estudios se han centrado principalmente en la producción de estos sonidos y en las técnicas instrumentales que conllevan, pero no ha habido estudios exhaustivos sobre la percepción tímbrica de estos sonidos. En nuestro grupo de trabajo se realizó un estudio musicológico previo [95] sobre los multifónicos del saxofón alto, donde se clasificaron estos sonidos en cuatro grandes clases. A partir de esta investigación y por lo mencionado sobre la falta del estudio tímbrico de estas sonoridades, se llevó adelante el siguiente estudio.

En el trabajo musicológico previo, se estudió un gran conjunto de multifónicos a partir de un enfoque tímbrico basado en la *escucha reducida* de Schaeffer [103]. Los resultados que utilizaremos en este trabajo corresponden a los sonidos estacionarios de los multifónicos y no se abordarán sus capacidades dinámicas. El estudio realizado permitió generar cuatro clases de multifónicos caracterizando los sonidos en base a sus técnicas interpretativas y los atributos perceptuales, definidos por Schaeffer, como el *grano interno*, la *cualidad de superficie* y la *iteración* [15]. Las clases son las siguientes: *Bicordios*, *Complejos*, *Multiarmónicos* y *Trémolos*. En la figura 2.3 podemos observar unos espectros de ejemplo para las cuatro clases de multifónicos. En el contexto del presente trabajo describiremos los multifónicos sólo en base al intervalo musical, el nivel de distorsión y el rango dinámico de ejecución².

- *Bicordios*. Están caracterizados principalmente por tener dos notas fundamentales formando un intervalo de tercera mayor, menor, o disminuida y un nivel de distorsión medio o bajo y son ejecutados con dinámicas entre *pp* y *mf* (figura 2.3.a).
- *Complejos*. Poseen un espectro muy inarmónico, con muchos productos de distorsión, un alto nivel de rugosidad y son ejecutados con dinámicas entre *mf* y *ff* (figura 2.3.b).
- *Multiarmónicos*. Cuentan con un espectro con dos componentes muy claras en inter-

²Para una descripción más completa sobre la ejecución de estos sonidos referirse al trabajo [95].

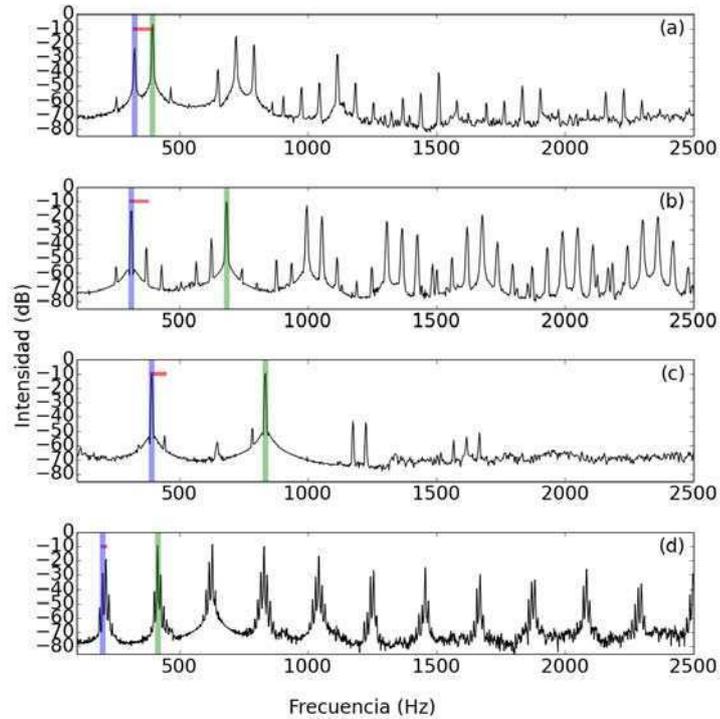


Figura 2.3: Cuatro espectros de tonos multifónicos del saxofón. (a) Bicordio, (b) Complejo, (c) Multiarmónico, (d) Trémolo. Con barras verticales se indican las frecuencias fundamentales que caracterizan el intervalo musical y con una barra horizontal la distancia en Hz entre bandas laterales, es decir la frecuencia de modulación (F_m).

valos de séptima y novena y muy pocas componentes en total, un nivel de distorsión prácticamente nulo y son ejecutados con dinámicas entre *ppp* y *mp* (figura 2.3.c).

- *Trémolos*. Están caracterizados principalmente por un intervalo de octava desafinada lo que genera batidos muy reconocibles, muchos productos de distorsión y son ejecutados con dinámicas entre *mp* y *f* (figura 2.3.d).

En los espectros se puede observar una organización de las componentes muy similar a la que se encuentra en tonos producidos por síntesis FM. Es muy común escuchar, con este tipo de síntesis, sonoridades inarmónicas similares a los multifónicos. Desde un punto de vista simplificado se puede pensar que muchas de las características de los multifónicos pueden ejemplificarse mediante síntesis por FM [39]. La separación en frecuencia entre los productos de distorsión (o bandas laterales) corresponde a la frecuencia de modulación

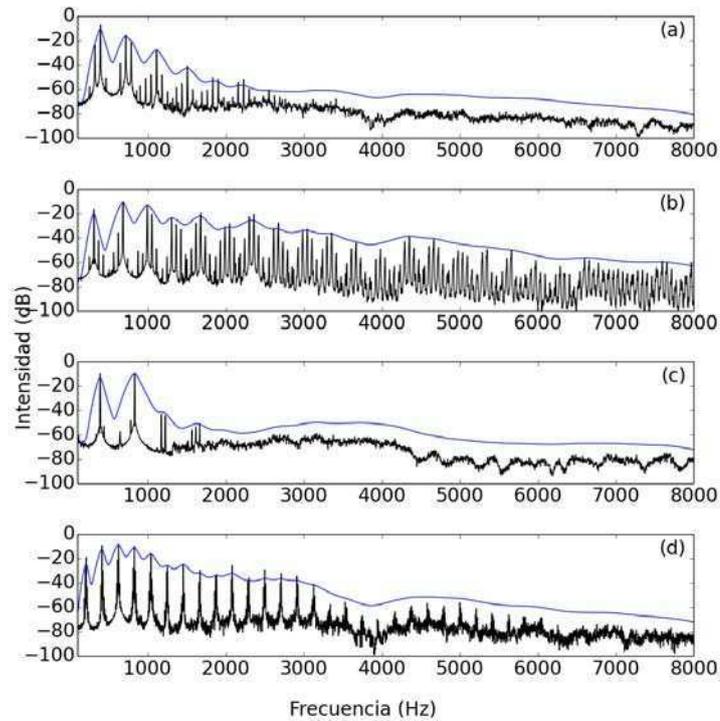


Figura 2.4: Cuatro espectros de tonos multifónicos del saxofón con sus correspondientes patrones de excitación. (a) Bicordio, (b) Complejo, (c) Multiarmónico, (d) Trémolo.

(f_m) y las frecuencias fundamentales corresponden a las portadoras.

A continuación, en la figura 2.4, se muestran los mismos multifónicos de ejemplo pero con sus correspondientes patrones de excitación (sección 1.3.2). Estos patrones, al igual que los espectros, sólo permiten representar el sonido de forma estacionaria. Sin embargo, el hecho de que a lo largo de todo el espectro se puede ver que las componentes caen dentro de las mismas bandas críticas, hace suponer que es necesaria una representación dinámica del sonido para estudiar los batidos, que justamente son la característica tímbrica más saliente de estos sonidos.

Para analizar las características dinámicas se debe utilizar una representación en tiempo y frecuencia como pueden ser los bancos de filtros *gammatones* (A.2). En la figura 2.5 podemos observar el cocleograma para el primer multifónico (2.3.a). En el panel superior de la figura se observa la forma de onda, la cual muestra una envolvente con modulaciones. En el panel lateral se muestra el patrón de excitación. En el panel central se observa el

cocleograma donde también se pueden ver presentes estas oscilaciones a lo largo de los canales auditivos³. En la figura 2.6 se muestran la forma de onda de los canales auditivos con mayor energía, los picos del patrón de excitación. En este caso se puede ver cómo los distintos canales auditivos tienen diferente comportamiento. En el canal del panel superior, la forma de onda corresponde a la interacción de las dos componentes que dan el intervalo de 3ra (ver figura 2.4.a), mientras que en los otros paneles la forma de onda corresponde a la interacción entre cada vez más armónicos y productos de distorsión. Se puede notar como la envolvente va tomando comportamientos más complejos en los canales de más alta frecuencia, lo que le da las características especiales al timbre del sonido.

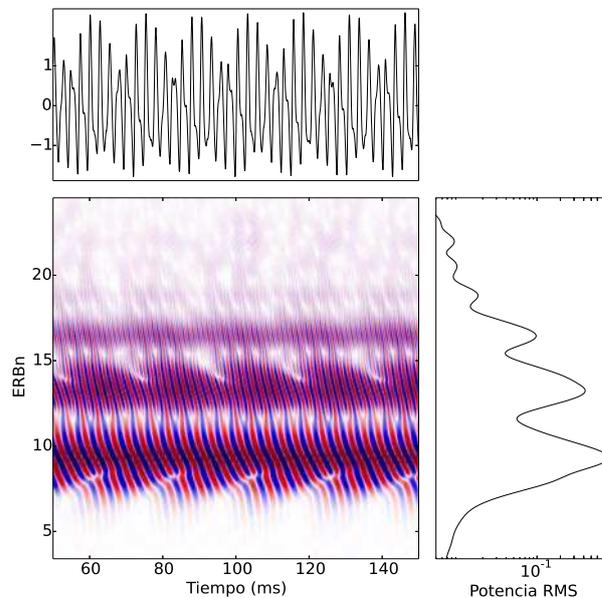


Figura 2.5: *Cocleograma de un multifónico Bicordio. Panel superior: forma de onda. Panel central: cocleograma. Panel lateral: patrón de excitación. La escala de colores del cocleograma está saturada para mostrar la señal de los canales con baja energía.*

Aunque es posible estudiar todos los multifónicos mediante este análisis detallado, en la siguiente sección los estudiaremos a partir de un conjunto reducido de atributos que capturan las características más importantes que se han mencionado.

³La escala de colores fue saturada para observar las oscilaciones en canales donde la energía era muy baja.

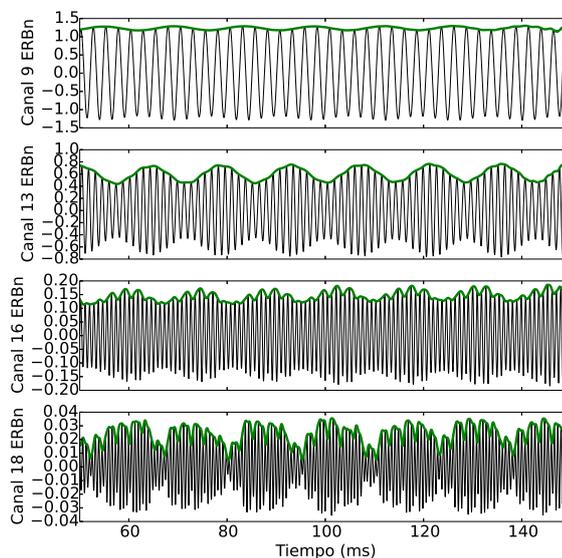


Figura 2.6: Formas de onda de los canales con mayor intensidad correspondientes al cocleograma de la figura 2.5

2.2. Análisis acústico de los multifónicos

2.2.1. Procedimientos

Se confeccionó una base de datos de 118 multifónicos a partir de grabaciones de un saxofón alto Selmer Super action 80 Serie III afinado en A4 440 Hz, con una boquilla Selmer Serie 80 C* y boquillas Vandorem 31/2. Las grabaciones fueron realizadas con un micrófono de medida (DBX RTA-M) y una interfaz de sonido externa Focusrite Sapphire a una frecuencia de muestreo de 48 kHz con una resolución de 24 bits. Los multifónicos fueron seleccionados y tocados uno por uno por Martín Proscia en una sala de 85 metros cuadrados, con tratamiento acústico ($T60 @1\text{kHz} = 0.3$). La duración de cada multifónico fue normalizada a 3 segundos utilizando las partes más estables y estacionarias de la grabación.

La base de datos quedó conformada con las siguientes cantidades de multifónicos por cada clase. 21 *Bicordios*, 39 *Complejos*, 34 *Multiarmónicos* y 24 *Trémolos*. Vale la pena notar que a pesar de que las clases están bien definidas hay un cierto grupo de multifónicos

que no se ajustan perfectamente en ninguna clase y hay sonoridades híbridas que se sitúan en los bordes entre las clases.

En base a lo descrito en la sección anterior, los multifónicos pueden ser estudiados en base a sus frecuencias fundamentales, productos de distorsión y frecuencia de modulación. Para cuantificar esto, se utilizaron los siguientes atributos acústicos:

1. Centroide espectral (CE)

Se computó el centroide espectral del espectro de potencias a partir de la siguiente fórmula.

$$CE = \frac{\sum_{n=0}^N f_n P_n}{\sum_{n=1}^N P_n}$$

Donde f_n y P_n son la frecuencia y el espectro de potencia del intervalo n de la transformada de Fourier discreta respectivamente.

2. Frecuencia de modulación (f_m)

La gran mayoría de los espectros tenía productos de distorsión organizados en forma de componentes separadas por una distancia fija (bandas laterales), que en este caso equivale a la frecuencia de modulación.

3. Rugosidad

La rugosidad se computó a partir de un modelo espectral que considera las rugosidades parciales entre todas los pares de componentes presentes en el espectro [104]. La descripción se encuentra en el Apéndice A.3.

4. Frecuencias fundamentales F1 and F2

Se consideraron las frecuencias fundamentales F1 y F2, a partir de las cuales se genera el intervalo musical más prominente y que determinan el contenido armónico del espectro, con $F2 > F1$. No fue necesario considerar más frecuencias para el grupo de multifónicos analizado.

5. Intervalo musical (IM) en semitonos

El intervalo se calculó entre las frecuencias F1 y F2

$$IM = 12 \log_2(F2/F1)$$

Es necesario realizar una aclaración con respecto a las intensidades del grupo de multifónicos utilizados. Algunos multifónicos, como los del grupo de los *Multiarmónicos* sólo pueden ser ejecutados a intensidades muy bajas, mientras que en el otro extremo, los *Complejos* sólo pueden realizarse a intensidades muy altas. Estas cuestiones hacen que las diferencias de intensidades en el conjunto de sonidos utilizados llegue a 30 dB. Esta clara diferencia hace que los multifónicos sean muy distinguibles en base a su intensidad, pero dado que esta no es un atributo tímbrico *per se*, se decidió no realizar un análisis especial sobre las intensidades. Sin embargo, como ya se ha mencionado, la intensidad tiene sus efectos en el nivel de distorsión y la rugosidad.

2.2.2. Resultados

En la figura 2.7, se muestran las distribuciones de los valores de los atributos acústicos para las cuatro clases de multifónicos y para el total de ellos. Para analizar qué atributos son capaces de discriminar las clases, realizamos un análisis ANOVA de una vía junto con un test *post hoc* de comparaciones múltiples de Tukey HSD ($\alpha = 0,05$). La hipótesis nula fue rechazada para todos los atributos de manera muy significativa ($p < 0,01$). Para la frecuencia fundamental F1, el test *post hoc*, mostró que los *Trémolos* son discriminados con respecto al resto de las clases ya que poseen la nota F1 más grave. Para la frecuencia fundamental F2, las clases de los *Bicordios* y los *Trémolos* se pueden discriminar de las clases de los *Complejos* y los *Multiarmónicos*. Para el intervalo musical, los *Bicordios* se logran discriminar del resto. Para el centroide espectral y la rugosidad, salvo la comparación entre los *Bicordios* con los *Multiarmónicos*, las clases logran discriminarse para todas las otras comparaciones. Finalmente, para la frecuencia de modulación se puede discriminar la clase de los *Trémolos* con el resto.

A continuación se estudió cómo se relacionan los atributos entre sí. En la tabla 2.1, se muestran los coeficientes de correlación entre los atributos. Este procedimiento permite

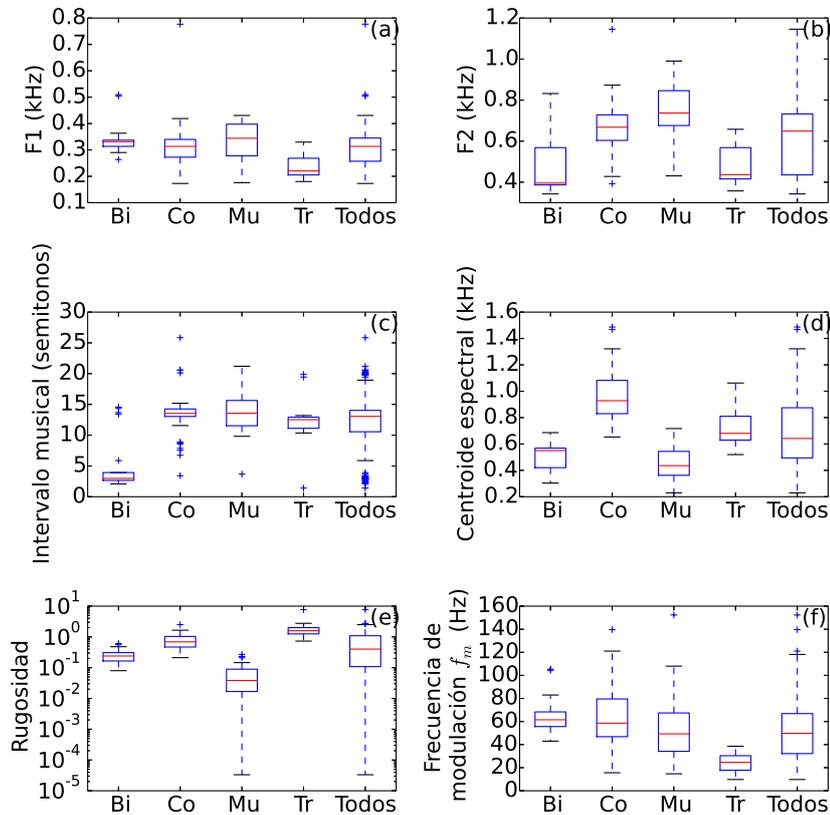


Figura 2.7: Diagramas de cajas para el conjunto de atributos acústicos utilizados separados por cada clase de multifónicos y el conjunto completo. (a) Primera frecuencia fundamental, (b) segunda frecuencia fundamental, (c) intervalo musical, (d) centroide espectral, (e) rugosidad y (f) frecuencia de modulación

	$F1$	$F2$	Intervalo	Centroide	Rugosidad	f_m
$F1$	1.00	0.51	-0.44	0.10	-0.40	0.33
$F2$	0.51	1.00	0.53	0.11	-0.35	0.34
Intervalo	-0.44	0.53	1.00	0.03	0.06	-0.03
Centroide	0.10	0.11	0.03	1.00	0.26	0.11
Rugosidad	-0.40	-0.35	0.06	0.26	1.00	-0.39
f_m	0.33	0.34	-0.03	0.11	-0.39	1.00

Cuadro 2.1: Tabla de coeficientes de correlación para el conjunto de atributos acústicos.

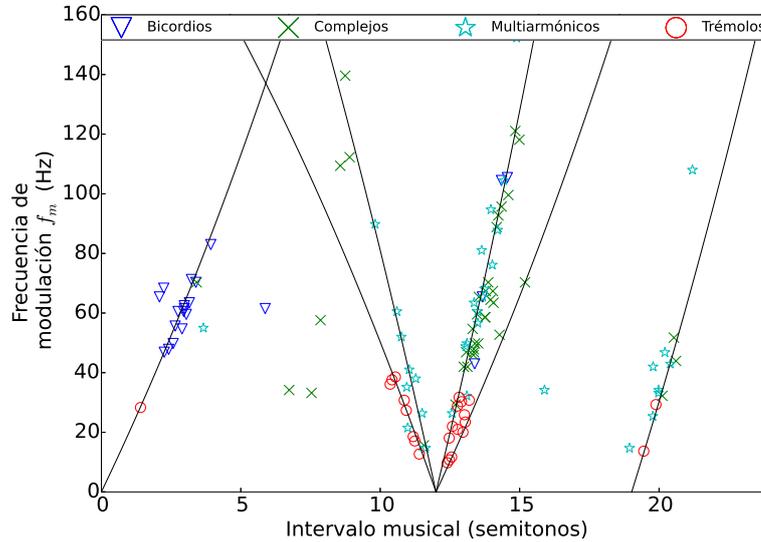


Figura 2.8: Atributos acústicos de los multifónicos: frecuencia de modulación vs intervalo musical.

ver si hay redundancia de información en los atributos a partir del grado de correlación. Algunos pares de atributos muestran cierto grado de correlación, como el intervalo musical con la frecuencia $F2$ y $F1$ de forma positiva y negativa respectivamente. Esto es esperable dada la definición del intervalo. Las otras correlaciones que se observan probablemente indiquen cierto tipo de relaciones internas, pero dado que el coeficiente de correlación sólo sirve para explicar estructuras lineales, es necesario otro tipo de análisis. Para entender esto con mayor profundidad, vamos a visualizar la información en gráficos con un atributo en función de otro.

En la figura 2.8 se representa la frecuencia de modulación en función del intervalo musical para las cuatro clases de multifónicos. Se puede observar un tipo de organización en forma de V asociado al hecho de que ciertos intervalos correlacionan con la frecuencia de modulación. Para comprender esto podemos notar que cerca del intervalo de 12 semitonos, es decir una octava, la frecuencia de modulación es baja. A medida que el intervalo se empieza a desafinar, los batidos y por ende la modulación, crecen. Las curvas que ajustan estos patrones están dadas por una relación con la siguiente forma: $f_m = kF1 + nF2$, manteniendo $F1$ fija y variando $F2$, y k y n son números enteros pequeños.

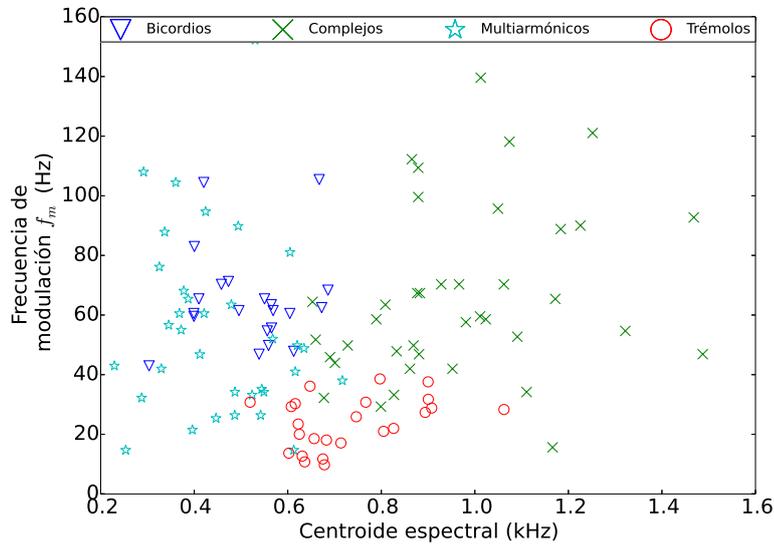


Figura 2.9: Atributos acústicos de los multifónicos: frecuencia de modulación vs centroide espectral.

En la figura 2.9 se muestra la frecuencia de modulación en función del centroide espectral. En este caso se puede observar a los sonidos de cada clase ocupando una región bien definida en el espacio. El centroide espectral logra discriminar bien entre los sonidos con menos energía en las frecuencias altas como los *Multiarmónicos* y los que tienen muchas componentes de alta frecuencia como los *Complejos*.

En la figura 2.10 podemos ver la rugosidad en función del intervalo musical. Esta figura, en comparación con la 2.8, no posee la organización precisa que alineaba a los sonidos, pero el atributo de la rugosidad agrupa muy bien los multifónicos de cada clase, ya que en cierta medida captura el nivel de distorsión relacionado con las componentes que caen dentro de una misma banda crítica.

Finalmente en la figura 2.11 podemos observar la rugosidad en función del centroide espectral. En este caso ambas magnitudes correlacionan levemente ya que los tonos con mayor rugosidad también tiene mayor cantidad de componentes y mayor energía en las altas frecuencias, pero al mismo tiempo se puede ver como los *Trémolos* son los que tienen el mayor nivel de rugosidad y con valores similares dentro de esa clase.

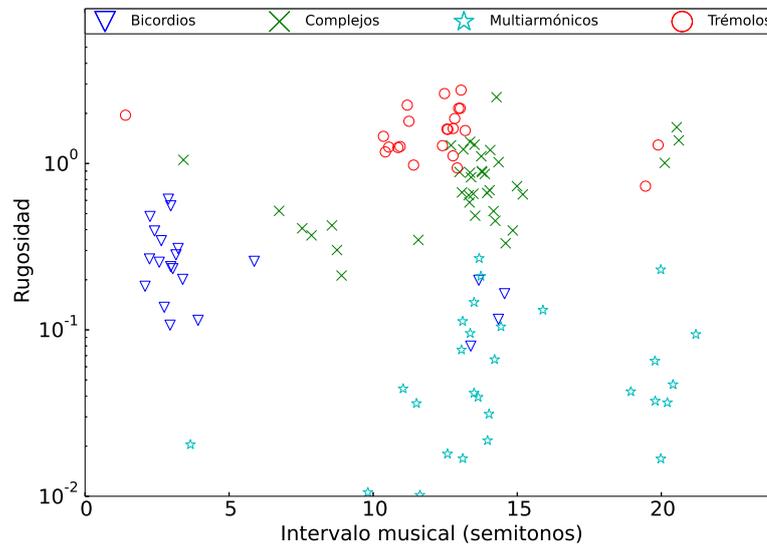


Figura 2.10: Atributos acústicos de los multifónicos: rugosidad vs intervalo musical.

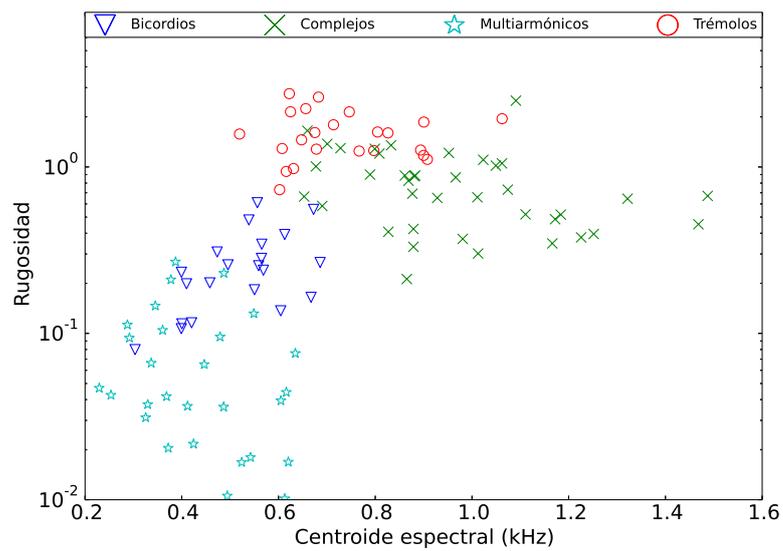


Figura 2.11: Atributos acústicos de los multifónicos: rugosidad vs centroide espectral.

2.2.3. Resumen

El conjunto de multifónicos, a pesar de tratarse a primera vista de un conjunto de sonidos muy variado y heterogéneo, posee ciertas características que hacen posible su organización y clasificación. Es así que el trabajo musicológico previo realizado por Martín Proscia supo captar ciertas esencias tímbricas en estos sonidos y generar una clasificación útil para los propósitos de la composición. A continuación de esto, el trabajo mostrado aquí, en base al estudio de los atributos acústicos, permitió cotejar cuáles fueron los factores utilizados para clasificar los sonidos. Al mismo tiempo, permitió considerar los orígenes físicos de la producción de los multifónicos en el saxofón y establecer cuáles son los fenómenos comunes a todos los multifónicos: el hecho de tener dos o más frecuencias fundamentales y un espectro denso como resultado de las intermodulaciones de los armónicos de estas fundamentales. Finalmente, las características de modulación de estos sonidos sugirieron que la rugosidad es un atributo acústico saliente para la percepción de estos sonidos. Mediante el análisis de los atributos acústicos se pudieron observar las diferencias entre cada clase de multifónicos. Con la información de estos atributos es posible generar un sistema de aprendizaje automático que dado un multifónico permite predecir a que clase pertenece [97]. Para expandir estos resultados y el estudio del timbre de estos sonidos, se realizó un experimento psicofísico de comparación de pares, que es expuesto a continuación.

2.3. Experimento de comparación de multifónicos

Se diseñó un experimento psicofísico de comparación de pares para obtener resultados perceptuales cuantitativos sobre el comportamiento tímbrico de los multifónicos y para poder construir cierta confianza con la clasificación propuesta por el trabajo musicológico. Este experimento se realizó mediante la técnica descrita en la sección 1.2 sobre comparación de pares de sonidos y el análisis por medio del escalamiento multidimensional (MDS).

2.3.1. Desarrollo experimental

Se seleccionaron 15 sonidos del conjunto de 118 como estímulos. Estos sonidos fueron representativos de las cuatro clases propuestas y fueron seleccionados para mostrar diferentes matices y frecuencias fundamentales de las muestras (cuatro *Bicordios*, cuatro *Multiarmónicos*, cuatro *Trémolos*, y tres *Complejos*). Los fragmentos fueron editados seleccionando un intervalo de 3 segundos de duración que fuese lo más estacionario posible. Las intensidades no fueron normalizadas debido a que diferentes sonidos tienen diversos niveles de ruido, los cuales podrían tomar mucho protagonismo si se aumenta la intensidad, y pretendimos preservar las características de una escucha realista. El nivel fue ajustado para restringir el rango de niveles sonoros entre 62 y 71 dB_{ASPL} , evitando generar un sesgo excesivo por las diferencias de intensidad. De todas formas, con el fin de estudiar los posibles sesgos, se insertaron algunas presentaciones con diferentes niveles sonoros para poder tomar como control en el experimento. Se consideraron todos los pares posibles por lo que los sujetos escucharon 225 presentaciones.

Cinco sujetos participaron del experimento con edades entre 25 y 35 años. Todos ellos fueron estudiantes o egresados de la Carrera con Medios Electroacústicos de la Universidad de Quilmes y fueron invitados a participar voluntariamente y manifestaron su consentimiento para participar por escrito. Los sujetos no estaban familiarizados con la clasificación propuesta, tenían experiencia en el campo de la música electroacústica contemporánea y reportaron no tener desórdenes auditivos.

El procedimiento fue el siguiente. En una primera etapa, los sujetos escuchaban el conjunto de los 15 sonidos para tener una primera idea de la variedad de sonidos que escucharían luego. Seguido a esto, se siguió con 6 comparaciones de pares de entrenamiento que no fueron tomados en cuenta para el análisis. Luego comenzaba el experimento en sí con los 225 pares de forma aleatoria más 23 pares para usar de control.

La tarea de los sujetos era juzgar el grado de similitud del par de sonidos utilizando la siguiente escala: 1 *nada similares*, 2 *poco similares*, 3 *algo similares*, 4 *muy similares*, 5 *prácticamente iguales*. El experimento fue realizado en MATLAB usando la biblioteca playrec en una Mac Mini con una interfaz de audio Focusrite Saffire. Los sonidos fueron

presentados a través de unos auriculares abiertos calibrados Sennheiser HD 600 en la misma habitación donde se hicieron las grabaciones.

Los datos obtenidos del experimento fueron analizados mediante MDS usando el algoritmo INDSCAL. Este tipo de MDS es utilizado cuando se quiere realizar un ajuste con múltiples respuestas, y permite encontrar una configuración espacial que trata de satisfacer las distancias perceptuales de todos los sujetos. Esto es realizado por el algoritmo asignando diferentes pesos a cada dimensión según cada sujetos [10].

2.3.2. Resultados

Se analizaron los controles que se hicieron sobre la intensidad, con un *t-test* y no se encontraron diferencias significativas entre las respuestas de parejas intensidades modificadas contra las parejas con intensidades sin modificar ($p = 0,64$).

A pesar de que participaron sólo 5 sujetos, las respuestas presentaron un alto grado de semejanza entre si, obteniéndose un coeficiente de correlación promedio entre las respuestas de 0,8. Asimismo las respuestas de cada sujeto no se vieron influenciadas por el orden de presentación de los sonidos.

Los resultados del MDS utilizando 2 dimensiones se exhiben en la figura 2.12. Para establecer si esta es una buena configuración se computa el parámetro denominado *stress* que indica la diferencia promedio normalizada entre las distancias en el espacio original y las distancias del espacio con escalamiento (Apéndice A.1). El *stress* de esta configuración fue de 0,18. Lo que primero que se puede observar es que los multifónicos de cada clase han permanecido agrupados. Esto refleja que los participantes juzgaron más similares entre sí los tonos de una misma clase en comparación a los de otra clase. El arreglo espacial permite observar las relaciones entre las clases. Por ejemplo, podemos ver a la clase de los *Bicordios* como la clase que se encuentra en una posición relativamente más central y más cerca del resto. También podemos observar que los *Complejos* y los *Trémolos* se encuentran alineados en la dimensión 1.

En la figura 2.13 se muestra la configuración MDS utilizando 3 dimensiones. El *stress* en este caso fue de 0,13. Las primeras dos dimensiones muestran una configuración similar a la situación 2D, pero la tercera dimensión muestra a los *Bicordios* en una posición más

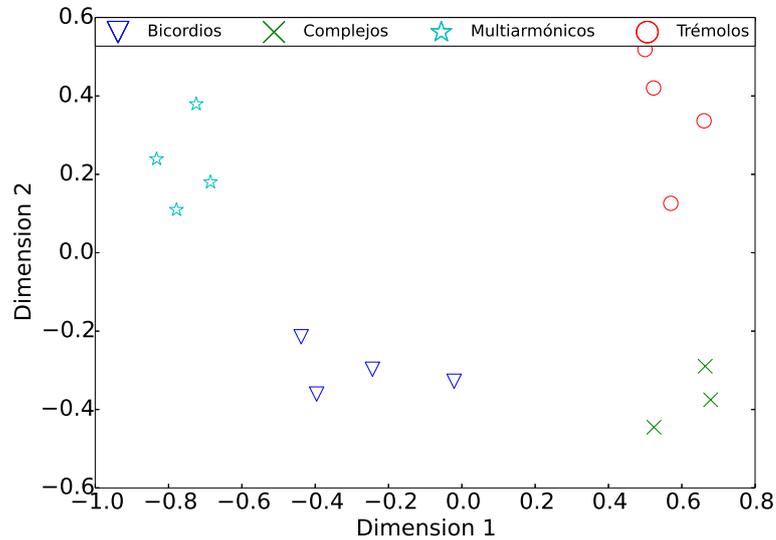


Figura 2.12: La configuración 2D MDS para los quince multifónicos utilizados en el experimento. Los marcadores indican las clases de multifónico: triángulos para los Bicordios, estrellas para los Multiarmónicos, círculos para los Trémolos y cruces para los Complejos.

alta separados del resto.

Para investigar las posibles interpretaciones de las dimensiones perceptuales entregadas por el MDS, computamos la correlación entre las dimensiones y los atributos acústicos utilizados en la sección 2.2.1. En la tabla 2.2 se muestran los valores de la correlación junto con un asterisco que indica si la correlación fue significativa ($\alpha = 0,05$). Para la configuración 2D, la primera dimensión correlaciona muy bien con la rugosidad y también con el centroide, y la segunda dimensión anticorrelaciona con la frecuencia fundamental F1 y la frecuencia de modulación. Para la configuración 3D, tenemos para la primera dimensión nuevamente una correlación con la rugosidad y el centroide, pero para la segunda dimensión se agrega una anticorrelación con la F2 y una correlación muy marcada con la rugosidad. La tercera dimensión anticorrelaciona principalmente con la F2 y el intervalo musical, lo cual explica porque los *Bicordios* se veían separados en esa dimensión en la figura 2.13.

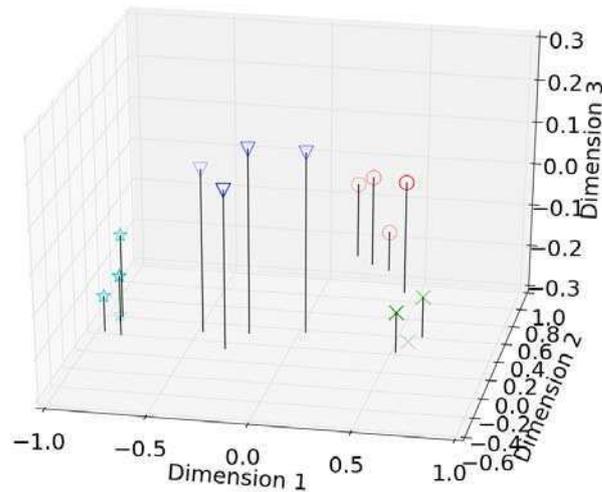


Figura 2.13: La configuración 3D MDS para los quince multifónicos utilizados en el experimento. Los marcadores indican las clases de multifónico: triángulos para los Bicordios, estrellas para los Multiarmónicos, círculos para los Trémolos y cruces para los Complejos.

	2D		3D		
	Dim 1	Dim 2	Dim 1	Dim 2	Dim 3
F1	-0.46	-0.62 *	-0.28	-0.78 *	0.26
F2	-0.25	-0.13	-0.23	-0.53 *	-0.68 *
f_m	-0.35	-0.66 *	-0.20	-0.78 *	0.08
Intervalo	0.18	0.51	0.04	0.26	-0.90 *
Centroide	0.69 *	-0.38	0.72 *	0.02	-0.54 *
Rugosidad	0.76 *	0.48	0.61 *	0.95 *	-0.08

Cuadro 2.2: Coeficientes de correlación entre las dimensiones MDS y los atributos acústicos para el conjunto de multifónicos del experimento. (f_m) frecuencia de modulación, (CE) centroide espectral, (F1) primera frecuencia fundamental (F2), segunda frecuencia fundamental. Las correlaciones significantes con tolerancia $\alpha = 0,05$ están marcadas con un asterisco

2.3.3. Resumen

Se realizó un experimento de comparación de pares de sonidos de multifónicos con el objetivo de estudiar las características tímbricas de un subconjunto de 15 multifónicos repartidos entre las cuatro clases. El experimento se realizó bajo condiciones controladas, con buen equipamiento y los participantes poseían un buen grado de entrenamiento auditivo. Al mismo tiempo se hicieron todas las comparaciones posibles con los 15 sonidos y se hicieron los controles suficientes para corroborar el criterio utilizado en las comparaciones.

Los resultados del escalamiento multidimensional MDS, permitieron observar tanto en 2 como 3 dimensiones cuáles fueron las dimensiones perceptuales utilizadas por los participantes. La configuración 2D se organizó básicamente a partir del centroide y la rugosidad en una dimensión y la frecuencia de modulación y $F1$ en la otra. Al aumentar las dimensiones, la configuración 3D mostró un nivel de *stress* más bajo e incluyó al intervalo musical como el atributo saliente en la tercera dimensión. Esto indica, que esta configuración, es una posible representación del espacio tímbrico perceptual utilizado por los sujetos participantes.

Podemos ver que a priori no es tan sencillo interpretar una dimensión puramente espectral y otra puramente temporal. Sin embargo, el hecho de que los atributos acústicos estén relacionados entre sí con diferentes tipos de estructuras, hace posible sacar algunas conclusiones extras. Para esto hay que analizar la rugosidad de los multifónicos. Por un lado la rugosidad es mayor cuando las componentes están muy cerca entre sí, es decir una f_m baja, pero por otro lado la rugosidad también depende de la cantidad de componentes, que en el caso de los multifónicos está relacionado con la cantidad de distorsión y el centroide. Esto hace que la rugosidad juegue tanto como un atributo espectral y temporal. A partir de este análisis podemos establecer, tanto en 2D como en 3D, a las primeras dimensiones como espectrales y a las segundas dimensiones como temporales.

Capítulo 3

Tonos de amplitud modulada sinusoidalmente

A partir del estudio de los sonidos multifónicos, pudimos apreciar que es posible generar una organización tímbrica en base a sonidos con modulaciones. Si bien el conjunto de sonidos multifónicos poseía una notable diversidad tímbrica, a pesar de provenir de un sólo instrumento y acústico, los multifónicos presentan una organización muy compleja de componentes e intervalos. A partir de esto, podría preguntarse si es posible generar una organización tímbrica similar con conjuntos de sonidos más sencillos, pero con características de modulación similares. Con esta motivación se desarrolló un experimento acerca de la percepción del timbre de tonos modulados en amplitud sinusoidalmente (tonos SAM). En este capítulo primero se da una introducción sobre los tonos SAM y luego se expone el desarrollo, resultados y análisis del experimento sobre la organización tímbrica de los tonos SAM.

3.1. Introducción

Los tonos SAM, a pesar de ser sonidos relativamente sencillos en su formulación y síntesis, tienen la posibilidad de generar un abanico tímbrico muy rico. Los diferentes tipos de perceptos que pueden evocar fueron desarrollados en la sección [1.4.1](#), a partir de la figura

1.11 que describe varias regiones del espacio de parámetros de los tonos SAM. Estos perceptos varían notablemente con la frecuencia de modulación. Para frecuencias de modulación bajas se genera la sensación de fluctuación de la intensidad, para frecuencias de modulación intermedias se da la sensación de rugosidad, y para frecuencias superiores se da la sensación de escuchar las componentes resueltas por separado. Al mismo tiempo, dependiendo de la frecuencia portadora y la frecuencia moduladora se puede generar una sensación de altura tonal, más baja que la frecuencia de la portadora (fenómeno de la fundamental ausente o altura virtual).

Los tonos SAM, pueden ser generados a partir de tres parámetros: la profundidad de modulación m , la frecuencia moduladora f_m y la frecuencia portadora f_c . Al variar la profundidad de modulación se puede realizar una transición entre un tono sinusoidal ($m = 0$) y un tono completamente modulado ($m = 1$). De ahora en más dejaremos fijo el valor en $m = 1$ para poder generar tono con modulaciones más salientes, por lo que usaremos la siguiente expresión para los tonos SAM:

$$SAM = [1 + \sin(2\pi f_m t)] \sin(2\pi f_c t) \quad (3.1)$$

De esta manera el conjunto de tono SAM utilizados queda representado en un espacio de 2 dimensiones, frecuencia moduladora f_m y frecuencia portadora f_c .

Al variar la frecuencia portadora f_c varía la nota o altura musical percibida de estos tonos. Si la frecuencia de modulación es muy baja, la altura es similar a la provocada por un tono puro, pero si la frecuencia de modulación es alta, entonces se puede generar una nota por el mecanismo de fundamental ausente [22]. Esto estaría indicando que el espacio tímbrico que se quiere estudiar tiene intrínsecamente variaciones en altura, que en principio están diferenciadas de la sensación tímbrica. A pesar de esto se consideró que hay suficiente interacción entre la frecuencias f_c y f_m como para generar sensaciones tímbricas que no son invariantes frente a cambios en f_c . Incluso existen estudios que indican que el rango invariante del timbre es de aproximadamente una octava [49], y en este experimento se realizarán comparaciones en un rango de más de tres octavas.

La posibilidad de generar un espacio de 2 dimensiones para controlar estos perceptos

planteó el siguiente interrogante. ¿Es posible reducir las dimensiones de un espacio perceptual solamente a las dimensiones que se utilizan para sintetizar el sonido? Dicho de otra manera, ¿Cómo se relacionan las distancias en el espacio de parámetros con las distancias perceptuales?

Para atacar estas preguntas, realizamos un experimento de comparación de pares de tonos SAM generados a partir de un espacio basado en las dimensiones f_m y f_c . A diferencia del experimento de los múltifónicos, donde se tuvo total control sobre los participantes y las condiciones del ambiente, en este caso se decidió hacer el experimento de manera masiva a través de Internet tomando ciertos criterios. Por un lado se consideró que los estímulos SAM eran lo bastante sencillos como para que la tarea de comparar y puntuar similitud pudiese ser realizada por un conjunto amplio de participantes, sabiendo que no se podrían controlar las condiciones de escucha y el estado de concentración de los sujetos. Por otro lado la posibilidad de llegar a un gran número de personas, permitió hacer un análisis sobre las respuestas y descartar los sujetos que hayan tenido una performance muy baja.

3.2. Método experimental

El experimento que se realizó es similar a los experimentos tímbricos de comparación de pares de sonidos. Dado un conjunto de sonidos, los participantes evalúan las similitudes tímbricas entre todas las combinaciones de pares de sonidos.

Plataforma Web

Para realizar el sitio web se desarrolló una plataforma para hacer experimentos utilizando HTML 5 y la *Web Audio* API para la generación y síntesis de sonido en tiempo real. Aunque en este experimento en particular los sonidos podrían haber sido generados previamente y almacenados como archivos, se decidió utilizar este mecanismo pensando en el desarrollo de un sistema general para la realización de experimentos psicofísicos *online*. El uso de *Web Audio* permite, entre otras cosas, crear sistemas de sonidos en módulos con disponibilidad de uso de filtros, FFT y convolución. El experimento de los tonos SAM esta disponible en <http://lapso.org/experimentos/sampairs>.

La escucha de estos sonidos es muy susceptible al tipo de parlantes y a los modos de resonancia de las salas, ya que son tonos hechos por sinusoides. Para tratar de minimizar las posibles variaciones que podía tener cada sujeto en su sistema de sonido, se decidió que el experimento requería el uso de auriculares. Aunque esto no garantiza respuestas planas, se asume que las fluctuaciones en las respuestas no son muy grandes comparadas al ancho de banda de los tonos presentados.

Condiciones experimentales

Se realizaron cuatro condiciones con distintos conjuntos de sonidos (figuras 3.1). En los cuatro casos se utilizaron 15 sonidos organizados en grupos y distribuidos de tal manera que se pudiesen escuchar varios tipos de perceptos. Se realizaron dos condiciones experimentales con grupos compactos y dos con grupos más abiertos.

Para esto se eligieron ciertas ubicaciones en el espacio de parámetros y se distribuyeron los sonidos en forma de triángulos equiláteros. La orientación de los triángulos fue elegida para que los sonidos no compartan la f_c o la f_m . Las condiciones 1 y 2, fueron tomadas como las principales, mientras que la 3 y 4 sirvieron como control de algunos parámetros. En la condición 3 se modificó la posición del racimo central mientras que en la condición 4 se rotaron las formas triangulares.

Desarrollo experimental

Al comienzo del experimento, se daba una etapa de entrenamiento donde los sujetos eran introducidos al tipo de sonidos que iban a escuchar y podían observar el espacio de parámetros f_m vs. f_c como el de la figura 3.2, donde se iban mostrando distintos ejemplos de sonidos y sus coordenadas. También, aprovechando las posibilidades de sintetizar sonido en tiempo real, luego de los ejemplos, los participantes debían explorar el espacio manualmente moviendo un cursor. Se realizó este entrenamiento guiado para generar familiaridad con los tonos para los participantes que no tenían experiencia en este tipo de sonoridades y para establecer un marco en común para la estimación de las similitudes.

Luego de la etapa de entrenamiento comenzaba la etapa de comparación de pares. Para esto se elegía al azar una de las cuatro condiciones experimentales. Con los 15 sonidos se

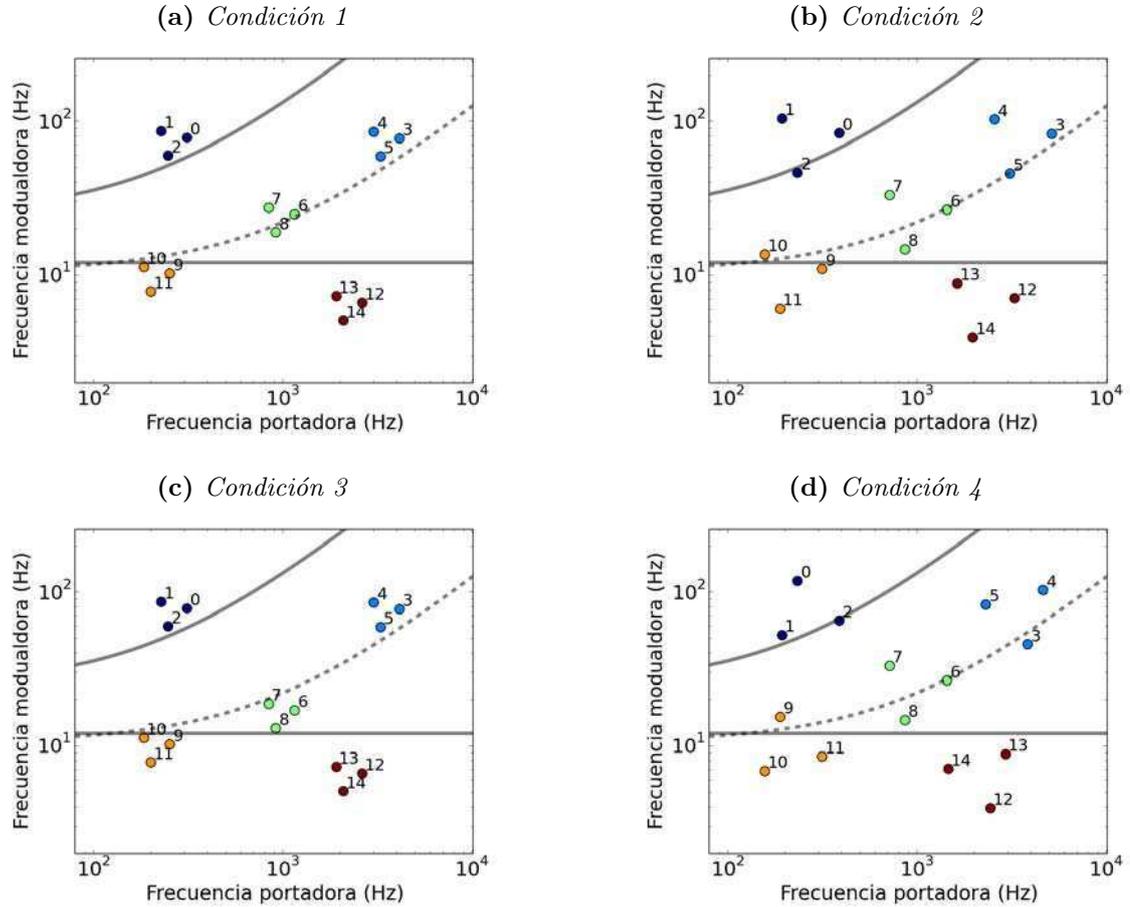


Figura 3.1: Coordenadas f_m y f_c de los conjuntos de sonidos dados por la ecuación 3.1 para las cuatro condiciones experimentales. La línea horizontal ($f_m = 11\text{Hz}$) indica aproximadamente el límite para el percepto de fluctuaciones. La curva entrecortada indica aproximadamente la zona de máxima rugosidad en función de f_c calculada con el modelo de [104] (Apéndice A.3), y la curva sólida indica el valor de ERB para cada f_c

generaron 105 pares que fueron presentados al azar y también se incluyeron, al comienzo, un conjunto de 10 pares extra. La tarea que debían realizar los participantes, requería estimar la distancia entre los sonidos en base a una escala de 11 puntos del 0 al 10, donde 0 correspondía a sonidos muy distintos o distantes y 10 a sonido muy parecidos o cercanos. Estas respuestas permiten generar configuraciones del espacio tímbrico perceptual mediante el análisis de MDS.

Una vez completadas las evaluaciones, los participantes debían completar un breve for-

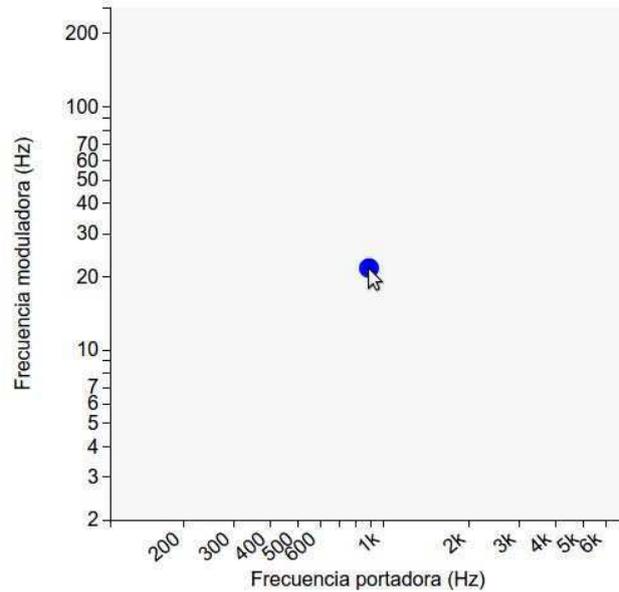


Figura 3.2: *Captura de la pantalla del ejemplo de exploración en el espacio de parámetros. El sujeto podía desplazarse con el puntero e ir escuchando en tiempo real los tonos dados por las sistema de coordenadas.*

mulario con información personal y establecer su experiencia previa. Se consultó la experiencia en los campos de la edición y producción musical, la ejecución de instrumentos y la experiencia en síntesis sonora. Para cada campo se dieron las opciones: (I) sin experiencia, (II) experiencia básica con dedicación esporádica y (III) gran experiencia con dedicación habitual.

Al finalizar, los participantes obtenían un resultado de cuán buena había sido su performance juzgando las distancias. Este puntaje se computó en base a la correlación entre las distancias perceptuales y las distancias en el espacio de parámetros. El espacio de parámetros utilizado para calcular estas distancias es equivalente al de las figuras 3.1, es decir, sus ejes fueron tomados en escala logarítmica.

3.3. Resultados

Presentaremos primero las estadísticas de los datos personales, experiencias en edición, ejecución y síntesis de los participantes en la tabla 3.1.

Se evaluó si las experiencias personales tuvieron influencia en la performance. Para esto

Condición	N	Género			Edades			Edición			Ejecución			Síntesis		
		F	M	Pr.	Min	Max	I	II	III	I	II	III	I	II	III	
1	24	8	16	30	19	46	9	8	6	4	9	10	9	11	3	
2	25	9	16	27	21	36	11	5	8	5	7	13	14	6	4	
3	21	7	14	27	18	41	12	5	3	8	5	7	13	4	4	
4	22	9	13	28	18	63	11	7	3	6	8	8	14	4	4	
Todos	92	33	59	28	18	63	43	25	20	23	29	38	50	25	15	

Cuadro 3.1: *Datos personales, experiencias en edición, ejecución y síntesis entregadas en el reporte por los participantes*

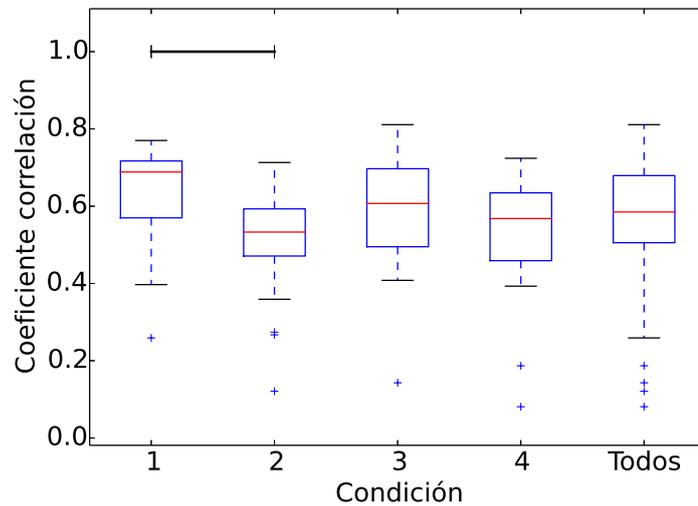


Figura 3.3: *Performances de los sujetos: distribuciones de los coeficiente de correlación entre las distancias perceptuales y las distancias del espacio de parámetros. Las barras horizontales indican las parejas de condiciones que presentaron diferencias significativas*

se hicieron los correspondientes tests estadísticos para evaluar si las performances agrupadas por niveles de experiencia I, II y III poseían diferencias significativas, pero estas no fueron observadas para ninguna de los tres tipos de experiencia (*Edición*, *Ejecución* y *Síntesis*).

Para establecer el valor de performance se utilizó el coeficiente de correlación entre las distancias perceptuales y las distancias en el espacio de parámetros. En la figura 3.3 se muestran las distribuciones de estos valores de correlación. A primera vista se puede observar que los valores de la correlación tienen comportamientos distintos dependiendo de la condición. Para la condiciones 1 y 3 (triángulos pequeños en figura 3.1), los valores de la

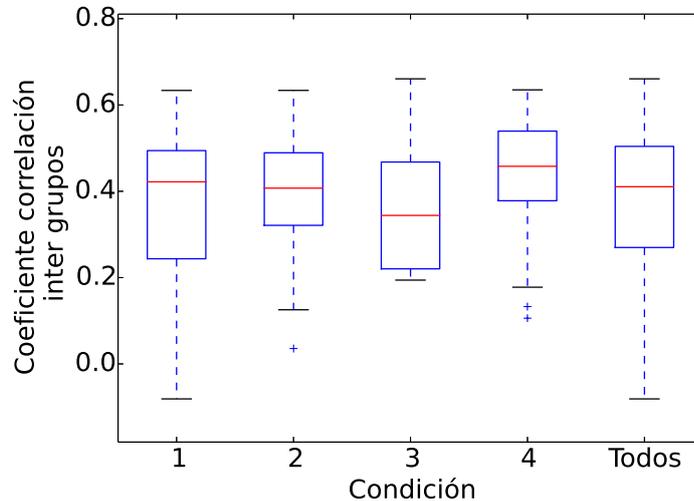


Figura 3.4: Performances de los sujetos: distribuciones de los coeficiente de correlación para las comparaciones de sonidos inter grupos.

correlación entregaron valores más altos que para la condiciones 2 y 4 (triángulos grandes en figura 3.1), siendo significativas las diferencias entre las condiciones 1 y 2 (indicado por una barra horizontal). Para estudiar este comportamiento, es necesario diferenciar entre las respuestas correspondientes a comparaciones dentro de un mismo grupo (o triángulo), y las que correspondían a comparaciones entre grupos. Al diferenciarlos, ocurre que los valores de la correlación se ven influenciados por el hecho de que las respuestas *intra* grupo son muy homogéneas para los casos de las condiciones 1 y 3. En la figura 3.4 se pueden ver las mismas correlaciones pero sólo para las comparaciones *inter* grupos, dando por resultando unas distribuciones mucho más homogéneas.

Para observar como se comportaron las comparaciones *intra* grupo, en vez de realizar una correlación, se realizó un cociente entre las distancias perceptuales y las distancias del espacio de parámetros. Primero se modificaron las escalas de las respuestas con una función lineal para que ajusten lo mejor posible con las distancias reales. Con este cambio de escala se simplifica el análisis del cociente de las distancias, ya que valores cercanos a 1 indicarían que las distancias perceptuales y las reales, están en concordancia.

En la figura 3.5 se muestran las distribuciones de estos cocientes diferenciando en los casos *intra* y *inter* grupos. Se puede ver en el panel inferior, que el cociente *inter* grupo

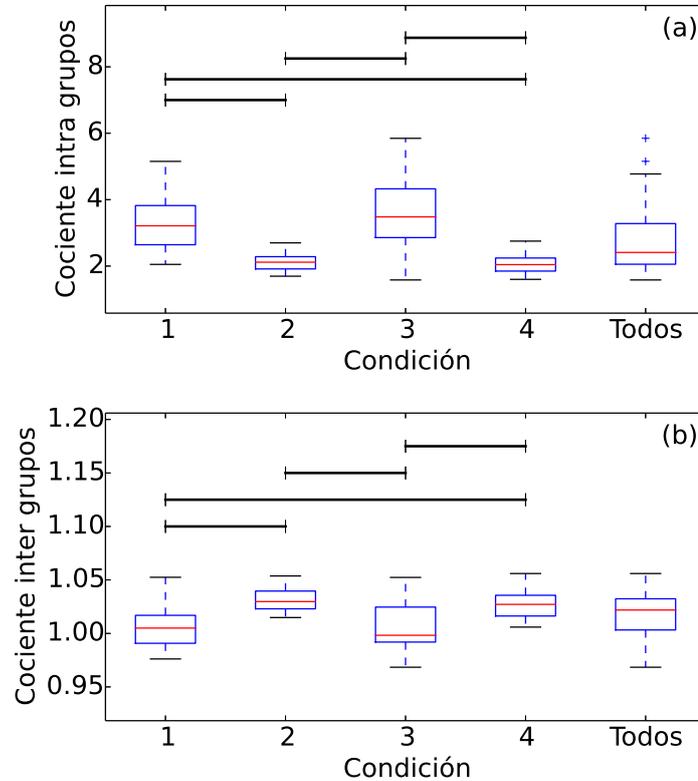


Figura 3.5: Cocientes entre distancias perceptuales y del espacio de parámetros: distribuciones de los cocientes para (a) distancias intra grupo y (b) distancias inter grupo. Las barras horizontales indican las parejas de condiciones que presentaron diferencias significativas

entregó valores cercanos a 1, por lo que se aprecia el escalamiento realizado, sin embargo para los cocientes *intra* grupo se ve que los valores son un tanto mayores que 1, por lo que las distancias perceptuales fueron mayores que las distancias del espacio. Esto indicaría que los sujetos sobreestimaron las distancias entre los sonidos más cercanos, sobre todo para las condiciones 1 y 3. Las barras horizontales indican si hubo diferencias significativas entre las condiciones, y puede verse que ocurre para todas las combinaciones de las condiciones 1 y 3 contra las 2 y 4. Estas diferencias indican que para las condiciones 1 y 3 donde los tonos se encontraban más próximos, hubo un efecto de proximidad que generó respuestas de distancias mayores en comparación con la de las condiciones 2 y 4.

Otra forma de estudiar las respuestas de los participantes, es analizando las distancias

entre los sonidos para cada dimensión física por separado y correlacionar las distancias perceptuales con las distancias en términos de la frecuencia portadora y moduladora. En la figura 3.6 se pueden ver las distribuciones para f_c y f_m y para cada condición. Los asteriscos indican si las diferencias entre las distribuciones de f_c y f_m fueron significativas. Puede observarse que las mayores correlaciones se dan con la frecuencia portadora f_c , ya que es más probable que los participantes hayan utilizado ese atributo de los sonidos como mayor pista para generar las estimaciones de las distancias. Sin embargo no es fácil observar si hubo algún comportamiento relevante relacionado con el tipo de condición, ya que las distribuciones parecen muy heterogéneas.

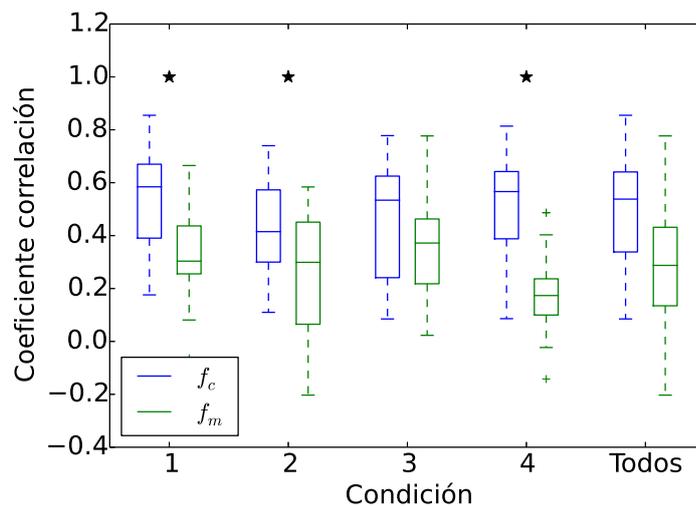


Figura 3.6: Correlaciones con las distancias por cada eje, f_c y f_m : distribuciones de los coeficiente de correlación entre las distancias perceptuales y las distancias en términos de f_c y f_m

Escalamiento multidimensional

De la misma manera que en el experimento de los multifónicos, se procedió a estudiar las distancias perceptuales mediante la técnica de escalamiento multidimensional (MDS). Para esto se utilizó el algoritmo INDSCAL que permite contemplar todas las respuestas de los participantes en una sola configuración geométrica. Se computaron configuraciones en 2 y 3 dimensiones. En la figura 3.7 se pueden ver las configuraciones 2D para las cuatro

condiciones. Lo más importante que se observa es que las configuraciones tienen una gran semejanza con las configuraciones originales (3.1), por lo que en términos generales se puede decir que las respuestas promedio de los participantes lograron capturar el espacio tímbrico que se había planteado.

A pesar de que el espacio original poseía sólo 2 dimensiones, es posible realizar una configuración MDS de 3 dimensiones para estudiar la existencia de alguna otra dimensión perceptual que haya influenciado la estimación de las distancias. En la figura 3.8 se muestran las configuraciones 3D. A primera vista puede observarse que las posiciones de los grupos verdes y algunos sonidos naranjas se encuentran más elevadas que el resto, lo que indicaría que en la tercera dimensión estos dos grupos están más cerca. Al mismo tiempo se puede ver que las posiciones de los puntos en las dimensiones 1 y 2, parecerían ser similares a las de la configuración 2D. Para ver esto con mayor claridad, se muestran sólo las dimensiones 1 y 2 de las configuraciones 3D en la figura 3.9, y puede observarse que las posiciones aunque respetan a grandes rasgos las posiciones del espacio de parámetros, se presentan algunas posiciones superpuestas.

Para cuantificar estas discrepancias, se calcularon los coeficientes de correlación entre las dimensiones de las configuraciones MDS y las posiciones de los sonidos en el espacio de parámetros (tabla 3.2). En el caso 2D, se puede ver una gran nivel de concordancia entre las configuraciones MDS y las posiciones originales de los sonidos para todas las condiciones y especialmente para las condiciones 2 y 3. Para el caso 3D, los valores son similares, pero mejora la correlación f_m de la condición 1. También se computó la correlación de la tercera dimensión (altura en figura 3.8) con los valores correspondientes de rugosidad de

Condición	2D		3D		
	Dim1 y f_c	Dim2 y f_m	Dim1 y f_c	Dim 2 y f_m	Dim 3 y Rugosidad
1	0.98	0.79	0.98	0.97	0.62
2	0.99	0.95	0.96	0.91	0.44
3	0.99	0.99	0.99	0.96	0.62
4	0.98	0.86	0.96	0.88	0.30

Cuadro 3.2: Correlaciones de las coordenadas del espacio de parámetros con las dimensiones de las configuraciones MDS

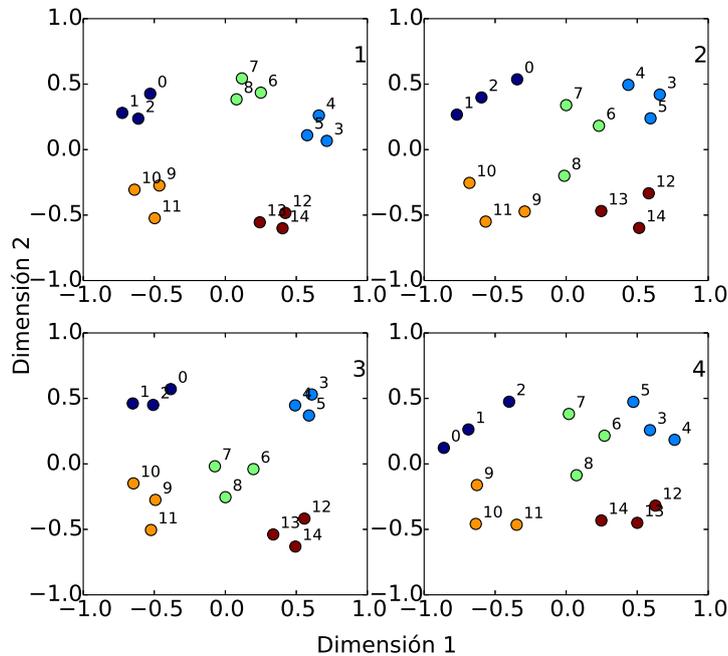


Figura 3.7: Configuraciones MDS 2D: resultados del escalamiento multidimensional 2D para cada condición.

cada sonido, obteniendo valores significativos para las condiciones 1 y 3. Esto indicaría que los participantes utilizaron la rugosidad como una dimensión perceptual, sin embargo es necesario estudiar si realmente la configuración 3D representa un mejor ajuste de los datos de las distancias. En la figura 3.10 se pueden observar la magnitud *Stress por sujeto*, que indica mejores ajustes para valores cercanos a 0. El asterisco indica diferencias significativas entre las distribuciones de *Stress por sujeto*, por lo que se concluye que la configuración 3D representa mejoró los datos pero sólo para las condiciones 1, 2 y 3.

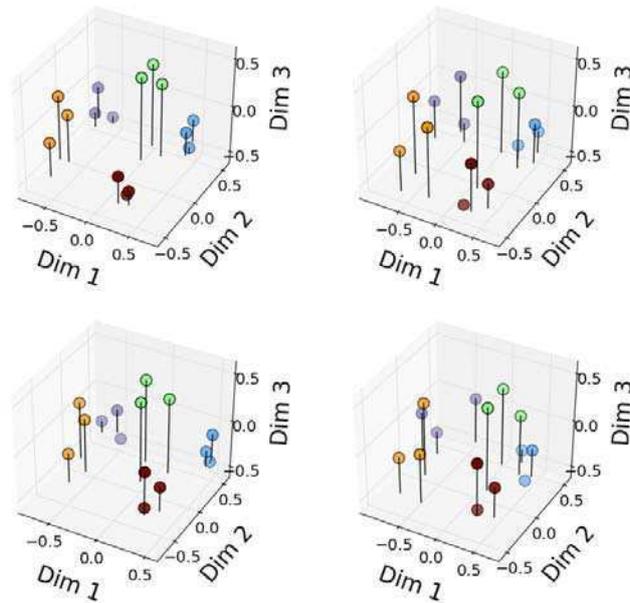


Figura 3.8: Configuraciones MDS 3D: resultados del escalamiento multidimensional 3D para cada condición.

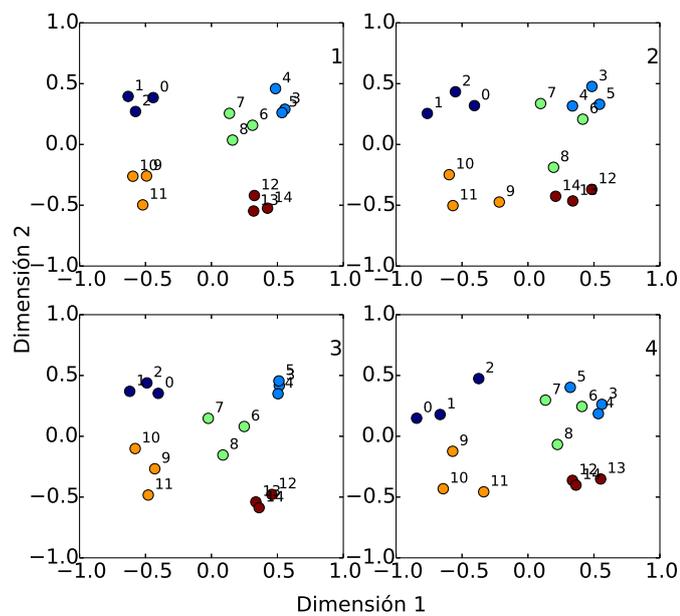


Figura 3.9: Configuraciones MDS 3D: proyección de la configuración de 3 dimensiones sobre las primeras 2 dimensiones.

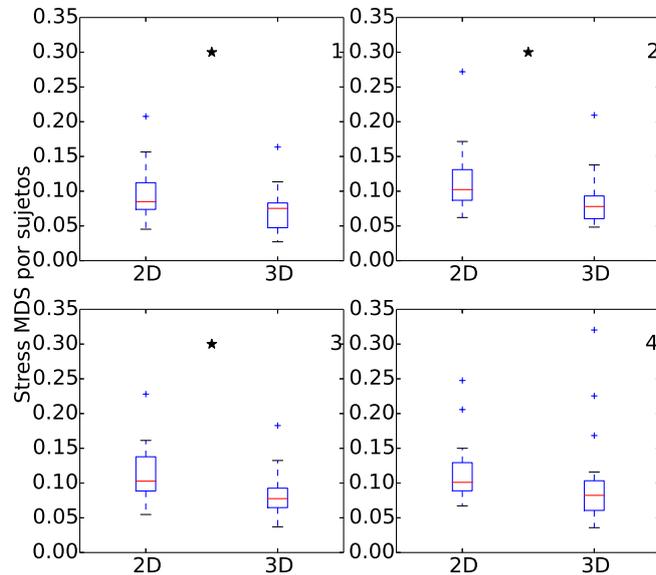


Figura 3.10: Distribuciones de los coeficientes de correlación entre las distancias evaluadas por los sujetos y las distancias del espacio de parámetros

3.4. Discusión

La realización de este experimento masivo permitió corroborar diferentes cuestiones. Desde el punto de vista técnico se pudo desarrollar un experimento web con una tecnología moderna como HTML 5 que permitió disponer de una interfaz intuitiva para hacer más accesible el experimento a un público heterogéneo. Con respecto al tipo de experimento y su relación con los participantes, podemos concluir que los participantes entendieron la consigna y pudieron desarrollar la tarea sin inconvenientes. Sin embargo las respuestas presentaron un nivel de dispersión importante, pero esto no impidió que puedan observarse las correlaciones entre las respuestas y los sonidos comparados. Con respecto a los análisis de los resultados, podemos establecer que las distintas condiciones presentaron comportamientos similares, encontrando las mayores diferencias para las condiciones de triángulos pequeños 1 y 3 con las de triángulos grandes 2 y 4. A partir de distinguir las comparaciones en *intra* e *inter* grupo, se observó que las distancias perceptuales para los pares de sonidos que son muy similares no son juzgadas en la misma escala que el resto de las distancias

en el espacio de parámetros. Es decir se da un efecto de proximidad, que permitiría a los sonidos agruparse con más fuerza. Finalmente las configuraciones de MDS indicaron que la correspondencia entre los espacios perceptuales y de parámetros fue muy alta, y sugirió que los participantes pudieron utilizar alguna pista relacionada a la rugosidad para las condiciones 1 y 3. Sin embargo no está claro si las diferencias en la tercera dimensión del MDS 3D, son un artefacto asociado a la forma que el MDS interpreta las diferencias entre las distancias *intra* e *inter* grupo.

Capítulo 4

Modelado de la periferia auditiva

Este capítulo está dedicado al desarrollo de un modelo de la periferia auditiva y su aplicación para analizar algunos aspectos relevantes a la percepción de tonos con modulación en amplitud, para complementar lo visto en los capítulos anteriores.

Uno de los propósitos de esta tesis fue desarrollar un modelo computacional eficiente de la periferia auditiva que permitiese calcular cocleogramas realistas de los estímulos utilizados en los experimentos psicofísicos, y derivar atributos vinculados a la percepción de tonos con modulación.

Este modelo se desarrolló luego de realizar una revisión exhaustiva (y en muchos casos realizando desde cero estos modelos o reutilizando código existente) de la enorme variedad de modelos de la mecánica coclear. En la elaboración del modelo final se priorizaron de forma simultánea la baja complejidad algorítmica, la eficiencia computacional, y el acuerdo con los datos publicados en la literatura. Como paso previo para este desarrollo se presentan en las secciones siguientes los aspectos más relevantes para el modelado de la periferia auditiva y los diferentes niveles de detalle en los cuales se puede describir la mecánica coclear (la cual sea quizás la etapa más crítica en la transducción del sonido a impulsos nerviosos en el sistema central). Luego se presentan los resultados numéricos del modelo para diferentes condiciones de algunos parámetros. Finalmente se muestran dos aplicaciones del modelo. En un caso para el cálculo de la rugosidad de forma temporal en tonos con batidos y tonos SAM. Y en otro para estudiar los procesos de la comparación de pares de

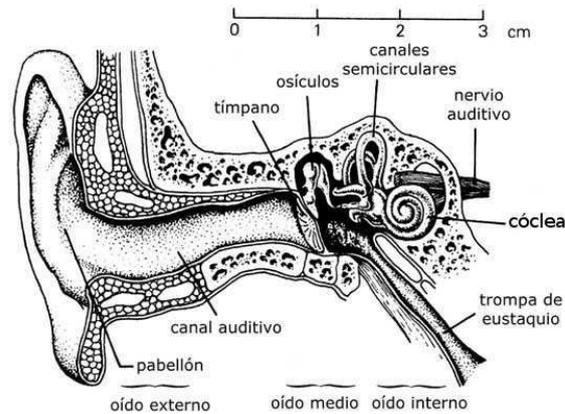


Figura 4.1: *Esquema del sistema auditivo periférico*

sonidos del experimento de tonos SAM.

4.1. Introducción

El sistema auditivo es el sistema sensorial que produce el sentido de la audición o escucha. Está dividido en una parte periférica y una parte central. La primera etapa de la percepción sonora ocurre en la periferia, e involucra la transducción del sonido en impulsos nerviosos. La periferia a su vez puede dividirse en tres secciones: oído externo, medio e interno.

4.1.1. Oído externo y medio

El sonido ingresa al sistema auditivo periférico (figura 4.1) a través de la oreja y el canal auditivo hasta llegar a la membrana timpánica, donde las ondas de presión son transformadas en vibraciones mecánicas. Luego, estas vibraciones son transmitidas dentro del oído medio por la cadena de huesecillos (martillo, yunque y estribo) hacia la ventana oval en el oído interno. El oído medio funciona como un adaptador de impedancias y logra un efecto transformador en la presión 24 dB, principalmente debido a las diferencias en áreas entre la membrana timpánica y la ventana oval [90].

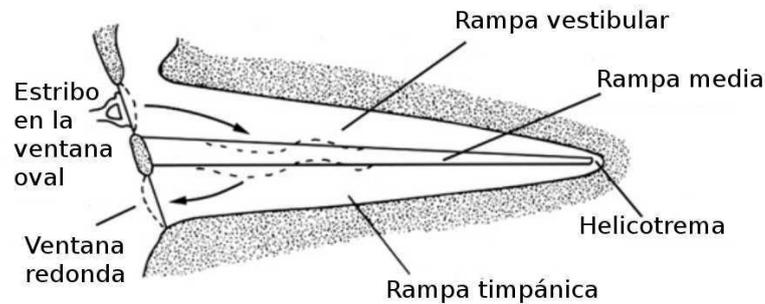


Figura 4.2: *Cóclea desenrollada*

4.1.2. Oído interno y la cóclea

El oído interno consiste en un conjunto complejo de cavidades llenas de fluidos que incluye a la cóclea y a los canales semicirculares.

La cóclea se encuentra enrollada de forma espiralada y cada vuelta es más pequeña que la anterior. En los humanos, la cóclea desenrollada tiene un largo aproximado de 35 mm, un ancho aproximado de 2 mm y suele girar dos vueltas y media. El comienzo de la cóclea se llama la base y el extremo final el ápex.

En la figura 4.2 se ilustra a la cóclea desenrollada y puede apreciarse que esta dividida en tres compartimentos conteniendo fluido, llamados rampas: la rampa timpánica, la rampa media (también llamado partición coclear) y la rampa vestibular. En la base de la cóclea las rampas están conectadas con el oído medio mediante dos ventanas. La ventana redonda es sensible al movimiento del fluido en la rampa timpánica, mientras que la ventana oval conecta el estribo directamente con el fluido en la rampa vestibular. En la región del ápex se encuentra el helicotrema, un pequeño orificio donde se juntan las rampas vestibular y timpánica. Los movimientos del fluido en bajas frecuencias pueden cruzar por este orificio, permitiendo que los cambios estáticos de la presión externa no perturben al sistema.

Parte de la separación entre la rampa timpánica y la partición coclear está dada por la membrana basilar (MB), una membrana elástica que tiene libertad para vibrar transversalmente. Las propiedades mecánicas de la MB varían a lo largo de su extensión, siendo más rígida y estrecha en la base y menos rígida y más amplia en el ápex. Esto permite que

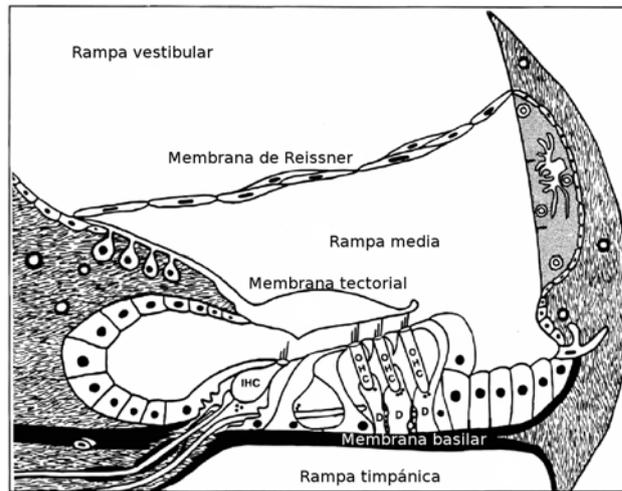


Figura 4.3: Estructura interna del órgano de Corti (dirección radial). OHC: células ciliadas externas; IHC: células ciliadas internas; D: células de Deiters.

las frecuencia de resonancia de la membrana varíen continuamente, permitiendo la segregación en frecuencias del sonido ingresante. Sobre la MB se encuentra el órgano de Corti, un epitelio de células altamente diferenciadas donde se encuentran las células sensitivas del sistema auditivo [90].

Órgano de Corti

El órgano de Corti (figura 4.3) es una compleja estructura que contiene dos tipos de células ciliadas mecánico-sensibles, las internas y las externas (IHC y OHC del inglés *inner hair cell* y *outer hair cell* respectivamente). Cada oído posee alrededor de 300 IHC y 900 OHC.

Los dos tipos de células cuentan con un racimo de cilios que está en contacto con el fluido. Las vibraciones que se propagan por el fluido perturban la posición de estos cilios y desencadena una serie de procesos fisiológicos dentro de las células [62]. Para el caso de las IHC, estos procesos tienen el propósito de traducir los estímulos acústicos en eventos sinápticos y los subsecuentes disparos neuronales. Para las OHC, estos procesos generan movimientos somáticos similares a un efecto piezoeléctrico sobre el cuerpo de la células que permiten amplificar activamente la señal [2].

Los movimientos de los cilios de las IHC modifican el potencial eléctrico de membrana

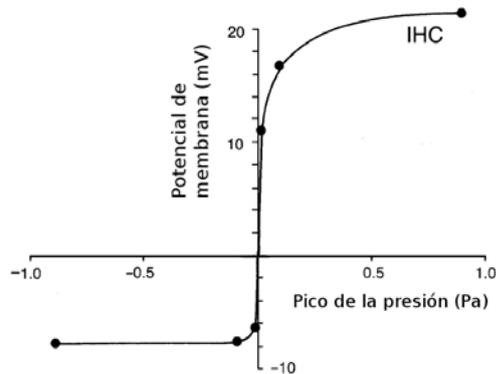


Figura 4.4: *Función de entrada-salida de las IHC. En el eje de las abscisas, se denotan los valores pico de la señal acústica de estímulo y, en el eje de las ordenadas, las variaciones del potencial de membrana. Las señales son traducidas por una función no lineal con saturación y asimetría. Figura adaptada de Dallos [16].*

de dichas células. Las vibraciones mecánicas que se propagan por el fluido mueven estos cilios y transducen las vibraciones en potencial eléctrico. En esta transducción ocurren dos fenómenos que modifican sustancialmente la señal. Por un lado, ocurre una saturación y rectificación de la señal debido a que los cilios se desplazan con mayor facilidad en una dirección más que en otra. En la figura 4.4 se puede ver la forma de la función de entrada-salida de las IHC, donde podemos destacar la asimetría y la saturación. Por otro lado, la capacidad de las membranas de las células de seguir variaciones en voltajes de alta frecuencia es limitada, por lo que la señal es filtrada con una característica pasa bajos con corte alrededor de los 3 kHz [108].

Dentro de la células, las variaciones del potencial eléctrico generan una cascada de efectos que conducen a la generación de disparos neuronales de la fibra auditiva. Este proceso puede pensarse como una conversión analógica-digital y es necesario que sea realizado de forma eficiente y robusta. Para este propósito está la sinapsis de cinta, una sinapsis altamente especializada para liberar vesículas rápidamente y con gran cantidad de neurotransmisores. Esta liberación tiene la cualidad de generar eventos post sinápticos con gran probabilidad de éxito, es decir, permite que se produzcan disparos neuronales que representan de forma robusta a la señal.

La sinapsis de cinta posee una marcada adaptación, lo que genera una gran liberación

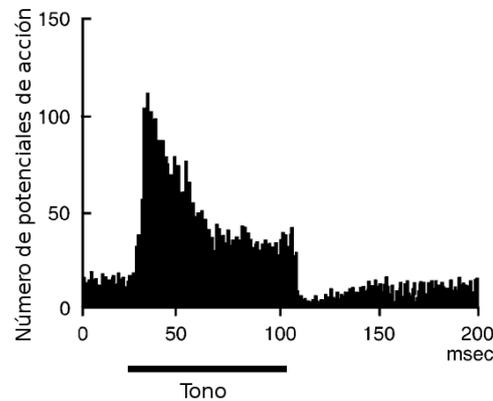


Figura 4.5: *Histograma post-estímulo mostrando los potenciales de acción en respuesta a un tono. Puede observarse la marcada adaptación luego del comienzo del tono y también una supresión al finalizar. Figura adaptada de Pickles [90].*

de vesículas en el comienzo de un sonido, una subsecuente disminución y una supresión al final (ver figura 4.5). Desde el punto de vista de la escucha, la adaptación permite al sistema auditivo detectar nuevos eventos sonoros sobre un continuo de sonidos de fondo, ya que funciona resaltando los ataques y la supresión, en cambio, genera un efecto de enmascaramiento temporal. Podemos hacer un paralelismo de este efecto de adaptación con el funcionamiento de un compresor de audio. El tiempo de ataque que se suele usar en un compresor, sería el equivalente al proceso de adaptación, y de manera análoga, el tiempo de relajación se correspondería con la supresión.

Toda la información del sonido es codificada en forma de disparos neuronales por las 30 mil fibras auditivas que conectan la cóclea con el núcleo coclear, el primer relevo en la ruta auditiva. La información es codificada de forma muy eficiente [105], aunque todavía se desconocen los mecanismos exactos que permiten esto y el proceso de decodificación posterior. Sin embargo, es sabido que a lo largo de la ruta auditiva se pueden encontrar mapas neuronales que describen diferentes aspectos del sonido en organizaciones espaciales de neuronas. Una de estas es la organización tonotópica, la cual está relacionada con la segregación de frecuencias que ocurre en la cóclea y se preserva en toda la ruta auditiva [99]. Por otro lado, las fibras auditivas pueden capturar las periodicidades de la envolvente temporal de los estímulos y hay evidencia de que esta información es decodificada en el colículo inferior generando una organización periodotópica.

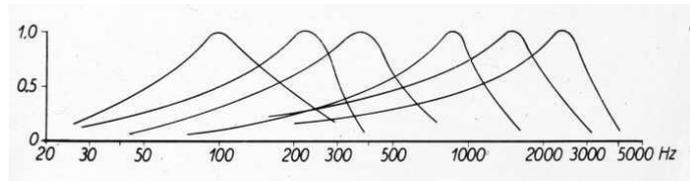


Figura 4.6: Observaciones de la onda viajera de von Békésy [6]

A continuación profundizaremos en la etapa del procesamiento auditivo relacionada a la mecánica coclear.

4.1.3. Mecánica coclear

La mecánica coclear estudia el funcionamiento de la cóclea en base a la interacción de un conjunto de elementos: la dinámica de los fluidos cocleares, la micromecánica del órgano de Corti y a la fisiología de las células ciliadas. La interacción de estos elementos generan la segregación en frecuencia y la amplificación del sonido dentro de la cóclea. Estos fenómenos pueden ser distinguidos entre pasivos y activos.

Cóclea pasiva

Se pueden estudiar los fenómenos pasivos a partir de una cóclea muerta, ya que los procesos activos necesitan que las OHC estén vivas. El comportamiento de la cóclea pasiva fue observado por von Bekesy [6] en 1928 a partir de cócleas de cadáveres humanos. Este trabajo identificó por primera vez la existencia de ondas viajeras sobre la MB y la segregación en frecuencias utilizando luz estroboscópica (figura 4.6).

Las ondas viajeras son el resultado del acoplamiento hidrodinámico entre el fluido coclear y la MB [19]. Este acoplamiento es el responsable de que las vibraciones de una sección de la MB se propaguen hacia otras, dado que el acoplamiento longitudinal de la MB en términos de su rigidez es mínimo comparado al que ejerce el fluido¹. Esto quiere decir que la MB puede verse como un arreglo de osciladores acoplados por la interacción con el fluido. Si las frecuencias de resonancia de estos osciladores forman un gradiente desde las frecuencias más altas en la base hasta las más bajas en el ápex, entonces es posible generar una

¹Recientemente, se han propuesto mecanismos alternativos para el acoplamiento entre las secciones, a través de la estructura en "Y" de la figura 4.10 [85]

onda viajera cuya posición de amplitud máxima depende de la frecuencia de estímulo. En este trabajo utilizaremos este enfoque de arreglo de osciladores para desarrollar el modelo.

Cóclea activa

Después de los hallazgos de von Békésy en la década de 1920, el descubrimiento del mecanismo activo tuvo que esperar hasta 1978 cuando Kemp [64] reportó por primera vez las emisiones otoacústicas, que son sonidos generados por la cóclea y evidencian su capacidad de generar energía de forma activa. En 1948 Gold ya habían hipotetizado la existencia de dicho mecanismo.

Hoy en día sabemos que la cóclea activa es la responsable de ciertos comportamientos del sistema auditivo que son fundamentales para la escucha:

- Sensibilidad:

Amplificaciones de hasta 60 dB para sonidos tenues, que permiten escuchar sonidos cercanos al límite impuesto por el ruido térmico dentro del oído.

- Selectividad en frecuencia:

La cóclea posee una alta selectividad en frecuencia. Mediante la medición de curvas de sintonía neuronales, es posible identificar cuánto tiene que ser modificada una frecuencia para que esta deje de ser percibida por una neurona. La pendiente de uno de los flancos de estas curvas es aproximadamente 300 dB_{SPL} por octava, lo que puede ser logrado mediante un filtro de orden 50^2 .

- Compresión:

Es posible escuchar sonidos en el rango de 0 a 120 dB_{SPL} , ya que dicho rango es comprimido en 30 dB (tres órdenes de magnitud) en términos de amplitud de las vibraciones membrana basilar. Esto sólo ocurre en la región de la cóclea que tiene la frecuencia característica del estímulo, y es el resultado de una no linealidad en el sistema activo.

²Cada orden atenúa -6 dB sobre la frecuencia de corte para un filtro continuo, $6\text{dB} \times 50 = 300$.

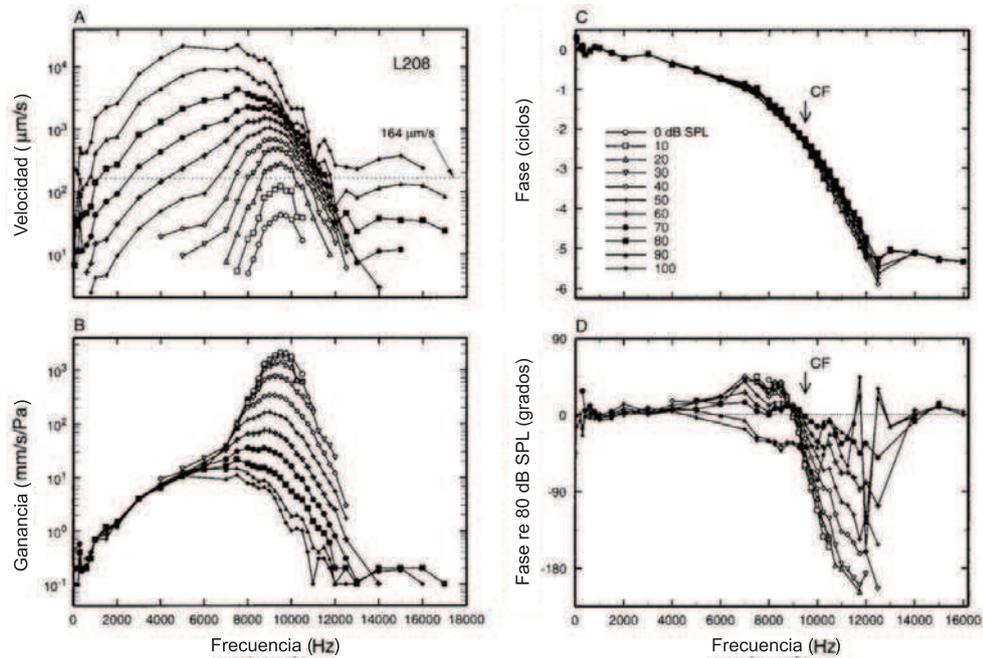


Figura 4.7: Las magnitudes y fases de respuestas de la MB en una cóclea de chinchilla. A: Magnitud de la velocidad en función de la frecuencia del estímulo e intensidades. B: Mismos datos del panel A, pero normalizados por la intensidad del estímulo. C: Respuestas de la fase del desplazamiento pico en dirección de la rampa timpánica, relativas al pico en el canal auditivo. D: Como en C, pero normalizadas a la fase del estímulo de 80 dB. Medidas realizadas con un vibrómetro láser y figura tomada del artículo Ruggero et al. [101].

- Productos de distorsión:

La no linealidad de la cóclea tiene como consecuencia la creación de nuevas frecuencias no presentes en el estímulo. Para un estímulo de dos frecuencias f_1 y f_2 , el producto más prominente suele darse en la frecuencia $2f_1 - f_2$.

- Emisiones otoacústicas:

El oído puede emitir sonido a partir de la energía que suministra el sistema activo mediante autooscilaciones que surgen de una inestabilidad. Las emisiones pueden ocurrir de manera espontánea o mediante estimulación.

En la figura 4.7, podemos observar algunos de estos comportamientos en mediciones realizadas in vivo en las cócleas de chinchillas a través de un vibrómetro láser [101]. Lo que se muestra en los paneles A y C son respuestas en magnitud y fase de la velocidad de

la membrana basilar en función de las frecuencias y de la intensidad. En los paneles B y D se muestran las mismas magnitudes pero normalizadas. Este tipo de mediciones se realiza observando las vibraciones para un determinado punto de la cóclea y realizando un barrido en la frecuencia del estímulo, en este caso el punto de observación posee una frecuencia característica (CF) de 9.5 kHz.

El nivel de sensibilidad puede verse en el panel B, donde se muestra la ganancia de la magnitud. Se observa la mayor ganancia para el estímulo de 0 dB_{SPL} debido a que para esas intensidades se da la mayor contribución de la parte activa. A medida que la intensidad del estímulo crece, la contribución de la parte activa decrece, siendo la ganancia cada vez menor hasta que, para intensidades altas, la respuesta es puramente pasiva. La compresión se evidencia cerca de la CF ya que se obtiene una ganancia dependiente de la intensidad del estímulo mientras que lejos de la CF la ganancia es constante. La respuesta de la fase en panel B, indica la cantidad de ciclos que la onda viajera recorre para los distintos estímulos.

En la figura 4.8 podemos ver una información similar a la de la figura previa pero con las intensidades en el eje de las abscisas. En esta representación podemos observar como la compresión ocurre principalmente para las frecuencias cercanas a la CF (10 kHz); en cambio, el comportamiento es lineal para las frecuencias más bajas (panel A) y más altas (panel B).

Otra forma de visualizar la información de estas respuestas es mediante las denominadas curvas de sintonía. Para cada frecuencia se busca la intensidad del estímulo que provoca una actividad umbral. En la figura 4.9 podemos observar las curvas de sintonía para la velocidad y desplazamientos de la MB junto con una curva de sintonía neuronal, las tres para una misma CF.

El proceso exacto por el cual las capacidades de las OHC junto con el resto de la mecánica de la cóclea logran tan notable amplificación y sensibilidad, todavía permanece elusivo [54, 43]

Ciertos trabajos plantean que todas estas características son distintas manifestaciones de un mismo fenómeno no lineal intrínseco a las OHC [54] y que pueden ser explicadas por la inestabilidad de una bifurcación de Hopf [28]. A partir de esta visión simplificada, pero general, se han realizado varios modelos de la mecánica coclear, que suelen denominarse

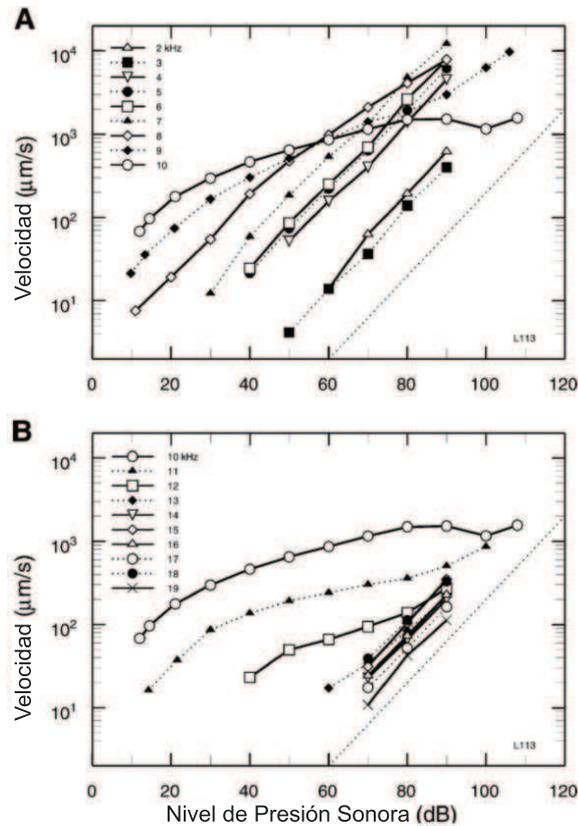


Figura 4.8: *Compresión: magnitud de las respuestas de velocidad de la MB en función de la intensidad y de la frecuencia de estímulo, para un sitio con $CF = 10$ kHz. Puede observarse la compresión para las frecuencias cercanas a la CF, y el comportamiento lineal para las frecuencias más alejadas. La línea punteada indica el comportamiento lineal 1 dB/1 dB. Figura adaptada de Ruggero et al. [101].*

como cóclea de Hopf [67, 27, 74].

Otro conjunto de trabajos considera que el mecanismo de amplificación está relacionado con algún tipo de retroalimentación que funciona ciclo por ciclo [21]. Para conseguir este resultado se han planteado diversos modelos. El modelo de Zweig [126] propone que la impedancia activa del órgano de Corti puede ser descrita si se incluyen componentes con retraso temporal. El modelo de Neely and Kim [86] plantea la micromecánica del órgano de Corti como un sistema de cuarto orden, lo que permite reducir el problema a dos cadenas de filtros de segundo orden, una pasiva y otra activa que interactúan entre sí. El modelo de Wen and Boahen [119] considera una interacción de alimentación lateral (en inglés *feedforward* y *feedbackward*), que a diferencia de los otros ejemplos permite interacciones no locales entre

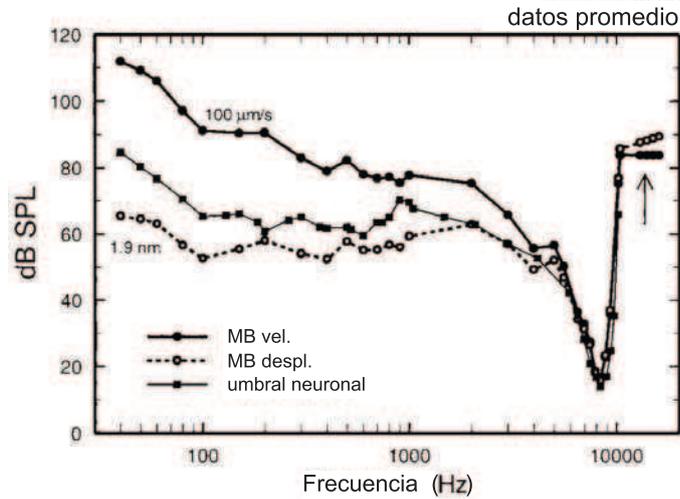


Figura 4.9: Curvas de sintonía de la velocidad y desplazamientos de la MB y actividad neuronal. Se puede observar la pronunciada pendiente del flanco derecho de la curva.

las distintas secciones de la cóclea y permite obtener una clara amplificación [34, 122]. La justificación de este acoplamiento lateral se debe a que los movimientos somáticos de las OHC podrían ejercer fuerzas en secciones adyacentes a donde se encuentran los cilios. En la figura 4.10 se observa un corte longitudinal de la cóclea donde se puede ver como las OHC y la fuerza que producen se dispone de forma oblicua con la membrana basilar.

Una diferencia entre los modelos de Hopf y los de retroalimentación es que los primeros son intrínsecamente no lineales, mientras que los segundos no. Sin embargo, para poder modelar la compresión en los modelos de retroalimentación, es necesario agregar una función

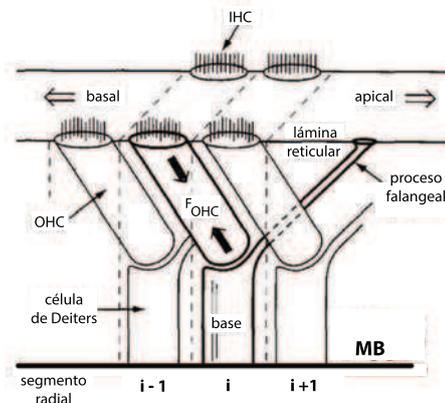


Figura 4.10: Forma de Y entre las OHC y el proceso falangeal de las células de Deiters

de saturación que limite la parte activa para grandes intensidades y sólo deje el accionar de la parte pasiva. El proceso de saturación es un fenómeno habitual en muchos procesos biológicos. En este caso ocurre en el movimiento somático de las OHC.

En nuestro trabajo de modelado utilizaremos el enfoque de alimentación o acoplamiento lateral de Wen and Boahen [119], ya que consigue amplificar las señales pero sin aumentar la complejidad algorítmica del modelo e incluiremos una función de saturación para obtener la compresión.

4.2. Modelo de la cóclea

Existen distintos tipos de modelos de la mecánica coclear que se centran en diferentes escalas y niveles de descripción. En general se puede trazar una relación entre el nivel de descripción y la eficiencia computacional de los modelos.

Una de las formas más eficientes de obtener un cocleograma es mediante un banco de filtros estimulado en paralelo como los descritos en la sección 1.3.2. Es posible diseñar filtros adaptativos para describir algunas de las propiedades no lineales de la cóclea como los casos de los filtros *dual resonance nonlinear* (DRNL) de Meddis et al. [80] y el *multiple band-pass non linearity* (MBPNL) de Goldstein [38]. Sin embargo, el uso de estos banco de filtros no permite contemplar de manera natural la propagación de productos de distorsión y de emisiones otoacústicas sobre la MB, dado que el banco es estimulado en paralelo y no hay acoplamiento entre los filtros.

En el otro extremo, existen modelos tridimensionales con mucho nivel descriptivo sobre la geometría de la cóclea, pero que sólo permiten simulaciones de corta duración y no sirven como una herramienta para el estudio del sonido [36].

Con un nivel de descripción medio podemos destacar una familia de modelos de una o dos dimensiones que permiten generar cocleogramas fácilmente y permiten el estudio de las emisiones otoacústicas y las inestabilidades dinámicas en la cóclea ([23], [75], [29], [115], [70])

Uno de nuestros objetivos consistió en desarrollar un modelo para procesar sonido y generar cocleogramas de manera realista y eficiente, por lo que la última familia de modelos

fue tomada en cuenta para nuestro desarrollo.

Con respecto al modelado de la IHC, la sinapsis de cinta y la fibra auditiva, también es posible encontrar modelos con diferentes niveles descriptivos. Podemos encontrar los modelos de conductancias que consideran los circuitos equivalentes de la célula y la dinámica del calcio en la liberación de neurotransmisor [107], y otros modelos fenomenológicos que reducen el comportamiento de la IHC a un procesamiento de señales más sencillo [124].

En nuestro trabajo optamos por utilizar el modelo de Zilany et al. [125] que no tiene una gran complejidad algorítmica y cuyo código se encuentra liberado para ser utilizado.

La exposición que se hará de nuestro modelo de la cóclea tratará de ser concisa y no se consideraran todos los enfoques o tipos de modelos dado que extendería bastante la sección. Para un tratado exhaustivo sobre los distintos enfoques para modelar la cóclea, referirse a los trabajos de Ni et al. [87], de Boer [19, 20, 21], Kern [67].

El modelo puede separarse en su parte pasiva y su parte activa. La primera está basada en un arreglo unidimensional de osciladores con acoplamiento hidrodinámico, mientras que la segunda está basada en una alimentación lateral. Para incluir la función de saturación se analizaron dos opciones, una en la amortiguación de los osciladores y otra en la parte activa.

4.2.1. Sistema pasivo: osciladores con acoplamiento hidrodinámico

El movimiento de la MB puede discretizarse como el movimiento de un arreglo de osciladores inmerso en un fluido (figura 4.11). Este esquema representa la cóclea desenrollada del mismo modo que la figura 4.2, con la simplificación de que los dos canales llenos de fluidos son combinados en uno solo. El fluido puede ser considerado incompresible y con una densidad similar a la del agua y cada oscilador representa una sección de MB.

Antes de exponer el sistema de osciladores, comenzamos el planteo en el caso continuo donde la expresión para el movimiento de la MB está dada por:

$$m \frac{\partial^2 h(x, t)}{\partial t^2} + \mu(x) \frac{\partial h(x, t)}{\partial t} + s(x) h(x, t) = -p(x, t) \quad (4.1)$$

Donde h representa el desplazamiento en la dirección y de la MB, p la presión del fluido,

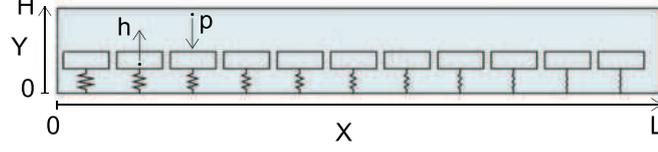


Figura 4.11: La cóclea como un arreglo de osciladores sumergidos en un fluido. L representa el largo y H la altura del canal, h el desplazamiento de un oscilador y p la presión del fluido

m es la masa, μ el amortiguamiento y s la rigidez, estas últimas tres son magnitudes por unidad de área. Todas las variables dependen de la posición x excepto la masa, que puede suponerse constante [26].

Según la denominada aproximación de onda larga [19], el acoplamiento hidrodinámico entre las distintas secciones de la MB viene dado por la siguiente expresión que relaciona la derivada segunda del campo de presión con la aceleración de la MB:

$$\frac{\partial^2 p(x, t)}{\partial x^2} = -\frac{2\rho}{H} \frac{\partial^2 h(x, t)}{\partial t^2} \quad (4.2)$$

Donde H es la altura del canal lleno de fluido y ρ es la densidad del fluido. Esta expresión puede derivarse de las ecuaciones de conservación del momento y de la masa del fluido para el problema de una cavidad unidimensional. La derivación se encuentra en el apéndice A.4.

Esta ecuación diferencial debe ser acompañada de las condiciones de borde para la presión en la ventana oval y en el helicotrema. Para la ventana oval se propone:

$$\left. \frac{\partial p(x, t)}{\partial x} \right|_{x=0} = -2\rho \frac{\partial^2 h(0, t)}{\partial t^2} - 2\rho \frac{\partial^2 w_s(t)}{\partial t^2} \quad (4.3)$$

Donde $w_s(t)$ representa el desplazamiento de la ventana oval, es decir, la excitación externa que proviene del oído medio. Por el otro lado en el helicotrema se propone que la presión se anule:

$$p(x, t)|_{x=L} = 0 \quad (4.4)$$

El sistema completo a resolver queda formado por las ecuaciones 4.1, 4.2, 4.3 y 4.4. Es posible obtener soluciones analíticas de este sistema en el dominio de la frecuencia

reemplazando las variables m , μ y s por formas sencillas o mediante el método WKBJ que permite obtener soluciones aproximadas [19]. En este trabajo sólo nos enfocaremos principalmente en las soluciones numéricas, sin embargo en el apéndice A.5 se muestra un análisis teórico de la relación de dispersión de la onda viajera que permite entender la naturaleza del sistema activo de alimentación lateral.

Para simplificar la notación suprimimos la dependencia con x y t . Comencemos reescribiendo la ecuación 4.1 para manipularla.

$$g = \mu \frac{\partial h}{\partial t} + kh \quad (4.5a)$$

$$\frac{\partial^2 h}{\partial t^2} = -\frac{p + g}{m} \quad (4.5b)$$

Para despejar la presión p reemplazamos esta expresión de $\frac{\partial^2 h}{\partial t^2}$ en la ecuación 4.2.

$$\frac{\partial^2 p}{\partial x^2} = -\frac{2\rho}{H} \frac{\partial^2 h}{\partial t^2} \quad (4.6a)$$

$$\frac{\partial^2 p}{\partial x^2} = \frac{2\rho}{mH} (p + g) \quad (4.6b)$$

$$\frac{\partial^2 p}{\partial x^2} - \frac{2\rho}{mH} p = \frac{2\rho}{mH} g \quad (4.6c)$$

$$\left(\frac{mH}{2\rho} \frac{\partial^2}{\partial x^2} - 1 \right) p = g \quad (4.6d)$$

A continuación realizamos la discretización espacial con diferencias finitas de la ecuaciones 4.5b y 4.6d:

$$\frac{\partial^2 h_i(t)}{\partial t^2} = \frac{-p_i(t) - g_i(t)}{m} \quad (4.7a)$$

$$\frac{mH}{2\rho} \frac{p_{i-1}(t) - 2p_i(t) + p_{i+1}(t)}{dx^2} - p_i(t) = g_i(t) \quad (4.7b)$$

$$(4.7c)$$

Y de forma equivalente para las condiciones de contorno 4.3 y 4.4, incorporando la

expresión de 4.5b.

$$\frac{m}{2\rho} \left(\frac{p_2(t) - p_1(t)}{dx} \right) - p_1(t) = g_1(t) - m \frac{\partial^2 w_s(t)}{\partial t^2} \quad (4.8a)$$

$$p_N(t) = 0 \quad (4.8b)$$

Donde $1 \leq i \leq N$, y $dx = L/N$. El sistema discretizado puede expresarse de forma matricial de la siguiente manera:

$$\frac{\partial^2 \mathbf{h}}{\partial t^2} = \frac{-\mathbf{p} - \mathbf{g}}{m} \quad (4.9)$$

$$(\mathbf{D} - \mathbf{I}) \cdot \mathbf{p} = \mathbf{g} + \mathbf{q} \quad (4.10)$$

$$\mathbf{C} \cdot \mathbf{p} = \mathbf{g} + \mathbf{q} \quad (4.11)$$

Donde \mathbf{D} representa la matriz del operador $\frac{\partial^2}{\partial x^2}$ y \mathbf{q} es el vector de las fuentes. El primer paso para resolver la evolución temporal es resolver el sistema lineal y despejar \mathbf{p} de la 4.11. Esto es posible invirtiendo la matriz \mathbf{C} .

$$\mathbf{p} = \mathbf{C}^{-1} \cdot (\mathbf{g} + \mathbf{q}) \quad (4.12)$$

Sin embargo el sistema lineal puede resolverse de manera eficiente sin la necesidad de invertir, ya que la matriz \mathbf{C} tiene forma tridiagonal:

$$\begin{bmatrix} -K-1 & K & & & & & & & 0 \\ A & B & A & & & & & & \\ 0 & A & B & A & & & & & \\ & & \ddots & \ddots & & & & & \\ & & & & A & B & A & & 0 \\ & & & & & A & B & A & \\ 0 & & & & & & & 0 & 1 \end{bmatrix} \times \begin{bmatrix} p_1(t) \\ p_2(t) \\ \vdots \\ p_{N-1}(t) \\ p_N(t) \end{bmatrix} = \begin{bmatrix} g_1(t) \\ g_2(t) \\ \vdots \\ g_{N-1}(t) \\ 0 \end{bmatrix} + \begin{bmatrix} m \frac{\partial^2 w_s(t)}{\partial t^2} \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad (4.13)$$

$$\lambda = \frac{mH}{2\rho} \quad A = \frac{\lambda}{dx^2} \quad B = \frac{-2\lambda}{dx^2} - 1 \quad K = \frac{\lambda}{Hdx}$$

Esta forma tridiagonal permite la eliminación gaussiana directa y el tiempo de cómputo para resolver el sistema lineal es de orden N .

Una vez obtenido el valor de la presión, se puede calcular la evolución temporal del sistema completo a partir de la ecuación 4.9. Para esto expresamos esta ecuación como un sistema de dos ecuaciones ordinarias de primer orden:

$$\frac{\partial \mathbf{h}}{\partial t} = \mathbf{v} \quad (4.14)$$

$$\frac{\partial \mathbf{v}}{\partial t} = \frac{-\mathbf{p} - \mathbf{g}}{m} \quad (4.15)$$

Finalmente se evoluciona temporalmente el sistema con algún método de integración como los métodos basados en *Runge-Kutta 4*.

Para comprender un poco más cómo se comporta el sistema final, podemos representarlo en una sola ecuación donde se aprecia explícitamente el acoplamiento entre osciladores. Modificaremos un poco la notación y las dimensiones del problema para obtener una fórmula más compacta y general.

$$\ddot{h}_i + \frac{\omega_i}{Q} \dot{h}_i + \omega_i^2 h_i = - \sum_{j=1}^N C_{ij}^{-1} \left(2\gamma_j \omega_j \dot{h}_j + \omega_j^2 h_j + \delta_{j,1} \ddot{w}_s \right) \quad (4.16)$$

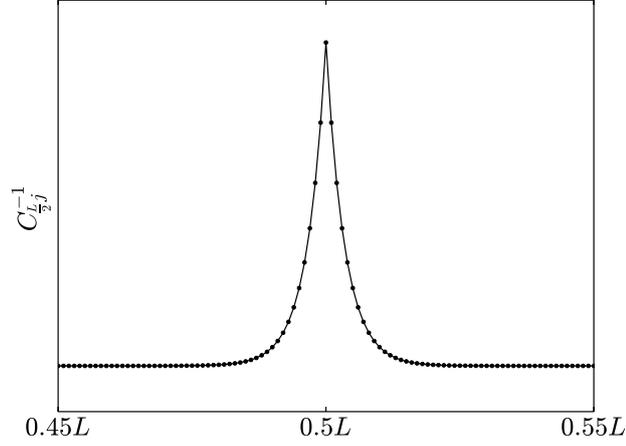


Figura 4.12: Fila de la matriz de acoplamiento C^{-1} , para $N = 300$ osciladores.

Donde $\omega = \sqrt{k/m}$ representa la frecuencia natural del oscilador, $Q = \sqrt{mk}/\mu$ es el factor de calidad y representa un factor inversamente proporcional al ancho de banda de la respuesta del oscilador y C^{-1} es la matriz de acoplamiento. En la figura 4.12 se puede ver un perfil o fila de esta matriz. Esto muestra cómo el movimiento de un oscilador es influenciado por sus vecinos cercanos. El perfil de acoplamiento está determinada exclusivamente por las propiedades del fluido y de la cavidad. Es posible resolver el problema hidrodinámico de la cavidad en 2 o 3 dimensiones y obtener nuevos perfiles y seguir utilizando la fórmula 4.16. En este trabajo decidimos utilizar la solución de 1 dimensión que se ha presentado, ya que es más eficiente de computar.

4.2.2. Sistema activo: alimentación lateral

El propósito del sistema activo en la cóclea es el de amplificar la onda viajera en la región donde la frecuencia del estímulo coincide con la resonancia de la MB. Para esto existen muchas estrategias de modelado [86, 75, 67] las cuales, por lo general, utilizan un segundo banco de filtros o arreglo de osciladores, aumentando el número de variables dinámicas del sistema. Dado que estamos buscando una forma eficiente de modelar la cóclea, es necesario mantener el número de variables lo más bajo posible, por lo que proponemos utilizar el modelo de alimentación lateral propuesto por [118] que no necesita de un segundo

arreglo de osciladores y conserva el número de variables.

El modelo de alimentación lateral (*feed-forward* y *feed-backward*) se basa en suponer que las OHC ejercen su fuerza en secciones aledañas a las que reciben estimulación. En la figura 4.10 se puede observar como la sección i de la MB puede recibir la fuerza de la OHC de la sección $i - 1$ a través de las células de Deiters. De manera análoga la prolongación falángica también puede transmitir la fuerza de la OHC de la sección $i + 1$. Esto permite expresar la fuerza de la OHC de la siguiente manera:

$$F_{OHC}(x) = \alpha s(x)(\beta h(x - d) - h(x + d)) \quad (4.17)$$

Donde d representa la distancia entre secciones, α representa el factor de la motilidad de la OHC, y β es el cociente entre la alimentación en dirección apical y la dirección basal. Podemos añadir esta expresión a la ecuación 4.1 y partir de ahí construir el sistema de osciladores con alimentación lateral.

$$m \frac{\partial^2 h(x, t)}{\partial t^2} + \mu(x) \frac{\partial h(x, t)}{\partial t} + s(x)h(x, t) = -p(x, t) + F_{OHC} \quad (4.18)$$

$$m \frac{\partial^2 h(x, t)}{\partial t^2} = -\mu(x) \frac{\partial h(x, t)}{\partial t} - s(x) (-\alpha\beta h(x - d) + h(x, t) + \alpha h(x + d)) - p(x, t) \quad (4.19)$$

Esta formulación permite que la onda viajera se alimente a sí misma con los desplazamientos de otras secciones. Si el valor de d es ajustado correctamente es posible lograr una alimentación positiva y amplificar a la onda viajera en la región de resonancia.

Hasta este punto el sistema funciona de forma lineal, por lo tanto no hay ni compresión ni productos de distorsión y es necesario incluir una no linealidad para obtener estos comportamientos. Propondremos dos formas distintas para lograr esto. Por un lado, una función de saturación para las fuerzas de las OHC siguiendo el modelo de Wen [118] y, por otro lado, una saturación en la amortiguación de la MB según el modelo de Duifhuis [26].

La primer forma utiliza una función tangente hiperbólica para saturar los desplazamientos de la membrana que generan F_{OHC} .

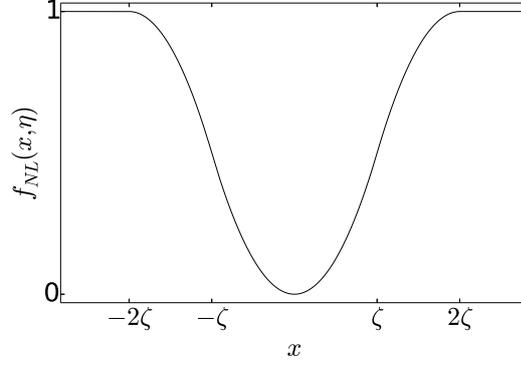


Figura 4.13: Función no lineal $f_{NL}(x, \eta)$ con saturación utilizada en el amortiguamiento

$$F_{OHC}(x) = \alpha s(x) \left[\nu \beta \tanh\left(\frac{h(x-d)}{\nu}\right) - \nu \tanh\left(\frac{h(x+d)}{\nu}\right) \right] \quad (4.20)$$

Donde ν permite ajustar el comienzo de la saturación. La segunda forma propone una amortiguación no lineal que depende de h , por lo que la ecuación del oscilador queda expresada de la siguiente forma:

$$m \frac{\partial^2 h(x, t)}{\partial t^2} + \mu(x) [1 + \gamma f_{NL}(h(x, t))] \frac{\partial h(x, t)}{\partial t} + s(x) h(x, t) = -p(x, t) + F_{OHC} \quad (4.21)$$

Donde la función f_{NL} debe cumplir ciertos requisitos tales como anularse para valores pequeños y saturar para valores más grandes [26] y γ permite justar el nivel máximo del amortiguamiento. En nuestro caso proponemos utilizar la siguiente función no lineal formada por funciones cuadráticas a tramos, la cual puede ser computada de manera eficiente (figura 4.13) :

$$f_{NL}(x, \eta) = \begin{cases} 1 & 2\zeta < |x| \\ 1 - \eta * (|x| - 2\zeta)^2 & \zeta < |x| < 2\zeta \\ \eta x^2 & |x| < \zeta \\ \zeta = \sqrt{\frac{1}{2\eta}} & \end{cases}$$

En la sección de resultados se muestran los diferentes comportamientos de los dos tipos

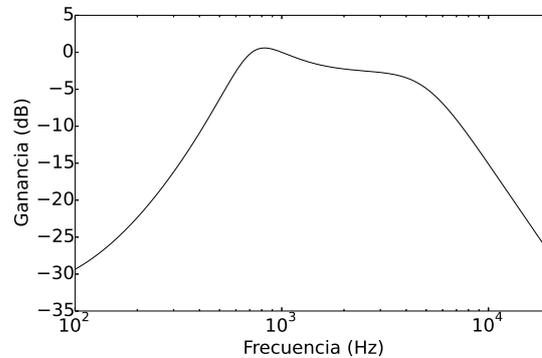


Figura 4.14: *Respuesta en frecuencia del oído medio, a partir de un filtro con dos polos y un cero. La ganancia está normalizada para la frecuencia de 1000 Hz*

de no linealidades propuestos. Ahora pasaremos a describir las otras etapas del modelo completo de la periferia auditiva, el oído medio, la IHC y la sinapsis.

4.2.3. Oído medio

Para poder alimentar el modelo de la cóclea es necesario incluir el procesamiento que ocurre en el oído medio. Esta etapa se suele modelar con un filtro pasa bandas que recorta ciertas frecuencias del sonido dejando sólo el rango audible. Para esto se utilizó un filtro lineal con dos polos y un cero, con una respuesta en frecuencia como la de la figura 4.14. El filtro fue adaptado de Tan and Carney [108].

4.2.4. Células ciliadas internas y sinapsis

Para incluir la IHC y la sinápsis se optó por utilizar el modelo de Zilany et al. [124], cuya implementación en software está liberada para su uso. Con respecto a la IHC, esta es modelada de manera simplificada sin utilizar un modelo celular de conductancias, mediante una función de entrada y salida similar a la de la figura 4.4 más un filtro pasa bajos con corte en 3000 Hz de séptimo orden.

La sinapsis, a diferencia de la IHC, es modelada de manera más compleja ya que se consideran dos mecanismos para el fenómeno de adaptación: una adaptación rápida con decaimiento exponencial y otra lenta con decaimiento con ley de potencias. Para modelar

la adaptación exponencial, se utiliza un sistema dinámico de tres reservorios con difusión, basado en el modelo de Westerman and Smith [121], mientras que para modelar la adaptación con ley de potencias, se utiliza una convolución con un *kernel* de ley de potencias [124].

4.3. Resultados numéricos

A continuación presentaremos algunas soluciones numéricas para ilustrar las distintas etapas de construcción del modelo que se mencionaron en la sección anterior: cóclea pasiva, cóclea activa, el fenómeno de compresión, la IHC y la sinapsis. El modelo de la cóclea fue implementado en C++ utilizando la biblioteca *odeint* [1] que permite utilizar métodos *Runge-Kutta* con paso variable. El modelo de la IHC fue implementado en Python y el de la sinapsis fue tomado del módulo de Python *cochlea* [100]. El modelo completo está disponible para su uso como módulo de Python en <https://github.com/pablorigera/cochlea>.

4.3.1. Cóclea

En la tabla 4.1 se muestra el conjunto de parámetros utilizado. La frecuencia natural de los osciladores es el único valor que depende de la posición y se eligió un rango que va desde 20 kHz en la base hasta 81 Hz en el ápex. El límite en bajas frecuencias fue impuesto por una limitación en el modelo de sinapsis utilizado que no está preparado para frecuencias más bajas.

En la figura 4.15 se muestra una respuesta de la cóclea pasiva donde se puede observar el desplazamiento de la MB para varios instantes mostrando la onda viajera para un estímulo de 1000 Hz. El eje de las abscisas representa la posición en la cóclea y está denotado en función de la frecuencia natural de los osciladores. Puede observarse que el máximo de la onda viajera no corresponde con la frecuencia del estímulo, esto ocurre por el alto grado de amortiguamiento que existe en la cóclea pasiva.

El parámetro α permite prender y apagar el sistema activo. En la figura 4.16 se muestra el valor cuadrático medio (RMS) y la fase de los desplazamientos de la MB, para el caso pasivo $\alpha = 0$ y activo $\alpha = 0,15$. El estímulo utilizado fue un tono sinusoidal de 1000 Hz

Parametros	Valor	Unidades
Parte pasiva		
Altura canal H	1	mm
Largo canal L	0.035	m
Densidad Fluido ρ	1000	Kg/m^3
Número de osciladores	300	
MB: Masa por unidad de área m	0.03	Kg/m^2
MB: Frecuencia natural $\omega(x)$	$2\pi \times 20000 \times 10^{-68,36x}$	$1/s$
MB: Factor $Q(x)$	5	
Parte activa		
Distancia acoplamiento lateral d	140	μm
Factor de motilidad de OHC α	0.15	
Cociente de alimentación β	0.3	
No linealidad		
Factor de amortiguación γ	2	
Parámetro amortiguación η	5×10^{11}	$1/m^2$
Parámetro saturación $F_{OHC} \nu$	1×10^{-8}	m

Cuadro 4.1: *Parámetros del modelo.*

e intensidad de 10 dB_{SPL} . Lo primero que se puede notar es que la posición del máximo de la respuesta se desplaza; para la cóclea activa el máximo está cerca del oscilador de frecuencia natural 1200 Hz, mientras que para la cóclea pasiva el máximo ocurre cerca de los 2080 Hz. Este corrimiento suele observarse en datos experimentales *post-mortem* [122]. Podemos explicar este fenómeno si tomamos en cuenta que el mecanismo activo de este modelo tiene el efecto de disminuir la amortiguación en la región de la resonancia, por lo que la onda pueda propagarse más y generar un pico más marcado.

Para observar el nivel de compresión es necesario estimular a diferentes intensidades. Primero mostraremos las simulaciones realizadas con la ecuación del oscilador 4.20 que utiliza a la función \tanh como función de saturación de la fuerza F_{OHC} . En la figura 4.17 se pueden observar las respuestas para distintas intensidades del estímulo. Para intensidades bajas se observa la respuesta del sistema activo, pero a medida que se aumenta la intensidad esta se va atenuando, el máximo se traslada hacia altas frecuencias y la respuesta vuelve a parecerse a la del sistema pasivo. Los comportamientos poco realistas que se generan para altas intensidades y bajas frecuencias son consecuencia del sistema activo amplificando

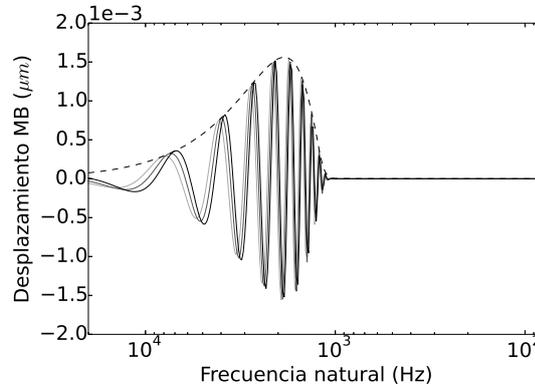


Figura 4.15: *Cóclea pasiva: Desplazamientos de la MB en varios instantes para un tono puro de 1000 Hz. Se puede observar la formación de una onda viajera debido al acoplamiento hidrodinámico y que se propaga desde las frecuencias altas hacia las bajas.*

ciertas ondas de bajas frecuencias que se generan debido al impulso que ocurre al comenzar el estímulo. En el panel inferior de la figura, se muestran las mismas curvas pero normalizadas. Esto representa la ganancia de las respuestas con respecto al estímulo. En este caso la ganancia máxima está dada por las diferencias entre las respuestas extremas, y es de aproximadamente 50 dB.

La característica más marcada de la compresión está dada por como las curvas se van compactando en el flanco derecho de la respuesta. Para ver esto en detalle en la figura 4.18 se muestran las respuestas en función de la intensidad para ciertas posiciones en la cóclea. Se puede ver cómo a medida que la intensidad aumenta las respuestas de las posiciones con frecuencias cercanas al estímulo (950, 1000 y 1200 Hz) muestran compresión, mientras que las frecuencias más alejadas muestran un comportamiento lineal.

A continuación mostramos los mismos resultados pero para la ecuación del oscilador 4.21 donde la amortiguación tiene una forma no lineal (figuras 4.19, 4.20). Los resultados son muy similares al anterior aunque se puede apreciar diferencias en la región del pico de resonancia para altas intensidades. En este caso las respuestas dejan de crecer a partir de cierta intensidad, lo que termina generando una compresión mucho más marcada. Este comportamiento se asemeja un poco más a los observados en experimentos con animales, por lo que esta no linealidad será utilizada de ahora en más, con el beneficio de que reduce

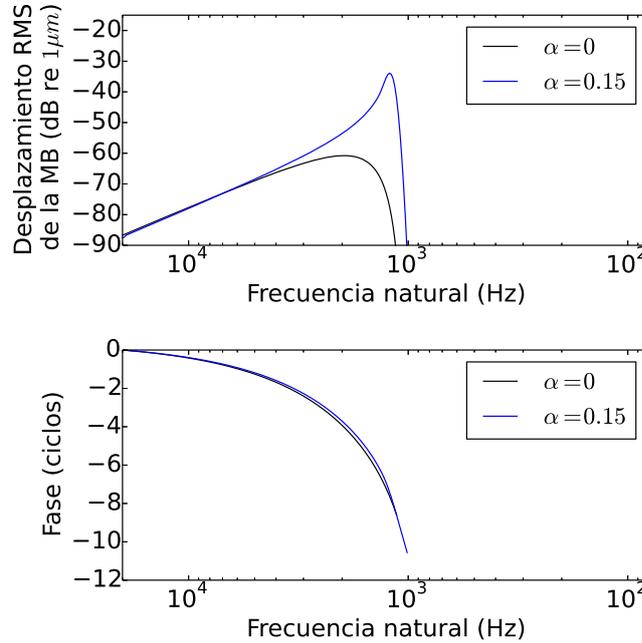


Figura 4.16: Respuesta RMS y fase de los desplazamientos de la MB en función de la frecuencia natural para un tono puro de 1000 Hz, 10 dB_{SPL} y duración de 0.3 segundos en la cóclea pasiva (línea negra) y activa (línea azul).

el tiempo de cálculos debido a la formulación a trozos.

Las respuestas del modelo que se han mostrado hasta aquí corresponden a estímulos para una frecuencia de 1000 Hz, sin embargo es posible extrapolar el comportamiento a otras frecuencias ya que para las regiones de la base y las centrales, esto corresponde solamente a una traslación rígida de los perfiles de respuestas. Sin embargo las respuestas para frecuencias muy bajas tienen algunos comportamientos de borde apreciables. En la figura 4.21 se pueden ver las respuestas para un estímulo de 20 Hz, que tiene una frecuencia más baja que la del último oscilador de la cóclea. Se observa que la onda viajera no llega hasta el final de la cóclea ya que la condición de borde impuesta para el ápex (presión = 0) impide que el último oscilador se mueva, lo cual forma reflexiones e interferencias en los patrones de la MB. También se puede observar que la falta de compresión, posiblemente porque el sistema activo no esté haciendo contribuciones debido a que la frecuencia del estímulo no resuena en ningún sitio de la MB.

Para observar como se comporta el modelo estimulado por varias frecuencias, a con-

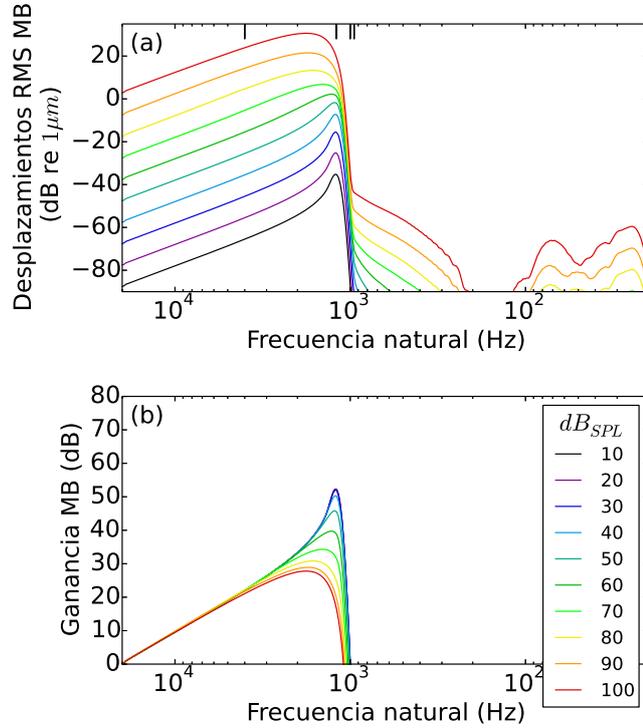


Figura 4.17: Modelo con saturación en la fuerza F_{OHC} . (a) Respuesta RMS de los desplazamientos de la MB en función de la frecuencia natural de los osciladores para un tono puro de 1000 Hz y 0.3 segundo de duración y para varias intensidades. (b) Ganancia de los desplazamientos de la MB a partir de normalizar las curvas del panel (a)

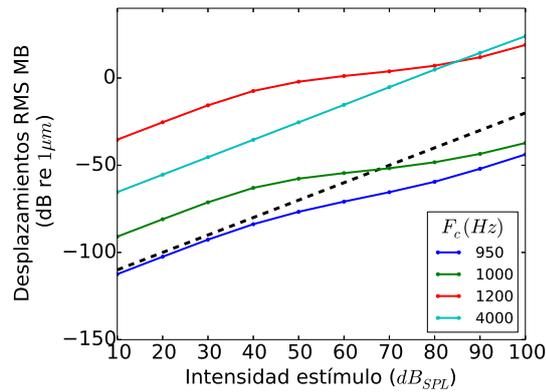


Figura 4.18: Respuesta RMS de los desplazamientos de la MB en función de la intensidad del estímulo para varias posiciones dentro de la cóclea. Se puede observar compresión para las frecuencias características cercanas a la del estímulo ($F_0 = 1000\text{Hz}$). La línea punteada tiene una pendiente de 1 dB/1 dB .

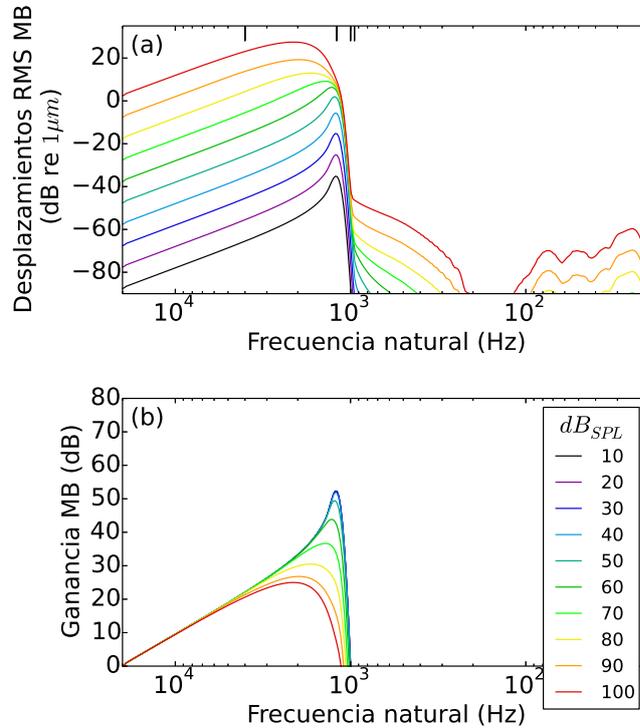


Figura 4.19: Modelo con amortiguación no lineal. (a) Respuesta RMS de los desplazamientos de la MB en función de la frecuencia natural de los osciladores para un tono puro de 1000 Hz y duración de 0.3 segundos y para varias intensidades. (b) Ganancia de los desplazamientos de la MB a partir de normalizar las curvas del panel (a)

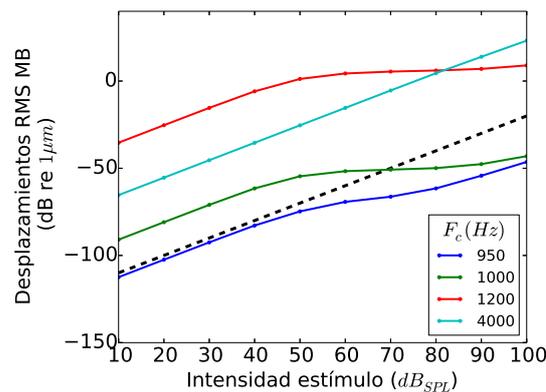


Figura 4.20: Respuesta RMS de los desplazamientos de la MB en función de la intensidad del estímulo para varias posiciones dentro de la cóclea. Se puede observar compresión para las frecuencias características cercanas a la del estímulo ($F_0 = 1000\text{Hz}$). La línea punteada tiene una pendiente de 1 dB/1 dB.

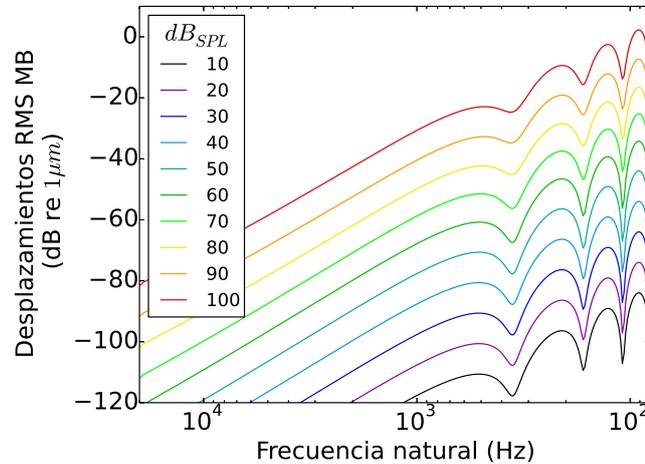


Figura 4.21: Respuestas RMS de los desplazamientos de la MB en función de la frecuencia natural para un tono puro de 20 Hz, para varias intensidades

tinuación mostramos los resultados de realizar un barrido en frecuencias sobre la cóclea y observar el comportamiento para una sola posición sobre la MB. Este protocolo es el mismo que se utiliza en los experimento *in vivo* con vibrómetros láser como el de la figura 4.8 realizado en chinchilla. En la figura 4.22 se muestran los resultados de dicho barridos en frecuencia y para facilitar la comparación con el resultado experimental se muestran las respuestas en términos de la velocidad de la MB y no los desplazamientos. Puede verse una gran similitud, por lo menos en términos cualitativos, entre las respuestas del modelo y las medidas experimentales.

4.3.2. Células ciliadas internas y sinapsis

Los desplazamientos de la MB son utilizados como entrada del modelo de IHC el cual genera las variaciones del potencial de membrana de dicha células, que luego son utilizadas de entrada para el modelo de sinapsis que genera como salida la tasa de disparos de la fibra auditiva. En la figura 4.23, podemos observar las respuestas de estas tres etapas para dos tonos, uno de 1000 Hz y otro de 5000 Hz. Se muestran las formas de ondas tomadas de los canales con la frecuencia característica (CF) de los estímulos. En los paneles superiores se muestra el desplazamiento de la MB. En los paneles del medio se muestra el potencial

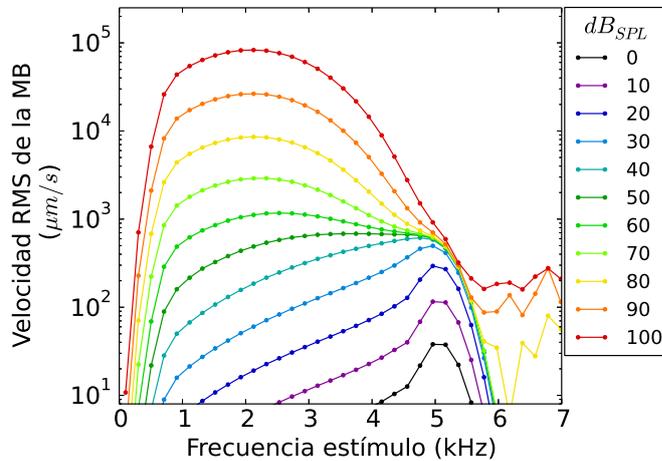


Figura 4.22: Respuestas RMS de las velocidades de la MB para el sitio ubicado a 6 mm de la base para varias intensidades

de membrana de la IHC, donde se puede apreciar como la señal es saturada y adquiere una componente de continua. Para el caso de 5000 Hz se puede apreciar como el filtrado de la IHC ha removido casi todas las componentes de alterna y lo que queda representa más bien a la envolvente del tono que a su componente oscilatoria. En los paneles inferiores se muestra la respuesta de la sinapsis que entrega la tasa de disparo que genera la fibra auditiva. En esta última, puede observarse la adaptación al comienzo del estímulo.

Con este modelo de periferia se pueden realizar cocleogramas realistas para analizar sonidos sintetizados o grabaciones. Los tiempos de cómputo dependen de la frecuencia de muestreo utilizada, para el caso de una frecuencia de 100 kHz, se realizaron simulaciones a una velocidad de 3.5x el tiempo real del estímulo para el modelo de cóclea y de 25x para el modelo con la IHC y la sinapsis. Es posible reducir los tiempos de cómputo del modelo de sinapsis si se implementa de forma distribuida. Estos tiempos corresponden a simulaciones realizadas con un procesador *Intel(R) Core(TM) i7-4790 CPU @ 3.60GHz*.

4.4. Resultados con tonos modulados

En esta sección estudiaremos como responde el modelo de la cóclea a tonos modulados en amplitud. Primero lo haremos para computar la rugosidad mediante un nuevo modelo

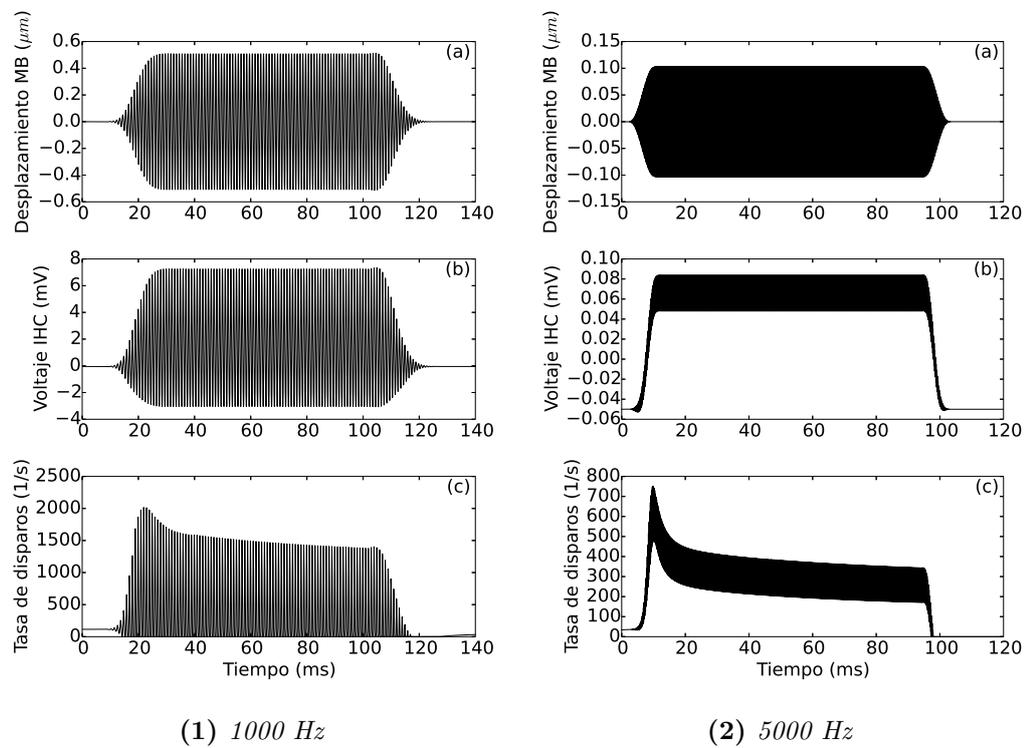


Figura 4.23: Respuestas del modelo de periferia auditiva para los sitios con frecuencia característica de 1000 y 5000 Hz respectivamente. (a) Movimientos de la MB. (b) Potencial de membrana de la IHC donde pueden verse los efectos de rectificación, saturación y filtrado sobre todo para el sitio de 5000 Hz. (c) Tasa de disparos producida por la sinapsis y la fibra auditiva, donde puede verse la adaptación luego del ataque.

basado en métodos temporales únicamente. Luego utilizaremos las respuestas de la cóclea para el conjunto de tonos SAM utilizados en el experimento psicofísico del capítulo 3 y propondremos un posible método para la generación de espacios tímbricos.

4.4.1. Modelo temporal de la rugosidad

Para explorar cómo responde el modelo de cóclea a las modulaciones en amplitud, comenzaremos observando el fenómeno de batidos. Estos ocurren cuando dos tonos están suficientemente cerca en frecuencia y generan interferencia en los desplazamientos de la MB. En la figura 4.24 podemos observar las respuestas del modelo a un par de tonos sinusoidales con frecuencias, $f_c - f_m/2$ y $f_c + f_m/2$, con $f_c = 1$ kHz y para tres valores de f_m : 10, 50 y 150 Hz. En el panel superior están las magnitudes de los desplazamientos de la MB y se puede observar cómo los dos tonos ocupan el mismo sitio de la MB salvo en el tono con f_m de 150 Hz. En el panel inferior se muestran las formas de onda de los sitios de la MB con mayor desplazamiento (puntos rojos en el panel superior) y la envolvente de estas formas de onda. Para la f_m de 150 Hz, podemos ver que dado que las componentes interactúan mucho menos, el nivel de modulación de la envolvente es menor y en el sitio del panel inferior no se presentan prácticamente modulaciones. Debemos notar que este análisis también podría ser hecho con la tasa de disparos de la fibra auditiva, pero en esta sección utilizaremos sólo el modelo de cóclea y la MB para simplificar los resultados.

Repasando los distintos perceptos que pueden ser evocados por tonos con modulación, sabemos que a medida que se va aumentando la frecuencia de modulación, para frecuencias bajas el sistema auditivo es capaz de seguir los cambios de la envolvente como cambios de intensidad y ocurre el percepto de fluctuaciones de la intensidad. Luego el sistema auditivo no puede seguir el ritmo de los cambios de la envolvente y se da el percepto de rugosidad. Finalmente, cuando la frecuencia de modulación es similar a la banda crítica, las componentes se empiezan a separar a la par que la rugosidad disminuye.

Es posible cuantificar estos perceptos con el modelo de la cóclea a partir de la información de las envolventes de cada sitio de la MB. En la sección 1.4.2 del primer capítulo se estableció la relación entre la rugosidad y la envolvente de una señal y se comentaron los distintos modelos espectrales y temporales que hay para calcularla. En esta sección utili-

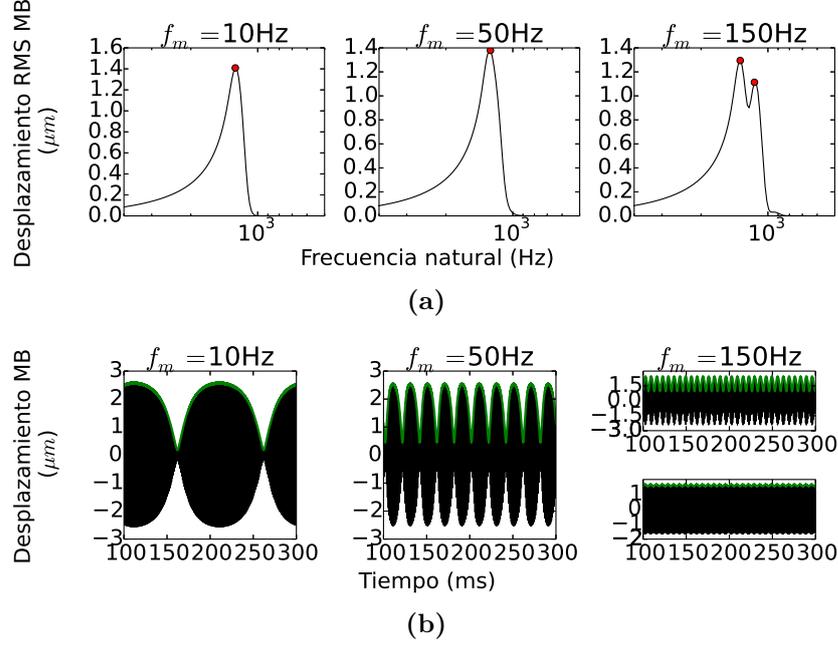


Figura 4.24: Respuestas del modelo de cóclea para tres tonos con batidos centrados en 1 kHz y con diferentes frecuencias de modulación: 10, 50 y 150 Hz. (a) Desplazamientos RMS de la MB en función de la frecuencia natural de los osciladores. (b) Desplazamientos de la MB en función del tiempo para los sitios con mayor actividad.

zaremos el modelo de la cóclea para calcular la rugosidad mediante un método puramente temporal.

Primero, recordemos que es posible calcular el valor la rugosidad a partir de la profundidad de modulación efectiva de la envolvente, siempre que la frecuencia de modulación esté dentro del rango donde se genera el percepción de rugosidad. Esto se puede hacer combinando las ecuaciones 1.4 y 1.5, reescritas a continuación en una sola expresión:

$$R = \alpha m_{ef}^p = \alpha \left(\frac{\sigma}{\mu} \right)^p \quad (4.22)$$

Donde R es la rugosidad, α un coeficiente de escalamiento, m_{ef} es la profundidad de modulación efectiva calculada a partir del valor medio μ y la desviación estándar σ de la envolvente y p es un exponente entre 1 y 2. Esta fórmula sólo establece que mientras más grande la modulación de la envolvente, mayor será la rugosidad, sin embargo no predice cómo varía rugosidad con la frecuencia moduladora. Incluyendo el modelo de la cóclea en

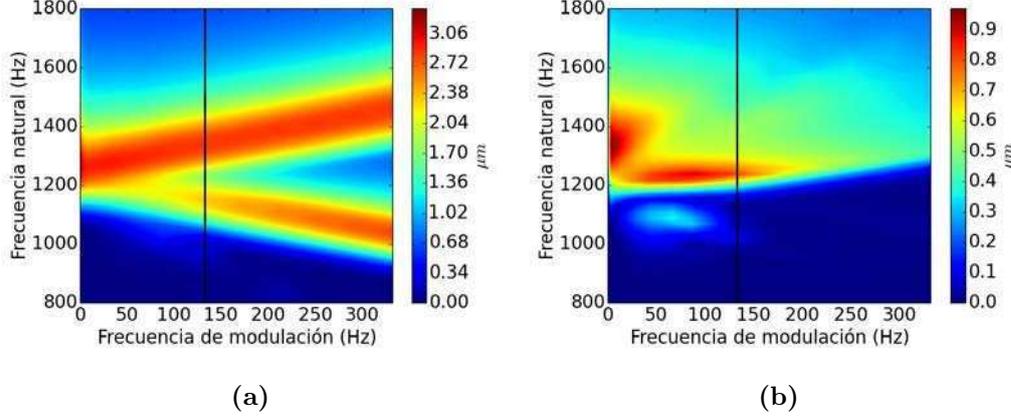


Figura 4.25: (a) Valor medio μ y (b) desviación estándar σ de la envolvente para cada sitio de la MB en función de la frecuencia de modulación para un tono con batidos. Las líneas verticales indican la banda crítica en ERB

el cómputo de la rugosidad, se puede limitar la rugosidad para frecuencias de modulación altas como se vio en la figura 4.24. En la siguiente expresión se agrega el cómputo de la profundidad de modulación para la envolvente de cada sitio de la MB:

$$R = \alpha \left(\frac{\sum_{i=1}^N \sigma_i}{\sum_{i=1}^N \mu_i} \right)^p \quad (4.23)$$

Donde el subíndice i indica el número de sitio o de oscilador del modelo de la cóclea. Comencemos aplicando este modelo a los tonos con batidos visualizando como se comportan σ_i y μ_i por separado. En la figura 4.25 podemos observar en (a) al valor medio y en (b) a la desviación estándar en función de la frecuencia de modulación para un tono de 1 kHz. Para las f_m bajas, podemos ver en el valor medio, que los dos tonos están fusionados en una misma región y luego a medida que aumenta la f_m se separan. Para la desviación estándar podemos observar cómo se va modificando el nivel de fluctuaciones de la envolvente en diferentes sitios de la MB, y cómo se atenúa una vez cruzada la banda crítica denotada por una línea vertical.

A continuación pasamos a promediar los valores entre todos los sitios de la MB. En la figura 4.26 podemos observar los promedios de μ y de σ y de la rugosidad $R = \left(\frac{\mu}{\sigma}\right)^2$. Aunque el valor de la rugosidad decae como se esperaría con la frecuencia de modulación porque las componentes dejan de interactuar, este no decae lo suficientemente rápido para

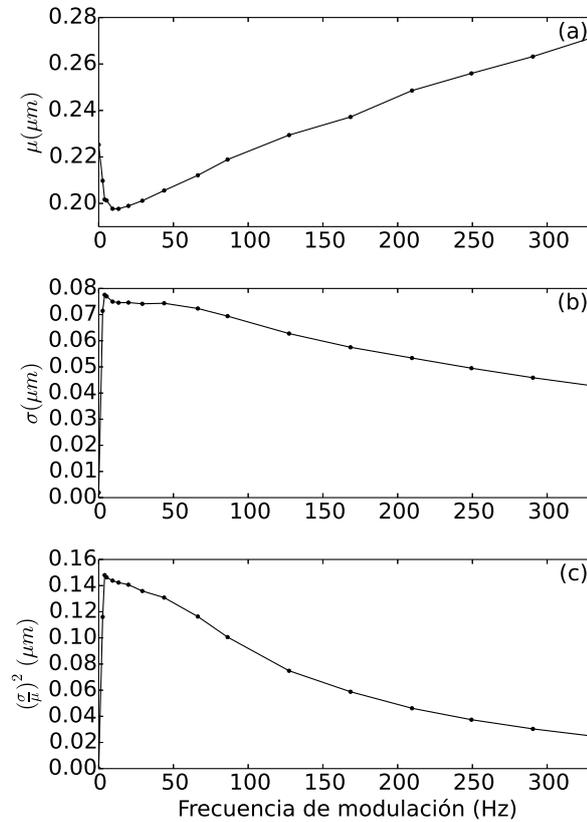


Figura 4.26: Promedios entre todos los canales: (a) valor medio μ , (b) desviación estándar σ y (c) el cociente $(\frac{\sigma}{\mu})^2$

bajas modulaciones. Esto puede explicarse dado que la envolvente no distingue entre los perceptos de fluctuaciones y rugosidad ya que es un fenómeno perceptual y no de la señal. Ya hemos mencionado la relación entre la rugosidad y la falta de velocidad de respuesta del sistema auditivo para seguir las variaciones rápidas de la envolvente. Podemos expresar esto de otra manera, suponiendo la existencia de una ventana de tiempo en la cual el sistema integra la señal auditiva para estimar su intensidad. Si la modulación es muy lenta es posible obtener un muestreo con suficiente precisión de la intensidad como para percibir su fluctuación, pero si la modulación es más rápida, el muestreo genera una medición más azarosa de la intensidad y se genera el percepto de rugosidad³.

A partir de este razonamiento, se incluye en el modelo un proceso de integración tem-

³ ¿aliasing?

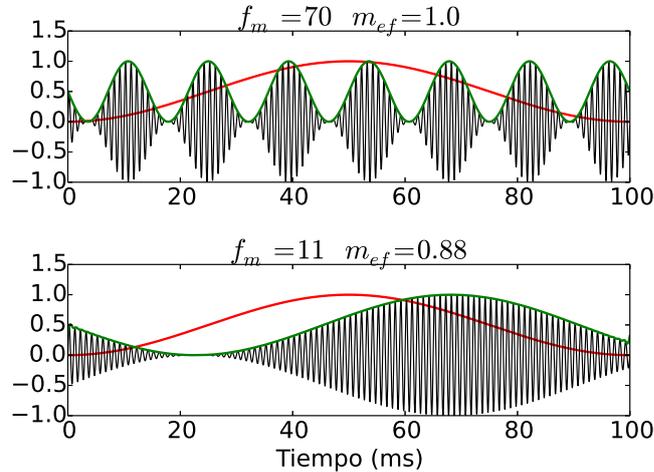


Figura 4.27: *Cómputo de la estadística de la envolvente con una ventana: (a) frecuencia de modulación de 70 Hz y ventana de 100 ms de ancho, el valor de la profundidad de modulación efectiva no se ve alterado (b) frecuencia de modulación de 11 Hz y ventana de 100 ms de ancho, el valor de la profundidad de modulación efectiva es menor que para frecuencias más altas.*

poral que calcula la desviación estándar y el valor medio mediante una ventana móvil. El funcionamiento de este proceso se ejemplifica en la figura 4.27. Este proceso es similar al que ocurre en el modelo de Daniel and Weber [17] con la diferencia de que en aquel modelo el efecto de filtrado de la ventana se aplica en el dominio de la frecuencia.

Con el agregado de este procesamiento, podemos replicar la medida de la rugosidad para los tonos con batidos en función de la frecuencia de modulación. En la figura 4.28 se muestran los resultados para nuestro modelo temporal, el modelo espectral de Sethares [104] descrito en el apéndice A.3 y los valores experimentales del trabajo de Miśkiewicz et al. [82]. El tamaño de la ventana fue de 25 ms; seleccionado para coincidir cualitativamente con el máximo de los resultados experimentales y los valores de los picos fueron normalizados a uno. Es importante notar que según el tamaño de la ventana es posible modificar la posición de la curva donde la rugosidad es máxima. En este caso se decidió ajustar con los resultados del experimento de Miśkiewicz et al. [82], debido a que son los resultados más actualizados y la posición de su máximo de rugosidad esta en concordancia con el 45% de la banda crítica (segunda línea negra) según lo visto en la sección 1.4.2.

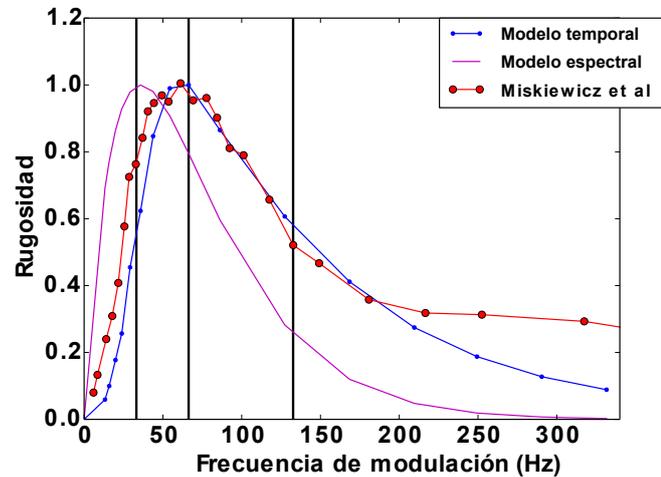


Figura 4.28: Rugosidad en función de la frecuencia de modulación para tonos con batidos y un tamaño de ventana de 25 ms. En tono azul, los valores del modelo temporal a partir de la respuesta de la cóclea, en tono magenta los valores del modelo espectral de Sethares [104] descrito en el apéndice A.3 y en círculos rojos valores experimentales del trabajo de Miśkiewicz et al. [82]. En líneas negras se muestran las fracciones de 25 %, 50 % y 100 % de la banda crítica en ERB

Por último mostraremos el cómputo de la rugosidad para los tonos SAM. En la figura 4.29, podemos observar las curvas de rugosidad, en azul para nuestro modelo temporal, en magenta para el modelo espectral de [104] y con círculos los trabajos experimentales de Aures [3] en rojo y Fastl and Zwicker [30] en verde. Es importante notar que se ha usado el mismo tamaño de ventana que en el caso de tonos con batidos por lo que la posición del pico es relativamente similar cerca del 50 % de la banda crítica. A pesar de que el ajuste no es perfecto se puede ver cierta correspondencia.

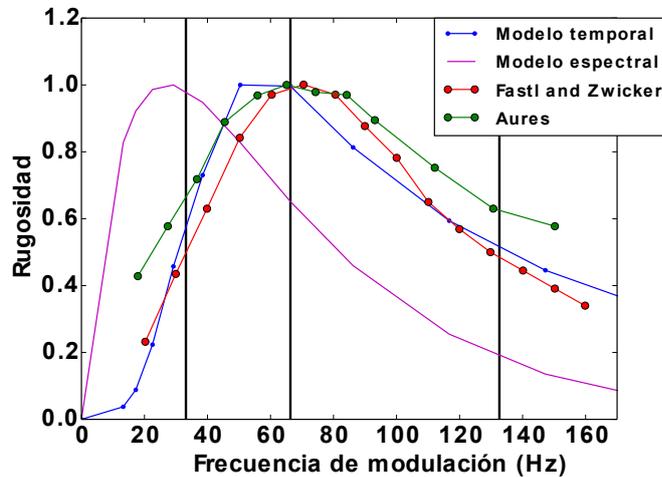


Figura 4.29: Rugosidad en función de la frecuencia de modulación para tonos SAM y un tamaño de ventana de 25 ms. En tono azul, los valores del modelo temporal a partir de la respuesta de la cóclea, en tono magenta los valores del modelo espectral de Sethares [104] descrito en el apéndice A.3, en círculos rojos valores experimentales del trabajo de Aures [3] y en círculos verdes de Fastl and Zwicker [30]. En líneas negras se muestran las fracciones de 25 %, 50 % y 100 % de la banda crítica en ERB

4.4.2. Espacios tímbricos de modulación

En el experimento del capítulo 3 los participantes juzgaron las similitudes entre un conjunto de tonos SAM. El proceso cognitivo responsable de las evaluaciones de estas similitudes es sin duda un proceso de alto nivel que involucra a la corteza auditiva y también al resto de las etapas de la vía auditiva. No es trivial suponer cuales fueron las dimensiones perceptuales utilizadas por los sujetos ya que cada participante puede haber utilizado diferentes pistas para estimar la distancia. A pesar de esto, los resultados del escalamiento multidimensional mostraron que en promedio los participantes utilizaron las dos dimensiones del espacio de configuraciones de los tonos SAM, la frecuencia portadora f_c y la frecuencia moduladora f_m . Por lo tanto, es posible suponer que en la estimación de las similitudes tuvo lugar algún mecanismo de factorización que permitió identificar por separado la frecuencia portadora de la moduladora. En esta sección proponemos un mecanismo para medir la frecuencia moduladora y la portadora a partir de respuestas prototípicas de grupos de neuronas que podrían extraer dicha información. Luego mediante

esta información podemos calcular la distancia perceptual entre los tonos.

En la periferia auditiva la información del sonido es segregada en el eje tonotópico y codificada por los disparos de la fibra auditiva mediante códigos de tasa de disparos y de enganche de fase. La organización tonotópica se preserva en toda la ruta auditiva, incluso en la corteza auditiva se han encontrado mapas neuronales con esta organización [96]. Los atributos perceptuales asociados al eje tonotópico, como el centroide espectral y la altura de tonos puros, son de los más reconocibles y salientes.

La información temporal del sonido se puede separar en lo que respecta a la envolvente y en la estructura fina de la fase. Ambas características son codificadas temporalmente en la fibra auditiva, la primera a través de la variación de la tasa de disparos y la segunda mediante el enganche de fase. Las codificaciones temporales son poco robustas ya que son susceptibles a sufrir modificaciones en los tiempos de los disparos en cada relevo de la ruta auditiva. Esto implica que son necesarios circuitos neuronales que transformen la codificación temporal a codificación en base a tasa de disparos. Los mecanismos exactos de esta transformación todavía permanecen elusivos [9].

Con respecto a la codificación de la frecuencia de modulación de tonos con modulaciones en amplitud, hay evidencia de la existencia de un banco de filtros sobre la frecuencia de modulación, que en paralelismo con el eje tonotópico se lo llama eje periodotópico. Este eje ha sido encontrado en forma de mapa neuronal en el colículo inferior del gato [72] y del jerbo [69]. Las características precisas de este banco de filtros todavía no son conocidas completamente, pero es razonable suponer que en alguna etapa de la ruta auditiva, existen grupos de neuronas que responden con preferencia a cierto rango de frecuencias de modulación. Existen trabajos de modelado que proponen circuitos neuronales capaces de extraer las frecuencias de modulación en base a tasa de disparos [52] y también trabajos que proponen de forma simplificada un banco de filtros que es alimentado directamente con las señales de la fibra auditiva [18].

A continuación pasaremos a medir la frecuencia de modulación y la frecuencia portadora con el modelo de la cóclea. Para la frecuencia portadora utilizaremos la información del eje tonotópico, mientras que para la frecuencia de modulación incluiremos un banco de filtros alimentado con la fibra auditiva que simula las respuestas de las neuronas del colículo

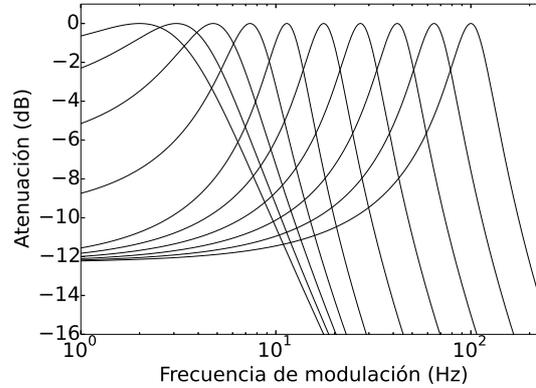


Figura 4.30: Banco de filtros sobre la frecuencia de modulación. Las frecuencias centrales de los filtros están espaciadas logarítmicamente. Para frecuencias centrales menores a 10 Hz el ancho de banda es 5 Hz, para frecuencias centrales mayores a 10 Hz, se fijó el valor de Q en 2. Adaptado de Dau et al. [18]

inferior sensibles a la frecuencia de modulación. En la figura 4.30 se puede ver el banco de filtros utilizado, que fue adaptado de Dau et al. [18]. Los filtros son resonadores de un polo complejo dados por la siguiente función de transferencia:

$$H(z) = \frac{1 - R}{1 - R e^{i2\pi f_c \Delta} z^{-1}}$$

Con $R = e^{-B\Delta}$, f_c la frecuencia central, B el ancho de banda y Δ el paso temporal. El ancho de banda de los filtros con frecuencia centrales menores a 10 Hz fue de 5 Hz, y luego para las frecuencias centrales mayores se fijó el Q en 2. Se usaron 10 filtros con frecuencias centrales espaciadas logarítmicamente entre 2 y 100 Hz.

En los resultados de esta sección utilizamos el modelo de la cóclea con sinapsis incluida, por lo tanto cada canal del modelo está representado por la tasa de disparos generada por la sinapsis. Las señales de cada canal son pasadas por el banco de filtros y su amplitud RMS es calculada. Luego, para condensar toda la información de modulación se promedian todos los canales. Este proceso genera, para cada estímulo, un vector con 10 elementos que tienen las respuestas del banco de filtro con la información de todos los canales.

Para el cálculo de la respuesta prototípica de la frecuencia portadora, proponemos una reducción del eje tonotópico en 10 bandas superpuestas, sumando la actividad de la sinapsis

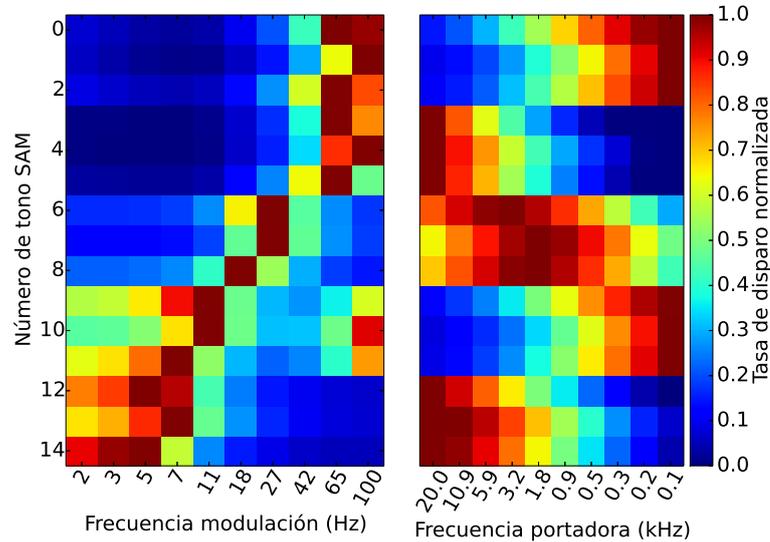


Figura 4.31: Población de neuronas del banco de filtros de modulación y de frecuencia portadora. Cada fila representa un sonido de los 15 sonidos del experimento con tonos SAM, y las columnas representan las respuestas de neuronas tuneadas a ciertas frecuencias de modulación o de portadora.

por cada banda. Este proceso junto con el del banco de filtros de modulación intentan simular la degradación de las curvas de sintonía que ocurre a medida que se avanza en la ruta auditiva.

En la figura 4.31 podemos observar los resultados de los vectores de modulación y de la portadora para los 15 sonidos de la condición 1 del experimento con tonos SAM (3.1a). Se puede ver como las dos representaciones capturan para los valores máximos, a la frecuencia de modulación y a la frecuencia portadora respectivamente.

Una de las cualidades de representar los tonos SAM de esta forma prototípica con pocas neuronas (10), es que permite que haya un solapamiento entre las respuestas de cada sonido. Esto permite calcular una distancia entre sonidos simplemente tomando la distancia euclídea entre estos vectores. Mediante esta idea, es posible simular un experimento de comparación de pares a partir de la distancia entre las representaciones prototípicas. Los resultados del experimento con tonos SAM eran el promedio de muchas respuestas, y cada participante podría haber usado diferentes pesos para las dimensiones perceptuales de modulación y de la portadora. Por este motivo, vamos a incluir un factor de ruido interno

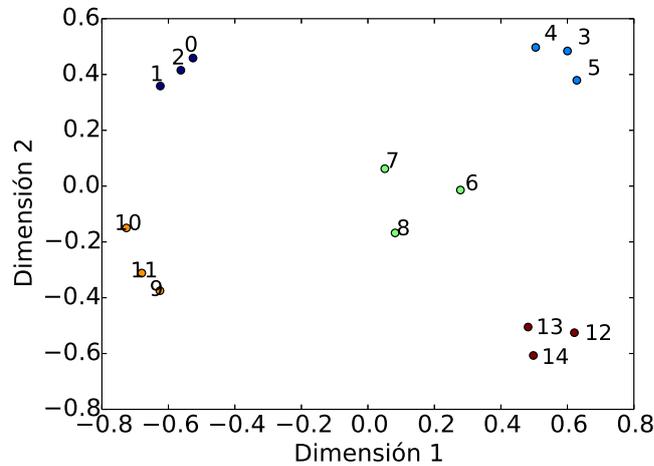


Figura 4.32: Configuración MDS INDSCAL para las representaciones prototípicas de la frecuencia moduladora y portadora para los 15 tonos SAM del experimento psicofísico.

por sujeto que regula el peso relativo entre el vector de modulación y de la portadora. En la figura 4.32 se puede ver la configuración MDS INDSCAL obtenida de simular el experimento para 25 sujetos con pesos relativos tomados de una distribución normal con desviación estándar 1 y media 1. La concordancia observada se debe al hecho de haber generado representación con baja definición y con solapamiento, en este caso 10 frecuencias centrales de modulación y 10 bandas del eje tonotópico.

4.5. Resumen y discusión

En este capítulo se presento un modelo dinámico de la cóclea basado en una cadena de osciladores con acoplamiento hidrodinámico y alimentación lateral. Se utilizó la aproximación de una cóclea unidimensional y con parámetros morfológicos constantes. A pesar de estas aproximaciones se lograron generar respuestas aceptables del movimiento de la membrana basilar. La inclusión de las no linealidades, en el término activo o en el término de amortiguación de los osciladores, lograron capturar los comportamientos de compresión para las respuestas con distintas intensidades, sin embargo se observó que la no linealidad en la amortiguación generó respuestas levemente más similares a las experimentales. Los parámetros elegidos para correr el modelo fueron seleccionados en base a lograr un compromiso de las respuestas frente a diferentes estímulos y situaciones. Queda pendiente el análisis exhaustivo del comportamiento del modelo en función de todos sus parámetros, ya sea mediante ajustes con datos experimentales o análisis de inestabilidades y comportamientos globales del sistema.

Una posibilidad para lograr mejores ajustes con los datos experimentales, recae en agregar o modificar las dependencias espaciales de ciertos parámetros. Muchos parámetros se han dejado constantes por simplicidad, pero es posible que se puedan obtener respuestas más realistas en todo el rango de frecuencia simplemente agregando una dependencia espacial.

Pudimos computar las respuestas de las IHC y la sinapsis, con resultados muy convincentes, gracias al modelo de Zilany et al. [124] que ya se encontraba implementado. Queda realizar un análisis sobre las ventajas de incluir la etapa de la IHC y la sinapsis para el análisis de sonidos, ya que en muchos casos es suficiente utilizar los desplazamientos de la MB. En este sentido, hay que profundizar la relación entre las respuestas de la membrana basilar y los patrones de excitación perceptuales.

Con respecto al modelo de rugosidad planteado, es necesario realizar ajustes para todo el rango de frecuencias portadoras, posiblemente utilizando largos de ventanas adecuados para cada sitio de la MB.

Finalmente, el agregado de uno de los procesos que ocurren más arriba en la ruta

auditiva, como es el caso de la codificación de la modulación en el colículo inferior, permitió, aunque de forma simplificada, simular el experimento de comparación de tonos SAM. Es necesario estudiar con mayor profundidad la relación entre los procesos cognitivos centrales que se involucran en la comparación de timbres y la información que se dispone en las primeras etapas de la ruta auditiva. También queda pendiente estudiar si es posible modelar las respuestas individuales de los participantes agregando más fuente de ruido interno.

Capítulo 5

Conclusiones generales

Esta tesis planteó el estudio de la percepción tímbrica de sonidos con modulaciones en amplitud. El trabajo comenzó a partir de la interacción con el saxofonista y compositor Martín Proscia que poseía un corpus de sonidos de multifónicos del saxofón. Este encuentro permitió notar que las herramientas clásicas para estudiar el timbre musical han sido desarrolladas en base a sonidos armónicos y no resultaban adecuadas para estudiar sonidos mas complejos espectralmente, como los multifónicos.

Por otra parte, existe un cúmulo de sonoridades propias de la música electroacústica y contemporánea que todavía no han sido estudiadas con herramientas analíticas de acústica y experimentos perceptuales, con algunas contadas excepciones [42]. Esta tesis propuso el abordaje sobre el timbre de un conjunto de estas sonoridades.

Los sonidos multifónicos estudiados habían sido previamente organizados por Martín Proscia en cuatro clases en base a sus cualidades tímbricas. Uno de los objetivos logrados en esta tesis fue el de dar un fundamento perceptivo y acústico de la clasificación propuesta originalmente basándose en un conjunto de atributos, entre los que se encontraban, la frecuencia fundamental, la frecuencia de modulación, el centroide espectral y la rugosidad. Para ganar confianza en esta clasificación se propuso un experimento de comparación de pares de sonidos donde los sujetos debían juzgar las similitudes tímbricas entre los multifónicos. El experimento fue realizado en un ambiente controlado con sujetos con experiencia en música electroacústica. Este experimento mostró que los sonidos que perte-

necían a una misma clase se mantuvieron agrupados en el espacio tímbrico perceptual. Al mismo tiempo, se pudieron interpretar las dimensiones tímbricas perceptuales en base a los atributos acústicos, encontrándose altos grados de correlación para el centroide espectral y la rugosidad.

El trabajo realizado con los multifónicos abrió nuevos interrogantes acerca del rol de las modulaciones como un atributo tímbrico relevante (algo escasamente estudiado en trabajos previos). Para contribuir a resolver esta cuestión partiendo de un caso mas sencillo y controlable se propuso entonces, como continuación de esta tesis, llevar a cabo un experimento de comparación de tonos modulados en amplitud sinusoidal mente (tonos SAM). Se decidió realizarlo de forma masiva mediante la web y una plataforma desarrollada de forma específica. La participación en el experimento fue muy buena, con más de 80 personas en 2 convocatorias.

Los estímulos utilizados en el experimento fueron tomados de un espacio de configuraciones de dos dimensiones: frecuencia portadora y frecuencia de modulación, dejando fija la profundidad de modulación en su valor máximo 1. Pudo observarse mediante el análisis MDS que las dimensiones perceptuales utilizadas por los sujetos fueron las mismas que las del espacio de configuración. Esto permitió concluir que los participantes, en promedio, pudieron comparar los sonidos factorizando perceptualmente la modulación y la frecuencia portadora. Por otro lado se evaluaron dos condiciones de grupos de sonidos: una condición con sonidos muy cercanos entre sí en el espacio de configuraciones y otra con sonidos más alejados. Los resultados indicaron que los sujetos utilizaron diferentes escalas para cada condición. Finalmente se pudo observar que los ajustes MDS de tres dimensiones fueron mejores para algunas condiciones, por lo que indicaría que el espacio tímbrico perceptual podría haber estado influenciado por otro atributo acústico como la rugosidad, aunque es necesario mayor análisis y mayor número de participantes para aclarar este punto.

De forma paralela al estudio experimental del timbre se desarrolló un modelo de la cóclea para poder estudiar el timbre a través de la actividad en la periferia auditiva. Los objetivos de esta tesis incluían el revelamiento de los modelos de la mecánica coclear, el desarrollo de un modelo reducido inspirado en la dinámica no lineal y la implementación de dicho modelo de forma eficiente. Todos estos objetivos fueron cumplidos logrando generar

un modelo versátil que permite modificar sus parámetros con comodidad, procesar sonidos de forma eficiente en tiempos de computo hasta solo 3 veces el tiempo real, generando además cocleogramas realistas. Se contrastaron los resultados de las simulaciones con los datos fisiológicos más consensuados de la literatura [98]. Si bien aun es necesario completar el modelo incorporando más datos fisiológicos y realizando alguna optimización de sus parámetros, se obtuvo en general un buen acuerdo tanto cualitativo como cuantitativamente.

Se realizaron algunas comparaciones con algunos datos fisiológicos pero es necesario completar esta etapa con más datos de fenómenos auditivos y realizar una optimización de los parámetros. Muchos de los parámetros utilizados fueron dejados constantes en el eje de la cóclea, pero para obtener mejores respuestas para todas las frecuencias es necesario introducir dependencias espaciales en ciertos parámetros como los del sistema activo y la no linealidad.

El modelo de la cóclea fue aplicado al estudio de tonos con modulación en amplitud. Por un lado se lo utilizó para generar un nuevo modelo para el cómputo de la rugosidad. Mientras que los modelos más utilizados para cuantificar la rugosidad son de carácter espectral o mixto, en este trabajo buscamos plantear un modelo puramente temporal. Para esto se utilizó la salida del modelo de la cóclea, y se computó, mediante una ventana móvil, la amplitud de las modulaciones de la envolvente para cada canal de la cóclea. La comparación de este modelo con datos experimentales, mostró que los resultados son promisorio, pero es necesario modificar el tamaño de la ventana utilizada en cada canal para lograr ajustes en todo el rango de frecuencias.

Finalmente se planteó un mecanismo posible como base para los procesos cognitivos de la comparación de timbres de sonidos con modulación. El modelado de la periferia con inclusión de un banco de filtros de modulación, permitió generar respuestas neuronales prototípicas que han sido observadas en el colículo inferior. Estas respuestas permitieron estimar la frecuencia de modulación de los tonos SAM y en conjunto con la información tonotópica entregar una representación compacta de cada sonido. Las respuestas prototípicas propuestas tuvieron baja resolución y gran solapamiento entre las bandas de frecuencias moduladoras o portadora, lo que permitió calcular la distancia euclídea entre las respuestas

y poder establecer una distancia entre los sonidos. Finalmente se simuló un experimento con varios participantes agregando un ruido interno que afecta el peso relativo entre las dimensiones de modulación y de la portadora, obteniendo buenos resultados.

Apéndice A

A.1. Escalamiento multidimensional

El análisis de escalamiento multidimensional (MDS) es una técnica de reducción de la dimensionalidad con la cual es posible representar datos de un espacio de muchas dimensiones, en un espacio de baja dimensión típicamente 2 o 3. Los datos deben estar representados por una matriz de distancias, o disimilitudes, y mediante un proceso de optimización se buscan las coordenadas que mejor representan estas distancias en un espacio de dimensión menor.

El análisis de MDS puede servir varios propósitos. Por un lado permite generar visualizaciones accesibles para estudiar conjuntos de datos que de otra manera sólo son representados por números. Por otro lado, en los experimentos cognitivos, también es una herramienta que permite descubrir las dimensiones perceptuales que pudieron ser utilizadas para juzgar las disimilitudes.

Existen muchas variantes de MDS. Esta el MDS métrico clásico, en el cual se utiliza una métrica euclídea, pero también es posible utilizar otras métricas para representar las distancias. Y también está el MDS no métrico, para los casos donde las disimilitudes no representan distancias sino ordenes en un rango.

El análisis MDS clásico es idéntico al análisis por componentes principales (PCA del inglés *principal component analysis*). Por lo tanto la esencia de la representación en baja dimensionalidad recae en filtrar las dimensiones que contribuyen menos a la varianza de los datos.

Para encontrar la configuración MDS que mejor satisface los datos, se trata de minimizar

una función de pérdida dada por la siguiente expresión:

$$\sigma(X) = \sum_{i < j} w_{ij} (d_{ij}(X) - \delta_{ij})^2 \quad (\text{A.1})$$

Esto representa, para una configuración dada X , el valor cuadrático medio de la diferencia pesada por w_{ij} entre las distancias $d_{ij}(X)$ y las disimilitudes δ_{ij} entre los puntos i y j .

Con una formulación similar se utiliza un coeficiente normalizado que simboliza la calidad del ajuste, denominado *Stress* o *Stress-1*:

$$\text{Stress-1} = \sqrt{\frac{\sum_{i < j} (d_{ij}(X) - \delta_{ij})^2}{\sum_{i < j} d_{ij}^2(X)}} \quad (\text{A.2})$$

También existen otras variantes de MDS relacionadas al tipo de ajuste y no a la métrica utilizada. En estos casos podemos encontrar el INDSCAL (del inglés *individual differences scaling*), el CLASCAL y el CONSCAL. En este trabajo utilizamos el INDSCAL, para una revisión de los otros métodos dirigirse a McAdams et al. [78]. El INDSCAL permite, a partir de varias fuentes de datos de disimilitudes, por ejemplo las respuestas de varios sujetos, encontrar una configuración geométrica que satisfaga en mayor medida a todos los datos. Esto lo consigue asignando diferentes pesos en cada dimensión para cada sujeto.

A.2. Cocleograma de banco de filtros *gammatones*

Los filtros *gammatones* fueron popularizados por Johannesma [59] como un modelo para la respuesta impulso de las fibras auditivas, estimados a partir de la técnica de correlación en reversa de los patrones de disparos. Los *gammatones* son un filtro pasa bandas cuya respuesta impulso $Gf_c(t)$ es el producto de una función *gamma* y un tono (de ahí el nombre "gammatone"):

$$g_{f_c}(t) = t^{N-1} e^{-2\pi t b(f_c)} \cos(2\pi f_c t + \phi) u(t)$$

Donde N es el orden del filtro, f_c es la frecuencia central en Hz, ϕ es la fase y $u(t)$ es la función escalón. La función $b(f_c)$ determina el ancho de banda para una dada frecuencia central. Para $f_c/b(f_c)$ suficientemente grande la respuesta en frecuencia del *gammatone* se puede aproximar por la siguiente expresión:

$$G(f) \approx \left[1 + \frac{j(f - f_c)}{b(f_c)} \right]^{-N} \quad (0 < f < \infty)$$

Por lo tanto puede verse que el filtro es simétrico alrededor de la frecuencia f_c . En el trabajo de Patterson et al. [88], mostraron que para un orden $N = 4$, el filtro *gammatone* permite ajustar muy bien los filtros auditivos determinados experimentalmente.

El ancho de banda de los filtros *gammatone* se suele ajustar de manera proporcional al ancho de banda rectangular equivalente (ERB) de los filtros auditivos, según Patterson et al. [88]:

$$b(f) = 1,013ERB(f)$$

La figura A.1 muestra la respuestas impulso y en frecuencia para ocho filtros *gammatone* que están espaciados equidistantes en la escala ERB_N .

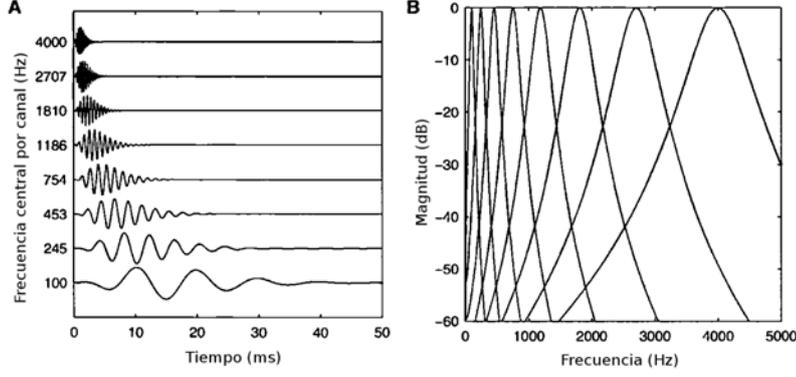


Figura A.1: Respuestas impulso y en frecuencia para ocho filtros gammatone, de diferentes frecuencias centrales

A.3. Modelo espectral de rugosidad

El modelo espectral de la rugosidad utilizado en este trabajo está basado en la formulación de Sethares [104] que calcula la rugosidad para un par de tonos puros de amplitudes y frecuencias A_1 , A_2 , y f_1 , f_2 .

$$R = 5A_{min} * A_{max} \left(e^{-b_1 s(f_{max} - f_{min})} - e^{-b_2 s(f_{max} - f_{min})} \right) \quad (\text{A.3})$$

$$s = 0,24 / (s_1 f_{min} + s_2) \quad (\text{A.4})$$

Donde $A_{min} = \min(A_1, A_2)$, $f_{max} = \max(f_1, f_2)$, $f_{min} = \min(f_1, f_2)$ y los coeficientes toman los siguientes valores $b_1 = 3,5$, $b_2 = 5,75$, $s_1 = 0,0207$, $s_2 = 18,96$.

Para el cómputo de la rugosidad total en un sonido con varias componentes se suman las rugosidades parciales entre todos los pares de componentes presentes en el sonido. A pesar de que este modelo no contempla las fases relativas de las componentes, los valores entregados de rugosidad, tienen buena correspondencia con experimentos perceptuales [114].

A.4. Acoplamiento hidrodinámico 1-D

El acoplamiento hidrodinámico es la principal forma de acoplamiento que hay entre las distintas secciones de la membrana basilar (MB) y permite la formación de la onda viajera. El acoplamiento lateral dado por la rigidez de la MB es despreciable en comparación al acoplamiento hidrodinámico. Para modelar este acoplamiento es necesario representar a la cóclea como una caja llena de fluido que está dividida por la partición coclear o MB para simplificar (figura A.2).

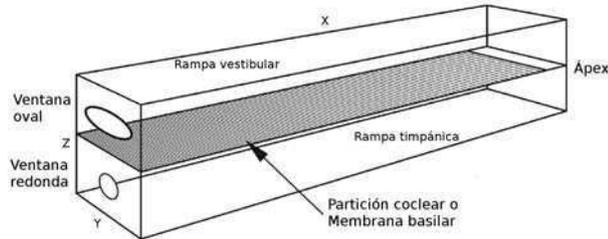


Figura A.2

Es posible resolver el problema hidrodinámico en 3 dimensiones, pero también en aproximaciones de 1 y 2 dimensiones [21]. Por simplicidad y eficiencia computacional utilizaremos la aproximación 1-D a partir del eje X. Esta aproximación se llama de "onda larga", ya que sólo tiene validez en la región lejos de la resonancia, donde la longitud de onda es grande, dado que en esa región la presión es homogénea en la dimensión Z. Cerca de la resonancia, cuando la longitud de onda es corta, la distribución de la presión varía en la dimensión Z de forma marcada, por lo que es necesario el desarrollo 2-D para un análisis con mayor exactitud. A pesar de esto los resultados de la aproximación 1-D son suficientemente válidos para realizar cocleogramas.

Por lo tanto consideremos las presiones del fluido unidimensionales p_v y p_t para las rampas vestibular y timpánica respectivamente. Dado que las ondas de presión que se propagan están conectadas al final de la cóclea por el helicotrema, es posible definir las variables de la presión simétrica y asimétrica.

$$p = p_- = \frac{1}{2}(p_v - p_t)$$

$$p_+ = \frac{1}{2}(p_v + p_t)$$

La onda de presión simétrica genera compresión sobre la MB y la antisimétrica genera movimientos transversales en la MB, y es la de interés. De esta manera se simplifican las dos cavidades de fluidos como una sola que ejerce el doble de la presión.

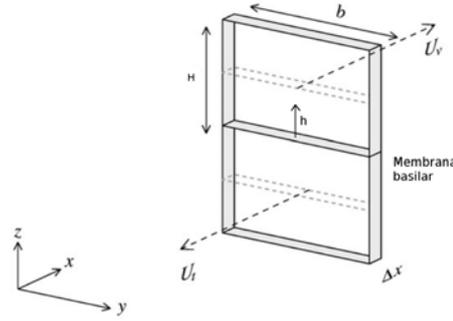


Figura A.3: Sección de la cóclea rectangular. La conservación de masa implica que las variaciones del flujo volumétrico U deben ser absorbidas por la velocidad de la MB.

El siguiente paso es establecer las leyes de conservación para el fluido. En el caso unidimensional, la ecuación de la conservación de la masa $\nabla \cdot \mathbf{u} = 0$ debe incluir el movimiento de la MB y no puede reducirse a $\partial u / \partial x = 0$. Esto implica que los cambios en la velocidad volumétrica del fluido U se reflejan en la velocidad de la MB $\partial h / \partial t$.

$$U(x + dx, t) - U(x, t) = \frac{\partial h(x, t)}{\partial t} b dx$$

$$Hb \frac{u(x + dx, t) - u(x, t)}{dx} = \frac{\partial h(x, t)}{\partial t} b$$

$$\frac{\partial u(x, t)}{\partial x} = \frac{1}{H} \frac{\partial h(x, t)}{\partial t} \quad (\text{A.5})$$

Donde b es el ancho de la cavidad, y H la altura. Luego, podemos expresar la conservación de momento lineal en una dimensión:

$$2\rho_0 \frac{\partial u(x, t)}{\partial t} = -\frac{\partial p(x, t)}{\partial x} \quad (\text{A.6})$$

Donde el factor 2 representa el hecho de que se consideran las presiones de las dos cavidades. Si se deriva a la ecuación A.5 con respecto al tiempo y a la ecuación A.6 con respecto a x , es posible igualar los términos con derivadas cruzadas:

$$\frac{\partial^2 u(x, t)}{\partial x \partial t} = \frac{1}{H} \frac{\partial^2 h(x, t)}{\partial t^2}$$

$$2\rho_0 \frac{\partial^2 u(x, t)}{\partial t \partial x} = -\frac{\partial^2 p(x, t)}{\partial x^2}$$

Igualando obtenemos:

$$\frac{\partial^2 p(x, t)}{\partial x^2} = \frac{-2\rho_0}{H} \frac{\partial^2 h(x, t)}{\partial t^2} \quad (\text{A.7})$$

Esta ecuación es la que gobierna el acoplamiento hidrodinámico. Puede entenderse como una fuerza sobre la MB que está relacionada a la concavidad de la onda de presión p .

A.5. Relación de dispersión del sistema activo

En la sección 4.2 se introdujo el modelado de la cóclea haciendo énfasis en las soluciones numéricas. Sin embargo, es posible establecer ciertos resultados teóricos si se consideraran soluciones armónicas del movimiento de la MB. Retomemos las ecuaciones principales que determinan el comportamiento de la MB.

$$m \frac{\partial^2 h(x, t)}{\partial t^2} + \mu(x) \frac{\partial h(x, t)}{\partial t} + s(x)h(x, t) = -p(x, t) \quad (\text{A.8})$$

$$\frac{\partial^2 p(x, t)}{\partial x^2} = -\frac{2\rho}{H} \frac{\partial^2 h(x, t)}{\partial t^2} \quad (\text{A.9})$$

La ecuación A.8 determina las características de los osciladores de la MB, mientras que la A.9 determina el acoplamiento entre los osciladores. Si se plantean soluciones armónica $h(x, t) = \tilde{h}(x)e^{i\omega t}$ y $p(x, t) = \tilde{p}(x)e^{i\omega t}$ podemos escribir la ecuación A.8 como la definición de la impedancia acústica específica de la MB $Z_{MB}(x)$:

$$Z_{MB} = -\frac{p(x)}{\dot{h}(x)} = -i\omega m + \mu(x) + \frac{s(x)}{i\omega} \quad (\text{A.10})$$

Luego la ecuación A.9 puede escribirse como:

$$\frac{\partial^2 p(x)}{\partial x^2} - \frac{2i\omega\rho}{HZ_{MB}(x)}p(x) = 0 \quad (\text{A.11})$$

$$(\text{A.12})$$

Es posible considerar una solución armónica también para la parte espacial de la presión $p(x) = p_0 e^{-ikx}$, y permite expresar la última ecuación como:

$$k(x)^2 = -\frac{2i\omega\rho}{HZ_{MB}(x)} \quad (\text{A.13})$$

La ecuación A.12 tiene cierta semejanza con la ecuación de Helmholtz, con la diferencia de que el término k^2 depende de x , por lo que no es posible encontrar soluciones analíticas

generales, salvo para ciertos casos. Sin embargo es posible analizar ciertos comportamientos del sistema a partir de la relación entre el vector de onda k y la impedancia, ecuación A.13.

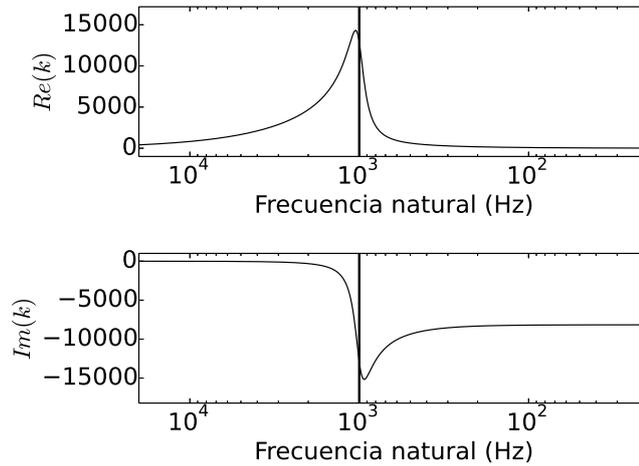


Figura A.4: Sistema pasivo: Parte real e imaginaria del vector de onda k en función de la frecuencia natural de los osciladores. La línea vertical indica la frecuencia del estímulo

Primero observemos en la figura A.4 las partes reales e imaginarias de k para un estímulo de 1000 Hz. Para la parte real, se puede observar cómo en la primera región, donde la impedancia está dominada por la constante elástica, el vector de onda crece hasta aproximadamente la posición del oscilador cuya frecuencia coincide con la del estímulo. Esto indica de forma recíproca el comportamiento de la longitud de onda, la cual se hace cada vez más pequeña hasta llegar a la región de resonancia. Luego de esta región, la parte real disminuye ya que la impedancia está dominada por la masa. Con respecto a la parte imaginaria, podemos ver que esta toma valores negativos los cuales crecen rápidamente cuando la impedancia está dominada por la amortiguación en la región de resonancia. Considerando la solución armónica de la presión $\tilde{p}(x) = p_0 e^{-ikx}$ podemos notar que la parte real de k gobierna el comportamiento oscilatorio, mientras que la parte imaginaria de k , es la que gobierna el comportamiento exponencial. Esto permite establecer que la parte imaginaria de k debe ser mayormente negativa para que las soluciones no diverjan. A continuación veremos como la inclusión del sistema activo afecta justamente este comportamiento.

La impedancia de la MB para el sistema activo puede escribirse a partir de la ecuación

ción 4.19 utilizando la aproximación $\tilde{h}(x_0) \approx e^{-ikx_0}$ sólo válida para un entorno pequeño alrededor de x_0 , lo que termina dando:

$$Z_{MB} = -i\omega m + \mu(x) + \frac{k(x)}{i\omega} \left(1 - \alpha\beta e^{ikd} + \alpha e^{-ikd} \right) \quad (\text{A.14})$$

Si reemplazamos esta impedancia en la ecuación A.13, podemos notar que queda una ecuación trascendente y no es posible despejar k analíticamente, pero si de forma numérica. En la figura A.5 se puede observar la parte real e imaginaria del vector de onda k a partir del sistema activo. Puede verse que las principales diferencias con el sistema pasivo se encuentran en la región de la resonancia y en la región apical. Para la parte imaginaria, puede verse una pequeña desviación hacia valores positivos en la región de la resonancia, lo que indicaría un comportamiento de expansión de la onda viajera en esa región.

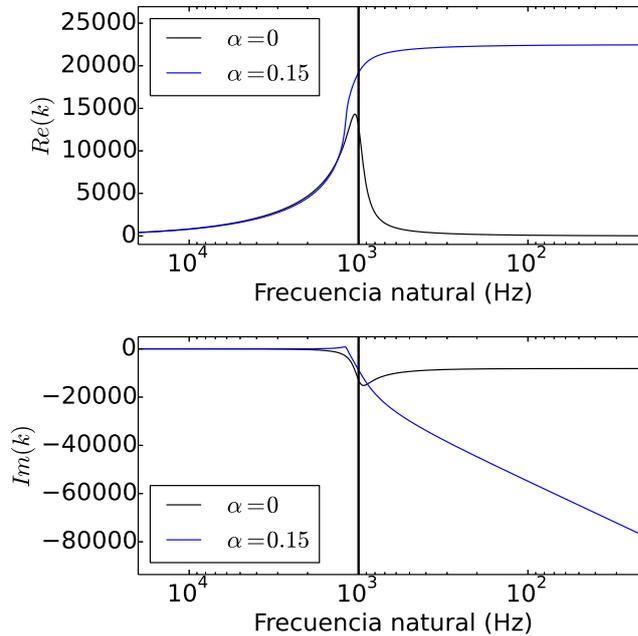


Figura A.5: Parte real e imaginaria del vector de onda k en función de la frecuencia natural de los osciladores, para el sistema pasivo (negro) y sistema activo (azul) con los siguientes valores $\alpha = 0,15$, $\beta = 0,3$. La línea vertical indica la frecuencia del estímulo

En la figura A.6 se muestra en detalle esta región y para varios valores de α . Podemos ver que hay un valor crítico donde ocurre una bifurcación y las curvas se despegan. Las

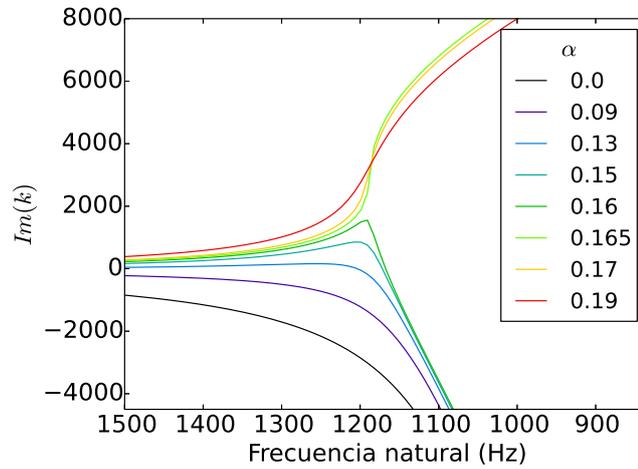


Figura A.6: Parte imaginaria del vector de onda k en función de la frecuencia natural de los osciladores, para el sistema activo y varios valores de α ($\beta = 0,3$).

curvas que permanezcan del lado positivo tendrán soluciones asociadas que terminaran divergiendo, por lo que el sistema será inestable. Sin embargo en el modelo completo con no linealidad, esto no ocurre ya que las divergencias son frenadas por los términos no lineales, pero las respuestas pueden quedar saturadas. El mecanismo del sistema activo para generar una inestabilidad puede ser utilizado para modelar las emisiones otoacústicas. Si el coeficiente α es modificado en una sola región de la cóclea se podrán generar inestabilidades locales, que terminan inyectando energía en el sistema y pueden propagar oscilaciones hacia la ventana oval.

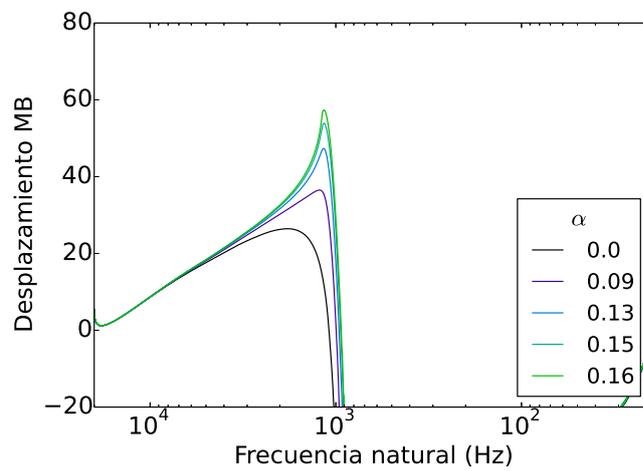


Figura A.7: Magnitud del desplazamiento de la MB para los valores de α utilizados en la figura A.6

Bibliografía

- [1] Karsten Ahnert and Mario Mulansky. Odeint-solving ordinary differential equations in c++. *arXiv preprint arXiv:1110.3397*, 2011. [93](#)
- [2] Jonathan Ashmore. Cochlear outer hair cell motility. *Physiological Reviews*, 88(1): 173–210, 2008. [74](#)
- [3] von W Aures. A procedure for calculating auditory roughness. *Acustica*, 58(5):268–281, 1985. [30](#), [107](#), [108](#)
- [4] Bruno Bartolozzi. *New sounds for woodwind*, volume 96. Oxford University Press, 1982. [37](#)
- [5] Arthur H Benade. *Fundamentals of Musical Acoustics: Second*. Courier Corporation, 2012. [35](#)
- [6] G Bi. Zur theorie des homes; die schwingungsform der basilarmembran. *Phys. Z*, 29: 793–810, 1928. [77](#)
- [7] Guy J Brown and Martin Cooke. Computational auditory scene analysis. *Computer Speech & Language*, 8(4):297–336, 1994. [20](#)
- [8] Anne Caclin, Stephen McAdams, Bennett K Smith, and Suzanne Winsberg. Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones). *The Journal of the Acoustical Society of America*, 118(1):471–482, 2005. [11](#)
- [9] Peter A Cariani. Neural timing nets. *Neural Networks*, 14(6):737–753, 2001. [109](#)

-
- [10] J Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multi-dimensional scaling via an n-way generalization of “eckart-young” decomposition. *Psychometrika*, 35(3):283–319, 1970. [50](#)
- [11] Gerard R Charbonneau. Timbre and the perceptual effects of three types of data reduction. *Computer Music Journal*, pages 10–19, 1981. [10](#)
- [12] Jer-Ming Chen, John Smith, and Joe Wolfe. Saxophone acoustics: introducing a compendium of impedance and sound spectra. *Acoustics Australia*, 37(1-19), 2009. [34](#), [35](#), [36](#)
- [13] Jer Ming Chen, John Smith, and Joe Wolfe. Saxophonists tune vocal tract resonances in advanced performance techniques. *The Journal of the Acoustical Society of America*, 129(1):415, January 2011. ISSN 1520-8524. doi: 10.1121/1.3514423. URL <http://www.ncbi.nlm.nih.gov/pubmed/21303021>. [35](#)
- [14] Taishih Chi, Yujie Gao, Matthew C Guyton, Powen Ru, and Shihab Shamma. Spectro-temporal modulation transfer functions and speech intelligibility. *The Journal of the Acoustical Society of America*, 106(5):2719–2732, 1999. [22](#)
- [15] Michel Chion and Pierre Schaeffer. *Guide des objets sonores*. Buchet/Chastel, 1983. [3](#), [37](#)
- [16] Peter Dallos. Neurobiology of cochlear inner and outer hair cells: intracellular recordings. *Hearing research*, 22(1):185–198, 1986. [75](#)
- [17] Peter Daniel and Reinhard Weber. Psychoacoustical roughness: Implementation of an optimized model. *Acta Acustica united with Acustica*, 83(1):113–123, 1997. [30](#), [106](#)
- [18] Torsten Dau, Birger Kollmeier, and Armin Kohlrausch. Modeling auditory processing of amplitude modulation. i. detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, 102(5):2892–2905, 1997. [109](#), [110](#)

- [19] E de Boer. Auditory physics. Physical principles in hearing theory. I. *Physics reports*, 62:87–174, 1980. URL <http://www.sciencedirect.com/science/article/pii/0370157380901003>. 77, 84, 85, 86
- [20] E de Boer. Auditory physics. Physical principles in hearing theory. II. *Physics Reports*, 1984. URL <http://www.sciencedirect.com/science/article/pii/037015738490108X>. 84
- [21] E de Boer. Auditory physics. Physical principles in hearing theory. III. *Physics Reports*, 203(3):125–231, May 1991. ISSN 03701573. doi: 10.1016/0370-1573(91)90068-W. URL <http://linkinghub.elsevier.com/retrieve/pii/037015739190068W>. 81, 84, 123
- [22] Alain De Cheveigne. Pitch perception models. In *Pitch*, pages 169–233. Springer, 2005. 56
- [23] RJ Diependaal and Hendrikus Duifhuis. Numerical methods for solving one-dimensional cochlear models in the time domain. *The Journal of the Acoustical Society of America*, (May 2013):1655–1666, 1987. URL <http://link.aip.org/link/?JASMAN/82/1655/1>. 83
- [24] Sophie Donnadieu. *Représentation mentale du timbre des sons complexes et effets de contexte*. PhD thesis, Paris 5, 1997. 4
- [25] Sophie Donnadieu. Mental representation of the timbre of complex sounds. In *Analysis, synthesis, and perception of musical sounds*, pages 272–319. Springer, 2007. 9
- [26] Hendrikus Duifhuis. *Cochlear Mechanics: Introduction to a Time Domain Analysis of the Nonlinear Cochlea*. Springer Science & Business Media, 2012. 85, 90, 91
- [27] TAJ Duke and Frank Jülicher. Active Traveling Wave in the Cochlea. *Physical Review Letters*, 90(15):1–4, April 2003. ISSN 0031-9007. doi: 10.1103/PhysRevLett.90.158101. URL <http://link.aps.org/doi/10.1103/PhysRevLett.90.158101>. 81

- [28] V M Eguíluz, M Ospeck, Y Choe, A J Hudspeth, and Marcelo O. Magnasco. Essential nonlinearities in hearing. *Physical review letters*, 84(22):5232–5, May 2000. ISSN 0031-9007. URL <http://www.ncbi.nlm.nih.gov/pubmed/14525401>. 80
- [29] Stephen J Elliott, Emery M Ku, and Ben Lineton. A state space model for cochlear mechanics. *The Journal of the Acoustical Society of America*, 122(5):2759–71, November 2007. ISSN 1520-8524. doi: 10.1121/1.2783125. URL <http://www.ncbi.nlm.nih.gov/pubmed/18189567>. 83
- [30] Hugo Fastl and Eberhard Zwicker. *Psychoacoustics: Facts and models*, volume 22. Springer Science & Business Media, 2007. 8, 18, 28, 29, 107, 108
- [31] Anne Faure. *Des sons aux mots, comment parle-t-on du timbre musical?* PhD thesis, Ecole des Hautes Etudes en Sciences Sociales (EHESS), 2000. 5
- [32] Harvey Fletcher. Auditory patterns. *Reviews of modern physics*, 12(1):47, 1940. 18
- [33] Maria N Geffen, Judit Gervain, Janet F Werker, and Marcelo O Magnasco. Auditory perception of self-similarity in water sounds. *Frontiers in integrative neuroscience*, 5(May):15, January 2011. ISSN 1662-5145. doi: 10.3389/fnint.2011.00015. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3095814&tool=pmcentrez&rendertype=abstract>. 10
- [34] CD Geisler and Chunning Sang. A cochlear model using feed-forward outer-hair-cell forces. *Hearing research*, 86:132–146, 1995. URL <http://www.sciencedirect.com/science/article/pii/037859559500064B>. 82
- [35] James Jerome Gibson. *The senses considered as perceptual systems*. 1966. 3
- [36] Edward Givelberg and Julian Bunn. A comprehensive three-dimensional model of the cochlea. *Journal of Computational Physics*, 191(2):377–391, 2003. 83
- [37] Richard J Gold, T y Pumphrey. Hearing. i. the cochlea as a frequency analyzer. *Proceedings of the Royal Society of London. Series B-Biological Sciences*, 135(881):462–491, 1948. 78

-
- [38] Julius L Goldstein. Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering. *Hearing research*, 49(1):39–60, 1990. 20, 83
- [39] R Gottfried. A more accurate notation for multiphonics using sideband ratios, 2008. 38
- [40] John M Grey and James A Moorer. Perceptual evaluations of synthesized musical instrument tones. *The Journal of the Acoustical Society of America*, 62(2):454–462, 1977. 10, 11
- [41] Thomas Grill. *Perceptually informed organization of textural sounds*. PhD thesis, Johannes Kepler University Linz, Austria, 2012. 15
- [42] Thomas Grill, Arthur Flexer, and Stuart Cunningham. Identification of perceptual qualities in textural sounds using the repertory grid method. In *6th Audio Mostly Conference (AM '11)*, pages 67–74, 2011. 5, 115
- [43] John J Guinan, Alec Salt, and Mary Ann Cheatham. Progress in cochlear physiology after békésy. *Hearing research*, 293(1):12–20, 2012. 80
- [44] Brian Gygi, Gary R Kidd, and Charle S Watson. Similarity and categorization of environmental sounds. *Perception & psychophysics*, 69(6):839–55, August 2007. ISSN 0031-5117. URL <http://www.ncbi.nlm.nih.gov/pubmed/18018965>. 11, 13, 15
- [45] John M Hajda. Relevant acoustical cues in the identification of western orchestral instrument tones. *The Journal of the Acoustical Society of America*, 102(5):3085–3085, 1997. 9
- [46] John M Hajda. The effect of amplitude and centroid trajectories on the timbre of percussive and nonpercussive orchestral instruments. In *Proc. 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America*, volume 3, pages 1887–1888, 1998. 8

- [47] John M Hajda. The effect of dynamic acoustical features on musical timbre. In *Analysis, synthesis, and perception of musical sounds*, pages 250–271. Springer, 2007. [9](#)
- [48] Stephen Handel. *Perceptual coherence: Hearing and seeing*. Oxford University Press New York, 2006. [4](#), [5](#)
- [49] Stephen Handel and Molly L Erickson. A rule of thumb: The bandwidth for timbre invariance is one octave. *Music Perception*, 19(1):121–126, 2001. [56](#)
- [50] Trevor Hastie, Robert Tibshirani, Jerome Friedman, T Hastie, J Friedman, and R Tibshirani. *The elements of statistical learning*, volume 2. Springer, 2009. [4](#)
- [51] Hermann LF Helmholtz. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Cambridge University Press, 2009. [27](#)
- [52] M J Hewitt and R Meddis. A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus. *The Journal of the Acoustical Society of America*, 95(4):2145–2159, 1994. ISSN 00014966. doi: 10.1121/1.408676. [109](#)
- [53] Christophe Hourdin, Gérard Charbonneau, and Tarek Moussa. A multidimensional scaling analysis of musical instruments’ time-varying spectra. *Computer Music Journal*, pages 40–55, 1997. [7](#)
- [54] AJ Hudspeth. Making an effort to listen: mechanical amplification in the ear. *Neuron*, 59(4):530–545, 2008. [80](#)
- [55] William Hutchinson and Leon Knopoff. The acoustic component of western consonance. *Journal of New Music Research*, 7(1):1–29, 1978. [27](#), [28](#)
- [56] American National Standards Institute. *Acoustical terminology ANSI S1.1-1994*. 1994. [2](#)
- [57] Toshio Irino and Roy D Patterson. A time-domain, level-dependent auditory filter: The gammachirp. *The Journal of the Acoustical Society of America*, 101(1):412–419, 1997. [20](#)

- [58] Paul Iverson and Carol L Krumhansl. Isolating the dynamic attributes of musical timbre). *The Journal of the Acoustical Society of America*, 94(5):2595–2603, 1993. [11](#)
- [59] Peter IM Johannesma. The pre-response stimulus ensemble of neurons in the cochlear nucleus. In *Symposium on Hearing Theory*, pages 58–69. IPO Eindhoven, Holland, 1972. [121](#)
- [60] P X Joris, C E Schreiner, and a Rees. Neural processing of amplitude-modulated sounds. *Physiological reviews*, 84(2):541–77, April 2004. ISSN 0031-9333. doi: 10.1152/physrev.00029.2003. URL <http://www.ncbi.nlm.nih.gov/pubmed/15044682>. [23](#), [25](#), [26](#)
- [61] RH Kay. Hearing of modulation in sounds. *Physiological Reviews*, 62(3):894–975, 1982. [22](#)
- [62] Piotr Kazmierczak and Ulrich Müller. Sensing sound: molecules that orchestrate mechanotransduction by hair cells. *Trends in neurosciences*, 35(4):220–229, 2012. [74](#)
- [63] Douglas H. Keefe and Bernice Laden. Correlation dimension of woodwind multiphonic tones. *The Journal of the Acoustical Society of America*, 90(October):1754–65, October 1991. ISSN 0001-4966. URL <http://link.aip.org/link/?JASMAN/90/1754/1http://www.ncbi.nlm.nih.gov/pubmed/1960272>. [36](#)
- [64] David T Kemp. Stimulated acoustic emissions from within the human auditory system. *The Journal of the Acoustical Society of America*, 64(5):1386–1391, 1978. [78](#)
- [65] Roger A Kendall and Edward C Carterette. Perceptual scaling of simultaneous wind instrument timbres. *Music Perception*, pages 369–404, 1991. [11](#)
- [66] Roger A Kendall, Edward C Carterette, and John M Hajda. Perceptual and acoustical features of natural and synthetic orchestral instrument tones. *Music Perception*, pages 327–363, 1999. [11](#)
- [67] Albert Kern. *A nonlinear biomorphic Hopf-Amplifier Model of the Cochlea*. PhD thesis, 2003. [81](#), [84](#), [89](#)

- [68] Daniel Kientzy. *Les sons multiples aux saxophones*. Editions Salabert, 1982. 37
- [69] B S Krishna and M N Semple. Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *Journal of neurophysiology*, 84: 255–273, 2000. ISSN 15221598. 109
- [70] Emery M Ku, Stephen J Elliott, and Ben Lineton. Limit cycle oscillations in a nonlinear state space model of the human cochlea. *The Journal of the Acoustical Society of America*, 126(2):739–750, August 2009. ISSN 00014966. doi: 10.1121/1.3158861. URL <http://www.ncbi.nlm.nih.gov/pubmed/19640040>. 83
- [71] S Lakatos. A common perceptual space for harmonic and percussive timbres. *Perception & psychophysics*, 62(7):1426–39, October 2000. ISSN 0031-5117. URL <http://www.ncbi.nlm.nih.gov/pubmed/11143454>. 11
- [72] Gerald Langner and Christoph E Schreiner. Periodicity coding in the inferior colliculus of the cat. i. neuronal mechanisms. *Journal of Neurophysiology*, 60(6):1799–1822, 1988. 23, 109
- [73] RF Lyon. Filter cascades as analogs of the cochlea. *Neuromorphic systems engineering*, (ID):3–19, 1998. URL http://link.springer.com/chapter/10.1007/978-0-585-28001-1_1. 17
- [74] Marcelo O. Magnasco. A Wave Traveling over a Hopf Instability Shapes the Cochlear Tuning Curve. *Physical Review Letters*, 90(5):1–4, February 2003. ISSN 0031-9007. doi: 10.1103/PhysRevLett.90.058101. URL <http://link.aps.org/doi/10.1103/PhysRevLett.90.058101>. 81
- [75] F Mammano and R Nobili. Biophysics of the cochlea: linear approximation. *The Journal of the Acoustical Society of America*, 93(June 1993):3320–3332, 1993. ISSN 00014966. doi: 10.1121/1.405716. 83, 89
- [76] Stephen McAdams. Segregation of concurrent sounds. i: Effects of frequency modulation coherence. *The Journal of the Acoustical Society of America*, 86(6):2148–2159, 1989. 22

- [77] Stephen McAdams. Musical timbre perception. In Diana Deutsch, editor, *The psychology of music*. Elsevier, 2013. 2, 9, 12
- [78] Stephen McAdams, Suzanne Winsberg, Donnadiou Sophie, Geert De Soete, and Jochen Krimphoff. Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological . . .*, pages 177–192, 1995. URL <http://link.springer.com/article/10.1007/BF00419633>. 7, 11, 12, 14, 120
- [79] Stephen McAdams, J W Beauchamp, and S Meneguzzi. Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. *The Journal of the Acoustical Society of America*, 105(2 Pt 1):882–97, February 1999. ISSN 0001-4966. URL <http://www.ncbi.nlm.nih.gov/pubmed/9972573>. 8, 10
- [80] Ray Meddis, Lowel P O’Mard, and Enrique A Lopez-Poveda. A computational algorithm for computing nonlinear auditory frequency selectivity. *The Journal of the Acoustical Society of America*, 109(6):2852–2861, 2001. 20, 83
- [81] James R Miller and Edward C Carterette. Perceptual space for musical structures. *The Journal of the Acoustical Society of America*, 58(3):711–720, 1975. 11
- [82] Andrzej Miśkiewicz, Andrzej Rakowski, and Teresa Rościszewska. Perceived roughness of two simultaneous pure tones. *Acta acustica united with acustica*, 92(2):331–336, 2006. 28, 106, 107
- [83] Brian C J Moore and Brian R. Glasberg. Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hearing research*, 28(2-3):209–25, January 1987. ISSN 0378-5955. URL <http://www.ncbi.nlm.nih.gov/pubmed/3654390>. 18, 19, 20
- [84] Brian CJ Moore. *An introduction to the psychology of hearing*. Brill, 2012. 17, 26
- [85] Jong-Hoon Nam. Microstructures in the organ of corti help outer hair cells form traveling waves along the cochlear coil. *Biophysical journal*, 106(11):2426–2433, 2014. 77

- [86] Stephen T Neely and D O Kim. A model for active elements in cochlear biomechanics. *The journal of the acoustical society of America*, pages 1472–1480, 1986. URL <http://link.aip.org/link/?JASMAN/79/1472/1>. 81, 89
- [87] Guangjian Ni, Stephen J Elliott, Mohammad Ayat, and Paul D Teal. Modelling cochlear mechanics. *BioMed research international*, 2014:150637, January 2014. ISSN 2314-6141. doi: 10.1155/2014/150637. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4130145&tool=pmcentrez&rendertype=abstract>. 84
- [88] RD Patterson, Ian Nimmo-Smith, John Holdsworth, and Peter Rice. An efficient auditory filterbank based on the gammatone function. In *a meeting of the IOC Speech Group on Auditory Modelling at RSRE*, volume 2, 1987. 20, 121
- [89] Roy D Patterson. Auditory filter shapes derived with noise stimuli. *The Journal of the Acoustical Society of America*, 59(3):640–654, 1976. 20
- [90] James O Pickles. *An introduction to the physiology of hearing*. Brill, 2012. 16, 17, 72, 74, 76
- [91] Reinier Plomp. Timbre as a multidimensional attribute of complex tones. *Frequency analysis and periodicity detection in hearing*, pages 397–414, 1970. 11
- [92] Reinier Plomp and Willem JM Levelt. Tonal consonance and critical bandwidth. *The journal of the Acoustical Society of America*, 38(4):548–560, 1965. 27, 28, 30
- [93] Daniel Pressnitzer. *Perception de rugosité psychoacoustique: D’un attribut élémentaire de l’audition à l’écoute musicale*. PhD thesis, Paris 6, 1998. 27, 28, 29
- [94] Daniel Pressnitzer, Stephen McAdams, Suzanne Winsberg, and Joshua Fineberg. Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception & psychophysics*, 62(1):66–80, 2000. 27
- [95] Martín Proscia. Acercamiento al saxofón multifónico. una perspectiva de estudio. *Revista del ISM*, 1(13):171–194, 2011. 37

- [96] Richard A Reale and Thomas J Imig. Tonotopic organization in auditory cortex of the cat. *Journal of Comparative Neurology*, 192(2):265–291, 1980. 109
- [97] Pablo E Riera, Martin Proscia, and Manuel C Eguia. A comparative study of saxophone multiphonics: Musical, psychophysical and spectral analysis. *Journal of New Music Research*, 43(2):202–213, 2014. 15, 48
- [98] Luis Robles and Mario A. Ruggero. Mechanics of the mammalian cochlea. *Physiological reviews*, 81(3):1305–52, July 2001. ISSN 0031-9333. URL <http://www.ncbi.nlm.nih.gov/pubmed/11427697>. 117
- [99] Gian Luca Romani, Samuel J Williamson, and Lloyd Kaufman. Tonotopic organization of the human auditory cortex. *Science*, 216(4552):1339–1340, 1982. 76
- [100] M. Rudnicki and W. Hemmert. Cochlea: inner ear models in python. 93
- [101] Mario A. Ruggero, S S Narayan, a N Temchin, and a Recio. Mechanical bases of frequency tuning and neural excitation at the base of the cochlea: comparison of basilar-membrane vibrations and auditory-nerve-fiber responses in chinchilla. *Proceedings of the National Academy of Sciences of the United States of America*, 97(22):11744–50, October 2000. ISSN 0027-8424. doi: 10.1073/pnas.97.22.11744. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=34344&tool=pmcentrez&rendertype=abstract>. 79, 81
- [102] Severine Samson, R J Zatorre, and JO Ramsay. Multidimensional scaling of synthetic musical timbre: Perception of spectral and temporal characteristics. *Canadian Journal of . . .*, pages 307–315, 1997. URL <http://psycnet.apa.org/journals/cep/51/4/307/>. 11, 12, 13
- [103] Pierre Schaeffer. *Traité des objets musicaux*. 1966. 3, 37
- [104] William A Sethares. Local consonance and the relationship between timbre and scale. *The Journal of the Acoustical Society of America*, 94(3):1218–1228, 1993. 27, 28, 42, 59, 106, 107, 108, 122

-
- [105] Evan C Smith and Michael S Lewicki. Efficient auditory coding. *Nature*, 439(7079): 978–982, 2006. 76
- [106] Michael Studdert-Kennedy, Alvin M Liberman, Katherine S Harris, and Franklin S Cooper. Motor theory of speech perception: A reply to lane’s critical review. 1970. 7
- [107] Christian J Sumner, Enrique A Lopez-Poveda, Lowell P O’Mard, and Ray Meddis. A revised model of the inner-hair cell and auditory-nerve complex. *The Journal of the Acoustical Society of America*, 111(5):2178–2188, 2002. 84
- [108] Qing Tan and Laurel H Carney. A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide. *The Journal of the Acoustical Society of America*, 114(4):2007, 2003. ISSN 00014966. doi: 10.1121/1.1608963. URL <http://scitation.aip.org/content/asa/journal/jasa/114/4/10.1121/1.1608963>. 75, 92
- [109] Damien Tardieu and Stephen McAdams. Perception of dyads of impulsive and sustained instrument sounds. *Music Perception*, 30(2):117–128, 2012. 11
- [110] Ernst Terhardt. On the perception of periodic sound fluctuations (roughness). *Acta Acustica united with Acustica*, 30(4):201–213, 1974. 30, 31
- [111] Amos Tversky. Features of similarity. *Psychological Review*, 84(4):327–352, 1977. ISSN 1939-1471(Electronic);0033-295X(Print). doi: 10.1037/0033-295X.84.4.327. 3
- [112] Kanzo Ueda. A hierarchical structure for adjectives describing timbre. *The Journal of the Acoustical Society of America*, 100(4):2751–2751, 1996. 5
- [113] Panteleimon Nestor Vassilakis. *Perceptual and physical properties of amplitude fluctuation and their musical significance*. PhD thesis, UNIVERSITY OF CALIFORNIA Los Angeles, 2001. 28
- [114] Pantelis N Vassilakis. Auditory roughness as means of musical expression. *Selected Reports in Ethnomusicology*, 12:119–144, 2005. 122

- [115] Sarah Verhulst, Torsten Dau, and Christopher a Shera. Nonlinear time-domain cochlear model for transient stimulation and human otoacoustic emission. *The Journal of the Acoustical Society of America*, 132(6):3842–8, December 2012. ISSN 1520-8524. doi: 10.1121/1.4763989. URL <http://www.ncbi.nlm.nih.gov/pubmed/23231114>. 83
- [116] Erich M von Hornbostel and Curt Sachs. Classification of musical instruments: Translated from the original german by anthony baines and klaus p. wachsmann. *The Galpin Society Journal*, pages 3–29, 1961. 4
- [117] Marcus Weiss and Giorgio Netti. *The techniques of saxophone playing*. Bärenreiter, 2010. 37
- [118] Bo Wen. *MODELING THE NONLINEAR ACTIVE COCHLEA: MATHEMATICS AND ANALOG VLSI*. PhD thesis, 2006. 89, 90
- [119] Bo Wen and Kwabena Boahen Kwabena Boahen. A linear cochlear model with active bi-directional coupling. *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*, 3:1–5, 2003. ISSN 1094-687X. doi: 10.1109/IEMBS.2003.1280129. 81, 83
- [120] David L Wessel. Timbre space as a musical control structure. *Computer music journal*, pages 45–52, 1979. 7, 11
- [121] LA Westerman and RL Smith. A diffusion model of the transient response of the cochlear inner hair cell synapse. *The Journal of the Acoustical Society of . . .*, 83(June 1988):2266–2276, 1988. URL <http://scitation.aip.org/content/asa/journal/jasa/83/6/10.1121/1.396357>. 93
- [122] Yong-Jin Yoon, Charles R Steele, and Sunil Puria. Feed-forward and feed-backward amplification model from cochlear cytoarchitecture: an interspecies comparison. *Biophysical journal*, 100(1):1–10, 2011. 82, 94
- [123] Xin Zhang and Zbigniew W Ras. Analysis of sound features for music timbre re-

- cognition. In *Multimedia and Ubiquitous Engineering, 2007. MUE'07. International Conference on*, pages 3–8. IEEE, 2007. 8
- [124] Muhammad S A Zilany, Ian C Bruce, Paul C Nelson, and Laurel H Carney. A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. *The Journal of the Acoustical Society of America*, 126(5):2390–412, November 2009. ISSN 1520-8524. doi: 10.1121/1.3238250. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2787068&tool=pmcentrez&rendertype=abstract>. 84, 92, 93, 113
- [125] Muhammad SA Zilany, Ian C Bruce, and Laurel H Carney. Updated parameters and expanded simulation options for a model of the auditory periphery. *The Journal of the Acoustical Society of America*, 135(1):283–286, 2014. 84
- [126] G Zweig. Finding the impedance of the organ of Corti. *The Journal of the Acoustical Society of America*, 89(3):1229–54, March 1991. ISSN 0001-4966. URL <http://www.ncbi.nlm.nih.gov/pubmed/2030212>. 81
- [127] Eberhard Zwicker and Ernst Terhardt. Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *The Journal of the Acoustical Society of America*, 68(5):1523–1525, 1980. 28

Agradecimientos

Agradecimientos

El primero de los agradecimientos corresponde a Manuel Eguía, el director de esta tesis, ya que me ha alentado a seguir este apasionante camino vinculando la ciencia y la música. Al mismo tiempo es responsable de haber fundado el ((LAPSo)), un espacio único en estas coordenadas, para la investigación en temas de música, acústica y neurociencia.

Recuerdo uno de los primero encuentros hace tiempo en el 2007, cuando Manuel me invitó a hablar sobre un proyecto que relacionaba disonancia musical y física cuántica (*Helmholtz vs Schrödinger*). Luego, otro gran hito en el 2011, con los llamadores de ángeles gigantes y aquella gran aventura en el puente.

Contar con el LAPSo para trabajar, experimentar y sobre todo participar desde el comienzo en su formación, con todo lo que eso implica, fue muy enriquecedor. Gracias Manuel.

También quiero agradecer a todas las personas detrás de las instituciones que tuvieron relación con este trabajo. En la Universidad Nacional de Quilmes a los docentes e integrantes de la Carrera de Música y Tecnología. En la Universidad de Buenos Aires a los docentes del Departamento de Física. Y por supuesto al Estado Argentino, al CONICET y a la educación pública.

Un agradecimiento muy especial a las personas que me ayudaron con correcciones sobre el manuscrito de esta tesis, Hernán Kerlleñevich, Alma Laprida, Ignacio Spiouzas. Sin la ayuda de ellos hubiese sido muy difícil.

Uno de los agradecimientos más importante es para mi madre Silvia, por todo lo in-

commensurable que ha hecho siempre para que las cosas funcionen y yo este donde estoy. Gracias a toda mi familia, mis abuelos y abuelas, tíos y tías, primos y primas, por toda la energía incondicional que me han dado.

Agradecimientos mayúsculos a Alma que me ha alimentado y curado durante el proceso de escritura de esta tesis.

Durante los años de la tesis tuve el gusto de trabajar en proyectos en conjunto con diversas personas que han influenciado de una u otra forma en mi trabajo. Gracias a Hernán Kerlleñevich por las interminables horas de felicidad de música neuronal con *pure data* y por su inspiración musical. Gracias a Martín Proscia por su compañerismo *académico* en el estudio de la acústica del saxofón y los multifónicos. Gracias a Marcos Trevisán por sumarme en el trabajo sobre los imitadores de voz, fue una muy buena experiencia. Gracias a Rafael Grimson por su generosidad y su genial forma de laburar. Gracias a Mercedes y Marina por sus espíritus e invitarme a participar de tantos proyectos musicales.

No pueden faltar los compañeros del laboratorio, las cataratas de humor y metegol. Gracias por la compañía a Ramiro Vergara, Ignacio Spiouzas, Esteban Calcagno, Ezequiel Abegú, Pablo Etchemendy y Alejo Alberti.

Y gracias a todos los amigos que siempre van a estar.