

Gustavo Basso

*Percepción*  
*auditiva*



Universidad  
Nacional  
de Quilmes  
Editorial

Gustavo Basso

# Percepción auditiva



Bernal, 2018

UNIVERSIDAD NACIONAL DE QUILMES

Rector  
Alejandro Villar

Vicerrector  
Alfredo Alfonso

Colección Música y ciencia  
Dirigida por Pablo Di Liscia

Basso, Gustavo

Percepción auditiva. - 1a ed. - Bernal: Universidad Nacional de Quilmes, 2017.

Libro digital, EPUB - (Música y ciencia / Pablo Di Liscia)

Archivo Digital: descarga

ISBN 978-987-558-413-6

1. Percepción Auditiva. 2. Articulación de Sonidos. 3. Música. I. Título.

CDD 780.77

Primera edición en papel, 2006

Primera edición e-book, 2018

© Gustavo Basso, 2006

© Universidad Nacional de Quilmes, 2006

Universidad Nacional de Quilmes

Roque Sáenz Peña 352

(B1876BXD) Bernal, Provincia de Buenos Aires

República Argentina

editorial.unq.edu.ar

editorial@unq.edu.ar

ISBN: 978-987-558-413-6 (e-book)

Queda hecho el depósito que marca la ley 11.723

*Hecho en Argentina*

# Índice

[Presentación, por Oscar Pablo Di Liscia](#)

[Prólogo](#)

[Capítulo I. Sistema auditivo periférico y fisiología de la audición](#)

[Capítulo II. Sonoridad](#)

[Capítulo III. El oído como analizador espectral](#)

[Capítulo IV. Procesamiento temporal de señales](#)

[Capítulo V. Percepción de la altura tonal](#)

[Capítulo VI. Timbre](#)

[Capítulo VII. Percepción auditiva del espacio](#)

[Capítulo VIII. Percepción de objetos auditivos](#)

[Capítulo IX. Percepción del habla](#)

[Apéndice I. Teorema de Fourier](#)

[Apéndice II. Principio acústico de indeterminación](#)

[Referencias bibliográficas](#)

*A Germán y a Florencia*

# Presentación

El arte sonoro, las modalidades y géneros de comunicación que involucran sonido, las técnicas y tecnologías de proceso de sonido y música y la investigación sobre percepción auditiva, se vienen desarrollando de manera interdependiente y significativa desde el siglo pasado. Cada una de las áreas mencionadas ha impregnado de dudas y certezas a las otras, al participar de un formidable proceso interactivo que continúa acelerándose en la actualidad.

Este libro presenta una exploración de los avances en el campo de la percepción auditiva. Su título tiene el grado de determinación y de generalidad que requieren tanto el dinamismo como la variedad del campo de estudio involucrado. Por otra parte, no puede negarse que su autor ha sido fiel a su triple condición de músico, científico y docente, al generar un texto que expone de manera muy didáctica y eficaz los resultados más significativos de la investigación en su área, en una constante realimentación con la reflexión sobre el arte sonoro.

Una de las cuestiones medulares de la investigación actual en esta área es la determinación del ámbito (físico-fisiológico) en el que se llevan a cabo los procesos de decodificación de la señal acústica. En otras palabras: ¿es el fabuloso diseño de nuestro oído interno y externo, o son los procesos neurales de alto nivel los responsables de que percibamos determinados eventos sonoros de determinada manera? ¿Y si son ambos, cómo se reparte la tarea y cómo se explica y modeliza su funcionamiento? Así es que, por ejemplo, algunos autores (véase Bregman, 1994) hacen referencia a procesos neurofisiológicos de “alto nivel” o de “bajo nivel” como responsables de determinadas tareas de decodificación y otros (Malham, 2000), coincidentemente, denominan “combinación oído-cerebro” al sistema auditivo en general. Y por citar un solo ejemplo de este trabajo, Basso escribe en el capítulo VI: “[...] la constancia del

timbre sólo puede ser explicada en función de complicados procesos en las etapas superiores del tratamiento neural”.

Los avances en esta dirección han sido tensados por la necesidad de involucrar dentro del estudio de la psicoacústica a secuencias sonoras complejas, que van más allá de los eventos individuales. Se involucran así, procesos de eventos sonoros de muy variada naturaleza y complejidad, como la percepción de secuencias de habla (capítulo IX), evaluación de recintos y localización de fuentes sonoras en movimiento (capítulo V), segregación y fusión de estratos sonoros simultáneos y sucesivos (capítulo VIII), o, finalmente, los estudios sobre el desplazamiento entre el tiempo físico y el fenoménico (capítulo IV).

Una de las dificultades que hubo de enfrentar el autor de este libro es la secuenciación de los contenidos, dada su interdependencia. Es así que, por ejemplo, en el capítulo referido a la “sonoridad” (capítulo II) se pregunta: “Pero, ¿es legítimo considerar a la sonoridad como un rasgo autónomo, que se puede estudiar de manera independiente, separada del resto de las características que definen un sonido real?”

Y qué decir del capítulo dedicado al “timbre” (capítulo VI), una de las nociones más interdependientes y cambiantes en toda la historia del sonido y la música. En su tratamiento, el autor pone en evidencia, además, el rol de los aportes de las técnicas y tecnologías de sonido propias de la electroacústica: “Los modelos de análisis por síntesis han demostrado ser eficaces para generar conocimiento científico y técnico, reducir la cantidad de información necesaria para expresar una señal acústica, producir variantes y modificaciones nunca experimentadas y controlar la prosodia musical en tiempo real. [...] Una descripción detallada de las técnicas empleadas contendría la historia de la electroacústica durante los últimos 60 años.”

Por todo lo antedicho, este libro constituye un trabajo valioso para explorar antes que para simplemente “leer”, para que el lector

realice múltiples recorridos en los que puede y debe encontrar variados puntos de entrada-salida y de relación, a despecho de la necesaria secuencia que impone el texto escrito.

Oscar Pablo Di Liscia

# Prólogo

Este libro fue concebido inicialmente como una publicación interna para un seminario de posgrado que se dictó en el año 2000 en la Universidad Nacional de Quilmes. Ese origen explica, en parte, el contenido y la secuencia de los capítulos. Su objetivo principal es el de exponer los elementos, principios y procesos que intervienen en la percepción del sonido. Nuestra intención no es presentar los temas como si formasen parte de un conjunto de conocimientos definitivos, sino integrar las dificultades propias de una disciplina en la que los mismos datos experimentales llevan muchas veces a la creación de modelos explicativos totalmente diferentes.

Quizá la confusión más habitual en cualquier texto sobre acústica tiene origen en la ambigüedad de la definición de sonido. Según el autor o el texto que tratemos, el término sonido se puede referir a un fenómeno físico –como una vibración o una onda en el espacio– o a un fenómeno psicológico. Aquí llamaremos sonido a la representación mental asociada a la percepción, y señal acústica al estímulo físico que lo origina. La psicoacústica tradicional estudia directamente la relación entre las señales acústicas de entrada y los sonidos, colocando los procesos intermedios en una especie de caja negra cuyo mecanismo no considera de su incumbencia. La fisiología de la percepción y la neurobiología, por su parte, se interesan especialmente por los procesos de transducción que operan entre la señal acústica y la representación mental asociada. En nuestro caso predomina la primera aproximación, pero no evitaremos mencionar los aspectos fisiológicos involucrados cuando lo creamos necesario. El primer capítulo, por ejemplo, está dedicado por completo a la fisiología de la audición.

Este texto está concebido como parte de un programa completo de acústica destinado a músicos e investigadores en temas relacionados con el sonido. Aunque los aportes de la psicología y de las

ciencias médicas resultan esenciales, nuestra aproximación al tema adopta decididamente la perspectiva específica de la ciencia acústica. Si bien se supone que el lector está familiarizado con los conceptos básicos de la acústica general, en la bibliografía al final del libro se mencionan algunas referencias para quien necesite refrescar alguno de ellos y en los apéndices se describen brevemente dos temas que no siempre aparecen en los textos de acústica habituales: el teorema de Fourier y el principio de indeterminación. Toda vez que fue posible se emplearon los términos técnicos en español y en algunos casos las expresiones se presentan también en el idioma de origen para facilitar la consulta en la bibliografía especializada.

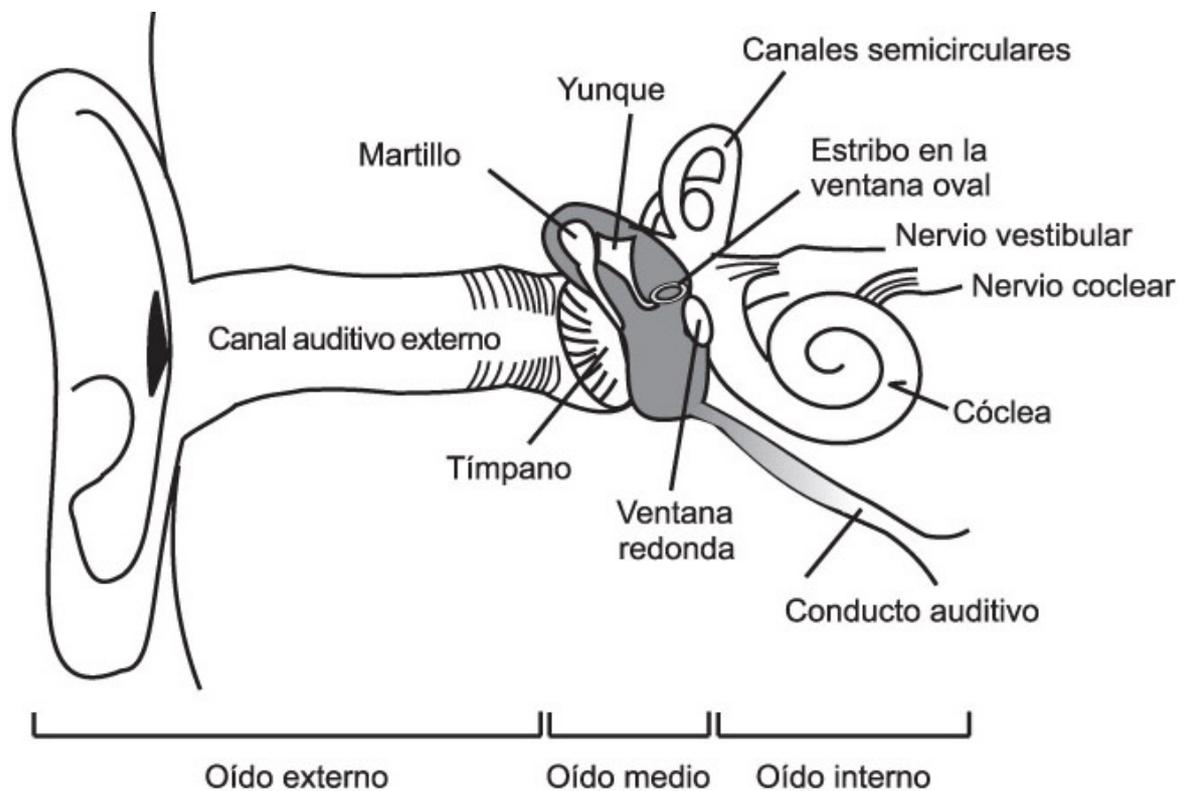
Mi más sincero agradecimiento al profesor Pablo Di Liscia –director de la colección Música y Ciencia– por el apoyo que le brindó al proyecto desde su inicio, al profesor Juan Pampin y al profesor Emanuel Bonnier –responsable de la realización de los gráficos. Mi gratitud a los integrantes del equipo de investigación en acústica de la Facultad de Bellas Artes de la Universidad Nacional de La Plata. Y a los colegas y alumnos que, directa o indirectamente, participaron en las charlas y en los innumerables intercambios de correo sobre percepción auditiva de los últimos años. Sin su estímulo este libro no hubiera sido posible.

La Plata, mayo de 2006

# Capítulo I

## Sistema auditivo periférico y fisiología de la audición

El oído –o sistema auditivo periférico– comienza en el pabellón auricular y llega hasta la cóclea. Su misión es convertir las vibraciones mecánicas en impulsos nerviosos para que sean procesados en el cerebro. Es muy pequeño y delicado y se halla protegido en el interior del hueso temporal. En la figura 1 se lo representa esquemáticamente.



**Figura 1.** Sistema auditivo periférico

Para su estudio se lo divide en tres partes: oído externo, oído medio y oído interno. Los canales semicirculares, que comparten algunas estructuras anatómicas con el oído interno, pertenecen en realidad al sentido del equilibrio.

En este capítulo nos detendremos extensamente en los procesos involucrados en la fisiología de la audición, cuyo objeto de estudio incluye la totalidad de los eventos físicos y neurales involucrados en la percepción del sonido, desde el sistema auditivo periférico hasta la corteza cerebral.[\[1\]](#)

En lo que sigue analizaremos anatómica y fisiológicamente cada una de las partes del sistema auditivo periférico.

## 1. Oído externo

El oído externo, junto a la cabeza y la parte superior del torso, forma parte del sistema receptor de vibraciones que acopla el tímpano con el campo acústico externo. Además de protegerlo ante posibles agresiones mecánicas, filtra y modifica las señales que le llegan del exterior y contribuye decisivamente en la tarea de localización espacial de las fuentes acústicas.

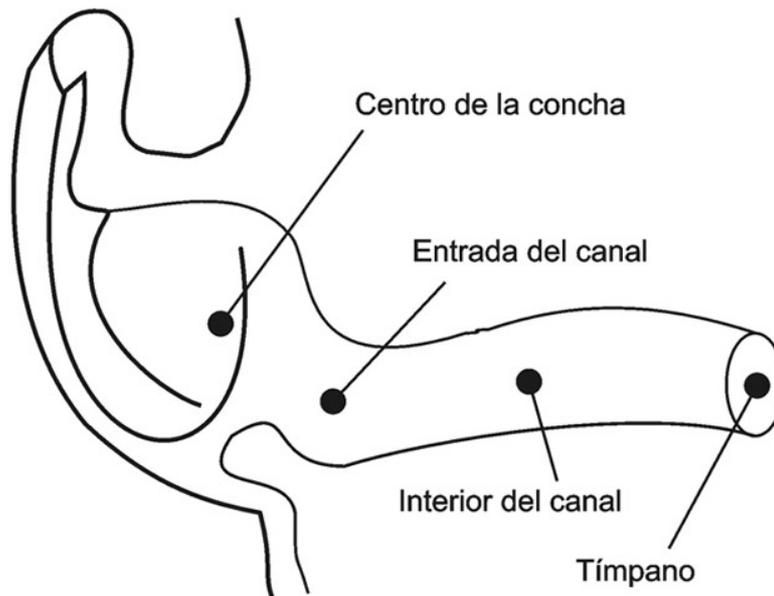
Una descripción anatómica del oído externo, de adentro hacia afuera, incluye el canal auditivo externo, la concha y el pabellón auricular. El canal auditivo externo es una estructura tubular irregular de aproximadamente 26 mm de largo por 7 mm de ancho y un volumen de 1 cm<sup>3</sup>, con un eje central sinuoso, que conecta el pabellón con el tímpano. Su pared está conformada por una porción cartilaginosa (móvil) y una ósea. La parte cartilaginosa, continuación de la estructura del pabellón auricular, es de mayor longitud que la ósea y está cubierta por piel con folículos pilosos, glándulas ceruminosas y sebáceas. En cambio, la piel que recubre la región ósea del canal auditivo externo, en la parte más profunda adherida al tímpano, es muy delgada.

La parte externa del canal auditivo externo se ensancha notablemente formando una cavidad en forma de taza llamada concha, con un volumen interno de aproximadamente 4 cm<sup>3</sup>. En la figura 2 se pueden apreciar una vista lateral y un corte del oído externo.

El pabellón auricular (la oreja) es una estructura cartilaginosa que actúa como un filtro direccional de las señales que llegan del exterior. Es muy diferente al de otros mamíferos, por ejemplo el de los felinos, que poseen pabellones cónicos que pueden enfocar hacia un punto del espacio de manera independiente por medio de músculos de control específicos.

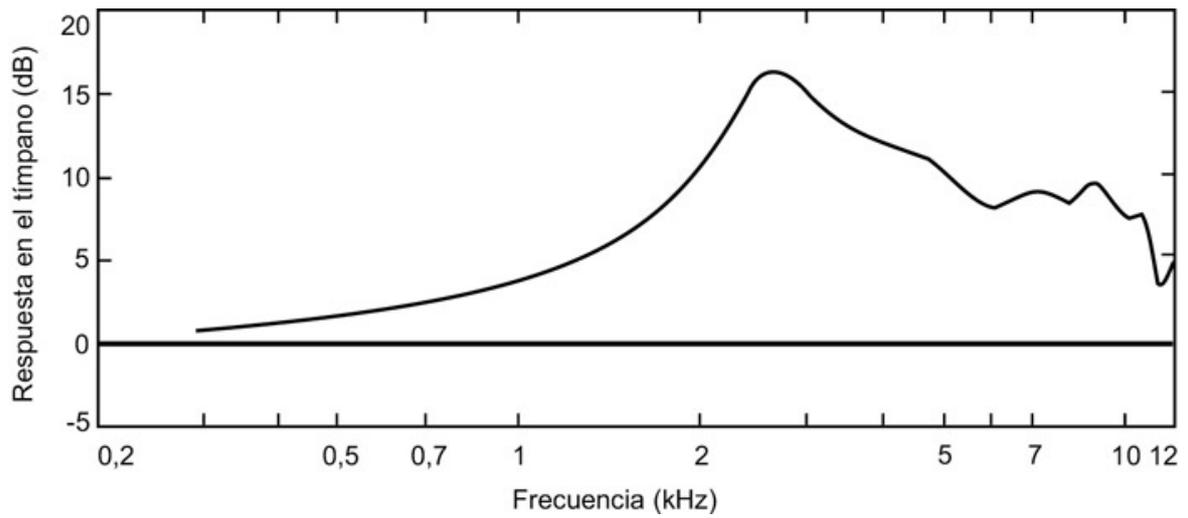
Los efectos del pabellón auricular (6 cm), de la cabeza (22 cm entre oídos) y de los hombros (15 cm de ancho) se vuelven significativos cuando las longitudes de onda son comparables con la dimensión de las estructuras. El sistema completo puede ser caracterizado comparando –o midiendo– la presión acústica en el campo libre, a la entrada del pabellón auricular, con la presión acústica en la superficie del tímpano. En la figura 3 se aprecia la acción de filtro espectral que produce el oído externo. La transferencia máxima, a una frecuencia de aproximadamente 2.800 Hz, se debe a la resonancia del canal auditivo externo.



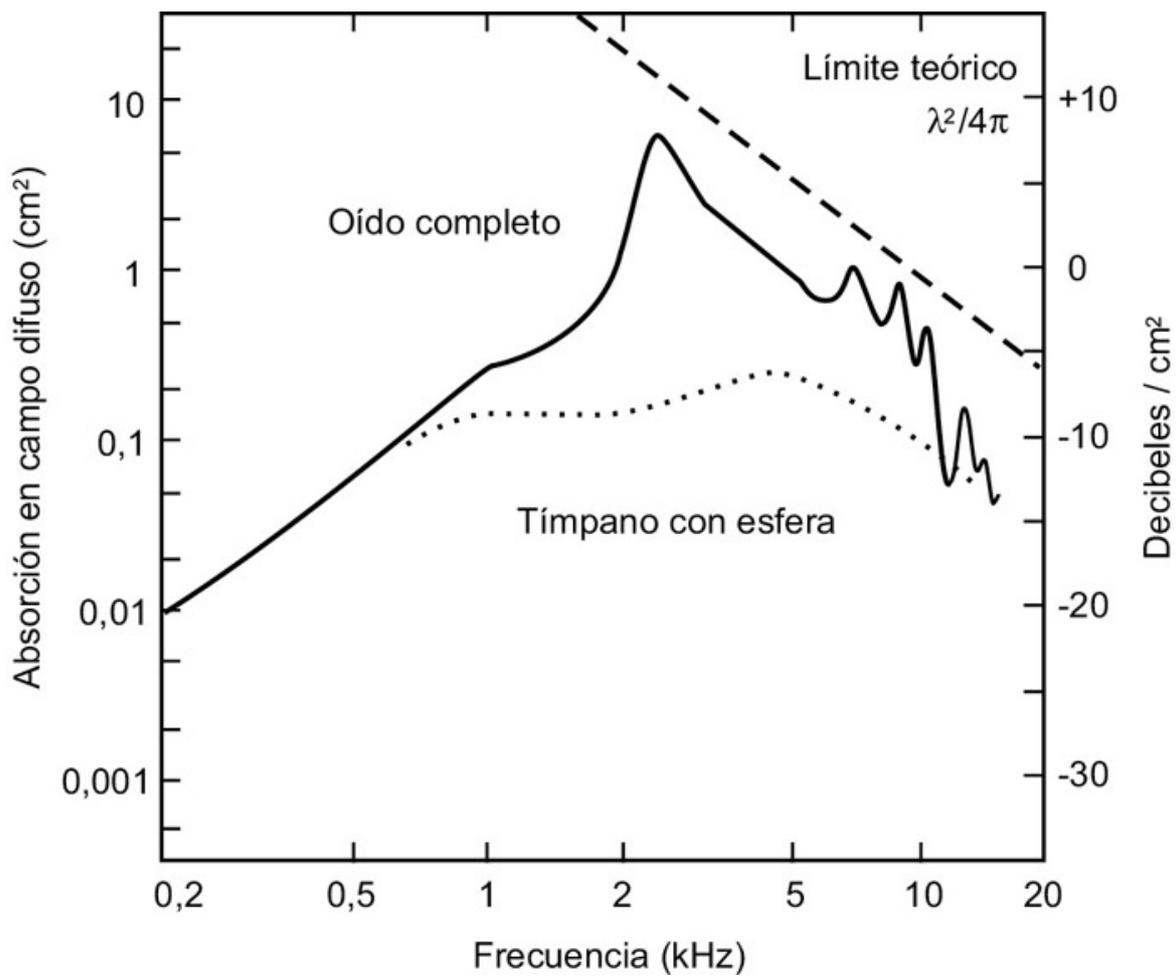


**Figura 2.** Vista lateral y corte esquemático superior del oído externo

La compleja forma de la curva se debe a la concurrencia de los fenómenos de reflexión, difracción y resonancia en diferentes partes del oído externo. Como curiosidad, en la figura 4 se reproduce el experimento de Goode, en el que se compara la transferencia de un oído normal con un modelo en el que el tímpano queda a ras de una esfera, como ocurre en el caso de algunos saurios (Goode, R. L. *et al.*, 1994): es evidente la compleja acción espectral del oído externo sobre las señales que le llegan del exterior. En el capítulo sobre percepción auditiva del espacio nos detendremos a estudiar con detalle este efecto y sus consecuencias.



**Figura 3.** Función de transferencia del oído externo ante un estímulo difuso de ruido blanco



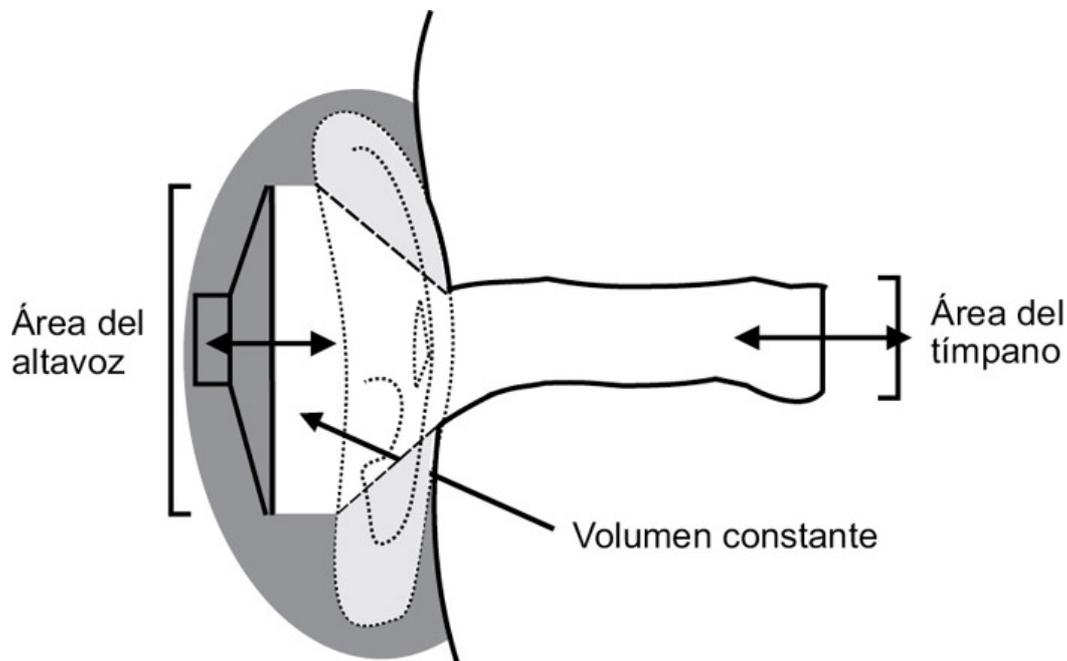
**Figura 4.** Comparación entre las funciones de transferencia de un oído completo y de un tímpano adosado a una esfera

## Audición con auriculares

Cuando los estímulos se presentan a través de auriculares las características del campo libre son reemplazadas por la interacción entre el auricular y el oído externo del oyente. Esta interacción incluye efectos propios de la mecánica de ondas que son muy difíciles de controlar, especialmente a altas frecuencias. Un método que permite acoplar con precisión el sistema consiste en calibrar los auriculares en su posición de uso en un oído real. Primero se mide, con un micrófono muy pequeño colocado dentro del canal auditivo, la presión sonora provocada por una fuente acústica de espectro plano alejada de la cabeza del oyente (campo libre). La señal en el micrófono corresponde exactamente a la *función de transferencia* – o HRTF– de ese oído externo en particular.<sup>[2]</sup> La curva de transferencia obtenida se reproduce en el oído ecualizando la salida del auricular en posición de uso. De este modo se puede lograr la misma respuesta espectral que en el campo libre para cada pareja auricular-oído. Como el procedimiento es complicado, se ofrecen cientos de curvas HRTF –que se pueden obtener, por ejemplo, en internet– que cada usuario escoge de acuerdo a su tipología particular.

Una de las características más notables de la audición a través de auriculares es la excelente respuesta en bajas frecuencias, imposible de conseguir con los mismos auriculares ubicados en el campo libre, lejos de la cabeza del oyente. Este efecto se explica por la diferencia en el acople acústico entre ambas situaciones. Con la fuente en el campo libre actúa el filtro del oído externo en toda su capacidad: la energía de baja frecuencia llega atenuada al tímpano pues difracta alrededor de la cabeza. Cuando se coloca un auricular se cierra la comunicación con el campo libre y se crea una cavidad estanca compuesta por el volumen interno del auricular más el

volumen interno del oído externo. En la figura 5 se puede ver este caso.



**Figura 5.** Acople entre un auricular y el oído externo

El volumen es ahora constante y posee únicamente dos superficies móviles, el altavoz del auricular y el tímpano. Cualquier desplazamiento neto del altavoz se va a transmitir íntegramente al tímpano amplificado por la relación entre sus áreas, 4:1 a favor del altavoz en un caso promedio. Dicha transmisión es independiente de la frecuencia, pues actúa como una prensa neumática y explica la notable eficiencia en bajas frecuencias del sistema. Lamentablemente, el dispositivo de ajuste estanco que permite tan buen rendimiento en graves es uno de los factores que convierte a los auriculares en causa potencial de hipoacusia por ruido, sobre todo si se los usa incorrectamente.[\[3\]](#)

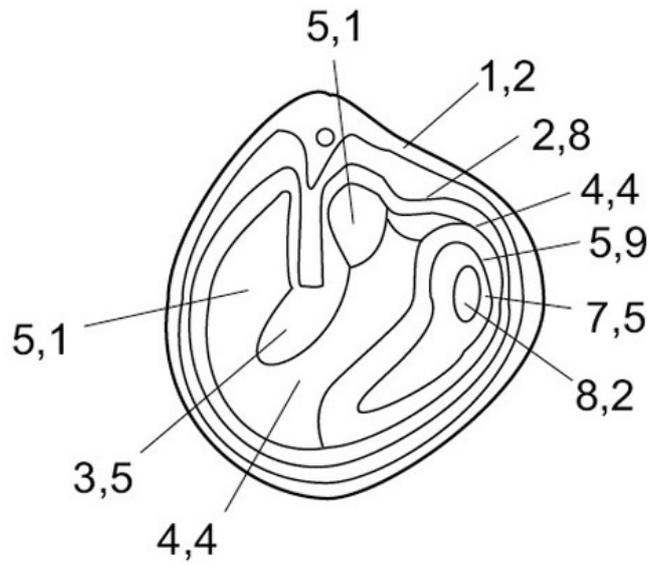
## 2. Oído medio

El oído medio permite acoplar las señales acústicas entre el canal auditivo y el oído interno. Ocupa una cavidad en el hueso tempo-

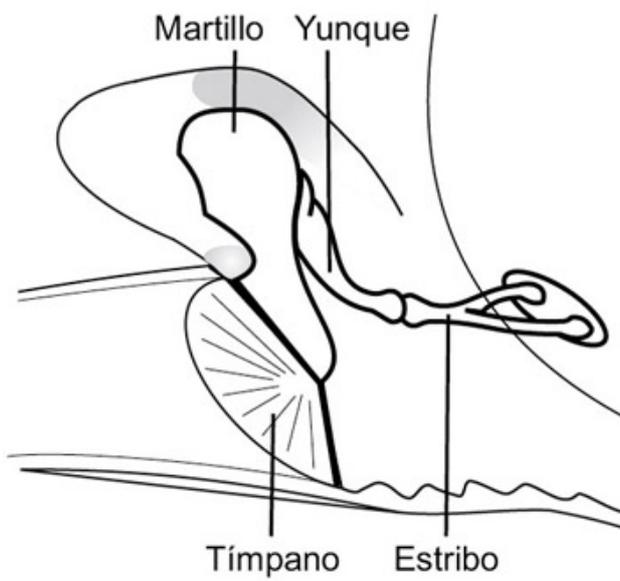
ral denominada caja timpánica y está constituido por el tímpano, la cadena de huesecillos –martillo, yunque y estribo– con sus ligamentos y músculos de control, la trompa de Eustaquio y las ventanas oval y redonda, que lo comunican con el oído interno.[\[4\]](#) Su principal función es la de permitir que las vibraciones acústicas lleguen a la cóclea con suficiente energía. Si las señales llegasen directamente al oído interno a través de la ventana oval, casi toda la energía se reflejaría a causa de la diferencia entre las propiedades mecánicas del medio aéreo exterior y del medio líquido de la cóclea. En los vertebrados terrestres el oído medio actúa como un sistema de acople que adapta ambos medios, permitiendo una efectiva transferencia de energía. Técnicamente, es un adaptador de impedancias y como tal lo analizaremos más adelante.

## **Tímpano**

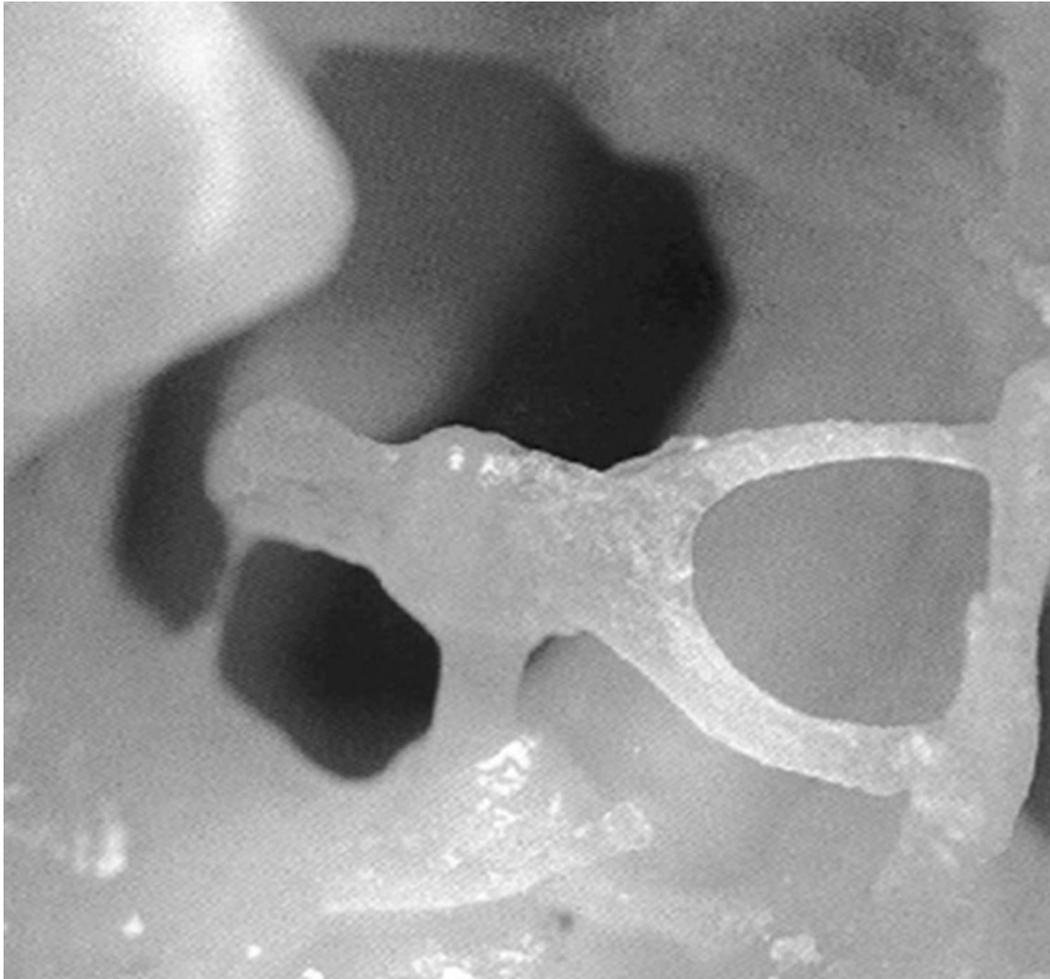
El tímpano es una membrana semitransparente con forma de cono con una superficie promedio de  $0,6 \text{ cm}^2$ . A causa de su geometría irregular y al anclaje solidario del manubrio del martillo, el tímpano vibra como una compleja membrana elástica no homogénea. En la figura 6 se aprecia su respuesta ante un estímulo sinusoidal de 525 Hz. Se destaca nítidamente la forma del área de contacto con el manubrio del martillo.



**Figura 6.** Vibración relativa del tímpano ante una excitación de 525 Hz y 121 dB



**Figura 7.** Cadena de huesecillos del oído medio



**Figura 8.** Fotografía del estribo

### **Cadena de huesecillos**

Las vibraciones del tímpano se transmiten hacia el oído interno a través de tres pequeños huesecillos: el martillo, el yunque y el estribo. El martillo posee una parte prominente, el manubrio, que se inserta entre las láminas de la membrana timpánica. A su vez, el martillo y el yunque están conectados entre sí y a la pared ósea por medio de ligamentos. El yunque se vincula con el estribo a través de otro ligamento, y el pie del estribo se conecta por medio de un ligamento anular a la ventana oval.

Dos músculos contribuyen a sostener los huesecillos y permiten controlar la rigidez de la cadena osicular. El tensor del tímpano es-

tá ligado al martillo y el estapedial al estribo. Tal como veremos, estos músculos funcionan como un control de ganancia variable del oído medio.

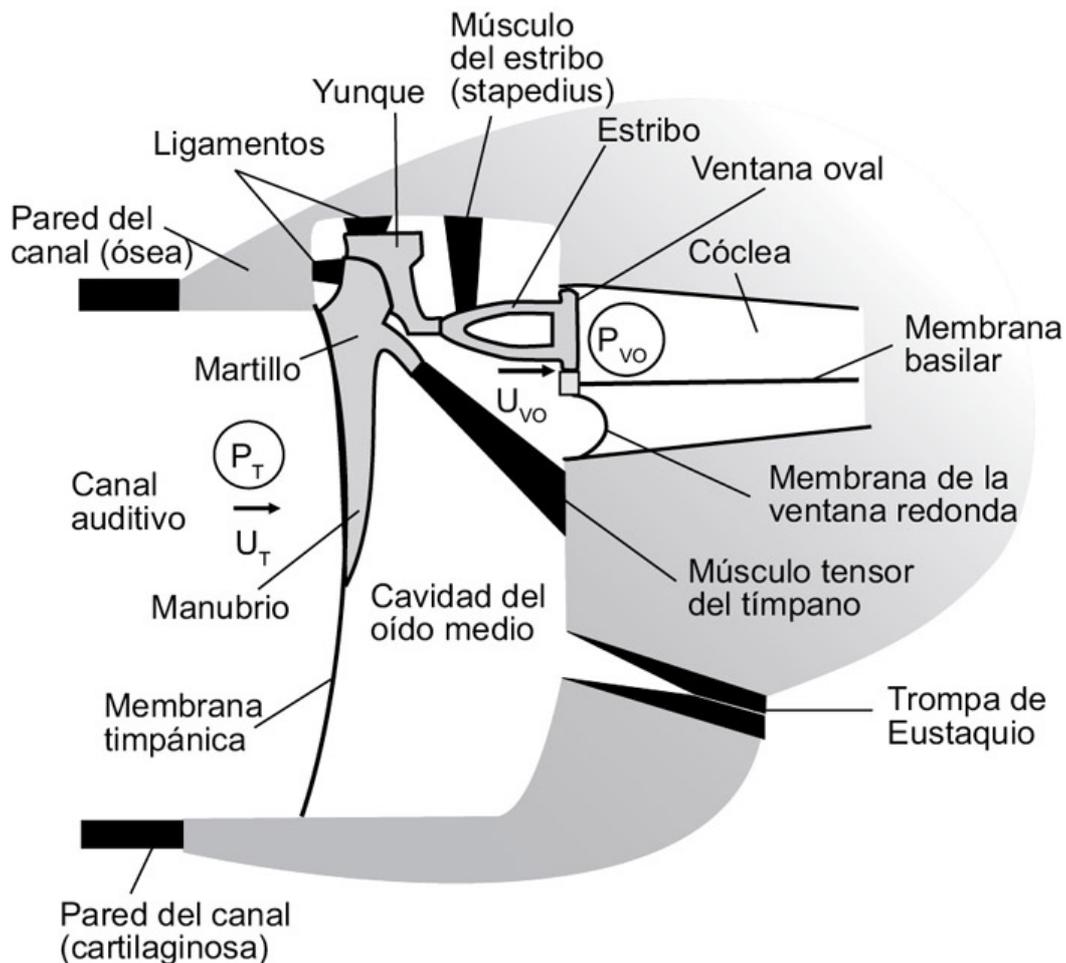
## **Cavidad timpánica y trompa de Eustaquio**

La trompa de Eustaquio, que conecta el oído medio con la cavidad nasofaríngea, es un conducto de paredes blandas que está cerrado la mayor parte del tiempo y que se abre al tragar, masticar o bostezar. Su función es la de permitir la ecualización de la presión estática a ambos lados del tímpano. Gracias a este mecanismo el aire en el interior ejerce la misma presión que la atmósfera desde el exterior. Se resuelve así la paradoja que inquiere por qué la presión atmosférica, de un valor superior a los 100 kPa, no se oye mientras que una señal acústica de sólo 20 Pa nos resulta de una intensidad intolerable. La respuesta es sencilla: la presión atmosférica, prácticamente estática, no mueve al tímpano pues se ejerce con igual valor en sus dos caras gracias a la acción de la trompa de Eustaquio. La presión dinámica, que vibra a frecuencias superiores a 16 Hz, se aplica con facilidad sobre la cara externa del tímpano pero llega con dificultad a la cara interna, aun con la trompa de Eustaquio abierta. La diferencia de presiones dinámicas se traduce en un movimiento efectivo del tímpano y de la cadena de huesecillos.

## **Función acústica del oído medio**

La función principal del oído medio es transmitir la mayor cantidad de energía desde el canal auditivo externo hasta la cóclea. Para lograr este fin la cadena de huesecillos debe adaptar la impedancia del sistema *tímpano-aire* con la del sistema *ventana oval-perilinfá*. Como ya señalamos, si las ondas acústicas pasasen del tímpano a la ventana oval, casi toda la energía se reflejaría por la gran diferencia de impedancia existente entre los dos medios de

propagación. La *impedancia* es una magnitud que permite evaluar la resistencia al movimiento de un sistema. En nuestro caso se verifica que la impedancia acústica en el tímpano es mucho menor que en el oído interno.

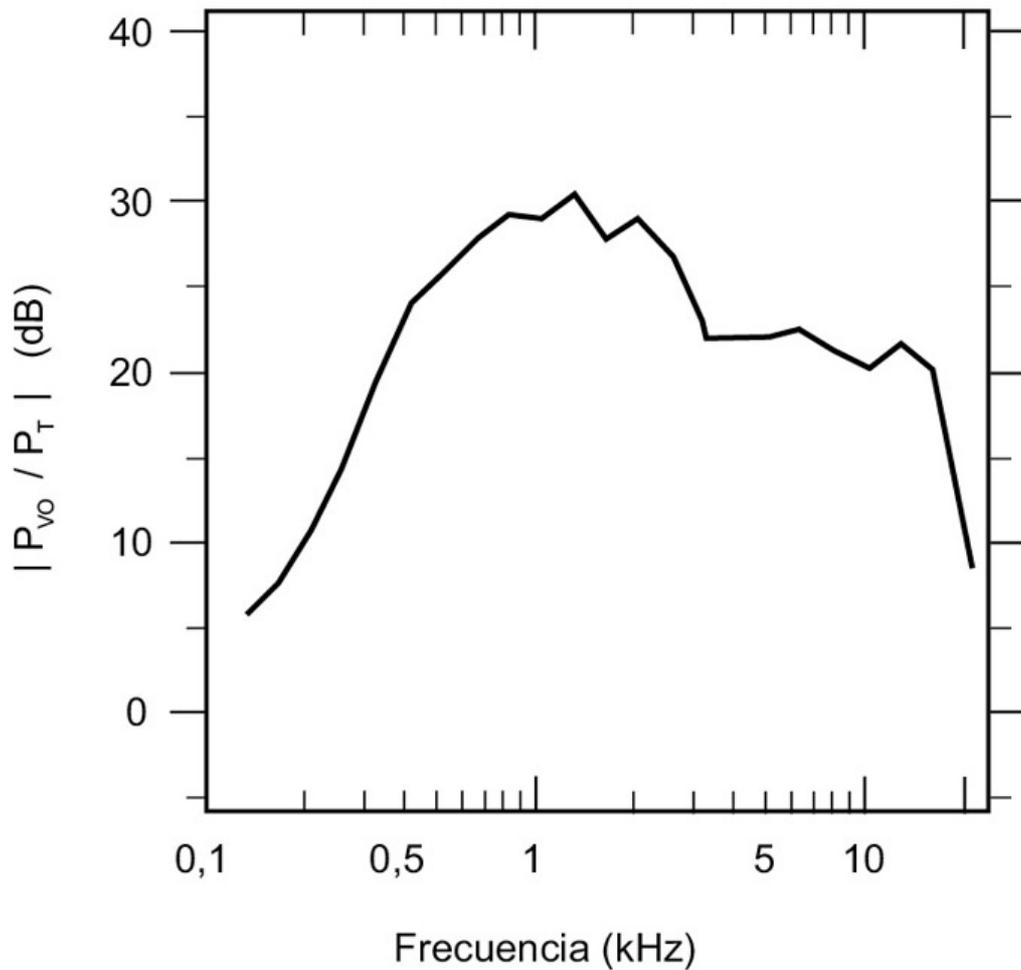


**Figura 9.** Esquema del sistema de transmisión del oído medio

Veamos el proceso en detalle. Con la excepción de las frecuencias más altas, por encima de 5 kHz, la dimensión transversal del canal auditivo externo es pequeña frente a las longitudes de onda y, por lo tanto, la presión sonora resulta aproximadamente constante sobre el área del tímpano. En este caso la interacción acústica puede describirse en términos de dos variables: la presión sonora en el tímpano  $p_T$  y su velocidad de volumen  $U_T$ . El movimiento del tím-

pano se transmite a través de la cadena de huesecillos a la ventana oval. Aquí también la interacción acústica puede describirse a partir de la presión sonora en la ventana oval  $p_{VO}$  y su velocidad de volumen  $U_{VO}$ . En la figura 9 se ve un esquema muy simplificado del sistema de transmisión.

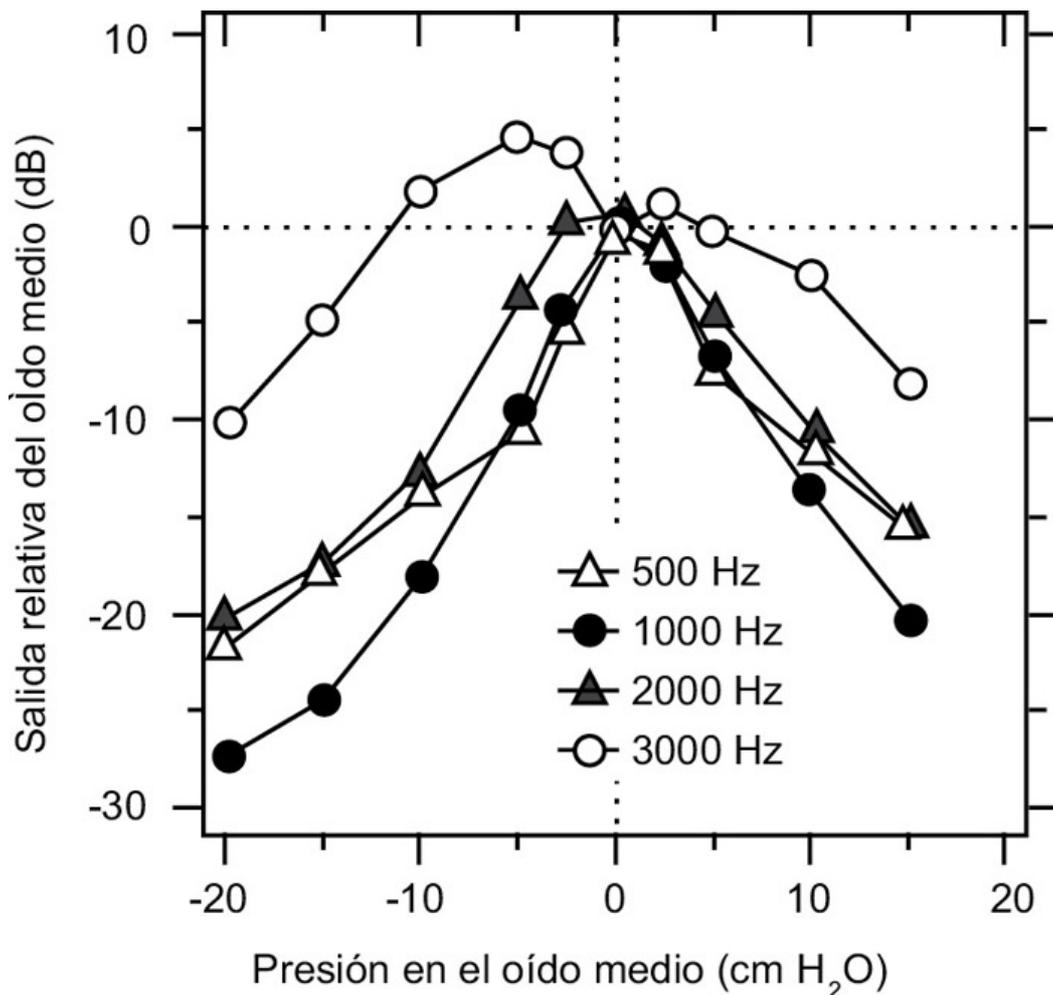
En una primera aproximación, la cadena de huesecillos actúa como un sistema de palancas. Para vibraciones de amplitud moderada, el estribo ejerce una fuerza 1,3 veces mayor sobre la ventana oval que la que ejerce el tímpano sobre el martillo. Como el área de la ventana oval –unos  $4 \text{ mm}^2$ – es mucho menor que la del tímpano –de alrededor de  $60 \text{ mm}^2$ – en ella puede aplicarse una presión hasta treinta veces mayor que en el tímpano. Y como la intensidad es función del cuadrado de la presión, la diferencia de intensidad puede llegar a ser de mil veces, equivalente a una ganancia de unos 30 dB. En realidad, el movimiento es más complicado que el de una simple rotación y el manubrio del martillo no se mueve como un cuerpo rígido a altas frecuencias. En la figura 10 se puede apreciar la curva real de transferencia de presión entre el tímpano y la ventana oval.



**Figura 10.** Función de transferencia de presión entre la ventana oval y el tímpano

En la función de transferencia se ve que se obtiene una ganancia de 20–30 dB en el rango medio. La variación espectral de la curva demuestra que la cadena de huesecillos no actúa como una palanca lineal.

En cuanto a la transferencia neta de potencia, se considera que en el oído medio se pierde alrededor del 50% de la energía: al menos la mitad de la potencia recibida por el tímpano no llega al oído interno.



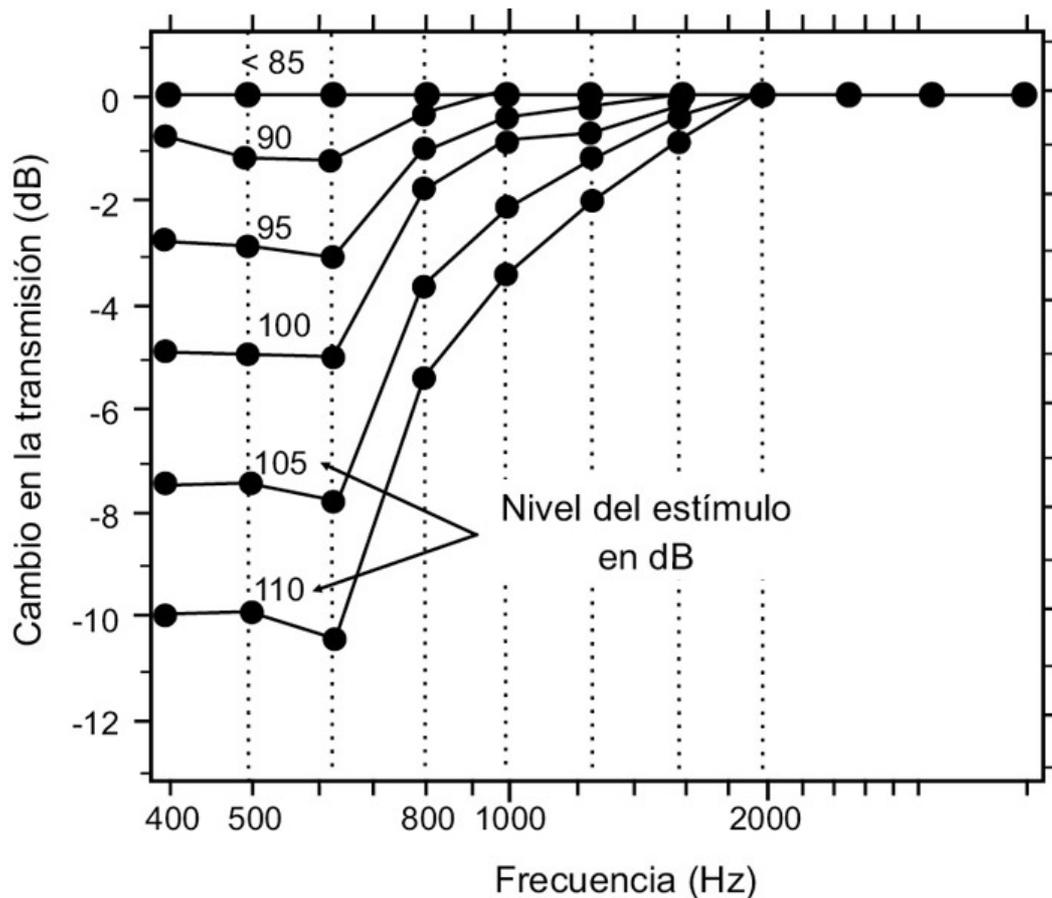
**Figura 11.** Variación de la transferencia de energía del oído medio en función de la presión estática

### Alteraciones en el sistema de transmisión del oído medio

Una alteración frecuente en la ganancia del sistema de transmisión se da cuando la trompa de Eustaquio permanece cerrada y la presión atmosférica varía rápidamente, por ejemplo en el interior de un avión mientras se presuriza la cabina, o al sumergirse a varios metros de profundidad en el agua. La diferencia neta de presión estática curva al tímpano, aumenta la tensión mecánica de la membrana y modifica la capacidad de transferencia de energía del oído medio. Por ejemplo, una diferencia de 1 kPa –que equivale a una columna de 10 cm de agua o a 0,01 atm– produce cambios de-

tectables en la sensibilidad auditiva. Como se puede apreciar en la figura 11, la profundidad del efecto depende de la frecuencia y genera la sensación de sordina que experimentamos al tener los oídos “tapados”.

Otra fuente de variación en la ganancia del sistema la aporta la acción de los músculos de oído medio. Estos músculos se contraen, en una acción denominada reflejo acústico, ante estímulos acústicos de gran intensidad, antes y durante la fonación y, en general, durante los movimientos corporales de mediana y gran amplitud. El estapedial es de acción rápida –el reflejo acústico aparece entre 10 y 20 ms después del estímulo– mientras que el tensor del tímpano responde más lentamente –tarda unos 100 ms en activarse. El retardo del sistema de protección lo vuelve ineficaz cuando el estímulo es de tipo impulsivo, por ejemplo el estallido de un petardo. Los dos músculos combinados logran atenuar la señal alrededor de 20 dB. Como se puede ver en la figura 12, el reflejo acústico no posee respuesta plana y su efecto es mayor a bajas frecuencias.

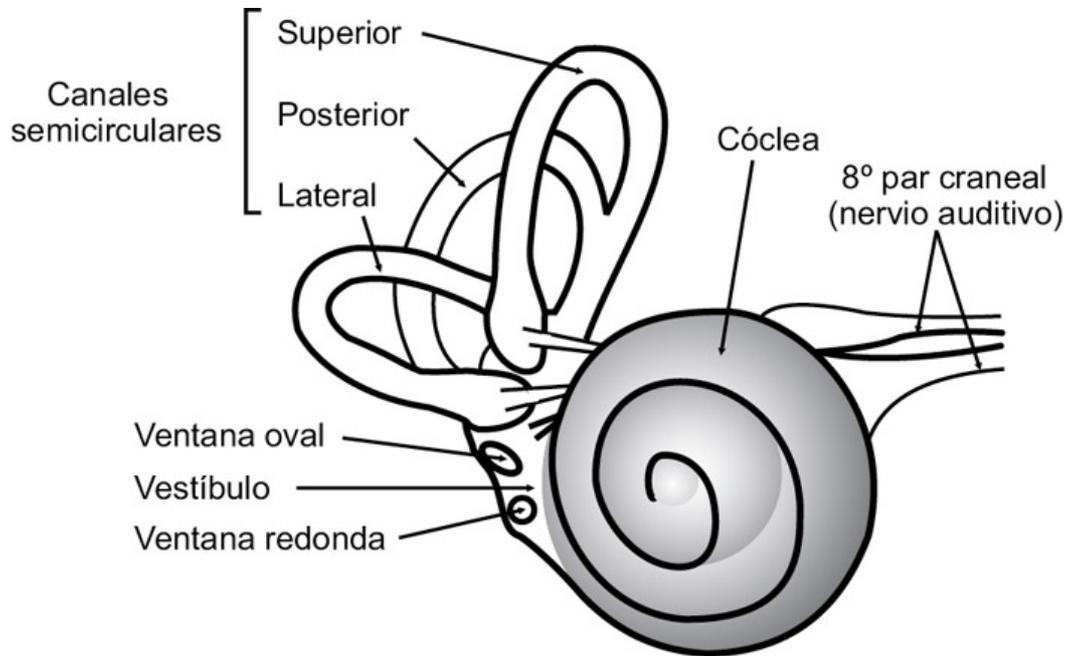


**Figura 12.** Reflejo acústico en función de la intensidad del estímulo

La supresión patológica del reflejo acústico tiene, entre otras, las siguientes consecuencias: 1) disminuye la discriminación del habla a altas intensidades, 2) incrementa la pérdida de audición ante estímulos prolongados de alta intensidad y 3) aumenta el enmascaramiento simultáneo. El daño audiológico provocado por el uso inadecuado de auriculares es similar al inducido por la supresión patológica del reflejo acústico.

Es interesante marcar que la función de transferencia desde el pabellón auricular hasta la ventana oval define casi completamente la inversa de la curva de sensibilidad auditiva en el ser humano (en la figura 30 de este capítulo se puede ver la curva de transferencia completa). A partir de este punto el proceso completo parece comportarse de manera lineal. Los estímulos que llegan por la

ventana oval producen los mismos patrones cocleares que los que llegan por vía ósea, aunque ambas funciones de transferencia difieren notablemente entre sí.



**Figura 13.** Esquema del laberinto óseo

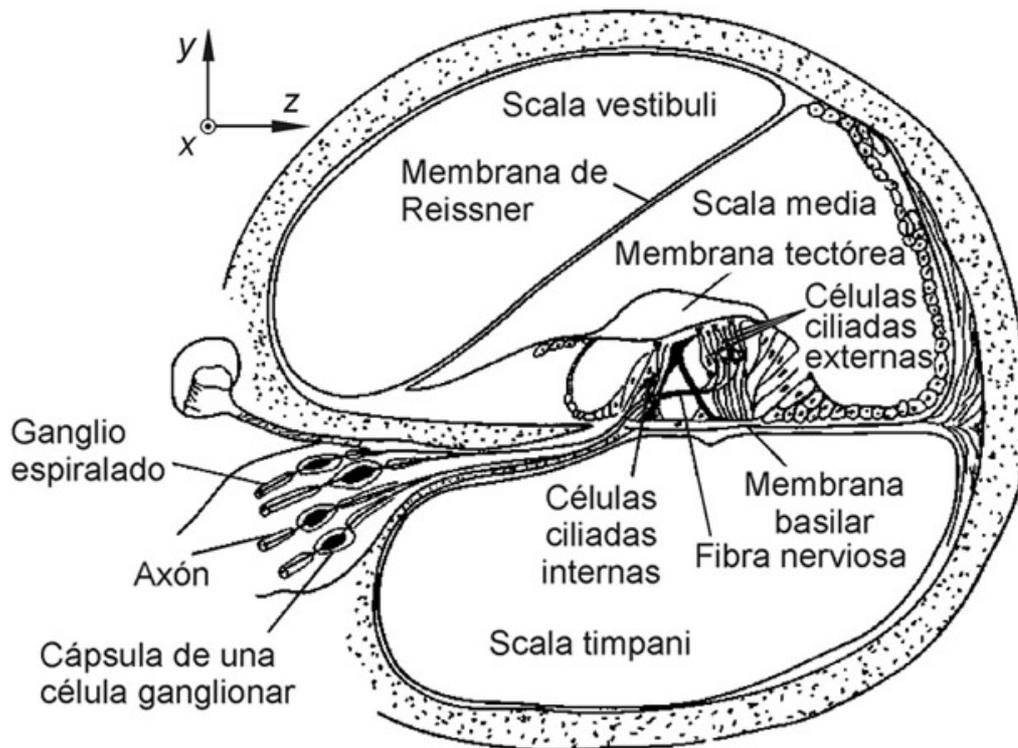
### 3. Oído interno

El oído interno es parte de una cavidad irregular del hueso temporal que recibe el nombre de laberinto óseo y comprende el vestíbulo, el caracol y los canales semicirculares. Sus paredes están formadas por hueso recubierto de epitelio y contiene un líquido acuoso parecido al líquido amniótico. Gran parte del espacio interior está ocupado por un sistema de finos conductos y sacos que constituyen una segunda capa, de paredes tapizadas por tejido conectivo blando, llamado laberinto membranoso. La porción del laberinto óseo que sigue a la ventana oval se denomina vestíbulo, que a su vez está dividido en dos partes, el utrículo y el sáculo. Mientras la base del estribo cubre casi completamente la ventana oval, la ventana redonda se halla recubierta por una delgada membrana de te-

jido conectivo. Un estrecho canal conduce del sáculo al conducto coclear, situado en el interior del caracol. El conducto coclear, también conocido como caracol membranoso, forma una espiral de dos vueltas y media de tejido conectivo que sigue el desarrollo de la espiral ósea del caracol.

El laberinto está inervado por el nervio auditivo –octavo par craneal– que en su interior se divide en dos ramas: el nervio coclear, que es el nervio vinculado a la audición, y el nervio vestibular, que pertenece al sentido del equilibrio.

En la porción central del laberinto están los canales semicirculares, que poseen células receptoras y cristales de calcio que responden a la gravedad y a la aceleración, informando al cerebro sobre la posición y el movimiento de la cabeza. Son tres conductos semicirculares orientados en cuadratura según las tres dimensiones espaciales. Al mover la cabeza se producen movimientos relativos entre el líquido y los cristales de calcio en suspensión que estimulan a las células sensibles que tapizan el tejido interior de los canales. Así, además de las estructuras propias del sentido de la audición, el laberinto óseo contiene también los órganos del sentido del equilibrio.

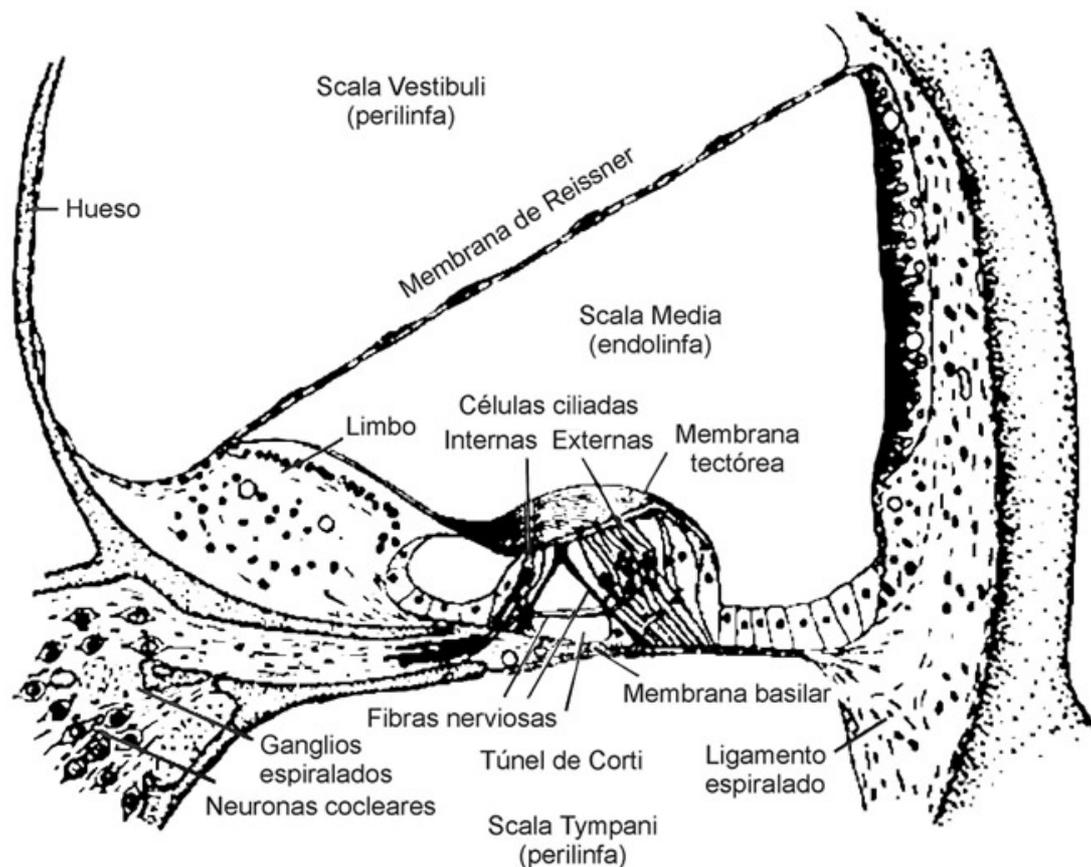


**Figura 14.** Corte transversal de la cóclea en su parte media

El tamaño de los elementos del oído interno es extraordinariamente pequeño. La base del caracol mide menos de 9 mm y el diámetro del conducto que forma la cóclea es de alrededor de 1 mm en su parte media, estrechándose a medida que se acerca al vértice. En su interior, el conducto coclear —o *scala media*— es aún más pequeño, ya que la cóclea está dividida en tres partes o *scalas* como se puede ver en la figura 14. La *membrana basilar* (MB) sigue la forma en espiral de la cóclea y forma el piso del conducto coclear. En su desarrollo describe dos vueltas y media y, totalmente extendida, mide alrededor de 35 mm de largo.

En la membrana basilar se encuentra el órgano de Corti, que contiene las células ciliadas responsables de convertir las vibraciones mecánicas en impulsos nerviosos. Durante mucho tiempo se creyó que la cóclea era únicamente un órgano receptor que proce-

saba las señales acústicas para transformarlas en potenciales eléctricos e impulsos nerviosos. Hoy sabemos, gracias a numerosos estudios, como los de Kemp y Zwicker, que también posee propiedades motoras que le otorgan la capacidad de actuar sobre las señales que llegan desde el exterior.



**Figura 15.** Detalle del conducto coclear

## Fisiología de la cóclea

El movimiento de la cadena de huesecillos del oído medio se aplica a través del estribo a la ventana oval. Como el laberinto está lleno de líquido y sus paredes son óseas, el conjunto resulta prácticamente incompresible. El movimiento del estribo sería poco menos que nulo de no existir la ventana redonda, que permite que el volumen de líquido en la cóclea se mantenga constante y que, por o

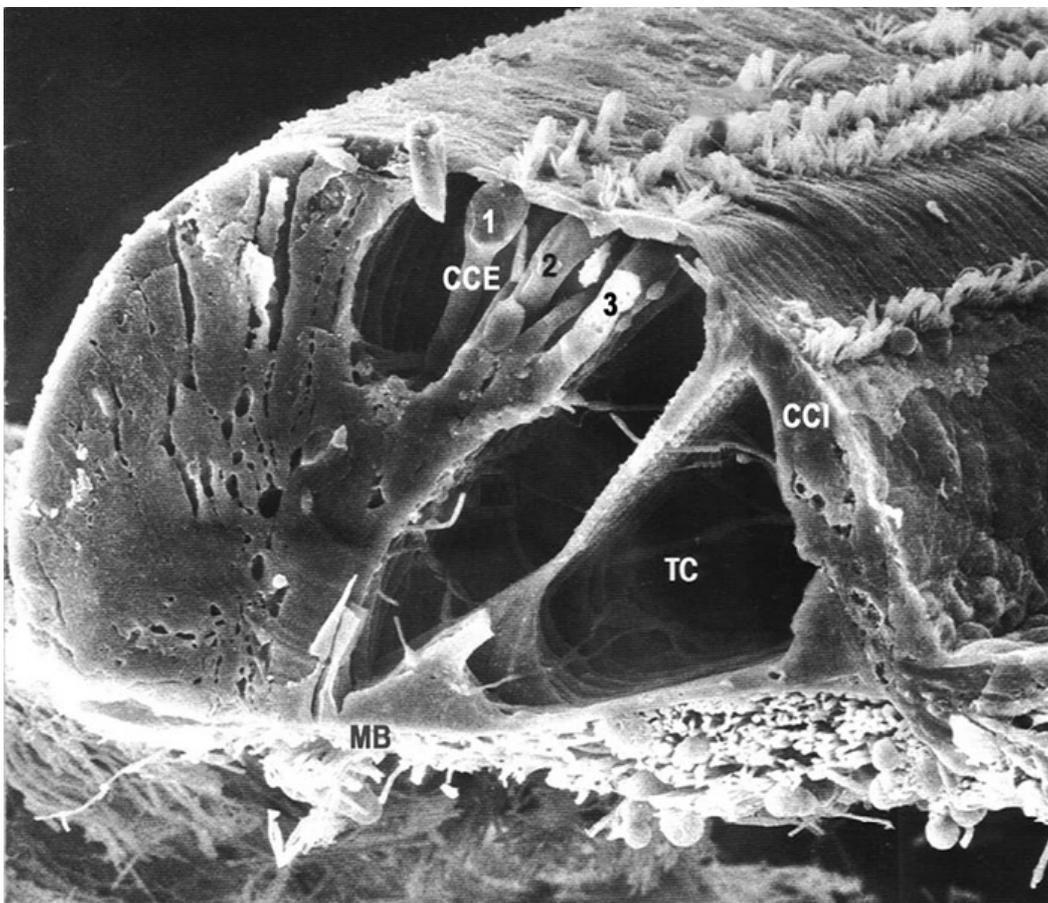
tanto, se compensen exactamente los movimientos netos de la ventana oval y de la ventana redonda. La energía acústica llega entonces a la cóclea que, como se puede ver en el corte de la figura 15, está dividida en tres compartimientos. La *scala tympani* comienza en la ventana oval y está llena de perilinfa, un líquido rico en sodio. La *scala vestibula* nace en la ventana redonda y se comunica con *scala tympani* a través de una abertura, el helicotrema, que permite el paso del líquido perilinfático en el extremo superior o apex del caracol. Entre ambos conductos se encuentra la *scala media* o conducto coclear, que está aislado de los anteriores por la membrana basilar y por la membrana de Reissner y que contiene endolinfa –un líquido viscoso rico en potasio. Como la membrana de Reissner es muy liviana y delgada no actúa mecánicamente sobre el movimiento de los fluidos cocleares. Su función principal es la de separar la endolinfa de la perilinfa para mantener una diferencia de potencial eléctrico de unos 80 mV entre ambos medios.

La membrana basilar es una membrana elástica en la que se ubica el órgano de Corti. Posee alrededor de 3.500 células ciliadas internas dispuestas en una fila y de tres a cinco filas con aproximadamente 25.000 células ciliadas externas en cada oído. Tal como veremos, las células internas son aferentes –envían información al cerebro a través del nervio auditivo– y las células externas son eferentes –reciben impulsos nerviosos desde el cerebro. La membrana tectórea cubre las estereocilias de las células ciliadas ayudando a su excitación mecánica. En la figura 16 se puede ver una imagen del órgano de Corti tomada con un microscopio electrónico de barrido.

### ***Órgano de Corti***

El órgano de Corti es una compleja estructura compuesta, entre otros elementos, por células nerviosas transductoras, células de soporte y fibras nerviosas.

Las células ciliadas son neuronas altamente especializadas y se disponen en dos arcos separados por el túnel de Corti. En el arco exterior se agrupan unas 25.000 células ciliadas externas, cada una con cerca de 140 estereocilias interconectadas. En el otro arco hay cerca de 3.500 células ciliadas internas, cada una con alrededor de 40 estereocilias interconectadas. La membrana tectórea cubre las células de Corti y se sabe que entra en contacto con las estereocilias de las células externas, pero actualmente se discute si también lo hace con las de las células internas.

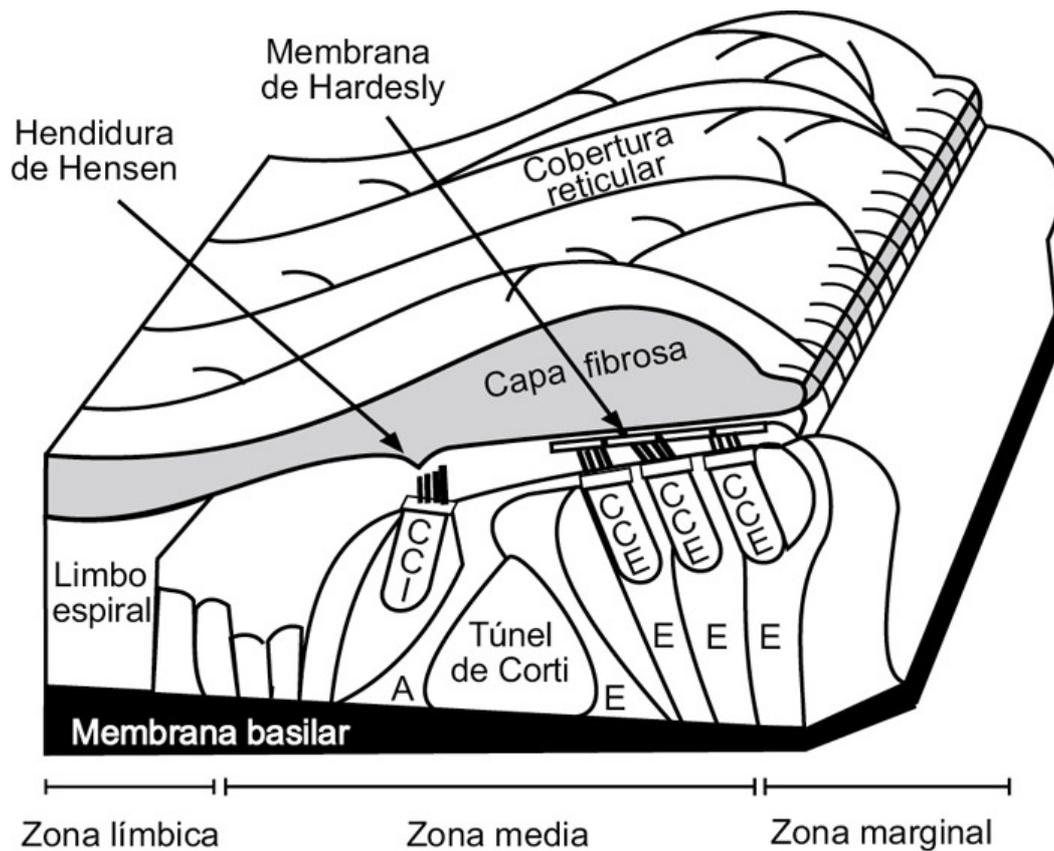


**Figura 16.** Imagen del órgano de Corti tomada con un microscopio electrónico de barrido. 1, 2 y 3: células ciliadas externas; CCI: células ciliadas internas; TC: túnel de Corti; MB: membrana basilar

Las células de soporte proveen apoyo estructural y metabólico a las células ciliadas. Se pueden dividir en dos grupos: el de las célu-

las con filamentos y microtúbulos que controlan la rigidez de la membrana basilar y que son en parte responsables de la distribución espacial de la frecuencia de resonancia a lo largo de la membrana, y el de las células que poseen orgánulos capaces de sintetizar las proteínas y de transportar los iones necesarios para el metabolismo del órgano de Corti.

El órgano de Corti se vincula con el sistema nervioso central por medio de dos tipos de fibras nerviosas: las aferentes llevan información desde el oído interno hacia el cerebro y las eferentes lo hacen en el sentido contrario, desde el cerebro hacia el oído interno. Existen dos tipos de fibras aferentes. Las del tipo I constituyen el 95% del total y contactan únicamente a las células ciliadas internas, mientras que las del tipo II –el 5% restante– se conectan con las células ciliadas externas. Cada célula interna se comunica con alrededor de 20 neuronas aferentes de tipo I y cada neurona aferente de tipo II se reparte entre varias células ciliadas externas. En cambio, las fibras eferentes que parten del cerebro se conectan sólo con las células ciliadas externas, aunque algunos investigadores suponen que pueden también actuar, aunque indirectamente, sobre las células internas.



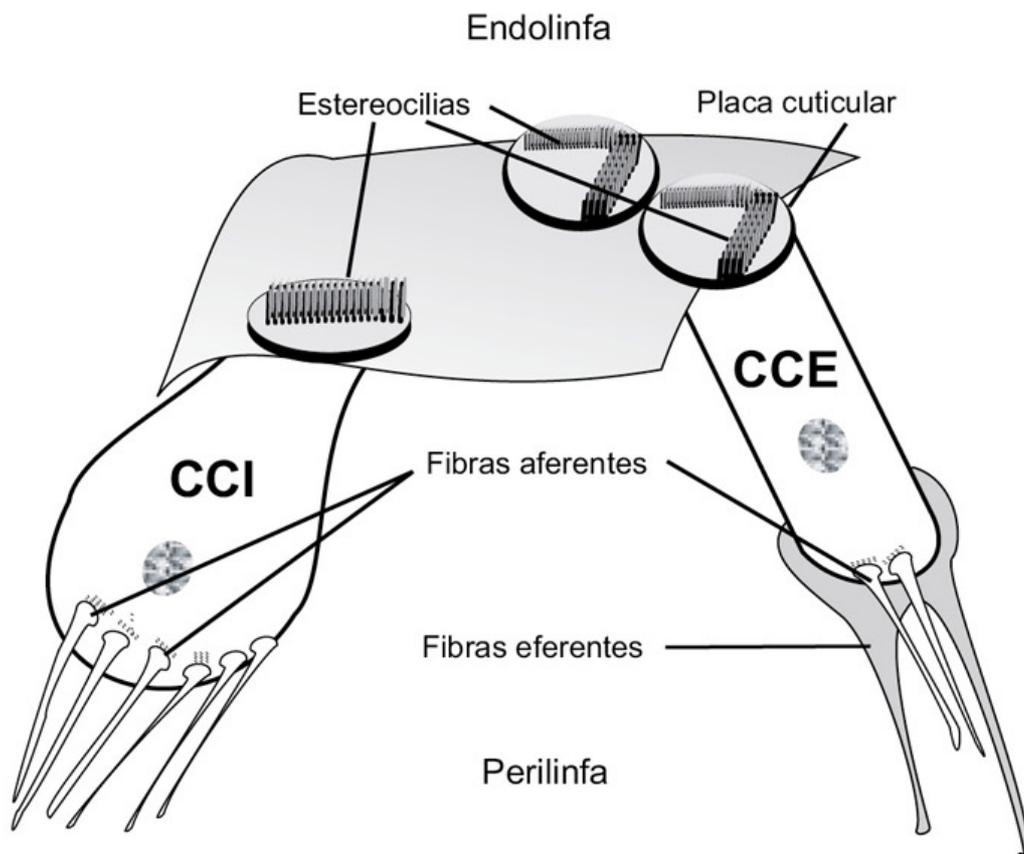
**Figura 17.** Células sensoriales y de soporte en el órgano de Corti. A: pilares internos; E: pilares externos

El punto de contacto a través del cual la información pasa de una neurona a otra se denomina sinapsis. La sinapsis entre una célula ciliada y una fibra aferente se caracteriza por un adelgazamiento de la membrana celular y la presencia de pequeñas vesículas en la célula ciliada que contienen las sustancias químicas necesarias para transmitir la señal. Por el contrario, en la sinapsis entre una fibra eferente y una célula ciliada las vesículas se encuentran en la fibra transmisora.

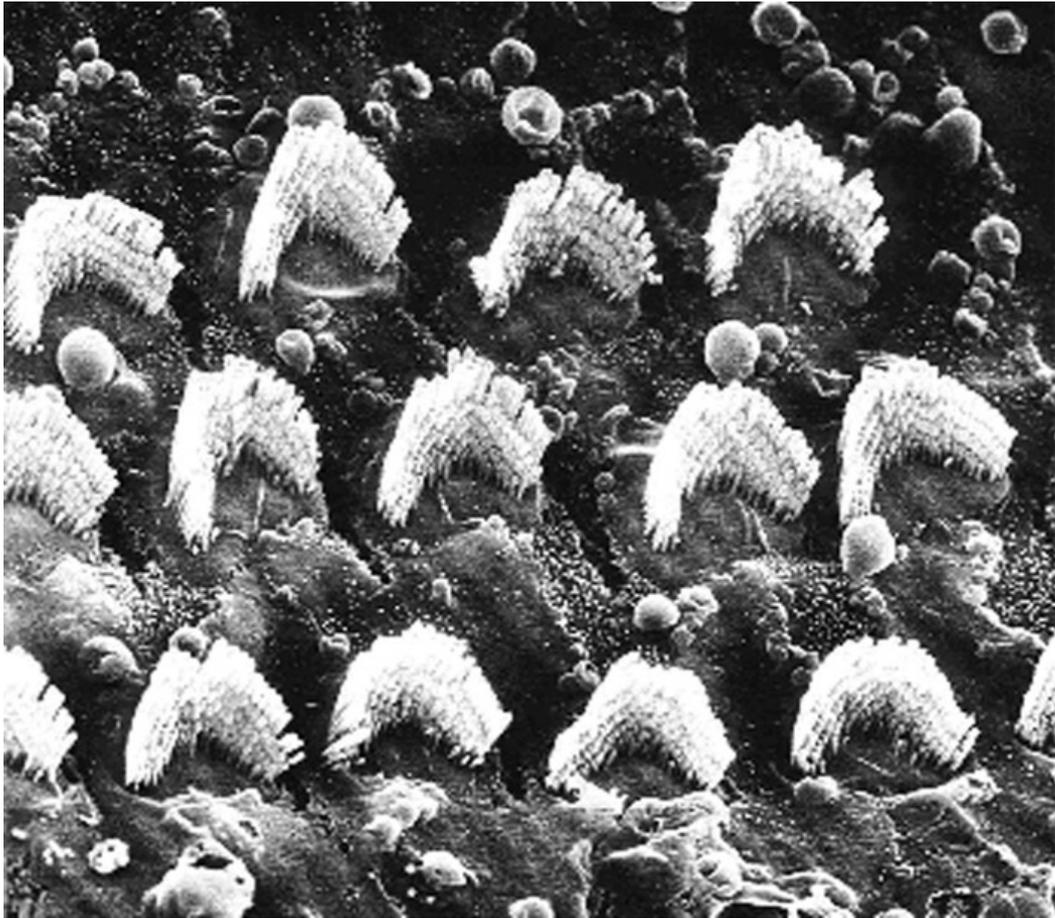
### *Células ciliadas*

Existen dos clases principales de células ciliadas, las internas y las externas, que difieren notablemente entre sí. Las células ciliadas internas tienen forma globular con el núcleo ubicado en el centro y

están soportadas lateralmente por células de sostén. La parte inferior de cada célula interna está en contacto directo con las fibras nerviosas que se proyectan hacia el sistema nervioso central. Como las fibras aferentes que parten de las células internas forman el 95% de las fibras aferentes del nervio auditivo, se supone que éstas actúan como los receptores primarios de la cóclea al convertir los movimientos mecánicos en impulsos neurales. Las fibras eferentes, que llevan información desde el cerebro hacia el órgano de Corti, están en contacto con las fibras aferentes pero no con las células ciliadas internas.



**Figura 18.** Células ciliadas internas y externas



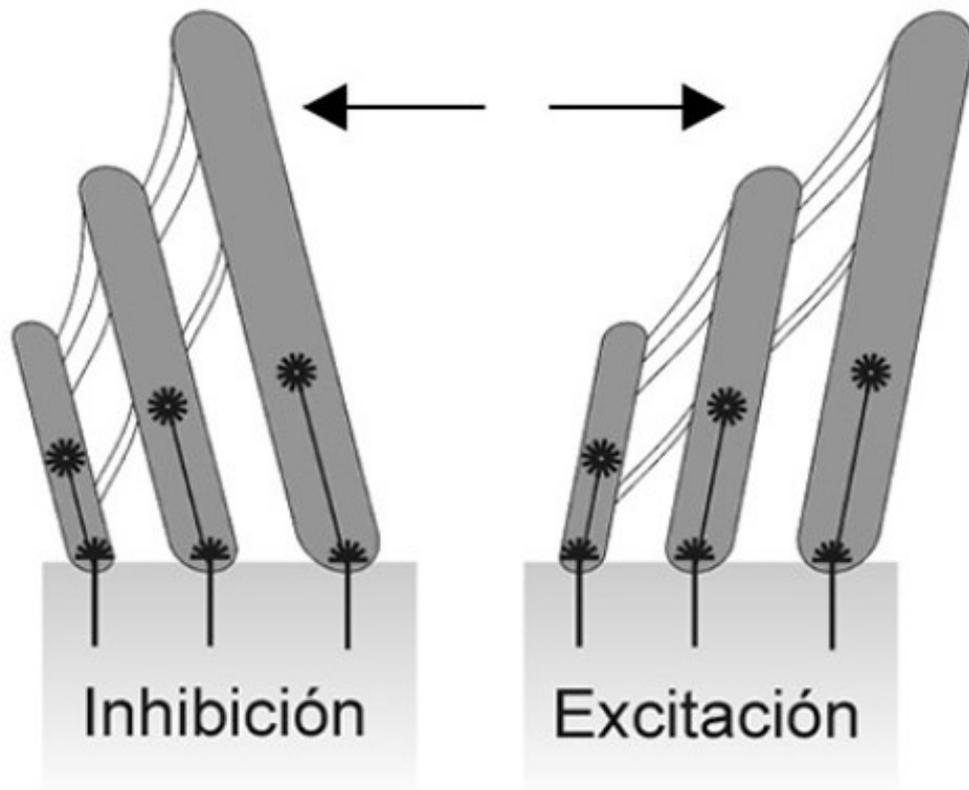
**Figura 19.** Tres filas de células ciliadas externas tomadas con un microscopio electrónico de barrido

Las células ciliadas externas son largas y de forma cilíndrica, con el núcleo situado en su parte inferior. Más de cuatro veces más numerosas que las células internas, forman de tres a cinco filas a lo largo de la membrana basilar. Se hallan en contacto directo tanto con fibras aferentes como eferentes. Durante la estimulación con tensión eléctrica, las células externas se acortan y alargan a una frecuencia estable. A raíz de esta capacidad motora se ha sugerido que son las responsables de los procesos cocleares activos, que incluyen la capacidad de sintonía de banda angosta y la generación de emisiones otoacústicas. Como existen cerca de 1.800 fibras nerviosas eferentes que llevan información hacia la cóclea desde el complejo olivar superior, se supone que en las primeras etapas del

análisis auditivo periférico actúan centros elevados de procesamiento y control.

### ***Proceso de transducción en las células ciliadas internas***

La perilinfa, rica en sodio, es similar a los fluidos que ocupan el espacio exterior a las células del organismo, mientras que la endolinfa, con una gran concentración de potasio, se asemeja al medio líquido interior de las células. Dada su ubicación en la membrana basilar, las células del órgano de Corti están expuestas a la endolinfa en su parte superior y a la perilinfa en la inferior, con una diferencia de potencial de alrededor de 80 mV entre ambos extremos. Durante la estimulación acústica el movimiento de la membrana basilar se transmite a través de diferentes clases de células de soporte, en unión solidaria con las células sensoriales. Las estereocilias de estas últimas se mueven entre la membrana basilar y la masa gelatinosa de la membrana tectórea, excitando mecánicamente las células ciliadas e iniciando el proceso de transducción entre vibración e impulso nervioso.



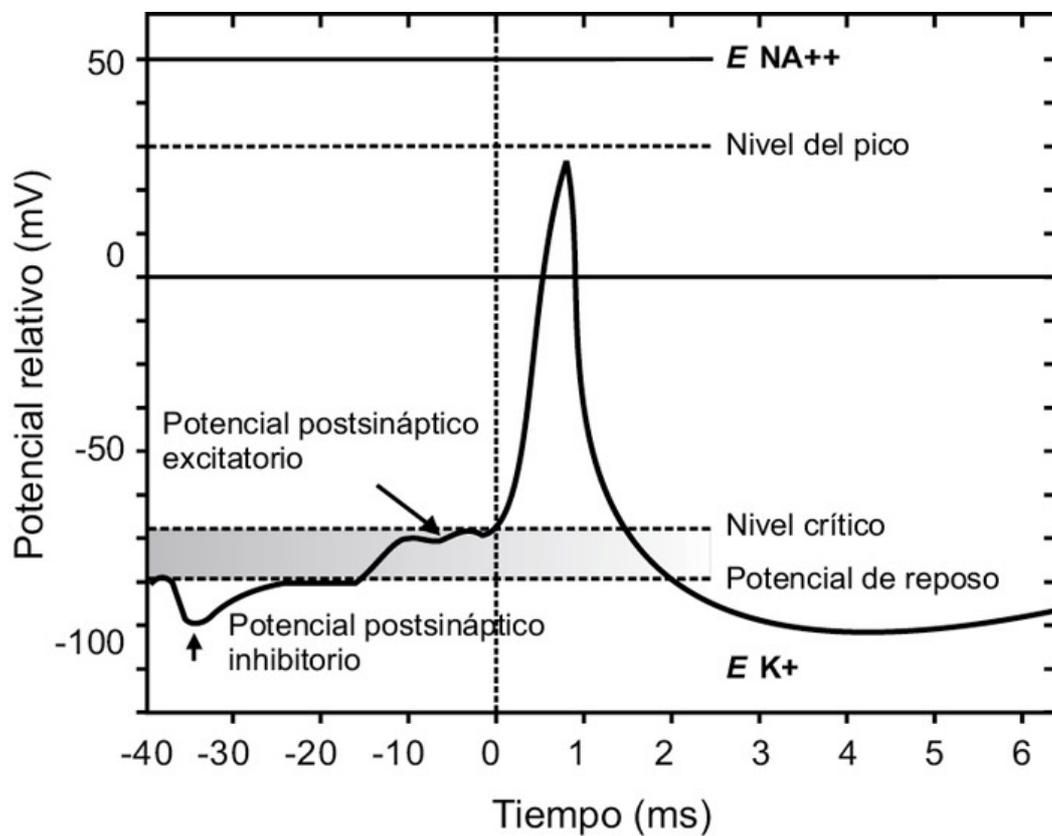
**Figura 20.** Deflexión de la estereocilias en una célula ciliada interna

Todas las estereocilias de una célula están unidas entre sí. Se supone que, como éstas son rígidas y pivotan sobre su base, el movimiento en una dirección abre canales selectivos de iones a través de los que fluyen iones de potasio hacia el interior y la célula se despolariza. La corriente iónica, que fluye desde la *escala media* hacia el interior de las células ciliadas internas, está modulada por la deflexión del haz de estereocilias en relación con un eje de máxima sensibilidad.

Los movimientos en dirección de la fila más alta provocan un aumento en la conductancia del sistema y los que van en sentido contrario la disminuyen, mientras que los movimientos ortogonales al eje no la alteran. La corriente iónica está constituida principalmente por iones positivos de potasio ( $K^+$ ) pues su concentración en la endolinfa de la *scala media* es muy alta. La transducción en

las células ciliadas es no lineal. Los desplazamientos simétricos del haz de estereocilias desde la posición de reposo, de la misma magnitud pero en sentidos opuestos, generan cambios de conductancia asimétricos.

Inmediatamente después de un disparo la célula ciliada interna entra en un *estado refractario*, en el cual se reajusta químicamente, de una duración aproximada de 1 ms. No podrá disparar nuevamente hasta haberlo superado. En la figura 21 se puede apreciar el comportamiento temporal de una célula de Corti durante un disparo.

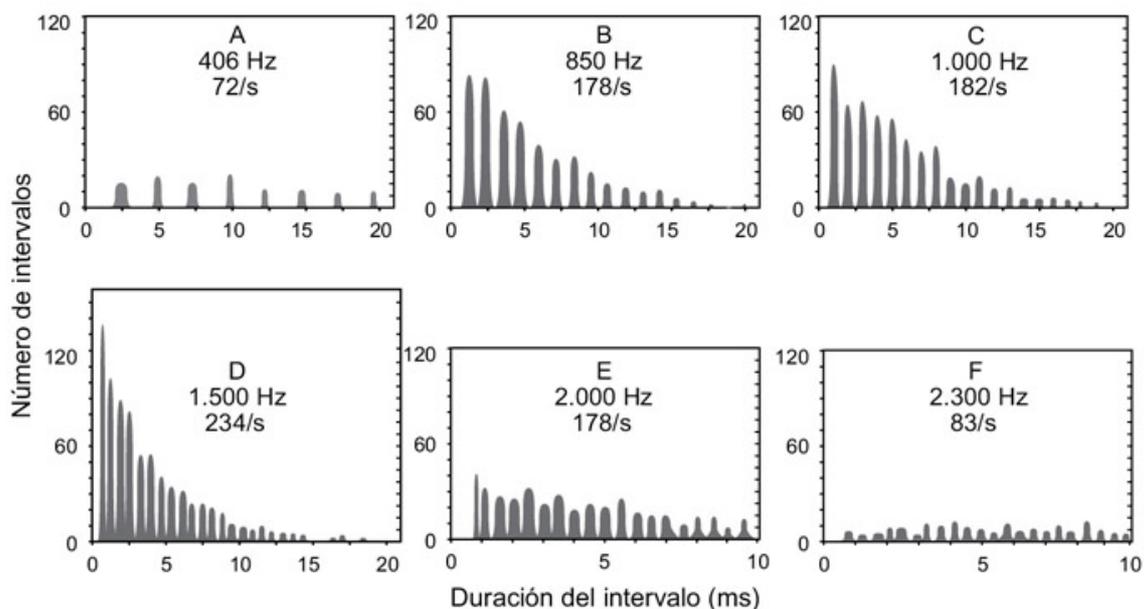


**Figura 21.** Diferencia relativa de potencial entre el interior y el exterior de una neurona especializada ante un potencial de acción variable

El período refractario de una célula es muy grande en relación con los períodos característicos de las señales acústicas: una sola célula

es incapaz de capturar la variación temporal, pues no puede seguir con suficiente velocidad sus ascensos y caídas. Sin embargo, las células ciliadas internas pueden responder ante ciertos valores de la forma de onda del estímulo. Si varias células disparan ante la misma señal a diferentes niveles o fases, el conjunto de la información permite reconstruir la forma de onda original aunque ésta posea una frecuencia muy elevada.

La información de fase del estímulo también es codificada en el patrón temporal de disparo de las células ciliadas internas. Los intervalos entre disparos son, aproximadamente, múltiplos enteros del período del estímulo. Una célula aislada no dispara necesariamente en cada ciclo del estímulo pero, cuando ocurre, lo hace siempre a la misma fase de la forma de onda. El patrón temporal de disparo de un grupo de neuronas contiene información clara sobre el período del estímulo porque existe siempre algún disparo en cada ciclo del mismo. En la figura 22 se ve que el intervalo entre disparos es siempre un múltiplo entero del período de la señal de entrada.



**Figura 22.** Intervalos interdisparo para una neurona auditiva ante estímulos de diferentes frecuencias

La sensibilidad a la fase desaparece por encima de 5 kHz. El límite superior no está determinado por la tasa máxima de disparo de las neuronas, sino por la precisión con que el impulso nervioso “engancha” una fase particular del estímulo. A 5 kHz la variabilidad estocástica de enganche es comparable con el período de la onda.

Muchos investigadores aceptan que nuestra habilidad para localizar sonidos depende en parte de la comparación de la información temporal de los dos oídos; pero la relevancia de la información temporal en el enmascaramiento y en la percepción de la altura tonal es todavía debatida acaloradamente, tal como veremos más adelante.

### ***Proceso de transducción en las células ciliadas externas***

En 1983 Brownell descubrió que las células ciliadas externas aisladas pueden cambiar de longitud ante un cambio en el potencial eléctrico de polarización (Brownell, W. E., 1985). A pesar de que el mecanismo no se conoce en detalle, se supone que el estímulo efectivo lo provee una variación en el potencial en la membrana celular y no, como en el caso de las células ciliadas internas, una corriente iónica. El cambio de longitud total se logra sumando los cambios geométricos de gran cantidad de estructuras sensomotoras independientes alojadas en el plasma de la membrana celular. La membrana de las células ciliadas externas posee gran cantidad de mitocondrias adyacentes que le permiten transformar las moléculas orgánicas para obtener la energía necesaria de acuerdo al ciclo de Krebs.

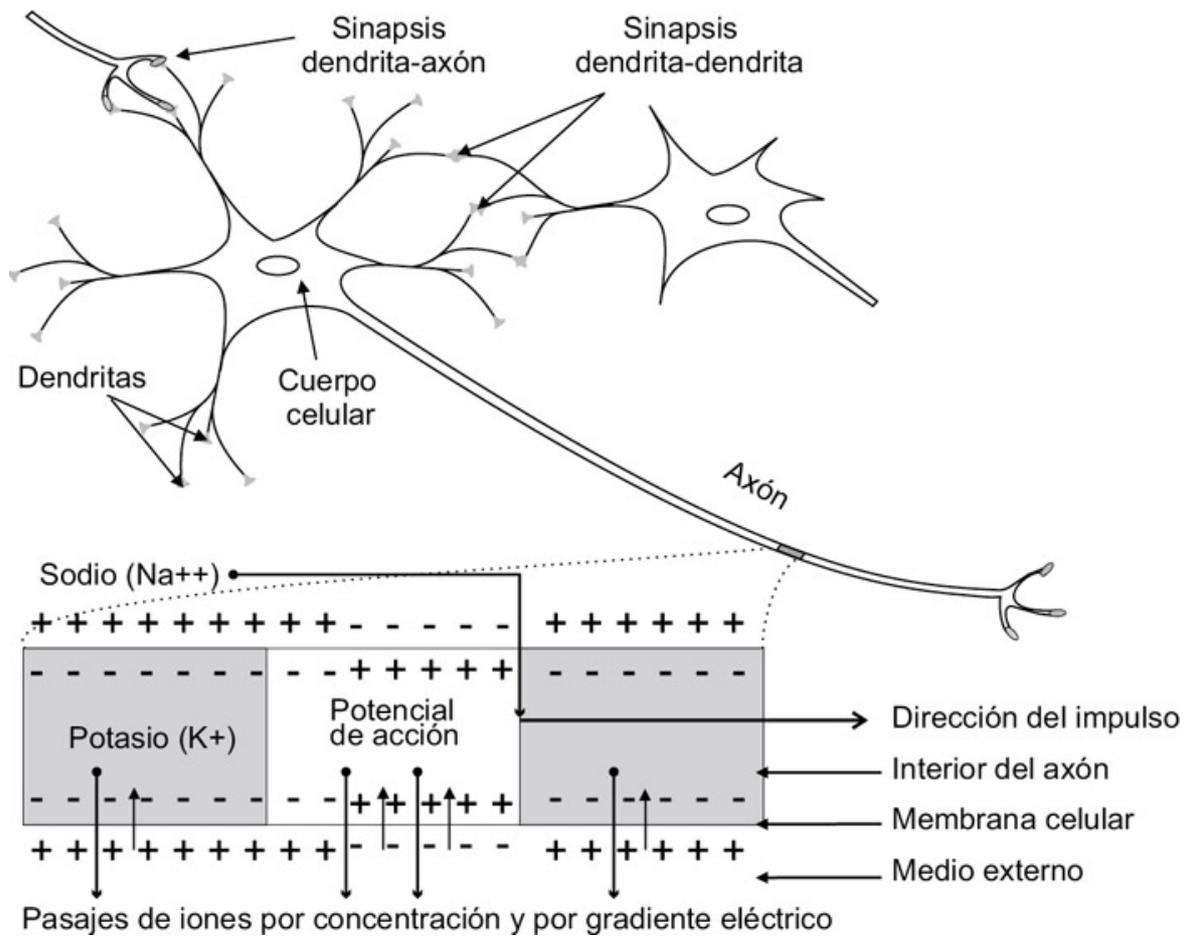
Las células ciliadas externas actúan como pequeños músculos que agregan energía mecánica a la onda progresiva en la membra-

na basilar. Se cree que la fracción de energía agregada puede ser controlada por los centros elevados de procesamiento a través de la información que viaja por las fibras eferentes del nervio auditivo. Actúan como una “resistencia negativa” que compensa la energía perdida en el proceso de vibración. Este control de ganancia también puede reducir la respuesta ante estímulos muy intensos, y actúa con mucha mayor velocidad que el reflejo del músculo estapedial. La realimentación positiva que tiene lugar permite sortear las limitaciones propias de un sistema pasivo impuestas por el principio de indeterminación acústico.[\[5\]](#)

Las células ciliadas externas afectan la transducción de las células ciliadas internas principalmente a través de la vibración inducida de la membrana basilar.

## **Características del nervio auditivo**

El nervio auditivo se divide en dos ramas anatómica y fisiológicamente diferentes, la rama vestibular –o nervio del equilibrio– y la rama coclear. Está formado principalmente por los axones de las neuronas conectadas con las células del órgano de Corti. Como la resistencia eléctrica a lo largo del eje es muy grande y la resistencia eléctrica de la membrana celular muy pequeña, resulta un conductor eléctrico ineficiente. En consecuencia, la carga positiva inyectada en el axón se disipa tras uno o dos milímetros de recorrido. Para que la señal recorra distancias mayores es necesario regenerar frecuentemente el potencial de acción a lo largo del camino a partir del intercambio iónico. Así, la información viaja a lo largo de los axones en forma de breves impulsos eléctricos que son el resultado del desplazamiento a través de la membrana celular de iones de sodio dotados de carga positiva. En la figura 23 se puede observar un esquema del proceso de transmisión.



**Figura 23.** Neurona típica y transmisión de un potencial de acción en el axón de una neurona

Un pulso nervioso que viaja por un axón es potenciado cada tanto a través de discontinuidades en las vainas mielínicas llamados nódulos de Ranvier. La necesidad de reforzar repetidamente la corriente eléctrica limita la velocidad máxima de viaje de los impulsos. Para axones de 2 mm de diámetro la velocidad de conducción es de 10 m/s, mientras que para axones de 20 mm llega a los 120 m/s.[\[6\]](#)

## 4. Teoría del lugar

Como los pulsos nerviosos que nacen en las células ciliadas internas no son lo suficientemente rápidos como para seguir la forma de onda exacta de una señal acústica (las ventanas que definen po-

seen anchos de banda limitados), deben existir necesariamente otros mecanismos que permitan codificar dichas señales. De todos los propuestos se destacan la teoría del lugar y las teorías temporales. Desarrollaremos aquí la primera de ellas y estudiaremos las teorías temporales en el capítulo dedicado a la altura tonal.

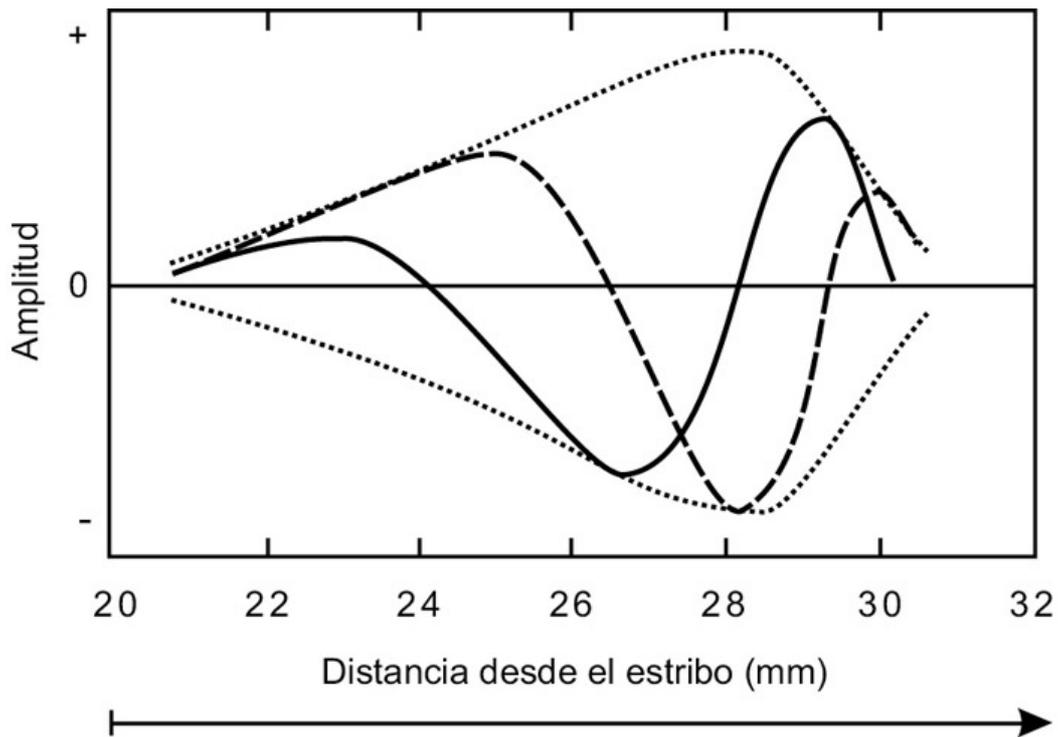
El descubrimiento de la mecánica de la membrana basilar, fundamento de la teoría del lugar –o teoría tonotópica–, se debe a Georg Von Békésy (Békésy, G. Von, 1960). Su magnífico trabajo, que le valió el Premio Nobel de Fisiología y Medicina en el año 1961, es el fundamento de casi todas las investigaciones posteriores sobre la fisiología de la audición. La teoría del lugar postula que cada punto de la membrana basilar vibra a una frecuencia particular, las más agudas en la zona cercana a la base de la cóclea y las más graves en la región próxima al apex, en una especie de mapeo frecuencia-espacio. [7]

Cuando vibra el estribo, la onda generada en la ventana oval se propaga rápidamente por el fluido coclear y llega al apex en apenas unos microsegundos. Pero también se propaga transversalmente a lo largo de la membrana basilar con una velocidad mucho menor, recorriendo su extensión en algunos milisegundos. La velocidad de propagación en la membrana basilar no es uniforme y varía en función de la distancia a la base de la cóclea. Cerca de la ventana oval la velocidad es rápida y la longitud de onda comparativamente grande. En el apex, por el contrario, la velocidad y la longitud de onda son mucho menores. La velocidad finita de la onda progresiva implica un retardo acumulado en función de la distancia desde el estribo que produce un corrimiento de fase, también acumulativo, de gran importancia en algunos procesos de transducción coclear.

Esta característica mecánica de la membrana basilar se debe a que funciona como una estructura de soporte con parámetros que se modifican gradualmente a lo largo de su extensión. Es más angosta, gruesa y rígida en su base y deviene progresivamente más

ancha, delgada y elástica a medida que se acerca a su extremo apical. La rigidez aumenta progresivamente hacia la base de la cóclea, mientras que su masa decrece en la misma dirección. Este gradiente de rigidez y masa permite la variación gradual de la frecuencia de resonancia mecánica de la membrana basilar. Recordemos que, para un sistema discreto, la *frecuencia de resonancia*  $f_R$  depende directamente de dichos parámetros a partir de la relación  $f_R = \sqrt{k/m}$ , en la que  $k$  representa la rigidez y  $m$  la masa del sistema. La variación de la elasticidad en función de la posición, en una razón de 1:100 entre ambos extremos, es la principal característica física que explica el mapeo frecuencia-espacio que tiene lugar en la cóclea. La estructura del órgano de Corti, a su vez, cambia continuamente a lo largo de su desarrollo. Cerca del apex, en la zona sensible a las bajas frecuencias, las células son más grandes y las estereocilias más largas que en la región cercana a la base, sensible a las altas frecuencias.

Si el estribo empuja muy lentamente a la ventana oval, la perilinfa pasa a través del helicotrema y el volumen total de líquido permanece constante al desplazarse hacia fuera la ventana redonda. Pero si empuja rápidamente a la ventana oval, la deflexión en la membrana basilar compensa el cambio de volumen y no existe flujo a través del helicotrema. En esencia, una onda de una frecuencia definida no puede viajar más allá de un determinado “punto de corte” en la cóclea. A medida que se aproxima a ese punto, la onda progresiva se hace más lenta, la amplitud aumenta y su energía es absorbida por el movimiento de la membrana basilar. En la figura 24 se aprecia una onda progresiva en la membrana basilar generada por un estímulo sinusoidal de 200 Hz.



**Figura 24.** Dos instantes diferentes de una onda progresiva generada por una excitación sinusoidal de 200 Hz

El gradiente espacial que define el comportamiento tonotópico de la membrana basilar se manifiesta también en la pendiente de la onda progresiva.

Se puede pensar en la membrana basilar como un filtro en el cual la respuesta máxima ante una señal sinusoidal se mueve hacia el extremo apical a medida que la frecuencia disminuye. En otras palabras, existe una conversión frecuencia-espacio, tal como se aprecia en la figura 25.

Al margen de la frecuencia del estímulo, la forma de la envolvente temporal de la vibración en cada punto de la membrana basilar sigue isomórficamente la de la vibración en la ventana oval. Ante estímulos sinusoidales toda la membrana basilar vibra de manera aproximadamente sinusoidal, a la frecuencia de la señal de entrada. Por ejemplo, si se aplica una senoide de 1.000 Hz, cada punto de la membrana que se mueve lo hace a esa frecuencia, con

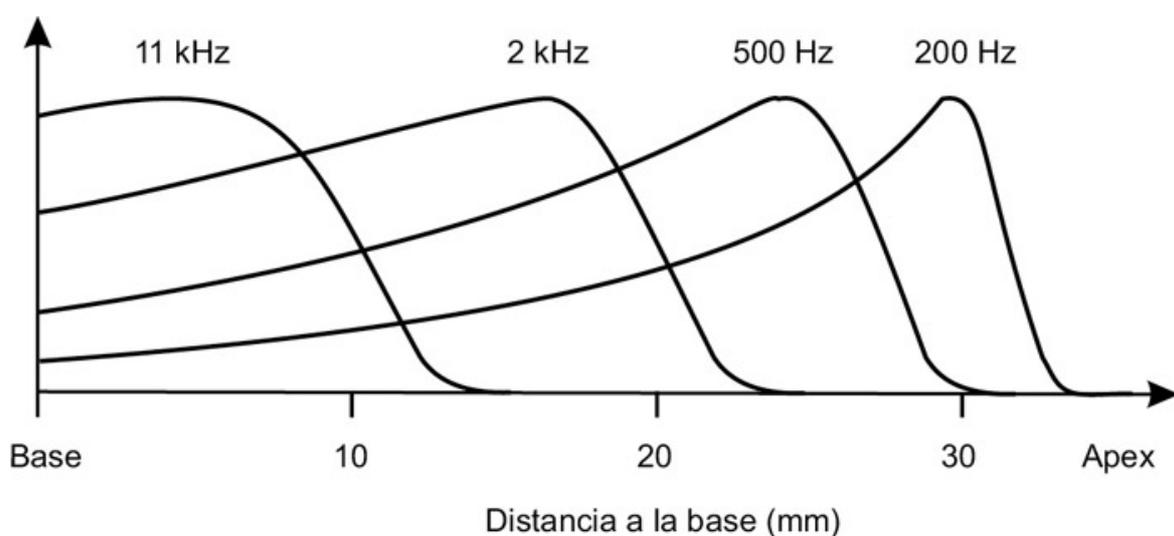
un desplazamiento máximo en el punto de resonancia correspondiente a 1.000 Hz.

El patrón de vibración de la membrana basilar no depende del punto sobre el cual se localiza la excitación: vibraciones que llegan a la cóclea desde los huesos de la cabeza generan respuestas similares a las que provienen de la ventana oval.

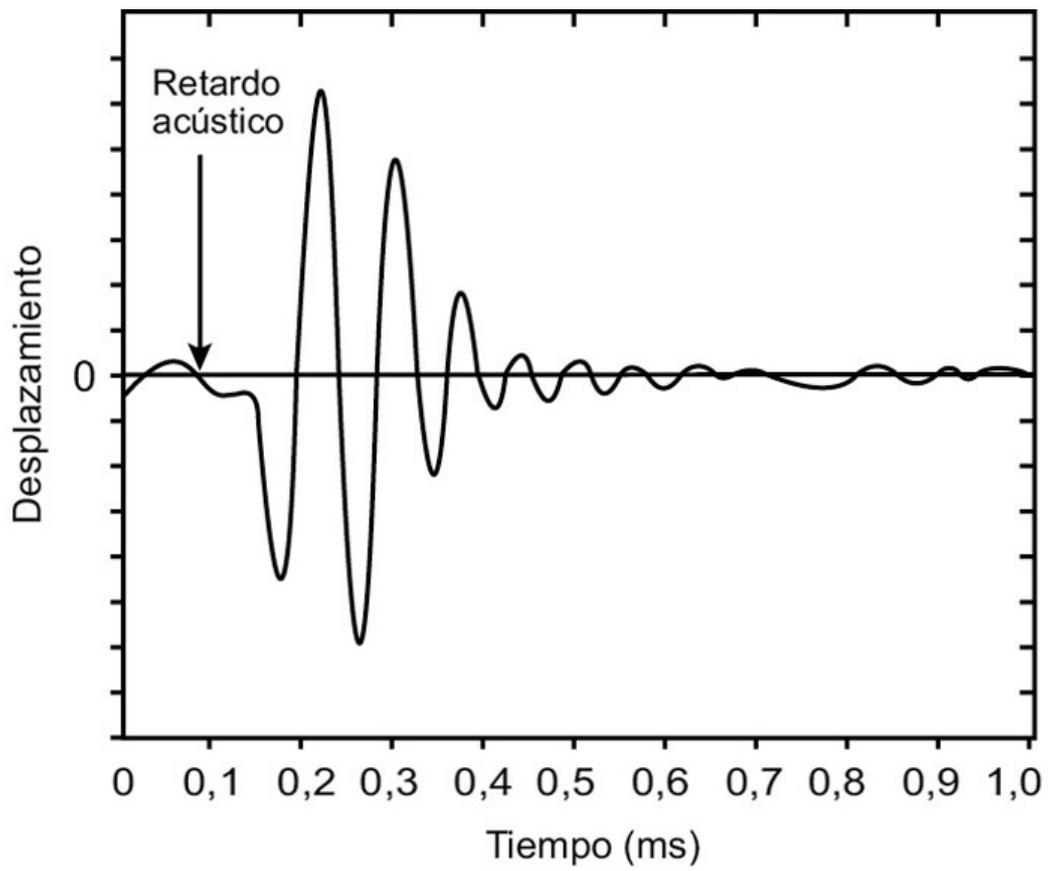
## Excitación de la membrana basilar con un impulso

Cuando se aplica un impulso al estribo aparece una perturbación cerca de la ventana oval que se propaga a lo largo de toda la membrana basilar. Cada punto de la membrana se comporta como un filtro de banda pasante cuya frecuencia central corresponde a la frecuencia de resonancia característica en ese punto. La forma de onda y el espectro varían continuamente con la posición, hecho que afecta nuestra habilidad para discriminar sonidos de carácter transitorio. [8]. En la figura 26 se puede ver la respuesta al impulso en un punto particular de la membrana basilar.

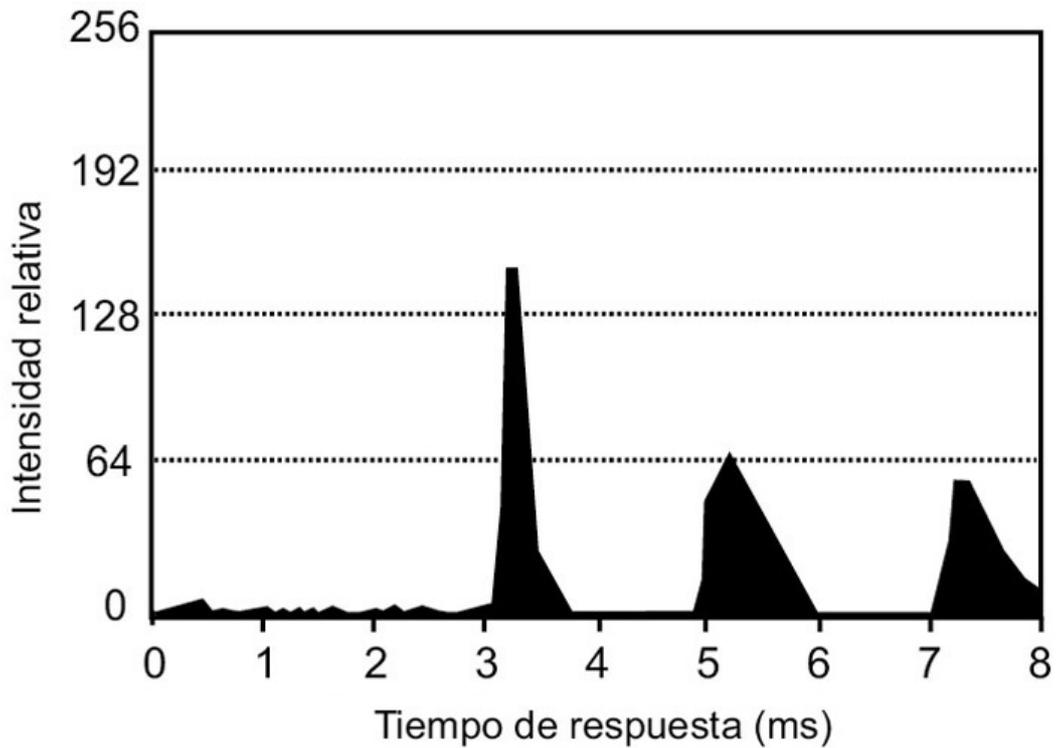
A su vez, la respuesta ante un clic de las células internas en un punto determinado de la membrana basilar se puede ver en la figura 27.



**Figura 25.** Resonancia de la membrana basilar en función de la frecuencia y de la distancia a la base de la cóclea



**Figura 26.** Respuesta mecánica ante un impulso del punto de la membrana basilar correspondiente a una frecuencia de resonancia de 10.000 Hz



**Figura 27.** Histograma posestímulo en un punto de la MB ante un clic

Los picos múltiples aparecen debido a las oscilaciones atenuadas del filtro basilar correspondiente: el intervalo temporal entre picos es el inverso de la frecuencia característica de la zona de la membrana basilar. Si la polaridad del clic se invierte, el patrón se desplaza en el tiempo de modo que los picos aparecen donde antes se localizaban los valles, efecto muy importante a la hora de localizar espacialmente una fuente acústica.

Estos hechos también se relacionan con la percepción de señales periódicas. Javel mostró que una porción de la actividad de las células ciliadas internas que responde a los armónicos de una señal periódica es sensible a la fase de repetición de la onda completa, que coincide con la fundamental de Fourier de la señal (Javel, E., 1980).

## 5. Comportamiento activo de la membrana basilar y emisiones otoacústicas

De acuerdo con la teoría del lugar clásica, tal como la enunció Von Békésy, la membrana basilar actúa como un analizador lineal de Fourier con un poder de resolución insuficiente para explicar la resolución en frecuencia del oído humano. En otras palabras, el ancho de banda de la resonancia de la membrana es mucho mayor que el que se mide en los experimentos psicoacústicos. Por ejemplo, de acuerdo con la teoría del lugar clásica, al excitar la membrana basilar a una frecuencia central de 1.000 Hz aparece una banda de 100 Hz de extensión, cien veces más grande que la resolución real, que es del orden de 1 Hz para dicha frecuencia central. La causa de esta enorme discrepancia es, lo sabemos ahora, la existencia de un mecanismo activo en el órgano de Corti que Von Békésy, al trabajar exclusivamente con cadáveres, no pudo descubrir.

En 1978 Kemp encontró que al estimular el tímpano de un individuo vivo con un impulso de baja intensidad se detecta un eco generado en el interior del oído (Kemp, D. T., 1978). Como estas señales poseen un retardo comprendido entre 5 y 60 ms se sugirió que tenían un origen coclear. Actualmente se las conoce como ecos cocleares, ecos de Kemp o emisiones otoacústicas evocadas. El nivel relativo de la reflexión es mayor a menores niveles de excitación, lo que revela una operación no lineal. En ocasiones, la energía reflejada supera a la energía incidente evidenciando un proceso activo de amplificación biológica. Las emisiones otoacústicas provienen de la actividad motora de las células ciliadas externas.

El descubrimiento de Kemp modificó profundamente el conocimiento que tenemos sobre la fisiología del oído. Los procesos cocleares activos posibilitan la sintonía de banda angosta que se observan en las experiencias psicoacústicas y que las teorías anteriores basadas en el oído como un receptor pasivo –como la teoría

clásica del lugar-, no podían explicar. El comportamiento activo del oído interno permite su asombrosa resolución en frecuencia, su gran sensibilidad ante estímulos débiles y explica el efecto de supresión de un tono por otro, temas que desarrollaremos más adelante.

La actividad coclear puede entenderse también como una relación alineal entre las señales mecánicas de entrada y salida. Numerosos trabajos en individuos vivos han demostrado que la respuesta de la membrana basilar es alineal, es decir que la salida no responde proporcionalmente a la entrada. A niveles de señal de entrada por debajo de 60 dB, un mecanismo biológico activo amplifica la respuesta coclear. A medida que el nivel crece la amplificación disminuye y, para altos niveles en la señal de entrada, cercanos a 90 dB, el mecanismo activo deja de operar y el sistema deviene lineal. Una prueba de la alinealidad coclear se verifica al excitar un oído con dos sinusoides de frecuencias  $f_1$  y  $f_2$ . El *eco coclear* que aparece coincide en frecuencia con el sonido de combinación principal  $2f_1 - f_2$ . Este resultado prueba la distorsión armónica o alineal que actúa durante el proceso de transducción (Basso, G., 2001a).

La alinealidad ocurre principalmente en los lugares de la membrana basilar próximos a la frecuencia de resonancia correspondiente. En un cadáver la respuesta es siempre lineal, tal como constató Von Békésy en sus experimentos, ya que desaparecen los procesos celulares activos aportados fundamentalmente por la motricidad de las células ciliadas externas. La alinealidad se pierde con el daño fisiológico y con la muerte: los ecos cocleares sólo se observan en sujetos en buena condición fisiológica. Están relacionados con procesos vulnerables y las hipoacusias temporarias o permanentes, ciertas drogas y otros trastornos fisiológicos los inhiben.

En aproximadamente el 30% de las personas aparecen emisiones otoacústicas espontáneas, que suceden en ausencia de estímulos acústicos externos. Los *tinnitus* o *acúsfenos* son, en esencia, emisiones otoacústicas espontáneas.

## 6. Funciones de transferencia del sistema auditivo periférico

Es posible reunir toda la información descrita hasta ahora para analizar cómo el sistema auditivo periférico modifica las señales acústicas que llegan a cada oído. El resultado final es la *función de transferencia del sistema auditivo periférico*, que relaciona la presión acústica de la fuente exterior  $p(t)$  con la vibración que llega hasta la membrana basilar  $S(t)$ .

Si  $h_d$  y  $h_i$  son las señales acústicas a la entrada de los pabellones auditivos derecho e izquierdo y  $g_d(t)$  y  $g_i(t)$  las respuestas al impulso del espacio acústico entre la fuente y las dos entradas auditivas,  $h_d$  y  $h_i$  se pueden obtener de las expresiones que siguen (el símbolo  $*$  denota convolución):

$$h_d(t) = p(t) * g_d(t)$$

$$h_i(t) = p(t) * g_i(t)$$

Una parte de las respuestas  $g_d(t)$  y  $g_i(t)$  está dada por las características del ambiente acústico exterior y otra parte por las funciones de transferencia definidas por los hombros, la forma y tamaño de la cabeza y el pabellón auricular. Los efectos de estas piezas anatómicas se vuelven significativos cuando las longitudes de onda son comparables con sus dimensiones a causa del fenómeno de difracción, tal como veremos en el capítulo VII. Para simplificar la exposición, obviaremos desde aquí las diferencias entre los oídos derecho e izquierdo y utilizaremos la transformada de Fourier

$H(\omega)$  para simbolizar la respuesta espectral de la entrada auditiva  $h(t)$ .

En el trayecto del estímulo interviene a continuación la función de transferencia del canal auditivo externo  $E(\omega)$ , que al combinarse con  $H(\omega)$  modifica la señal de entrada de la manera que se ve en la figura 28.

En seguida aparece la acción del sistema de palancas y filtros del oído medio, cuya función de transferencia promedio,  $C(\omega)$ , se aprecia en la figura 29.

Si se combinan las tres funciones de transferencia anteriores se obtiene la función de transferencia total entre la fuente acústica y la entrada de la cóclea  $S(\omega)$ :[\[9\]](#).

$$S(\omega) = H(\omega) E(\omega) C(\omega)$$

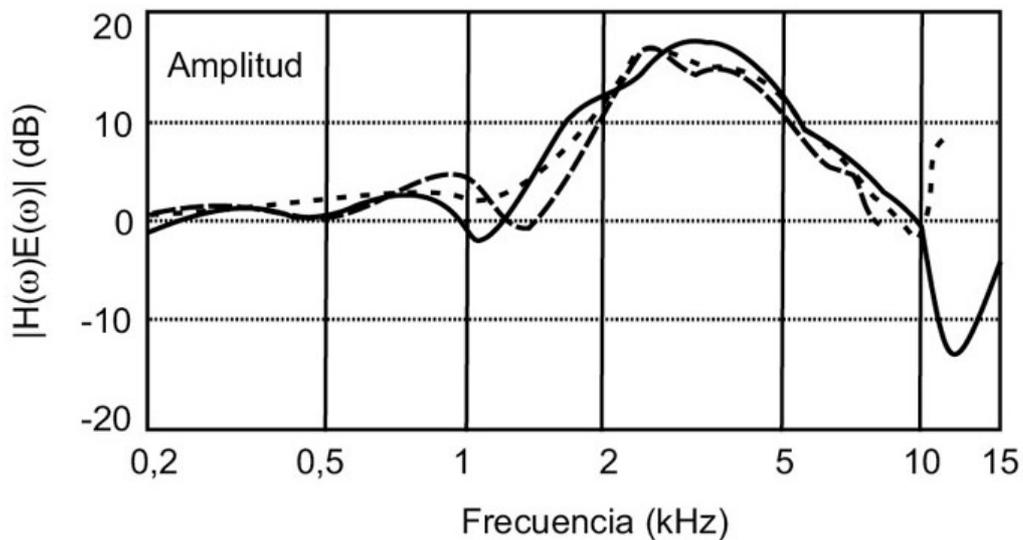
En la figura 30 se compara a  $S(\omega)$  con la sensibilidad promedio del oído humano.

El notable ajuste entre la función  $S(\omega)$  con la curva de sensibilidad del oído permite afirmar que el proceso que sigue luego de la llegada de la señal a la membrana basilar es, al menos para intensidades medias, esencialmente lineal. Risset afirma: “En un artículo teórico, Yang, Wang y Shamma (1992) demostraron que un modelo fisiológico plausible de la periferia auditiva puede ser invertido y usado para regenerar la entrada original completamente. Este resultado sugiere que el sistema auditivo no pierde información acústica en las etapas tempranas del proceso” (en Deutsch, D., 1999).

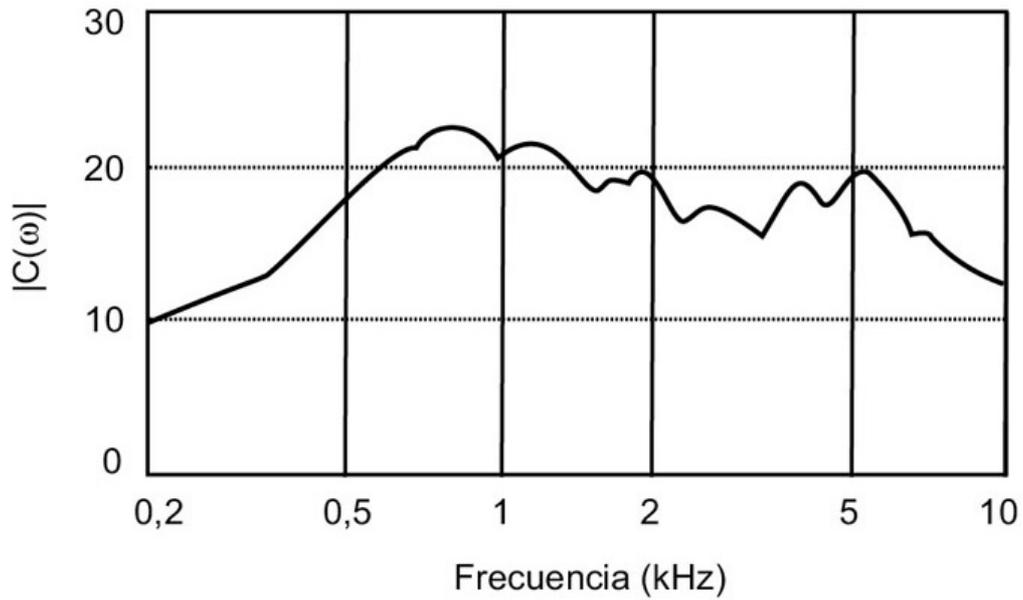
## **7. Respuestas neurales a altos niveles en el sistema auditivo**

Existe evidencia de que concurren grupos de detectores neurales altamente especializados en los niveles superiores del sistema auditivo. Aparentemente, en la corteza cerebral se analiza algo más

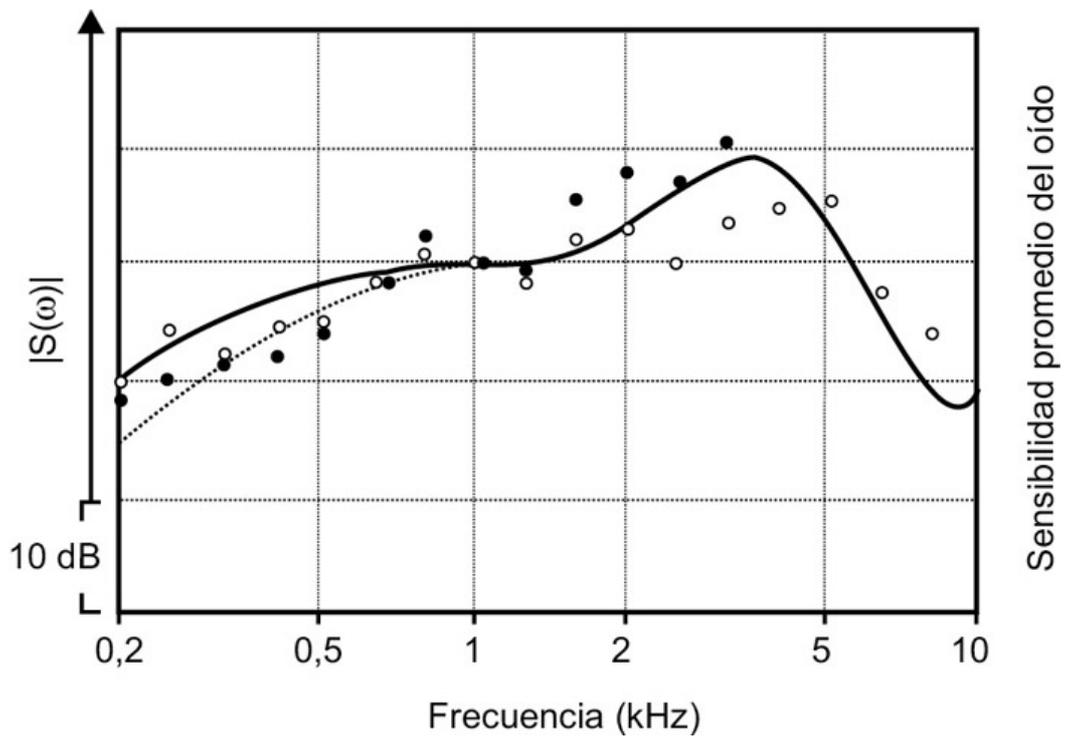
que la frecuencia o la intensidad de las señales que le llegan. Algunas neuronas corticales parecen responder directamente a los transitorios de la señal de entrada; otras reaccionan ante cambios unidireccionales de frecuencia (se las denomina “frequency sweep detectors” en la literatura especializada); otras son sensibles a la tasa de modulación (importante quizá para la inteligibilidad del habla); otras a los datos necesarios para la localización espacial, etcétera. La mayoría de los investigadores opina que la tonotopía de la membrana basilar llega hasta el cerebro, pero no cree que la relación isomórfica se conserve plenamente y propone un mapeo múltiple entre ambos sistemas. En todos los casos, el mapeo ocurre entre un sistema unidimensional, la cóclea, y otro bidimensional, la corteza cerebral. Aunque nos representamos al cerebro como una estructura tridimensional, quizá sea más cercano a la realidad pensarlo como una estructura bidimensional –la materia gris– plegada en numerosas convoluciones.



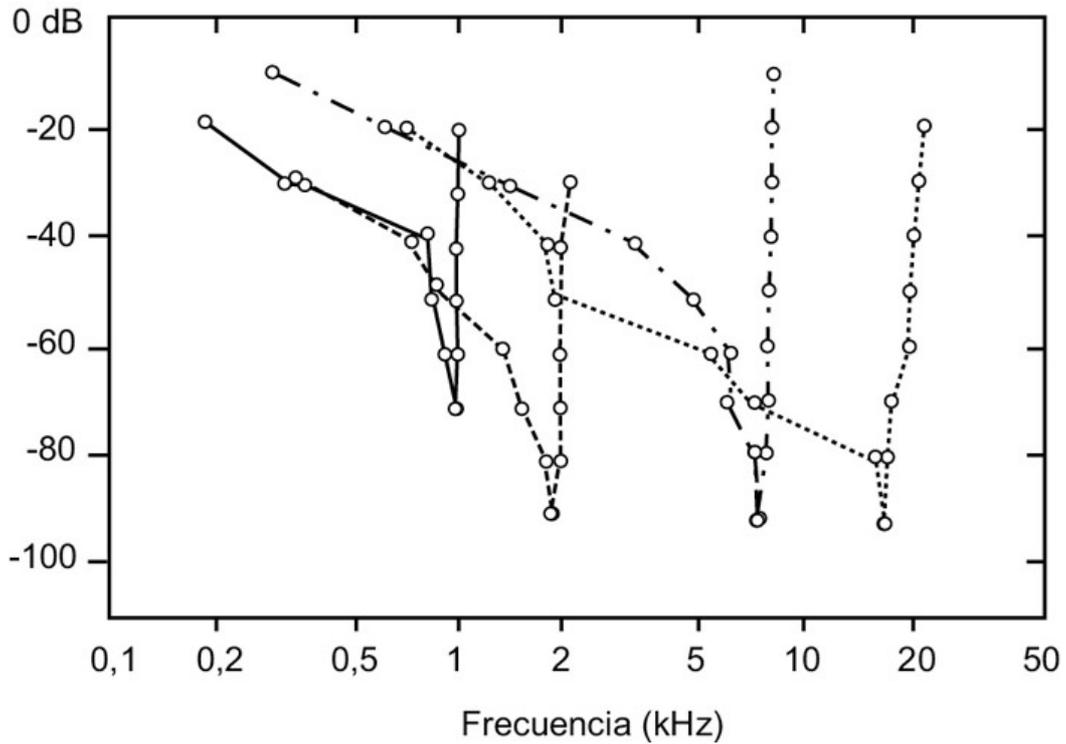
**Figura 28.** Función de transferencia combinada  $H(\omega) E(\omega)$



**Figura 29.** Función de transferencia del oído medio  $C(\omega)$

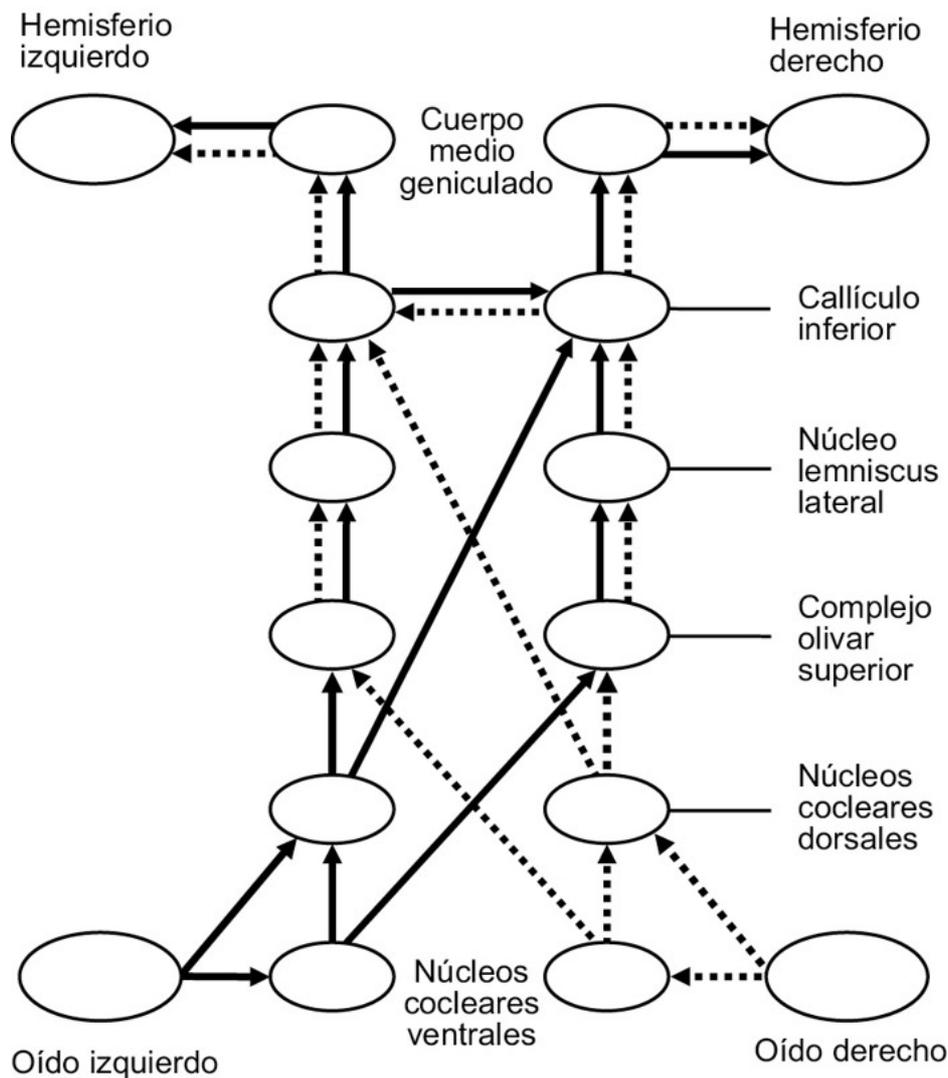


**Figura 30.** Función de transferencia entre el campo externo y la cóclea  $S(\omega)$  (*puntos*) comparada con la sensibilidad promedio del oído humano (*línea*)



**Figura 31.** Umbrales de respuesta de diferentes fibras simples ante diferentes estímulos sinusoidales. La segunda pendiente es del orden de 1.000 dB/octava

Lo que sí se puede afirmar es que la información acústica llega al sistema auditivo central con un grado de detalle extremadamente fino. Por ejemplo, distinguimos diferencias de apenas  $10 \mu\text{s}$  entre las señales que nos llegan a los dos oídos; tomando en cuenta que la duración de los pulsos mismos es de  $100 \mu\text{s}$ , diez veces más largos, la precisión del sistema resulta poco menos que increíble.



**Figura 32.** Recorrido esquemático del flujo de información entre el oído externo y la corteza cerebral

Las propiedades de respuesta en frecuencia de las células ciliadas internas, asistidas por la realimentación positiva provista por las células ciliadas externas, se muestran en la figura 31 para cuatro frecuencias centrales distintas. La segunda pendiente es del orden de 1.000 dB/octava, valor consistente con nuestra capacidad de discriminación diferencial de aproximadamente el 0,3% de la frecuencia central correspondiente.

Se define así el máximo nivel de detalle espectral posible a la entrada del sistema de transducción. De acuerdo con las experien-

cias psicoacústicas en laboratorio, este nivel de detalle se conserva intacto a lo largo de toda la cadena sensorial.

El recorrido esquematizado de la información auditiva, desde que abandona el oído hasta que llega a la corteza cerebral, se puede ver en la figura 32.

Aunque el recorrido y la naturaleza de la información acústica distan de ser comprendidos en profundidad, algunos autores han aplicado esquemas similares al anterior para crear modelos completos de percepción auditiva. Por ejemplo, Yoichi Ando postula un modelo en el que aparece un mecanismo precortical de extracción de la autocorrelación de las señales, una posterior correlación cruzada interaural entre las vías derecha e izquierda y la especialización de los hemisferios cerebrales para el proceso de factores espaciales o temporales. En la figura 33 se puede ver un gráfico del modelo de Yoichi Ando (Ando, Y., 1985, 1998).

Para ciertas tareas complejas, como las involucradas en el “efecto de cafetería”,[\[10\]](#) es necesaria la interacción de ambos hemisferios: el izquierdo –temporal– tiene a su cargo el procesamiento de la palabra, mientras que el derecho –espacial– procesa la localización de las fuentes. El modelo de Ando se ha empleado con éxito en el diseño de salas para música no amplificadas.[\[11\]](#)

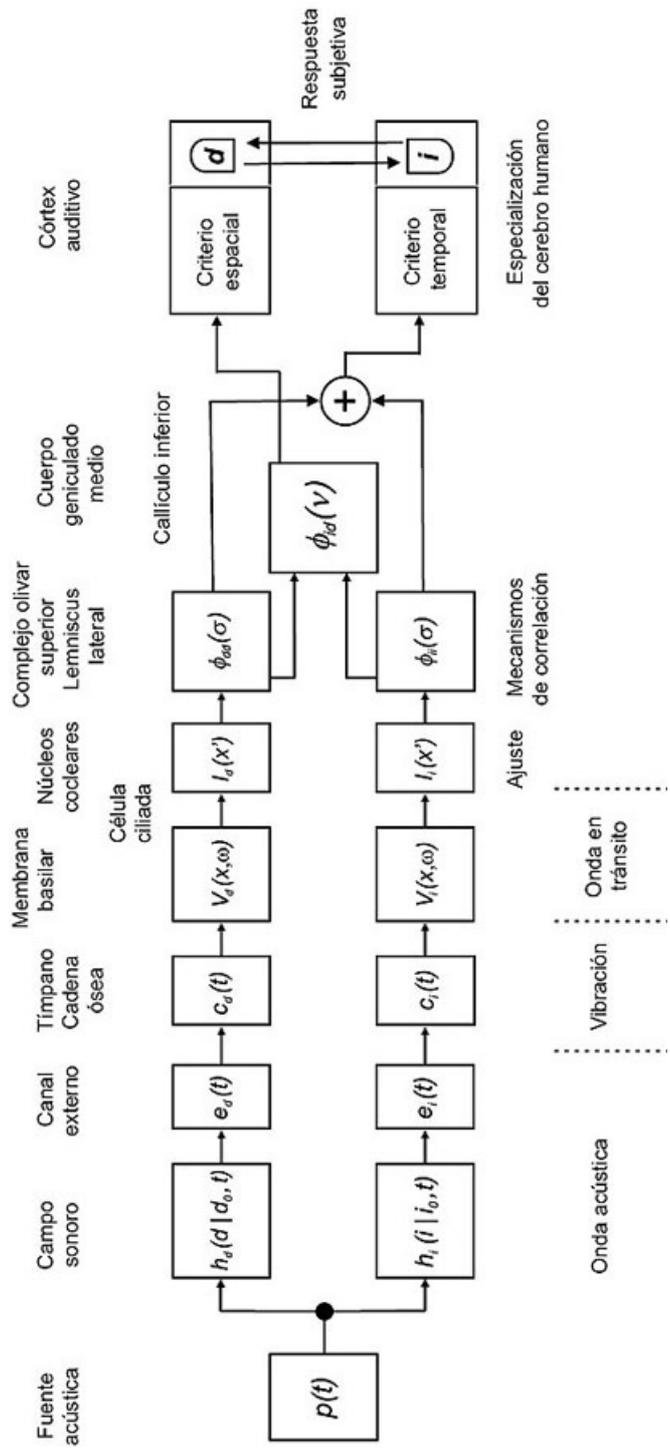


Figura 33. Modelo de Yoichi Ando

## Notas

- [1](#) La parte de la acústica que estudia la fisiología de la audición se denomina *fisioacústica*.
- [2](#) *Head related transfer functions* (HRTF). Este tema será estudiado en detalle en el capítulo VII.
- [3](#) Algunos de los factores que aumentan el riesgo de hipoacusia son el uso de auriculares en ambientes con gran nivel de ruido de fondo, por ejemplo la vía pública, y el empleo de amplificadores de audio de baja calidad, con deriva estática y alto porcentaje de “rumble” (ruido de muy baja frecuencia).
- [4](#) Esta estructura es similar a la de casi todos los vertebrados terrestres. Las especies que no poseen membrana timpánica, como las salamandras y las serpientes, son la excepción.
- [5](#) Algo similar ocurre cuando un percusionista “apaga” el sonido de un timbal.
- [6](#) Le toma cerca de un quinto de segundo a un mensaje llegar desde la cabeza de un organista hasta sus pies –en un pasaje rápido éste debe compensar este retardo indicándole a sus pies qué hacer antes que suene la nota.
- [7](#) En el teclado de un piano se realiza un mapeo espacio-frecuencia: cada punto del espacio –cada tecla– corresponde a una frecuencia diferente. En esta analogía la membrana basilar vendría a ser una especie de anti-piano.
- [8](#) El comportamiento de la membrana basilar en este caso es similar al definido por la *transformada wavelet* (Basso, G., 2001a).
- [9](#) La relación expuesta es válida sólo para fuentes acústicas localizadas frente al oyente. En el capítulo VII se analizará el caso general.
- [10](#) Efecto que describe la capacidad de comprender el habla en ambientes de gran contaminación por ruido. Donald Hall lo denominó “cocktail party effect”.
- [11](#) Por ejemplo, en el Auditorio de Kirishima, en Japón, diseñado por Yoichi Ando.

## Capítulo II

### Sonoridad

Es razonable relacionar la sonoridad —una magnitud perceptual— con la intensidad de la señal física que la origina. El mecanismo parece claro: una onda acústica produce cierta presión dinámica en el aire a la entrada del oído externo y genera movimientos de amplitud proporcional en el tímpano, en los huesos del oído medio, en la ventana oval y en la membrana basilar. El movimiento de la membrana basilar, a su vez, implica la estimulación de las células ciliadas y la aparición de una cantidad de impulsos nerviosos viajando por el nervio auditivo. La relación causal *intensidad física* → sonoridad se insinúa con aparente claridad. Pero, ¿es legítimo considerar a la sonoridad como un rasgo autónomo, que se puede estudiar de manera independiente, separada del resto de las características que definen un sonido real? Es prácticamente una perogrullada decir que todos los sonidos que nos resultan relevantes están caracterizados por diferencias en la intensidad del espectro a lo largo del tiempo, pues la percepción de la intensidad del estímulo acústico está presente en todos los aspectos de la función auditiva. Pero la intensidad en sí misma, aunque resulte vital para la función superior de extraer información de los sonidos del ambiente, no es un rasgo de interés primario para la audición.

Aclarado lo anterior, debemos reconocer que el estudio autónomo de la percepción de la intensidad es uno de los temas tradicionales de la investigación acústica y ha generado gran cantidad de información relevante. Durante los últimos diez años se ha incrementado notablemente nuestro conocimiento sobre una de sus principales características, que es la forma en que el sistema auditivo representa o codifica la intensidad de las señales acústicas.

En lo que sigue describiremos el estado actual de los estudios sobre el tema, tratando de no perder de vista nuestro objetivo cen-

tral –a cumplirse unos capítulos más adelante: integrar cada uno de los rasgos en una imagen global y completa del sonido tal como efectivamente lo percibimos.

Por el momento, podemos adoptar la siguiente definición de sonoridad, que nos va a servir para comenzar su estudio: “La sonoridad es el rasgo de la sensación auditiva en términos del cual los sonidos pueden ser ordenados en una escala que se extiende del silencio a lo muy sonoro”.[\[1\]](#)

En esta definición, la sonoridad aparece como un atributo general del sonido y se refiere, ante todo, a la sonoridad global de un objeto acústico. Más adelante la ampliaremos para analizar la sonoridad de los diferentes elementos que componen dichos objetos.

## **1. Relación entre amplitud e intensidad**

En el caso de una onda sinusoidal, la amplitud no es más que una manifestación de la cantidad de energía presente en el sistema. Como esta energía no se puede medir directamente y es muy difícil determinar la amplitud del movimiento de las moléculas, se acostumbra medir algún parámetro más accesible. De todos los parámetros relacionados con la energía de la onda, el más sencillo de evaluar es la presión dinámica ( $p$ ), que puede ser medida empleando directamente los micrófonos de presión clásicos. La presión es una magnitud que vincula una fuerza con la superficie sobre la que se aplica, y la unidad que le corresponde es el pascal –que equivale a una fuerza de un newton aplicada sobre una superficie de un metro cuadrado. Una onda acústica que se propaga en el aire está formada por una sucesión de zonas de presión por encima ( $p$  positiva) y por debajo ( $p$  negativa) de la presión atmosférica media. La presión atmosférica normal, a nivel del mar, es de aproximadamente 101.300 Pa y se mantiene relativamente constante durante intervalos de tiempo breves. [\[2\]](#)

### **Decibeles**

Se sabe que para obtener incrementos de sonoridad en pasos perceptualmente iguales los incrementos en la presión del estímulo deben ser cada vez mayores. O, con mayor precisión, para percibir un aumento de sonoridad constante hay que aumentar el estímulo un porcentaje también constante. A mediados del siglo XIX Weber y Fechner[3] condensaron matemáticamente este comportamiento en dos leyes que permitieron la creación de una unidad logarítmica, el decibel, mucho más adecuada para describir la relación entre la energía del estímulo físico y la sonoridad resultante.[4] De aquí se obtiene una de las magnitudes más usadas en acústica, el nivel de presión sonora (NPS),[5] cuya unidad es el decibel (dB):

$$NPS = 20 \log_{10} \frac{p}{p_{ref}} [dB]; \text{ con } p_{ref} = 20 \mu Pa$$

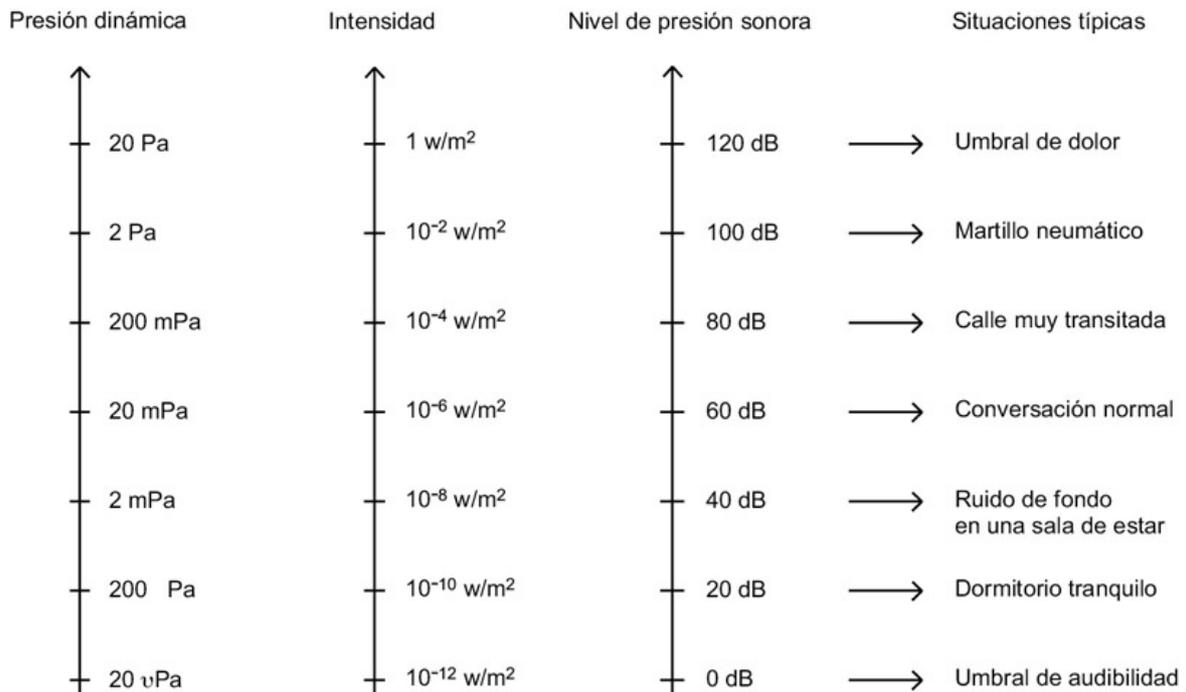
En esta relación  $p$  es la presión acústica en un punto del espacio y  $p_{ref}$  es una referencia convencional estandarizada. Hay varios elementos a destacar en esta ecuación: en primer término, el nivel compara dos presiones dinámicas entre sí, indicando cuál de ellas es mayor y en qué proporción. En segundo lugar, para poder comparar dos niveles de presión en dB éstos deben haber sido tomados según una misma presión de referencia ( $p_{ref}$ ). El estándar internacional estipula que la referencia debe coincidir con el umbral de audibilidad a 1.000 Hz, que corresponde a una presión de 20  $\mu Pa$ . Así se consigue vincular de manera directa una unidad física, el decibel, con una de las principales características de la audición humana. Por último, el enorme rango físico presente en los estímulos acústicos, en el que la mayor presión audible supera un millón de veces a la menor, se reduce considerablemente al ser expresado en decibeles. En la figura 1 se aprecia la relación entre la presión dinámica de la señal, expresada en pascales, y el nivel de presión sonora en decibeles. Para que el lector pueda vincular estos valores

con los sonidos cotidianos, en la cuarta columna se hace referencia a los contextos acústicos con los que habitualmente se asocian. Se observa que el valor de 0 dB corresponde a una presión de 20  $\mu\text{Pa}$ , el umbral de audibilidad, y no a presión cero. Por su parte el umbral de dolor ronda los 120 dB, aunque este límite varía apreciablemente según la sensibilidad de cada individuo.

## Potencia e intensidad acústica

Toda onda acústica nace en una fuente cuya función es transferir energía al medio. La cantidad de energía (E) que entrega por unidad de tiempo se denomina potencia (P) y constituye una de sus características más importantes. La potencia se mide en watts (W) y se vincula con la energía a través de la siguiente relación:

$$P = \Delta E / \Delta t \text{ [W]} = [\text{Joule}]/[\text{s}]$$



**Figura 1.** Relación entre presión dinámica y nivel de presión sonora en decibeles

Es decir que la potencia describe la variación de energía  $\Delta E$  que ocurre en el tiempo  $\Delta t$ .[\[6\]](#) En otros términos, es el trabajo físico ejercido por la fuente durante el tiempo  $\Delta t$ . La potencia acústica máxima que entregan algunos instrumentos musicales se puede apreciar en la siguiente tabla (los valores son un promedio entre varias mediciones):

<i>Fuente acústica</i>	<i>Potencia acústica máxima</i>
Cantante (tenor)	0,04 W
Flauta	0,05 W
Trompeta	0,2 W
Trombón	6 W
Orquesta sinfónica (fff)	60 W

La energía total que radia la fuente se propaga en muchas direcciones. En particular, a nosotros nos va a interesar la que atraviesa determinada área (por ejemplo, el tímpano de una persona). Otra magnitud, la intensidad acústica ( $I$ ), da cuenta de la potencia que atraviesa un área  $A$ :

$$I = P / A \text{ [W/ m}^2\text{]}$$

Es posible demostrar que la intensidad acústica de una onda progresiva plana es proporcional al cuadrado de su presión dinámica:

$$I = \frac{p^2}{\rho_0 c} \text{ [W/m}^2\text{]}$$

en la que  $\rho_0$  es la densidad media del aire y  $c$  es la velocidad de propagación del sonido en ese medio. En la figura 1 se pueden ob-

servar los valores que toma la intensidad acústica ante algunas fuentes acústicas cotidianas.

A partir de la intensidad  $I$  es posible definir el nivel de intensidad acústica (NI): [7]

$$NI = 10 \log_{10} \frac{I}{I_{ref}} [dB]; \text{ con } I_{ref} = 10^{-12} \text{ W / m}^2$$

Para el caso de ondas planas o esféricas progresivas el valor del nivel de intensidad NI coincide aproximadamente con el del nivel de presión sonora NPS.

Cualquiera de las magnitudes anteriores (amplitud, presión, NPS, intensidad y NI) puede deducirse de las demás: todas se refieren, en última instancia, a la cantidad de energía asociada a una onda acústica.

## 2. Percepción de la sonoridad

La sonoridad es una magnitud perceptual y como tal no puede medirse directamente a partir de sensores físicos. Puesto que todas las técnicas indirectas de medición aportan sus propios presupuestos y limitaciones, nuestra discusión incluirá necesariamente una descripción de la metodología aplicada en cada uno de los casos expuestos. Quizá la menos controvertida de las técnicas de medición sea la de comparación de sonoridades, en la que un oyente debe variar la intensidad de un estímulo hasta conseguir igualarlo en sonoridad con otro estándar de intensidad fija. En este método el oído se usa como un detector de cero y es capaz de revelar cómo otros parámetros de la señal, tales como la frecuencia o el ancho de banda, afectan la sonoridad. También sirve para detectar la manera en que alguna característica particular del oyente, como una pérdida de sensibilidad auditiva, influye sobre su percepción de la sonoridad.

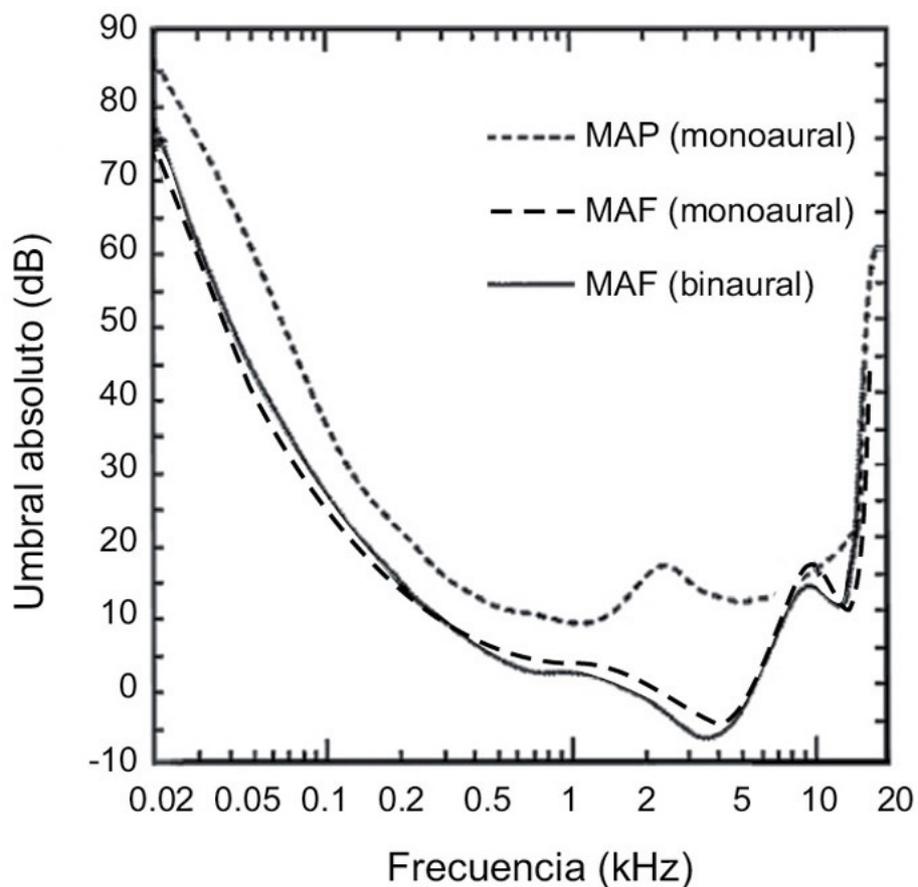
## Umbrales absolutos

El umbral absoluto de intensidad –la intensidad a partir de la cual comenzamos a percibir sonido– fue determinado por primera vez en la segunda mitad del siglo XIX. Se lo midió a una frecuencia central de 1.000 Hz y se supuso que el oído humano mantenía un comportamiento parejo a lo largo de todo el rango audible. Hoy sabemos que, para una señal sinusoidal de 1.000 Hz, la mayoría de las personas con oído sano comienza a oír cuando la intensidad alcanza unos  $10^{-12}$  W/m<sup>2</sup>. Esta intensidad equivale a una presión sonora de 20 millonésimas de pascal (20  $\mu$ Pa), o a una amplitud de movimiento de  $10^{-11}$  m (o 0,1 Å ¡un décimo del diámetro del átomo de hidrógeno!). El oído humano es increíblemente sensible a la intensidad acústica, pues si pudiésemos percibir intensidades aún menores oíríamos el ruido provocado por la agitación térmica de las moléculas del aire. Cuando comparamos nuestros oídos con los de algunos animales nocturnos, como los gatos o las lechuzas, no debemos buscar en una mayor sensibilidad absoluta la razón de sus excelentes capacidades auditivas (no al menos a 1.000 Hz).

La variación del umbral de audición en función de la frecuencia del estímulo se obtiene habitualmente a partir de dos métodos: el de la presión mínima audible (MAP), que se evalúa con auriculares, y el del mínimo campo audible (MAF), que se mide con altavoces y que incluye los efectos de filtro del canal auditivo externo y del pabellón auricular (de gran importancia para definir los filtros HRTF). [8] La figura 2 permite apreciar, junto a estos umbrales, el correspondiente al mínimo campo audible binaural (MAF binaural), que es alrededor de 2 dB menor que el monoaural.

Se consideran sujetos con audición normal aquellos que presentan umbrales de hasta 20 dB por encima de la curva anterior. Para frecuencias superiores a los 1.000 Hz las diferencias interindividuales son apreciables.

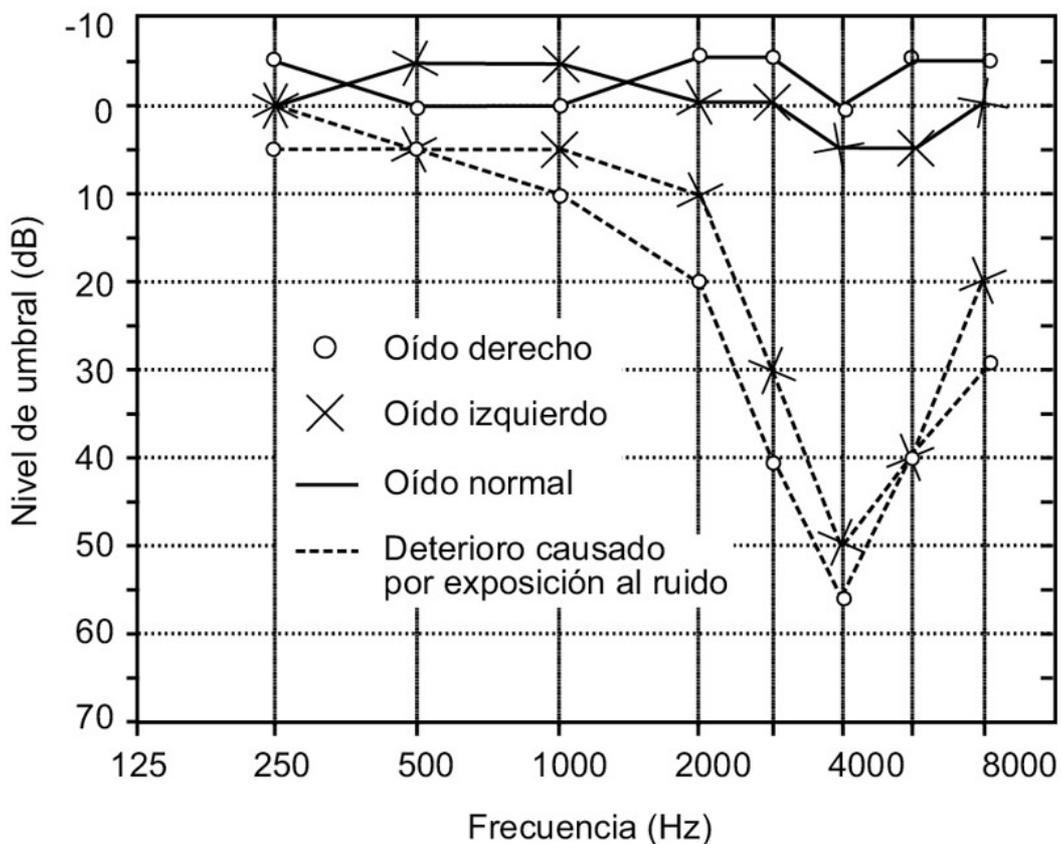
Es notable la semejanza de forma que existe entre la sensibilidad del oído derivada de los filtros del sistema auditivo periférico (figura 30 del capítulo I) y la inversa de la curva umbral que acabamos de presentar. Además de proporcionar la identificación puntual de la causa de cada resonancia y de cada factor de ponderación espectral, la semejanza aludida permite suponer una codificación espectral “plana” en los niveles superiores de procesamiento neural, a partir del flujo de información que asciende por el nervio auditivo. Así, podemos inferir que tanto la sensibilidad máxima del oído alrededor de los 3.000 Hz, como la característica de filtro pasa altos que presenta hasta los 5.000 Hz, se originan principalmente en la resonancia acústica del conducto auditivo externo.



**Figura 2.** Mínimas intensidades acústicas audibles (umbrales MAP y MAF)

En cuanto a los límites de audibilidad para bajas frecuencias, podemos apuntar que Whittle midió en 1972 los umbrales entre 3,5 y 50 Hz y concluyó que forman un continuum con las curvas a frecuencias medias (a 3,5 Hz se necesitan 120 dB para llegar al umbral) (Whittle, L. S. et al., 1972). En 1974 Jonson sugirió, no obstante, que para sonidos de menos de 16 Hz lo que se oye es el producto de la distorsión armónica del oído (Johnson y Gierke, 1974). El límite para situaciones normales parece ser de alrededor de 16 Hz, frecuencia cercana a la más baja que produce una sensación clara de altura tonal.

En la práctica clínica los umbrales se toman con relación a los considerados “estadísticamente normales” y se miden en niveles de umbral (HL): un umbral de 40 dB HL a 100 Hz significa que está 40 dB por encima del umbral normal a dicha frecuencia. En el audiograma de la figura 3 se pueden ver los niveles de umbral HL para ambos oídos.



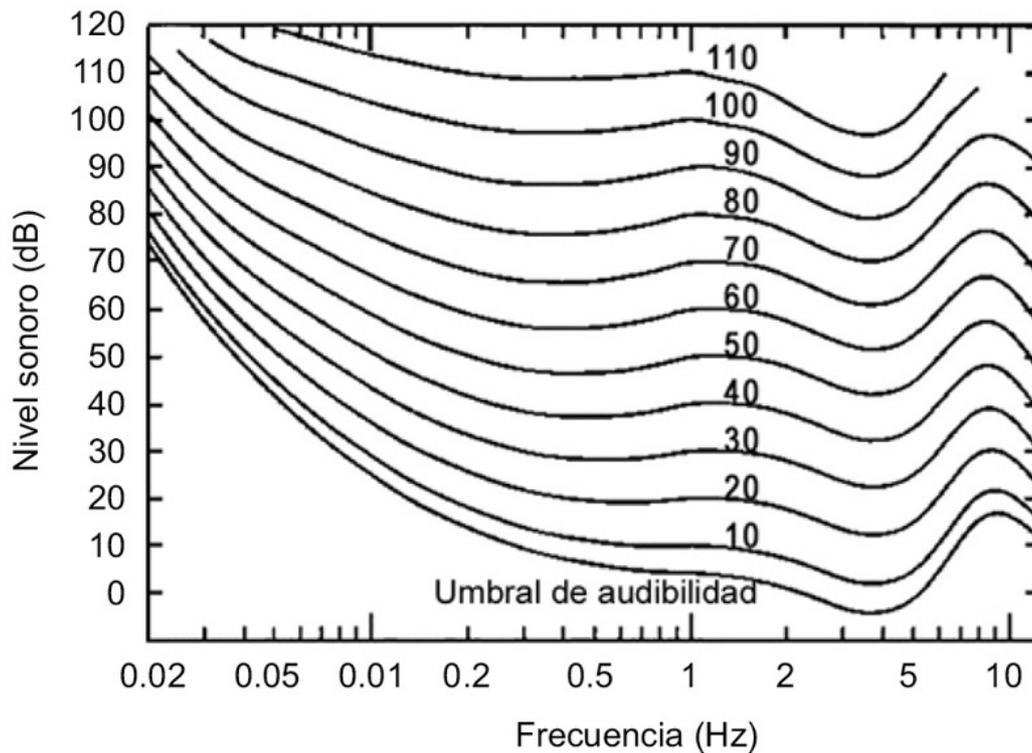
**Figura 3.** Audiograma de un sujeto con oído sano y de otro con hipoacusia provocada por ruido

La sensibilidad del oído humano a la intensidad es destacable tanto en términos absolutos como en términos del rango de valores al que responde (del orden del billón de veces entre la máxima y la mínima intensidad perceptible, o del millón si lo medimos en presiones acústicas). [9] El motivo de semejante rango hay que buscarlo en el conjunto de sonidos habituales en los tiempos en los que estas capacidades se fijaron biológicamente. La gran sensibilidad ante estímulos minúsculos, del orden del decibel, nos permitía oír por ejemplo el andar distante de un depredador o de una pequeña presa. En aquellos tiempos también existían estímulos de gran energía, como el sonido de un trueno o de un grito a corta distancia. La estrategia que nos permitió conservar una sensibilidad auditiva extrema sin que por ello aumente excesivamente la fragilidad del sistema es compleja y aún no se conoce completamente. Entre los factores de importancia que intervienen está, sin duda, el efecto de lupa propio de la compresión logarítmica que opera entre el estímulo y la respuesta: nuestra sensibilidad es máxima a pequeñas intensidades y se va reduciendo a medida que la intensidad aumenta. Así, para estímulos cercanos al umbral de dolor, el conjunto de mecanismos periféricos y centrales que se describieron en el capítulo I atenúan considerablemente la señal que nos llega del exterior.

### **Contornos de igual sonoridad**

Existen dos técnicas que permiten medir la sonoridad: una es la estimación de magnitud, que intenta determinar valores absolutos de sonoridad; la otra consiste en medir el nivel de sonoridad en relación con una senoide de 1.000 Hz. Las curvas de contornos de igual sonoridad de la figura 4 se obtienen a partir de este último

método y resultan algo diferentes según el laboratorio en el cual fueron determinadas (la forma exacta no debe tomarse literalmente). Pueden ser reproducidas con sencillez al pedirle a un oyente que iguale la sonoridad de una señal sinusoidal, de frecuencia e intensidad variable, con la de un tono estándar de 1.000 Hz a N decibels. Obvia decir que son un promedio elaborado sobre la base de las respuestas de cientos o miles de sujetos.



**Figura 4.** Curvas isófonas (Fletcher-Munson)

Las curvas de igual sonoridad —o isófonas— son una descripción de la variación de la sonoridad en función de la frecuencia del estímulo (Fletcher y Munson, 1933). El nivel de sonoridad (medido en fones) del sonido de prueba es numéricamente igual a la intensidad NI (medida en dB) del tono de referencia de 1.000 Hz. Así, por ejemplo, todo sonido que posea la misma sonoridad que una sinusoide de 40 dB a 1.000 Hz, tendrá un nivel de sonoridad de 40 fones. Un contorno de igual sonoridad es una línea que une los nive-

les de los estímulos –de diferentes frecuencias– que presentan el mismo nivel de sonoridad en fones.

La cantidad de información útil que aportan estas curvas es muy grande. Mencionemos algunos ejemplos:

- La forma de las curvas es, para bajos niveles de sonoridad, parecida a la del umbral de la figura 2: la sensibilidad decrece a medida que nos alejamos de la zona central del espectro hacia las altas y bajas frecuencias. Es necesario aplicar más energía en la región grave que en la media para obtener el mismo resultado auditivo.
- Para sonoridades intermedias, entre 60 y 70 fones, es menor la diferencia de sensibilidad auditiva en función de la frecuencia.
- Las curvas de nivel, para altas sonoridades, son prácticamente planas: nuestro oído responde de manera similar en todo el rango audible a consecuencia de la atenuación provista por los mecanismos de protección auditiva.
- Existe una zona de máxima sensibilidad entre los 2.500 y los 3.500 Hz. Tiene una gran importancia pues el estímulo que ocupe esa banda va a ser percibido con mayor facilidad que los que ocupen otras regiones del espectro. Como ya sabemos, este máximo de sensibilidad se origina en la resonancia del conducto auditivo externo.
- Si se genera una señal sinusoidal que recorra completamente una de las curvas de igual sonoridad se oye un glissando perfecto, de sonoridad constante.

Una consecuencia de las curvas isófonas, muy común en el campo de la música, es el control de sonoridad –o loudness– que poseen algunos preamplificadores de audio. Las grabaciones de piezas de música se realizan por lo general a niveles altos, donde la sensibili-

dad del oído es plana. Si se reproducen a niveles menores cambia el balance espectral al cambiar nuestra sensibilidad auditiva: oímos debilitadas las regiones más graves y agudas del espectro. Para corregir este efecto algunos equipos de audio vienen equipados con un control de sonoridad que refuerza los extremos del espectro a bajos niveles de amplificación. La salida del preamplificador con el control de sonoridad activado sigue de modo más o menos feliz la curva umbral de audibilidad.

## **Escala absoluta de sonoridad**

El nivel de sonoridad en fones todavía se refiere, sin embargo, a la señal física. Sólo proporciona una escala comparativa y no permite describir la sonoridad propiamente dicha. ¿Cuánto más sonoro es un sonido de 70 fones que uno de 50 fones? Con los elementos hasta aquí evaluados no se puede dar respuesta a esta pregunta. Hace falta definir una unidad auténticamente perceptual, que permita medir la magnitud de las percepciones en sí mismas y no con relación a referencias externas.

## **La ley psicofísica**

Este problema se encaró inicialmente en la segunda mitad del siglo XIX. Los fisiólogos alemanes Weber y Fechner propusieron una ley psicofísica que pretendía vincular los estímulos con las percepciones a través de una relación común a todos los sentidos. Esta ley fue considerada adecuada y se la reprodujo en la mayoría de los textos especializados hasta no hace mucho tiempo. Aunque dio lugar a una variedad de técnicas y unidades de medición entre las que se encuentra el decibel, en la mayoría de los casos no es ni aproximadamente correcta. La definición más difundida de la ley psicofísica establece que la magnitud psicológica percibida es proporcional al logaritmo del estímulo físico. Es decir, que multiplicaciones similares en la magnitud del estímulo generan adiciones proporcionales en la respuesta perceptual, transformando una escala

geométrica en una escala aritmética. En el caso del sonido, lo anterior significa simplemente que nuestros juicios sobre la sonoridad deben seguir exactamente una escala logarítmica similar a la pautada en la escala de decibeles: un estímulo de 80 dB debería tener una sonoridad doble que uno de 40 dB. O que a cada adición de 10 dB le debería corresponder un paso siempre igual en sonoridad. Como veremos a continuación, nuevas mediciones psicoacústicas refutaron estas afirmaciones.

## **La sonoridad**

A pesar de que las técnicas que emplean al oído como un detector de cero –como la de comparación de sonoridades– han probado ser útiles en la determinación de los factores que afectan la sonoridad, no proveen una medida directa de la sensación percibida. Uno de los procedimientos empleados para resolver el problema fue aplicar la técnica de estimación de magnitudes, en la que el oyente debe simplemente asignarle un número a cada sonido de diferente intensidad. Una alternativa consiste en emplear la técnica de producción de magnitudes en la que, a la inversa, el oyente debe ajustar la intensidad de un estímulo de acuerdo a un número dado de antemano. Durante la década de 1950 Stevens empleó ambos métodos para desarrollar sus escalas de sonoridad, que describen la relación entre sonoridad –perceptual– e intensidad –física– (Stevens, S. S., 1959). Los problemas metodológicos que encontró fueron muy grandes pues resulta difícil cuantificar la sonoridad que es, como ya sabemos, una magnitud enteramente perceptual. Stevens tuvo que definir una nueva unidad, el son, para cuantificar directamente la sonoridad. El son es, finalmente, una unidad lineal: un sonido de 2 sones posee el doble de sonoridad que uno de 1 son, y uno de 3 sones el triple. Al igual que la altura, la sonoridad es una magnitud perceptual y debería definirse a partir de una unidad de referencia estable. Pero como la estabilidad en el campo de las sen-

saciones es muy poco confiable, se toma una referencia física externa: una señal 1.000 Hz y 40 dB (40 fones) produce por definición un sonido que posee una sonoridad de 1 son (para la altura también se toma una referencia física externa: 440 Hz definen el La en música).[\[10\]](#)

A partir de los trabajos de Stevens se encontró que la relación estímulo-respuesta sigue aproximadamente una ley de potencia, diferente a la ley logarítmica de los psicofísicos. Propuso la siguiente relación entre la sonoridad (S) y la intensidad (I):

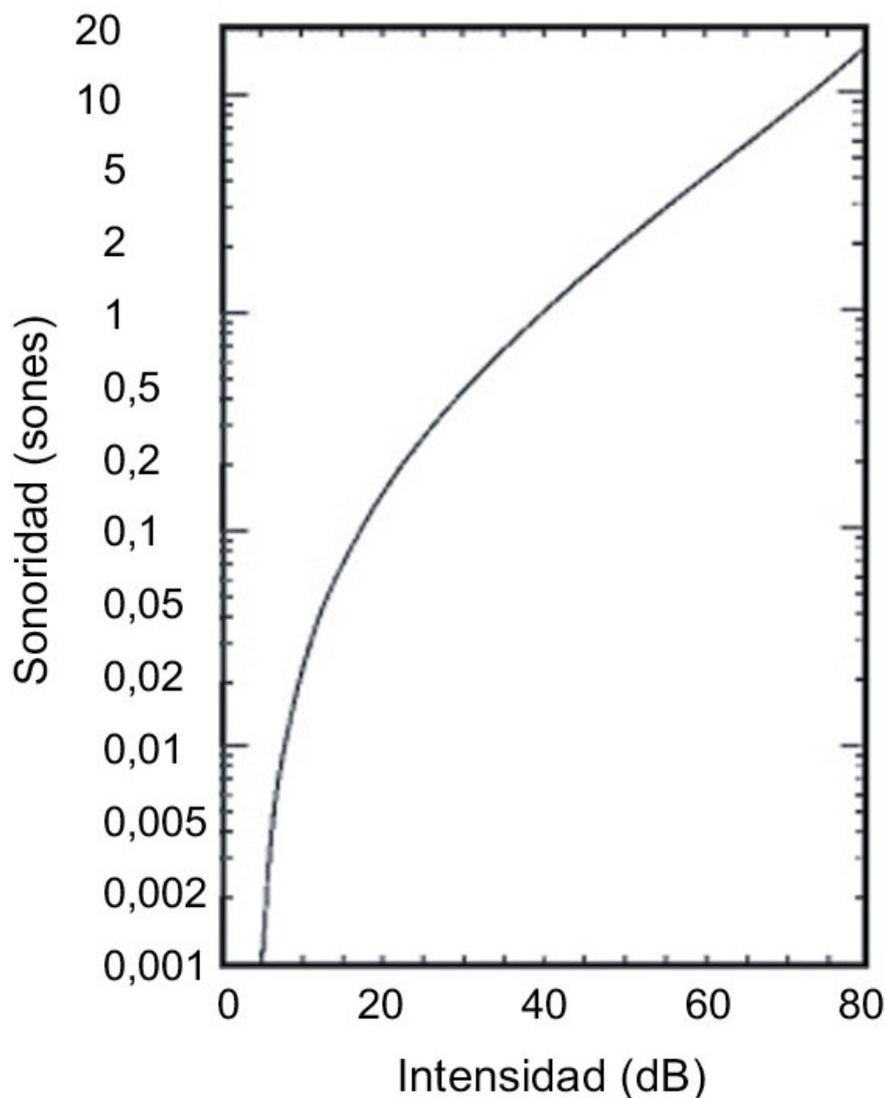
$$S = k I^{0,3} \text{ [sones]}$$

El valor de la constante k depende de las condiciones experimentales y de las unidades empleadas. El gráfico de la figura 5 muestra la relación de Stevens (no hay que olvidar que es consecuencia de experimentos realizados con ondas sinusoidales y no con estímulos complejos). Se puede observar que a un aumento de 10 dB en el estímulo le corresponde una duplicación de la sonoridad.

Aunque la relación anterior ha sido ratificada para señales mayores a 40 dB, se le hacen varias críticas: existen diferencias individuales muy grandes, y el resultado se obtuvo a partir de medias de poblaciones muy grandes y dispersas. Warren encontró que la duplicación de la sonoridad corresponde a un aumento de la intensidad de 6 dB, en vez de los 10 dB de Stevens (Warren, R. M., 1970). Además, aunque las técnicas de estimación y de producción de magnitudes tienen la ventaja de permitir la obtención de una relación directa entre un parámetro físico y un rasgo perceptual, a diferencia de la técnica de comparación de magnitudes, ambas asumen que los oyentes pueden usar los números para calificar sus percepciones de un modo consistente. Existen varias razones que permiten suponer que esto último no siempre es así, que los oyentes polarizan y sesgan sus juicios de estimación acerca de la sonoridad que perciben. Sin embargo, asumiendo que el riesgo existe y

extremando los cuidados metodológicos, la técnica de estimación de magnitudes es usada ampliamente para medir la sonoridad.

Existe una objeción aún más importante: algunos autores desaprueban el hecho de preguntar por la sonoridad en sí misma. De ordinario, evaluamos ciertas características de las fuentes acústicas como un todo que involucra al contexto, a la perspectiva y a muchos otros factores circunstanciales. La estimación aislada de la sonoridad parece una abstracción muy difícil de realizar. Helmholtz escribió: “[...] estamos muy bien entrenados para evaluar los objetos que nos rodean, pero fallamos completamente cuando intentamos observar las sensaciones per se [...]” (Helmholtz, H. von, 1885).[\[11\]](#)



**Figura 5.** Relación de Stevens entre intensidad y sonoridad

## **Medidores de nivel sonoro**

Los instrumentos de medición más tradicionales en acústica son los medidores de nivel sonoro,[\[12\]](#) que permiten determinar el nivel de presión sonora. Por lo general, tienen un banco de filtros en la etapa de entrada que permite medir la señal acústica tal como llega al micrófono –respuesta plana o lineal– o compensada según cierto patrón espectral. Así, un medidor de nivel sonoro en compensación “A” sigue de manera aproximada la curva de respuesta del oído para un nivel de sonoridad de 30 fones, en compensación “B” la de 70 fones, y en compensación “C” es prácticamente lineal. Casi toda la normativa relacionada con acústica ambiental está especificada en decibeles A. Por ejemplo, el nivel equivalente de ruido aceptable en zonas residenciales no debería superar los 65 dBA en horarios de actividad diurna ni los 50 dBA durante la noche.

Si se toman dos lecturas de una misma señal, una en dB lineales y otra en dBA, por lo general la última dará un valor inferior pues se filtran las bandas graves del espectro de ruido.

Los medidores de nivel sonoro, sin embargo, sólo son confiables cuando se los emplea para medir señales continuas de larga duración. Las respuestas perceptuales ante señales impulsivas o muy cortas no corresponden a la magnitud que indican. Los medidores tampoco proveen lecturas satisfactorias de la sonoridad de estímulos compuestos muy separados en frecuencia. Y, por supuesto, no están calibrados en sones, la unidad que debe emplearse para evaluar la sonoridad propiamente dicha.

## **Sumario de unidades relacionadas con la sonoridad**

El camino que nos llevó de la amplitud de las ondas en el aire a los sones es complejo y se podría suponer de una dificultad innece-

saria. Sin embargo, la auténtica complejidad está dada por la multiplicidad de mecanismos que actúan en la codificación auditiva de la intensidad. Para ilustrar esta situación volvamos a las curvas isófonas de la figura 4: se las esquematizó en dos dimensiones –nótese que no aparece la dimensión sonoridad–, pero en realidad deberían haber sido representadas en tres dimensiones. En la figura 7 se puede ver una porción de la representación tridimensional, más adecuada pero mucho más difícil de leer.

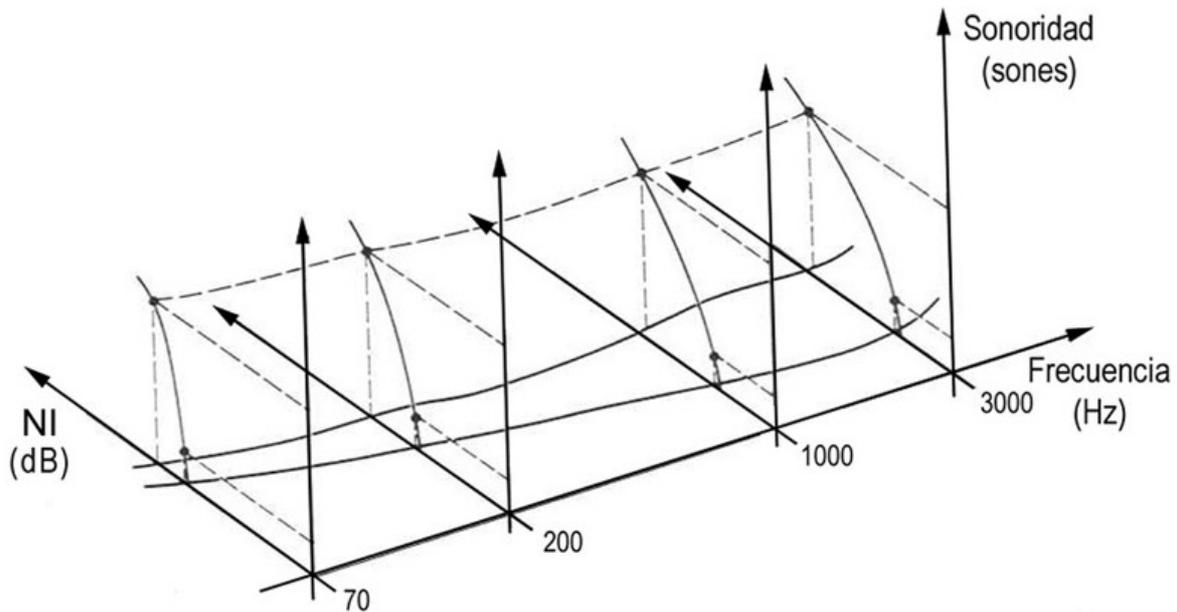
Para sintetizar el camino que llevamos recorrido, la siguiente tabla enumera las diferentes unidades relacionadas con la sonoridad, con sus unidades y características particulares.

**Tabla 1.** Magnitudes relacionadas con la sonoridad (para ondas acústicas planas)

<i>Magnitud</i>	<i>Naturaleza</i>	<i>Unidad</i>	<i>Rango audible</i>
Amplitud	Física	Metro	$10^{-11}$ m - $10^{-5}$ m
Presión	Física	Pascal	20 $\mu$ Pa - 20 Pa
Nivel de presión sonora (NPS)	Física	dB	0 dB - 120 dB
Intensidad	Física	W/m <sup>2</sup>	$10^{-12}$ W/m <sup>2</sup> - 1 W/m <sup>2</sup>
Nivel de intensidad (NI)	Física	dB	0 dB - 120 dB
Nivel de sonoridad	Híbrida	Fon	0 fon - 120 fones
Sonoridad	Psicológica	Son	0 son - 250 sones



**Figura 6.** Medidor de nivel sonoro



**Figura 7.** Representación tridimensional de las curvas isófonas

### 3. Discriminación de la intensidad

La discriminación en intensidad se refiere a la habilidad del sistema auditivo para detectar diferencias entre la intensidad de dos o más estímulos. Durante el siglo XX se realizaron numerosos experimentos con el fin de medir los límites de este significativo aspecto de la audición, que adquirió especial importancia con la llegada del audio digital.

#### Limén de intensidad

Varias definiciones de la diferencia apenas perceptible (DAP o JND [\[13\]](#) en intensidad, también llamada limén de intensidad ( $\Delta L$ ), fueron empleadas en el pasado. En el presente las más usuales son:

- 1)  $\Delta L = \Delta I / I$  conocida como fracción de Weber

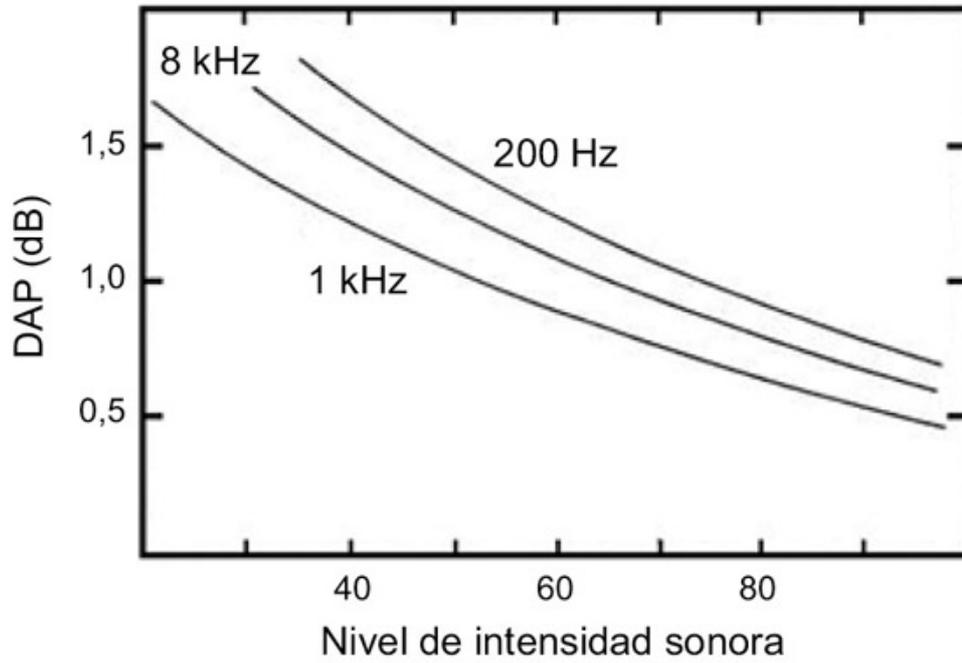
- 2)  $\Delta L = 10 \log (\Delta I/I)$  la fracción de Weber definida en dB
- 3)  $\Delta L = 10 \log \{(I + \Delta I) / I\}$  en dB.

En todas ellas  $I$  es la intensidad de la señal de base (el pedestal) y  $\Delta I$  la intensidad del menor incremento detectable por el oyente.

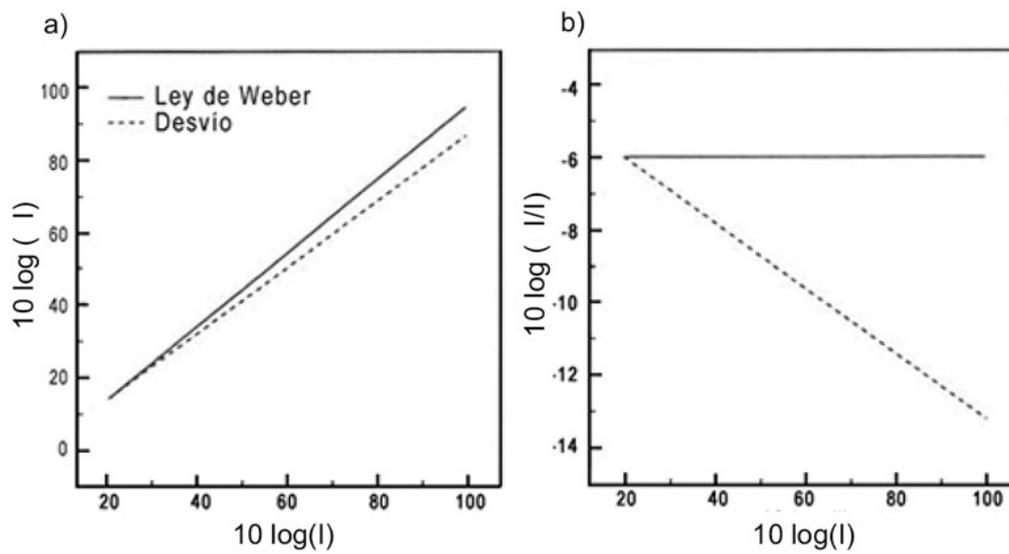
Para medir el limen de intensidad se emplea con frecuencia el método de elección forzada entre dos alternativas (2 AFC), en la que cada sujeto debe optar obligatoriamente por alguno de los sonidos que oye, aún cuando le parezcan indistinguibles. El umbral de discriminación se obtiene cuando la diferencia  $\Delta I$  en el estímulo produce el 75% de respuestas correctas. Esta técnica se aplica a partir de tres métodos principales: método de detección de modulación, de detección de incrementos y de discriminación de estímulos pulsantes. En la detección de modulación los oyentes deben notar la presencia de una pequeña modulación en amplitud en una señal dada. En la detección de incrementos y en la discriminación de estímulos pulsantes deben percibir un cambio en la intensidad de un estímulo estándar (el pedestal).

## **La Ley de Weber**

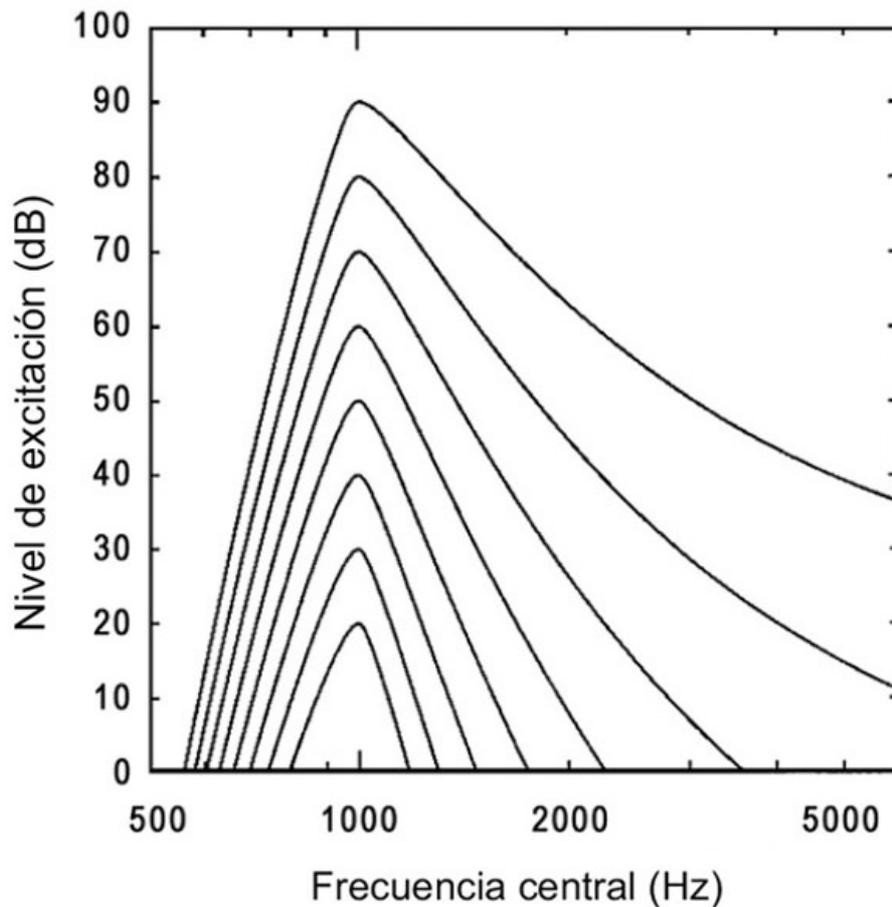
Para ruido de banda ancha se comprobó que el mínimo cambio de intensidad detectable es una fracción constante de la intensidad del estímulo. En otras palabras, la razón  $\Delta I/I$  se mantiene constante. Éste es un buen ejemplo particular de la conocida Ley de Weber, que establece que nuestro sistema perceptivo discrimina cambios porcentuales constantes. La relación  $\Delta I/I$  se llama fracción de Weber y es habitualmente expresada en dB como  $\Delta L = 10 \log (\Delta I/I)$ . En el rango de 20 a 100 dB el valor de  $\Delta L$  permanece entre 0,5 y 1 dB. Cerca del umbral de audibilidad la fracción de Weber aumenta drásticamente, en particular a frecuencias por debajo de los 200 Hz.



**Figura 8.** DAP en intensidad



**Figura 9.** Ley de Weber y desvío de la Ley de Weber para estímulos sinusoidales



**Figura 10.** Ensanchamiento de la zona de la membrana basilar que explicaría el desvío de la Ley de Weber para estímulos sinusoidales

En contraste con los resultados anteriores para ruido de banda ancha, la fracción de Weber sufre un desvío cuando el estímulo se construye con sinusoides: la discriminación mejora al aumentar el nivel del estímulo. A 80 dB  $\Delta L$  llega a ser de 0,3 dB, por debajo de la Ley de Weber. En la figura 9a se puede ver que el gráfico de la relación  $10 \log (\Delta I)$  en función de  $10 \log (I)$  presenta una pendiente de 0,9 en vez de 1, tal como correspondería a la fracción de Weber exacta. De acuerdo con McGill y Goldberg, la diferencia quizá se deba al ensanchamiento de la zona afectada de la membrana basilar a causa del incremento en la intensidad del estímulo sinusoidal (McGill y Goldberg, 1968).

El desvío –aumento en la discriminación– que sufre la Ley de Weber para estímulos sinusoidales puros parece relacionarse con el patrón de enmascaramiento: al aumentar el nivel sonoro se activan otros canales espectrales que contribuyen a la sonoridad general. Este ensanchamiento de la zona de influencia en la membrana basilar se puede apreciar en la figura 10.

## **Ley de Fechner para la sonoridad**

Como ya sabemos, la sonoridad es el correlato perceptual de la intensidad física, al menos en una primera aproximación. Podemos suponer, por lo tanto, la existencia de alguna relación entre la discriminación de la sonoridad y la habilidad de los oyentes para detectar diferencias de intensidad. La Ley de Fechner va un paso más allá proponiendo que la fracción de Weber,  $\Delta I/I$ , está asociada a un incremento constante de la sonoridad  $\Delta S$  (Fechner, G., 1860):

$$\Delta S = k \Delta I/I$$

en la que  $k$  es una constante. Aunque esta relación fue defendida durante más de un siglo, se ha comprobado que es incorrecta, tal como advertimos antes, en este mismo capítulo (Zwislocki y Jordan, 1986).

## **4. Otros factores que intervienen en la percepción de la sonoridad**

### **Integración temporal**

Garner y Miller encontraron que el umbral de audición de la sonoridad corresponde más a una cantidad constante de energía que a una cantidad constante de potencia –o de intensidad–, al menos sobre cierto rango de duraciones (Garner y Miller, 1947). Tiempo después, Henning descubrió que para estímulos de corta duración

la energía del menor incremento detectable ( $\Delta E$ ) se mantiene constante, contra la creencia generalizada de la constancia de la intensidad.[\[14\]](#)

En 1995 Moore comprobó que para duraciones menores a 200 ms la intensidad necesaria para la detección del estímulo se incrementa a medida que la duración disminuye (Moore, B., 1995). Pareciera que el oído integrara la energía de la señal:

$$I \times t = \text{cte.}$$

En la que  $I$  es el umbral de intensidad para un pulso de duración  $t$ . En la práctica la siguiente expresión da mejores resultados:

$$(I - I_L) \times t = I_L \times \tau = \text{cte.}$$

donde  $I_L$  es el umbral para un pulso de larga duración y  $\tau$  es una constante que representa el “tiempo de integración” del sistema auditivo. La relación anterior es válida sólo para intensidades cercanas al umbral. Lejos del mismo no se puede asegurar la forma que adopta la integración de la energía. Solamente se puede decir que, para una intensidad dada, la sonoridad crece hasta los 200 ms de duración. Para menos de 80 ms de duración se puede afirmar que, aproximadamente, energía constante = sonoridad constante. Aquí la fracción de Weber disminuye 3 dB cada vez que se duplica la duración del estímulo. Más allá de una cierta duración crítica  $\Delta t$  la fracción de Weber deviene constante. Para duraciones mayores a 500 ms los umbrales son independientes de la duración.

El valor crítico  $\Delta t$  es función de la frecuencia: decrece al aumentar esta última y toma valores que van desde 100 ms a 250 Hz hasta 10 ms a 4 kHz. Al parecer, integramos mejor los impulsos de alta frecuencia.

En el órgano de tubos no se puede modificar la intensidad de cada sonido porque las teclas sólo abren o cierran las válvulas que

distribuyen el flujo de aire. Sin embargo, la integración temporal a energía constante le permite al organista modular agógicamente la sonoridad, actuando únicamente sobre la duración –al menos en los pasajes veloces.

## Factores espectrales

En un experimento desarrollado en la década de 1950, Zwicker, Flottorp y Stevens le pidieron a un grupo de oyentes que igualasen la sonoridad de una banda de ruido con la de un tono sinusoidal estándar (Zwicker, Flottorp y Stevens, 1957). Los autores concluyeron que si la sonoridad en cada banda crítica sigue la ley de la potencia  $S = k I^{0,3}$ , al distribuir la misma intensidad de la señal en  $n$  bandas críticas se debe obtener un incremento en la sonoridad total simplemente porque:

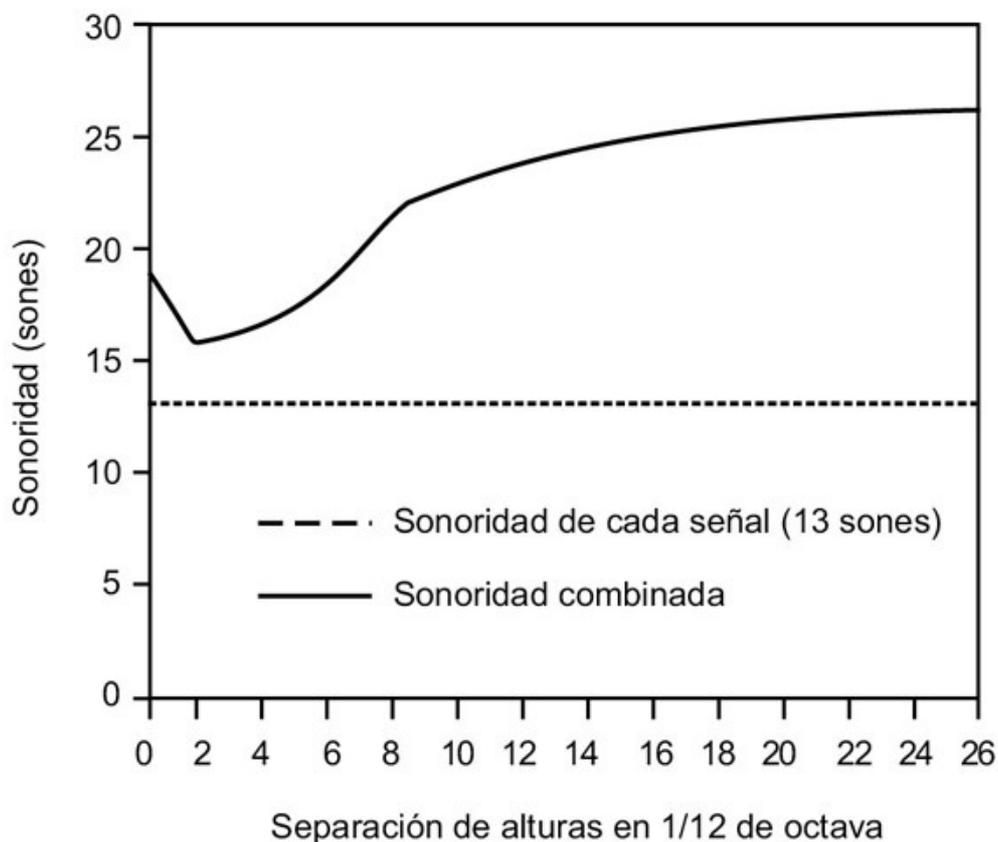
$$k I^{0,3} < nk (I/n)^{0,3}, \text{ para } n > 1$$

En la actualidad se realiza el siguiente análisis para calcular la sonoridad total de un conjunto de estímulos sinusoidales combinados.

1. Nunca se suman los niveles de presión sonora o los niveles de intensidad directamente: dos estímulos de 80 dB jamás generan juntos 160 dB.
2. Si se desean sumar señales de frecuencias cercanas –ambas dentro de la misma banda crítica– se deben sumar en primer término las intensidades eficaces, pues la información de fase no es relevante.[\[15\]](#) Con este resultado se puede obtener el nivel de intensidad en dB. A continuación se extrae de las curvas isófonas (figura 4) el nivel de sonoridad en fones y, por fin, de la relación de Stevens (figura 5) se obtiene la sonoridad total. Por ejemplo, si se suman dos señales de 100 y 120 Hz con el mismo nivel de intensidad de 80 dB, el resultado da un NI

de 83 dB, equivalentes a 77 fones que corresponden aproximadamente a una sonoridad de 12 sones.

3. Si se desea sumar señales de la misma frecuencia el procedimiento es similar, pero en primer lugar deben sumarse las amplitudes porque aquí la información de fase resulta decisiva al aparecer el fenómeno de interferencia de ondas: dos señales de 80 dB de la misma frecuencia pueden dar lugar a una combinación cuyo nivel tenga entre 0 y 86 dB (Basso, G., 2001a).
4. Si la diferencia de frecuencia entre los estímulos supera una banda crítica, en la práctica es usual sumar directamente las sonoridades en sones. Sin embargo, el escenario es algo más complicado, y sigue la curva de la figura 11.
5. Si los estímulos están muy separados en frecuencia, en más de tres octavas, se comportan independientemente uno del otro y no puede hablarse con propiedad de una verdadera suma de sonoridades.



**Figura 11.** Sonoridad combinada de dos sinusoides de 13 sones cada una. La frecuencia de la señal base es de 300 Hz

Un ejemplo nos permitirá comprender cómo se estructura la sonoridad total en un sonido periódico complejo. Consideremos una onda diente de sierra con un nivel de intensidad de 100 dB y una periodicidad de 200 Hz. Las frecuencias, niveles de intensidad y sonoridades de los primeros ocho armónicos serán entonces:

f1 = 200 Hz	NI = 98 dB	Sonoridad = 70 sones
f2 = 400 Hz	NI = 92 dB	Sonoridad = 45 sones
f3 = 600 Hz	NI = 88,5 dB	Sonoridad = 40 sones
f4 = 800 Hz	NI = 86 dB	Sonoridad = 35 sones
f5 = 1.000 Hz	NI = 84 dB	Sonoridad = 30 sones
f6 = 1.200 Hz	NI = 82,5 dB	Sonoridad = 20 sones
f7 = 1.400 Hz	NI = 81 dB	Sonoridad = 18 sones
f8 = 1.600 Hz	NI = 80 dB	Sonoridad = 15 sones

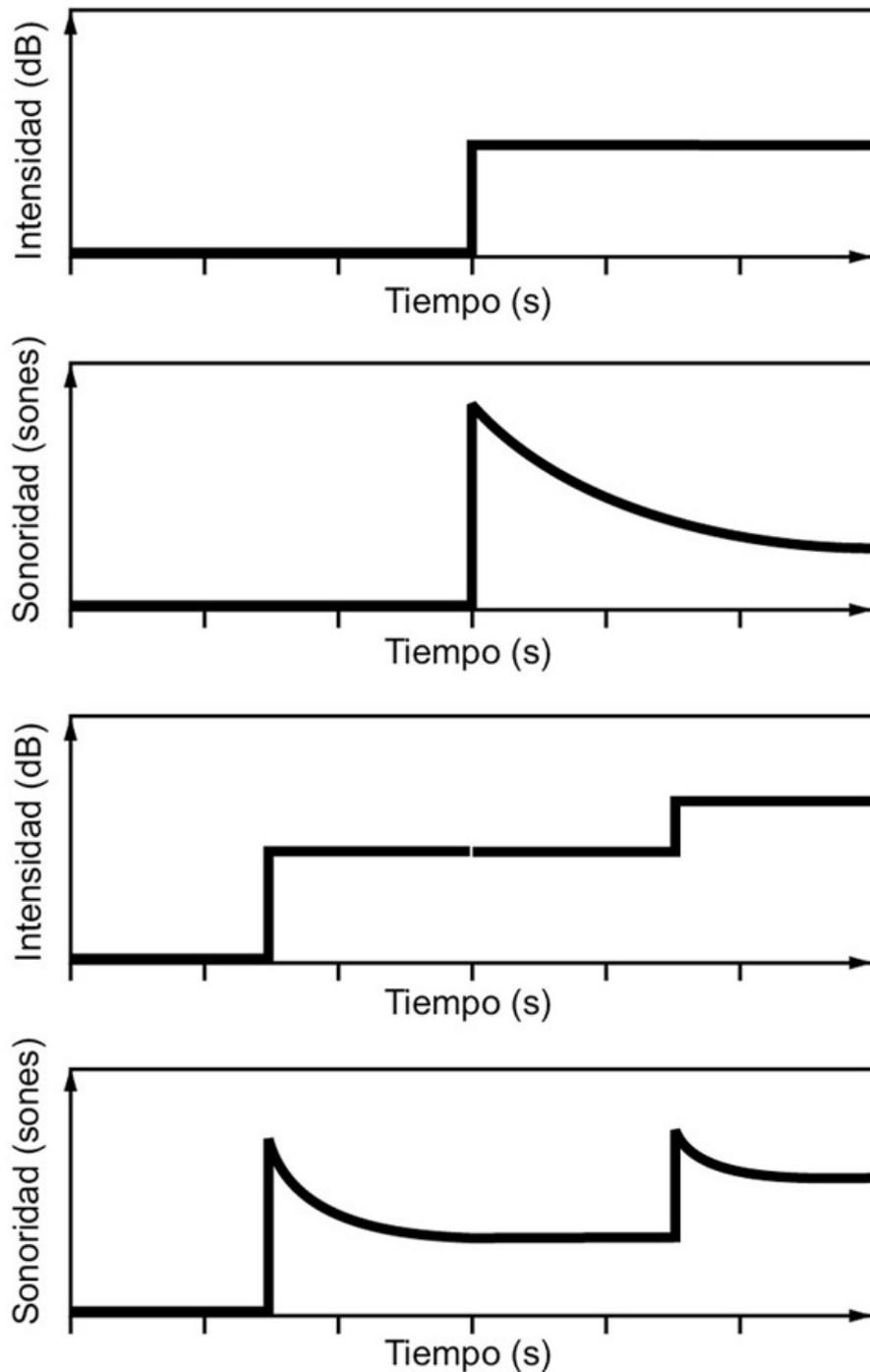
La sonoridad total se obtiene al sumar las sonoridades de los primeros cinco armónicos de acuerdo con la figura 11 –la distancia entre ellos supera el ancho de banda crítico–, agregando luego la ocasionada por los de mayor orden que caen dentro de bandas críticas solapadas. En el ejemplo se suman las intensidades de los armónicos 6, 7 y 8 que contribuyen con 21 sones al conjunto. La suma completa alcanza unos 200 sones. En comparación, una sinusoide de 200 Hz con un nivel de intensidad de 100 dB determina una sonoridad de apenas 80 sones. Se entiende claramente aquí la causa del mayor rendimiento sonoro asociado a una señal periódica compleja. Éste es uno de los motivos que explica el escaso empleo de señales sinusoidales en cualquier sistema de comunicación acústica.

## **Cambios en la sensibilidad auditiva provocados por sobrestimulación**

### ***Adaptación auditiva***

Una propiedad común a todos los sentidos consiste en que la respuesta neural ante estímulos estables de larga duración decae rápidamente, luego de su aparición, hasta alcanzar un cierto “estado de equilibrio”. Esta adaptación auditiva resulta de la interacción entre la energía de respuesta del oyente frente a un estímulo estacionario y la energía necesaria para mantenerlo en actividad. La contraparte psicológica es una declinación en la sonoridad aparente hasta que ésta alcanza un valor constante.

Sorprende, sin embargo, que la sonoridad ante estímulos de media o alta intensidad no parezca decrecer en el tiempo. Sólo los estímulos dentro de los primeros 40 dB por encima del umbral presentan una apreciable adaptación auditiva. En la figura 12 se muestra de manera esquemática esta propiedad de la percepción auditiva.



**Figura 12.** Adaptación auditiva (gráficos esquemáticos)

Las señales de alta frecuencia se adaptan más profundamente que las de baja frecuencia, y los sonidos estables en el tiempo lo hacen mejor que los sonidos modulados. Si cumplen ambas condiciones, es decir, estímulos agudos y estables, los sonidos pueden llegar a

desaparecer completamente luego de una exposición prolongada. Este efecto no ocurre con estímulos de baja frecuencia.

Las mediciones típicas de cambio de umbral por adaptación se realizan comparando las sonoridades en el ataque y tres minutos más tarde. En condiciones favorables, la adaptación puede alcanzar una disminución de sonoridad correspondiente a un cambio de -20 dB en el estímulo.

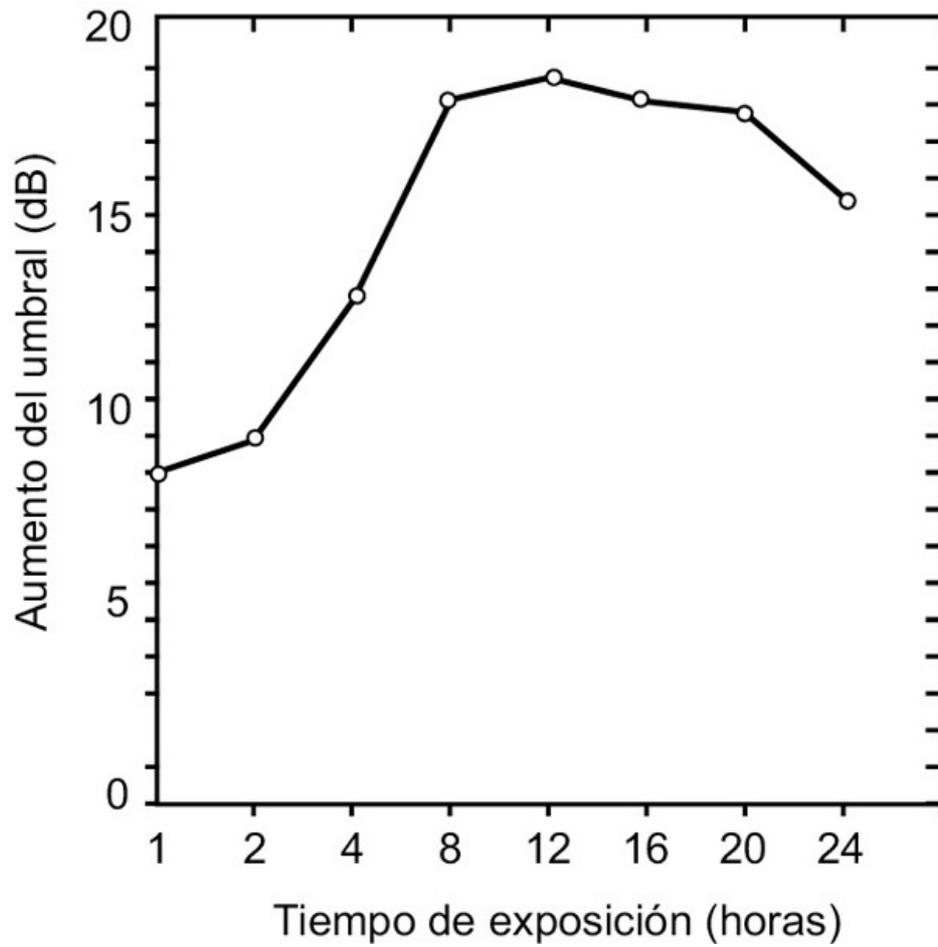
En la adaptación patológica la sensación auditiva desaparece completamente ante un estímulo continuo al cabo de unos segundos.

### *Fatiga auditiva*

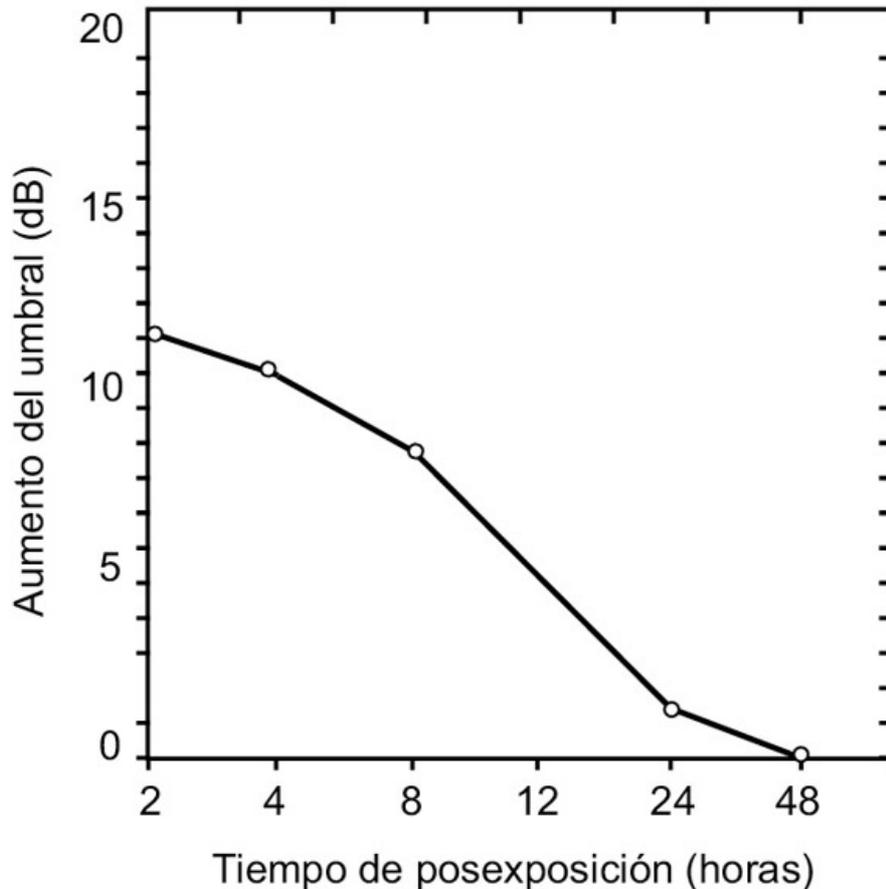
La fatiga auditiva es la pérdida temporal de sensibilidad que aparece cuando nos exponemos a un sonido muy intenso y se la mide luego de remover el estímulo que la causa. Generalmente está caracterizada por el desplazamiento temporario del umbral (TTS)[16].

La fatiga auditiva es más acusada a altas frecuencias. El máximo TTS ocurre entre 4 y 6 kHz y está relacionado con el nivel de presión acústica del ruido –a mayor nivel en dB mayor incremento del umbral–, con la frecuencia y el espectro del mismo –cuanto mayor sea la frecuencia, hasta llegar a 6.000 Hz, mayor será el incremento del umbral– y con el patrón temporal del ruido –los ruidos intermitentes producen menos pérdidas temporarias que los continuos.

Como ejemplo, podemos observar en las figuras 13 y 14 el deterioro y la posterior recuperación de la sensibilidad auditiva causada por un ruido de un ancho de banda de una octava, centrado en 4.000 Hz y con nivel de presión sonora de 85 dB.



**Figura 13.** Desarrollo de las pérdidas transitorias de audición



**Figura 14.** Recuperación del umbral una vez interrumpida la fuente acústica

Observamos que la sensibilidad auditiva se recupera totalmente sólo 48 horas después de la supresión de la fuente de ruido. Si una persona, por ejemplo, reactiva la fuente a las ocho horas de haberla suprimido, su umbral desde el reinicio será 10 dB mayor que el normal a causa de la fatiga auditiva, obligándola a aumentar 10 dB la potencia de la fuente acústica para alcanzar el nivel deseado: su oído verá multiplicada por diez la energía recibida. Este proceso puede realimentarse varias veces hasta alcanzar niveles capaces de provocar daños no reversibles. Algunos de los mecanismos fisiológicos involucrados en la pérdida temporaria de la audición son los mismos que producen la pérdida permanente de la capacidad auditiva. La diferencia es sólo una cuestión de grado.

En la actualidad (normas ISO) se consideran seguros niveles de hasta 85 dBA, ocho horas por día. A partir de este valor a cada incremento de +3 dB en la intensidad del estímulo le corresponde una disminución a la mitad del tiempo de exposición permitido. [17] Por ejemplo, si se aplica esta regla una persona podrá exponerse sin riesgo auditivo hasta dos horas por día a un nivel continuo de 91 dB. En algunas normas se estipula un nivel máximo, por ejemplo 115 dBA, que no debe superarse en ninguna circunstancia.

### *Incremento de la sonoridad*

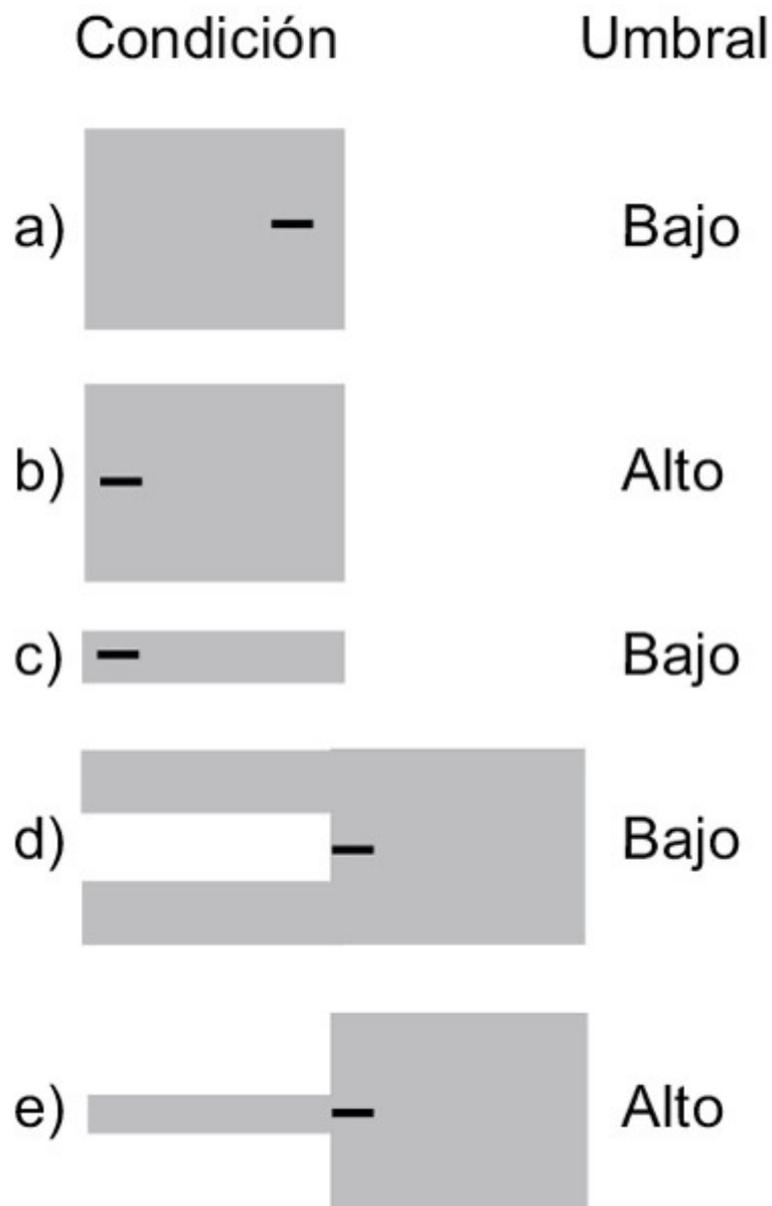
En contraste con la reducción de la sonoridad producida por sonidos presentes simultáneamente con el sonido de prueba (enmascaramiento), los sonidos que aparecen antes del de prueba pueden a veces producir un incremento en su sonoridad. Si el primer estímulo es, por ejemplo, 20 dB más intenso que el segundo, los experimentos muestran que este último puede incrementar su sonoridad hasta la equivalente a un aumento en intensidad de 15 dB. El efecto decrece al aumentar el intervalo temporal entre ambas señales y desaparece luego de unos pocos cientos de milisegundos. El efecto es mayor si los dos estímulos tienen la misma frecuencia y si son expuestos al mismo oído (o a los mismos oídos).

### **Detección de señales sinusoidales en presencia de ruido**

Para analizar señales sinusoidales a las que se le agrega una banda de ruido se emplea una metodología similar a la que se usa para estudiar la discriminación en intensidad entre sinusoides.

Un efecto temporal relacionado es el de desborde (overshoot). [18] El umbral necesario para detectar una señal breve es mayor si se la presenta cerca del ataque o la extinción del ruido enmascarante. El efecto aumenta entre 10 a 15 dB cuando el ruido enmascarante cubre un gran ancho de banda y la señal es de alta fre-

cuencia. Las componentes en frecuencia del ruido por encima de la señal son importantes para producir el efecto.



**Figura 15.** Sonogramas esquemáticos del efecto de overshoot. a) y b): situación básica; c) a e): efectos del ancho de banda del ruido y del retardo en el ataque

El efecto de overshoot muestra que el umbral para un breve estímulo sinusoidal iniciado luego del ataque de un ruido puede ser 10 dB mayor que si se lo coloca, por ejemplo, 200 ms después del ini-

cio del ruido enmascarante (Zwicker, E., 1965). Algunos experimentos posteriores señalaron que, para que ocurra un gran efecto de overshoot, el ruido enmascarante debe contener energía a frecuencias alejadas de la frecuencia de la señal, y que esta energía debe aparecer un breve instante antes del ataque de la señal (McFadden, D., 1989). La figura 15 sintetiza los trabajos de McFadden.

Los mecanismos que dan lugar a estos resultados no se comprenden todavía con claridad, aunque sin duda están vinculados con el principio de indeterminación acústico (véase el Apéndice II).

## **Patologías auditivas**

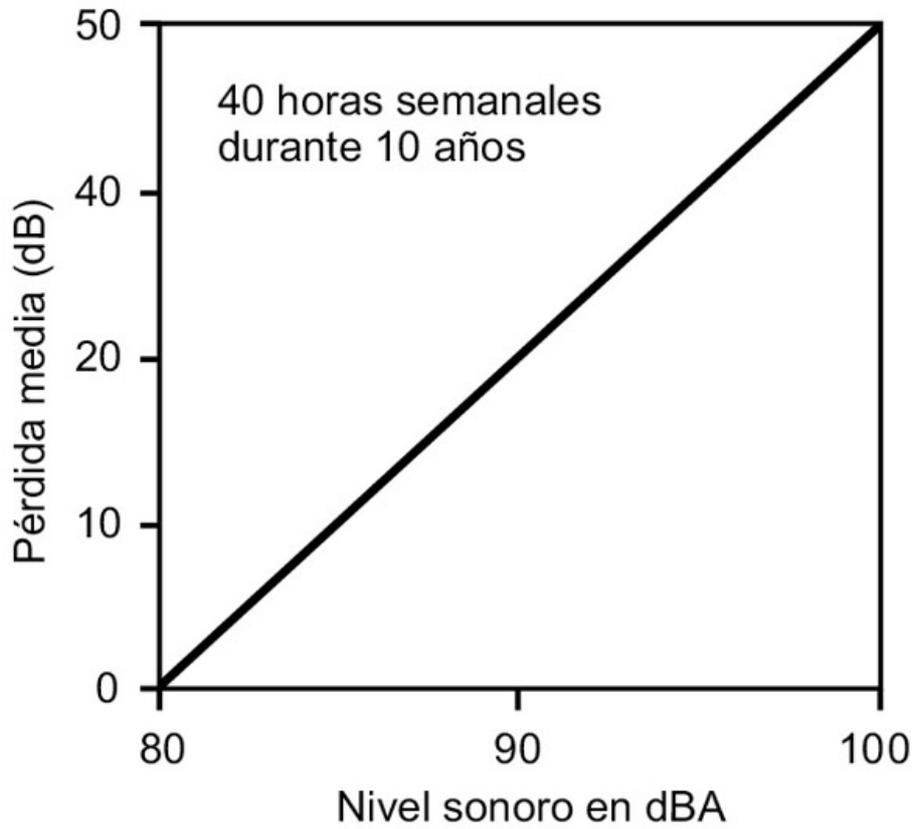
Los efectos del ruido sobre el sistema auditivo humano se pueden dividir en desplazamientos temporarios del umbral de audibilidad (pérdidas temporarias o reversibles, TTS) y en desplazamientos permanentes de dicho umbral (pérdidas permanentes o irreversibles, PTS).[\[19\]](#) Estos últimos pueden ser causados por la destrucción inmediata de las estructuras físicas del oído, en cuyo caso estamos en presencia de un trauma acústico, o por el efecto acumulativo de sucesivas exposiciones durante un período relativamente prolongado.

### ***Pérdida permanente de sensibilidad***

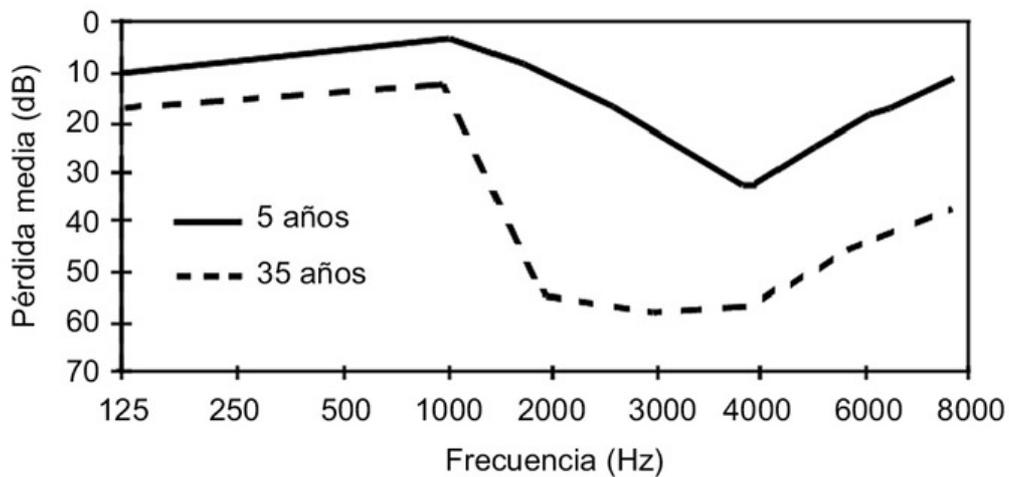
Las pérdidas auditivas permanentes pueden dividirse en dos grandes grupos: las conductivas y las sensoneurales (o perceptuales). Las conductivas llevan a un aumento en el umbral más o menos parejo en función de la frecuencia, mientras que las sensoneurales son más complejas.

El conocimiento que poseemos del desplazamiento permanente del umbral de audibilidad o PTS se basa en estudios realizados en industrias y en experimentos con animales. De ellos se desprende que existe una relación aproximadamente lineal entre las pérdidas

de sensibilidad y el nivel sonoro de la exposición por encima de los 85 dBA, tal como se observa en la figura 16. Se considera que los niveles inferiores a 80 dBA no producen efectos nocivos.



**Figura 16.** Pérdida permanente de audición para  $f = 4.000$  Hz



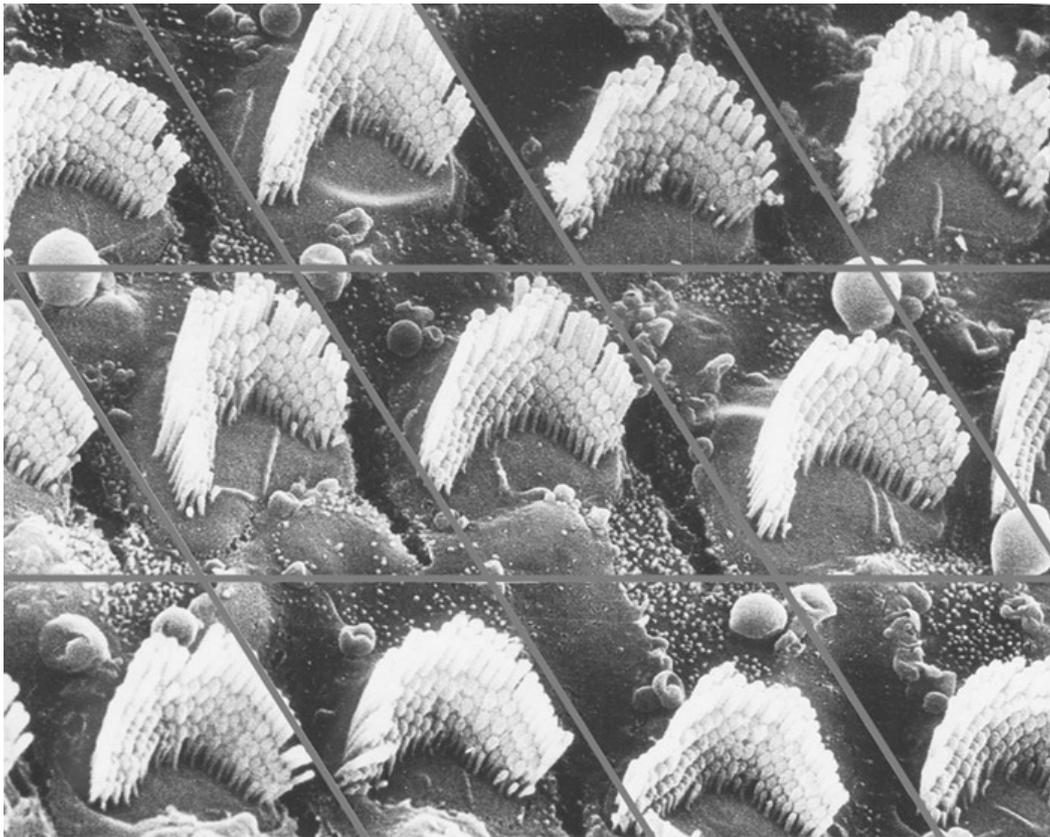
**Figura 17.** Pérdida permanente de audición en función del período de exposición

La relación entre el tiempo de exposición y la variación espectral de las pérdidas es más compleja. Es común que aparezca durante los primeros años una disminución de la sensibilidad alrededor de los 4.000 Hz, ampliándose con el correr del tiempo hasta ocupar la banda comprendida entre los 3.000 y los 6.000 Hz. Este proceso se muestra en la figura 17.

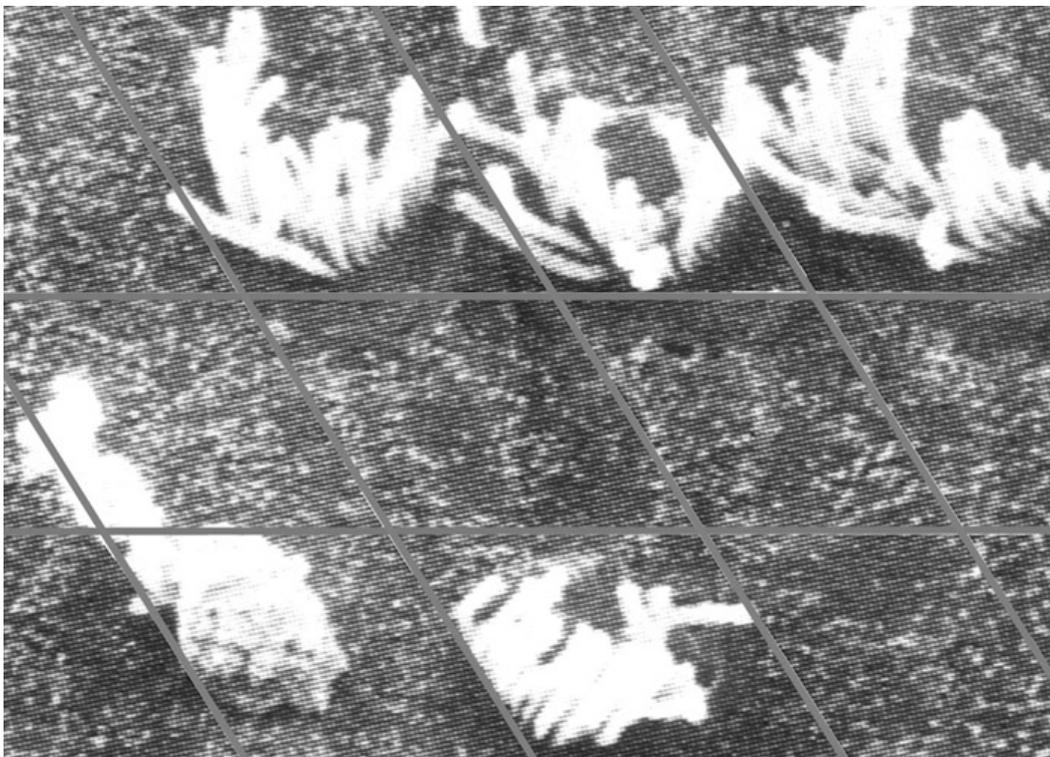
Vemos que las frecuencias más altas (entre 3.000 y 4.000 Hz) son afectadas en los primeros años de exposición, mientras que las menores (entre 1.000 y 2.000 Hz) son alcanzadas con pérdidas significativas al cabo de 20 años o más.

En los gráficos de las figuras 16 y 17 se ve con claridad que el oído es más sensible, y por lo tanto más frágil, en la región de altas frecuencias que en la de bajas. Este hecho es el causante de la detección tardía de la hipoacusia provocada por ruido, dado que la inteligibilidad de la palabra hablada depende de la energía en frecuencias relativamente bajas: las personas acuden al médico sólo cuando tienen dificultades en la comunicación oral, etapa en la que el proceso ya manifiesta un gran avance.

En las figuras 18 y 19 se pueden ver microfotografías de la membrana basilar en las que se aprecia el deterioro que sufren las células de Corti a causa del ruido.



**Figura 18.** Células de Corti en un oído sano



**Figura 19.** Células de Corti en un oído dañado por ruido

### ***Reclutamiento auditivo***

El reclutamiento auditivo ocurre cuando el umbral de audibilidad de una persona está por encima del normal, pero su audición es normal a altos niveles de sonoridad. En el sujeto con reclutamiento, la sonoridad aumenta más rápidamente con la intensidad de lo que ocurre en un oyente con oído promedio –al menos en las frecuencias afectadas. Se podría asimilar a una suerte de compresión del rango de sonoridad.

Moore relaciona el reclutamiento auditivo con un daño en las células ciliadas externas de la membrana basilar, que impide la correcta codificación del rango dinámico del estímulo acústico (Moore, B., 1997).

## **5. Codificación de la intensidad**

El estudio de la codificación de la intensidad tiene por objetivo explicar la forma que tiene el oído de informarle al cerebro sobre cuán intenso es un estímulo. O, más específicamente, por la manera en que se configura la representación mental de la intensidad física de una señal a partir de la actividad de las células nerviosas del sistema auditivo. Se ha verificado que la codificación de la intensidad está vinculada directamente con nuestra habilidad para discriminar entre estímulos de diferente intensidad. Si podemos distinguir entre dos sonidos, la diferencia debe estar representada en todos los niveles de la cadena auditiva, desde la cóclea hasta los centros superiores de decisión. Los experimentos de discriminación de la intensidad son, por lo tanto, la herramienta base para crear modelos de codificación de la intensidad.

Un primer modelo, muy simplificado, establece que la intensidad de un estímulo sinusoidal es codificada simplemente por la ta-

sa de disparo de las fibras de Corti sintonizadas a la frecuencia del estímulo: a mayor intensidad, mayor cantidad de disparos por unidad de tiempo. Este modelo, sin embargo, presenta varios inconvenientes.

## **El problema del rango dinámico**

Como hemos visto, el sistema auditivo humano es capaz de detectar diferencias en la intensidad de los estímulos sobre un gran rango dinámico, tan grande como 120 dB en oyentes con oído sano. La mayoría de las células nerviosas auditivas, aquellas que presentan una tasa de disparo relativamente alta, tienen bajos umbrales de excitación y rangos dinámicos comparativamente pequeños. Casi todas se saturan al llegar a una intensidad de aproximadamente 60 dB cuando son estimuladas a sus frecuencias características. La proporción de fibras en el nervio auditivo que presentan un amplio rango dinámico (60 dB o más) es muy baja, de alrededor del 10% del total. Por lo tanto, se supone que interviene más de un mecanismo diferente en la discriminación de la intensidad. Los cambios en intensidad podrían ser codificados por cambios de la tasa de descarga de las neuronas que ocupan el centro del patrón de excitación, por una extensión del patrón de excitación (más neuronas entran en operación), o por una combinación entre ambos mecanismos. La sensibilidad a la fase también parece intervenir en el proceso.

La importante Ley de Weber para ruido de banda ancha, que continúa actuando para altas intensidades, debe provenir de la operación del sistema auditivo, no de propiedades estadísticas del estímulo. Pero no se puede explicar a partir de la información codificada por una sola neurona por los motivos antes apuntados. Es necesario encontrar un modelo que combine la tasa de disparo de un número pequeño de neuronas, cuyos umbrales y cambios dinámicos estén escalonados de manera tal que permitan cubrir el ran-

go dinámico total del sistema auditivo. La información de las neuronas que responden a frecuencias características similares podría combinarse y operar en un “canal” independiente. La ley de Weber parece responder a cada uno de estos canales, sintonizado a un rango de frecuencias dado y en paralelo con los demás.

Buscando una respuesta a este problema, Carlyon y Moore sugirieron que en la codificación periférica intervienen neuronas de diferente umbral de disparo: a intensidades moderadas intervienen las neuronas de bajo umbral, luego éstas se saturan y son reemplazadas paulatinamente por las de alto umbral de disparo (Carlyon y Moore, 1984). En la misma dirección, Delgute indicó que son pocas las neuronas necesarias para dar cuenta de la discriminación en intensidad (Delgute, B., 1987). Según este autor, el número requerido es de aproximadamente 100 neuronas. Si se extiende esta conclusión a la totalidad de las células que ocupan la cóclea, la discriminación debería ser mucho mayor que la que tenemos (podría ser de menos de 0,1 dB a 1.000 Hz). De ser cierta esta hipótesis, la limitación no dependería de la información que transporta el nervio auditivo sino de los niveles centrales de procesamiento.

### ***Codificación por extensión del patrón de excitación***

Una explicación posible para la aparente paradoja es que, al menos para estímulos de banda angosta, la información sobre la intensidad se obtiene también de las células auditivas sintonizadas a frecuencias por encima y por debajo de la frecuencia del estímulo. A pesar de que las células correspondientes a la frecuencia del estímulo se saturan, las que tienen frecuencias centrales diferentes reciben menos excitación y continúan dentro del rango de funcionamiento. Se ha sugerido que estas células “fuera de frecuencia” son las responsables de la codificación a altas intensidades (Zwicker y Fastl, 1990). Sin embargo, los resultados de los experimentos sobre enmascaramiento demostraron que, a pesar de que la extensión la-

teral de la zona de excitación puede ayudar a discriminar a altas intensidades, el sistema auditivo puede codificar la intensidad sobre un gran rango dinámico empleando sólo la información proveniente de una pequeña banda de frecuencias centrales en la membrana basilar.

### ***El papel de la supresión y de la adaptación***

Una posible solución sería que el sistema auditivo pudiera ajustar el rango de operación de las células de acuerdo al nivel de entrada, de modo de mantenerlas siempre en la zona de actividad, lejos de la saturación. Estamos familiarizados con algo similar, el control automático de ganancia en la vista, que nos permite ver bien tanto a pleno sol como en la relativa oscuridad de un cine. En el caso de la vista la recuperación es bastante lenta y necesitamos un buen tiempo para adaptar nuestra sensibilidad al pasar de una situación lumínica a otra diferente. En el caso de la audición, sin embargo, tenemos que ser capaces de procesar sonidos que varían dramáticamente en intensidad sobre un intervalo de tiempo muy breve: necesitamos disponer del rango dinámico completo “al instante”.[\[20\]](#)

La supresión es un proceso no lineal en el cual una intensa excitación en una región de la membrana basilar reduce la excitación en las regiones vecinas. Al tener una constante de tiempo muy pequeña ocurre solamente cuando se presentan simultáneamente el estímulo supresor y el suprimido. Es posible que el mecanismo que produce la supresión actúe también como un control automático de ganancia para estímulos de banda ancha, y ayude a dar cuenta del gran rango dinámico del sistema auditivo.

### ***Codificación por sincronización neural***

Carlyon y Moore sugirieron que, en ciertas circunstancias, la intensidad puede ser codificada por la estructura de seguimiento de fase en las células auditivas: un incremento en la intensidad de un

estímulo sinusoidal en presencia de ruido puede generar un aumento de la cantidad de células sincronizadas con la estructura fina del estímulo, diferente de la del ruido (Carlyon y Moore, 1984). Los experimentos mostraron, sin embargo, que este mecanismo por sí solo no puede ser responsable del gran rango dinámico presente en el sistema auditivo.

### ***Modelos basados en funciones intensidad-tasa de disparo***

Una hipótesis diferente establece que las células de Corti con altas tasas de disparo son responsables de la codificación a bajas intensidades, mientras que las células de Corti con bajas tasas de disparo se destinan a codificar los estímulos a altas intensidades, por encima de la saturación de las anteriores. Los modelos basados en esta hipótesis se denominan modelos de doble población. La principal dificultad que encuentran es la de explicar la constancia de la fracción de Weber a lo largo de casi todo el rango dinámico. Esta dificultad se incrementa al constatarse que hay muchas más células con altas que con bajas tasas de disparo.

En uno de estos modelos de doble población se simularon las codificaciones basándose en el número de células necesarias para cubrir el rango dinámico completo a una frecuencia dada: con sólo diez células combinadas es posible imitar al sistema auditivo (Viemester, N. F., 1988). Sin embargo, poseemos más de 50 para la misma función. Es decir, tenemos a la salida del nervio auditivo información más que suficiente para codificar los estímulos sobre un gran rango de intensidades. Quizá la pregunta no deba ser, entonces, por qué es tan buena la codificación a altas intensidades, sino por qué no es mejor de lo que es a bajas intensidades.

### **Límites a la discriminación en intensidad**

Hemos trazado los caminos que pueden servir para representar la intensidad en el sistema auditivo periférico. Ahora consideraremos el caso de los límites impuestos por las transformaciones que ocurren en los centros de procesamiento neural. En esta etapa debe incorporarse una descripción del mecanismo que permite que las intensidades pueden ser almacenadas en la memoria y de cómo las limitaciones de la memoria afectan los resultados en las tareas de discriminación.

### *Limitaciones periféricas y centrales*

Los experimentos realizados para determinar las características de la discriminación de la intensidad, para excitaciones que abarcan una sola frecuencia central en la membrana basilar, arrojan resultados contradictorios. Mientras los modelos de doble población sugieren que la discriminación está caracterizada por la fracción de Weber, los modelos fisiológicos predicen mejor discriminación a bajas y medias intensidades que a grandes intensidades.

La riqueza de la información presente en el nervio auditivo y la falla de los modelos de codificación basados en las propiedades de las fibras auditivas, en su intento por explicar el éxito de la Ley de Weber, implican que en algún punto del proceso neural central no se hace un uso completo de la información que circula a niveles periféricos (Carlyon y Moore, 1984). Presumiblemente, esta limitación central determina el desempeño de los oyentes en las tareas de discriminación de la intensidad, y evita que el sistema auditivo tenga un rendimiento superior a bajas intensidades.

### *Memoria de la intensidad*

Muchos experimentos de discriminación de la intensidad emplean secuencias en las que dos intervalos de observación están separados por un intervalo inter estímulo (IIE). La ausencia de un efecto sustancial de este intervalo en las tareas de discriminación citadas

sustenta la hipótesis de que los oyentes emplean la memoria de largo término para hacer las comparaciones entre el primero y el segundo de los estímulos. Un almacenamiento en la memoria de corto término debería decaer en el tiempo, generando una gran dependencia con la duración del intervalo inter estímulo. Como un ejemplo extremo, Pollack empleó un IIE de 24 horas y prácticamente no encontró deterioro en la tarea de discriminación de la intensidad (Pollack, I., 1953).

La memoria de largo término puede ser anulada variando al azar, entre pruebas, la intensidad de los estímulos en los intervalos de observación. Cuando esto ocurre, el éxito en las tareas de discriminación de la intensidad baja notablemente al aumentar el IIE. Por ejemplo, la fracción de Weber aumenta de 2 a 5 dB cuando el IIE crece de 250 ms a 8 s. Este cambio se debe, presumiblemente, al deterioro del almacenamiento en la memoria de corto término.

## **Codificación de rasgos y codificación contextual**

Durlach y Braida distinguieron dos modos diferentes de operación de la memoria relacionada con la discriminación de la intensidad: la codificación de rasgos y la codificación contextual (Durlach y Braida, 1969). En la primera son almacenadas las sensaciones directas –inmediatas– producidas por el estímulo, que tienden a decaer en el tiempo (son afectadas por la duración del IIE). En el modo de codificación contextual, por otra parte, la intensidad es codificada en relación con una intensidad de referencia externa o con relación a una referencia interna “fuerte”, como el umbral de audibilidad o el umbral de molestia.

En circunstancias normales la memoria de rasgos es suficientemente rica como para ser relativamente inmune a la degradación con el tiempo. Pero si la información en la memoria de rasgos es reducida gradualmente, el sistema se apoya cada vez más en la codificación contextual –aunque ésta no es efectiva a intensidades

medias debido a la distancia que separa al estímulo de los límites perceptuales de referencia.

Es evidente que el sistema auditivo necesita codificar la intensidad de modo relativo en muchas ocasiones, simplemente porque los objetos auditivos (como los fonemas en el habla) se definen por su intensidad relativa, ya sea entre componentes espectrales o entre eventos sucesivos. La intensidad absoluta no afecta la identidad de los objetos auditivos en la mayoría de los casos –juzgamos que es el mismo perro el que ladra a 2 o a 50 m de distancia. Es lógico pensar, por lo tanto, que el sistema auditivo deba extremar su eficiencia para las comparaciones entre estímulos cercanos en el tiempo antes que para los separados por más de unos cientos de milisegundos.

## 6. Modelos de percepción de la sonoridad

Se han elaborado varios modelos formales que intentan describir cuantitativamente la percepción de la sonoridad. Zwicker desarrolló un modelo de sonoridad basado en la distribución del patrón de excitación en la membrana basilar (Zwicker y Fastl, 1999). En su modelo el estímulo es convertido en una sonoridad específica por banda crítica ( $N'$ ) a través de una ley de potencia:

$$N' = C E^\alpha$$

En la que  $N'$  es la sonoridad por banda crítica,  $E$  es la intensidad de la excitación, y  $C$  y  $\alpha$  son constantes con  $\alpha < 1$ .

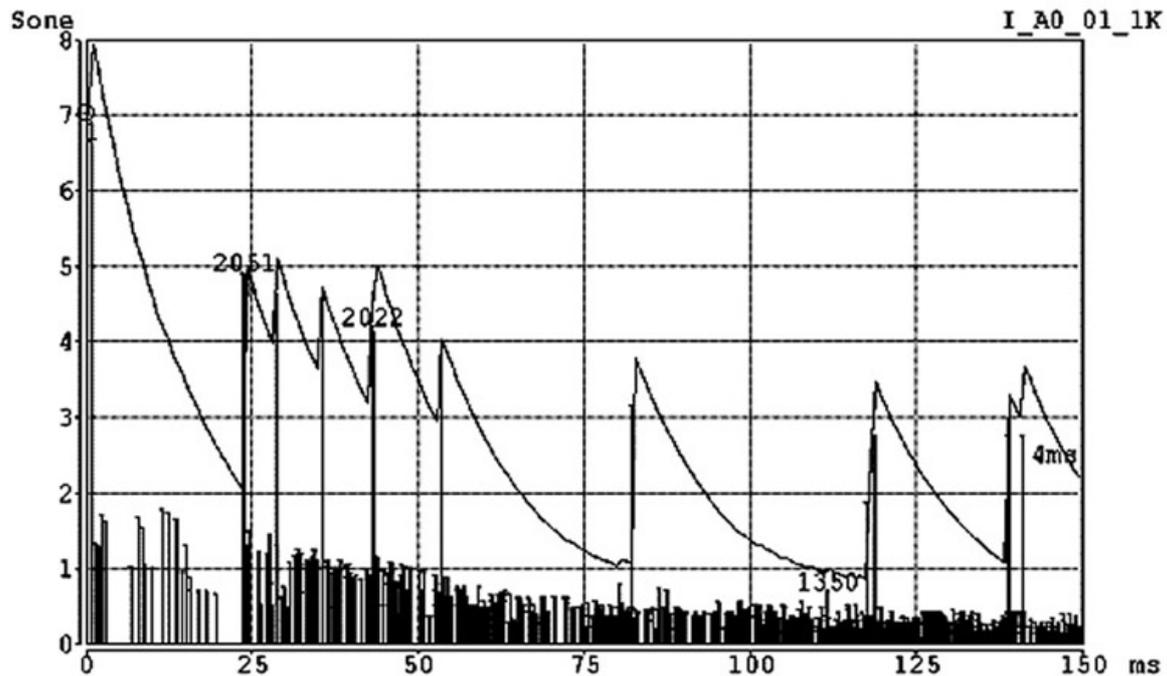


**Figura 20.** Modelo de Moore en bloques para calcular la sonoridad

Este modelo fue modificado por Moore y Glasberg, quienes asumen que, por debajo de 1 kHz, la forma del filtro auditivo es igual al contorno isofónico correspondiente a 100 fones, y que por encima de 1 kHz la forma del filtro está determinada por la inversión de la curva umbral de audibilidad (Moore y Glasberg, 1986).[\[21\]](#) La excitación es convertida en sonoridad específica por banda crítica a partir de la siguiente expresión:

$$N' = C [(I \text{ señal})^\alpha - (I \text{ umbral})^\alpha]$$

Aquí  $I$  señal e  $I$  umbral son las intensidades correspondientes a la excitación y al umbral definido en el párrafo anterior. Nótese que si  $I$  señal =  $I$  umbral la sonoridad específica es nula, y que cerca del umbral un pequeño cambio en la excitación provoca un cambio proporcionalmente grande en la sonoridad. La sonoridad global se calcula integrando las sonoridades específicas positivas a lo largo del patrón total de excitación. El éxito de este modelo al predecir la variación de la sonoridad en función de cambios en frecuencia, intensidad y ancho de banda apoya la hipótesis que establece que la sonoridad está relacionada con la selectividad en frecuencia del sistema auditivo periférico, y no directamente con la intensidad física del estímulo acústico.



**Figura 21.** Ecograma en sonos simulado para una butaca en la platea del Teatro Argentino de La Plata. La fuente acústica emitió un impulso de Dirac virtual\*

\* En este caso se utilizó el paquete CATT- Acoustic v7.2.

En la figura 20 se puede ver el modelo en bloques de Brian Moore en su versión de 1997 (Moore, B., 1997). La sonoridad total es proporcional al área bajo un patrón de sonoridad específico. Este esquema se ha incorporado a ciertos medidores de sonoridad –no de intensidad o presión acústica–, y de acuerdo con los informes parece que funciona razonablemente bien.

Algunos paquetes de software para acústica de salas emplean modelos similares para brindar información sobre predicción de sonoridad y con ellos es posible calcular los ecogramas en una escala de sonos. En la figura 21 se aprecia un ejemplo de esos ecogramas.

# Notas

- [1](#) Algunos autores llaman “intensidad subjetiva” a la sonoridad. Por ejemplo, véase el artículo de Plack y Carlyon en el que emplean los términos loudness y subjective intensity (Moore, B., 1995).
- [2](#) Son los 1.013 hPa (hectopascales) que informa el servicio meteorológico como presión atmosférica normal.
- [3](#) Ernst Heinrich Weber (1795-1878) y Gustav Theodor Fechner (1801-1887), fisiólogos alemanes pioneros de la psicología experimental y creadores de la psicofísica.
- [4](#) Un poco más adelante vamos a volver sobre este punto.
- [5](#) En inglés se lo denomina sound pressure level (SPL).
- [6](#) La potencia instantánea se calcula a partir de la derivada de la energía con respecto al tiempo:  $P_i = \partial E / \partial t$
- [7](#) Sound intensity level (IL) en inglés.
- [8](#) Funciones de transferencia relativas a la cabeza (Head Related Transfer Functions).
- [9](#) En números, una relación de  $10^{12}/1$  en intensidad, o de  $10^6/1$  en presión acústica.
- [10](#) El contexto es aquí de gran importancia. 40 fones originan un sonido de 1 son sólo si en los instantes previos el oyente se encuentra en un ambiente extremadamente silencioso.
- [11](#) Traducción de GB. Hermann Ludwig Ferdinand von Helmholtz (1821-1894) fue un científico alemán de destacada actuación en el campo de la fisiología, la óptica, la acústica y la electrodinámica.
- [12](#) Conocidos popularmente como “decibelímetros”.
- [13](#) JND, del inglés just noticeable difference.
- [14](#) G. B. Henning (1970). Este hecho ya había sido notado por Exner en 1876: “Tanto los umbrales absolutos en sonoridad como la sonoridad en sí misma dependen de la duración del estímulo.”
- [15](#) Las bandas críticas se van a analizar en detalle en el capítulo III. Por ahora, nos alcanza con saber que corresponden aproximadamente al 15% de la frecuencia central considerada.
- [16](#) Del inglés Temporary Threshold Shift.
- [17](#) Ya no se emplea la “regla de los 5 dB” propiciada por ciertas industrias.
- [18](#) La palabra “desborde” no es una traducción literal de “overshoot” (exceder, rebasar, sobrepasar), pero creemos que describe mejor el fenómeno aludido.
- [19](#) Del inglés Permanent Threshold Shift.
- [20](#) Plack y Carlyon dan un muy buen ejemplo: podemos oír el grito: “¡El psicópata está bajando al corredor!”, y a continuación los débiles sonidos de sus pasos en la escalera (Moore, B., 1995).

[21](#) En Moore, B. (1995).

# Capítulo III

## El oído como analizador espectral

En los capítulos precedentes mencionamos muchas veces las nociones de banda crítica, análisis en frecuencia, discriminación espectral y enmascaramiento. Ha llegado el momento de analizar estos conceptos con detalle.

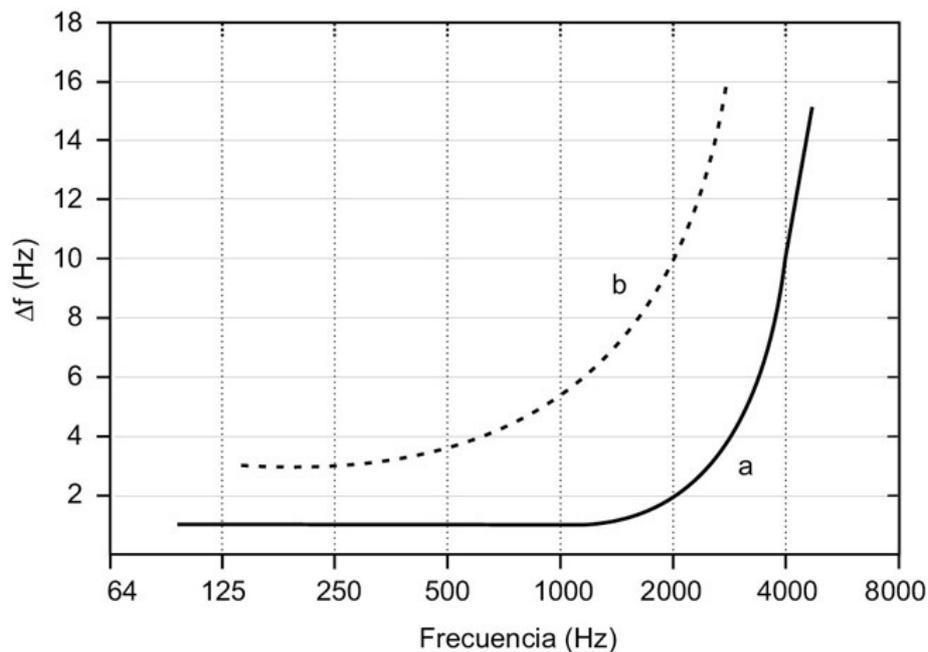
La habilidad para separar o resolver los componentes espectrales de una señal compleja es consecuencia de la capacidad de análisis en frecuencia del sistema auditivo. Por ejemplo, si dos flautas afinadas a diferentes frecuencias se soplan simultáneamente, es posible oír al mismo tiempo dos sonidos diferentes. Se sabe que el análisis en frecuencia depende en gran medida de la exploración espectral que ocurre directamente en la cóclea. Toda señal compleja, como la que genera la voz humana o un instrumento de música, es procesada en esta temprana etapa de la cadena auditiva. Los componentes sinusoidales o de banda angosta son separados, codificados y transmitidos de modo independiente por el nervio auditivo. Más aún, esta etapa temprana del análisis no puede obviarse: todas las señales sufren necesariamente dicho proceso. Por esto, la percepción de un sonido como un todo coherente depende del rearmado que tiene lugar en algún punto superior del sistema nervioso.

### 1. Diferencia apenas perceptible (DAP) en frecuencia

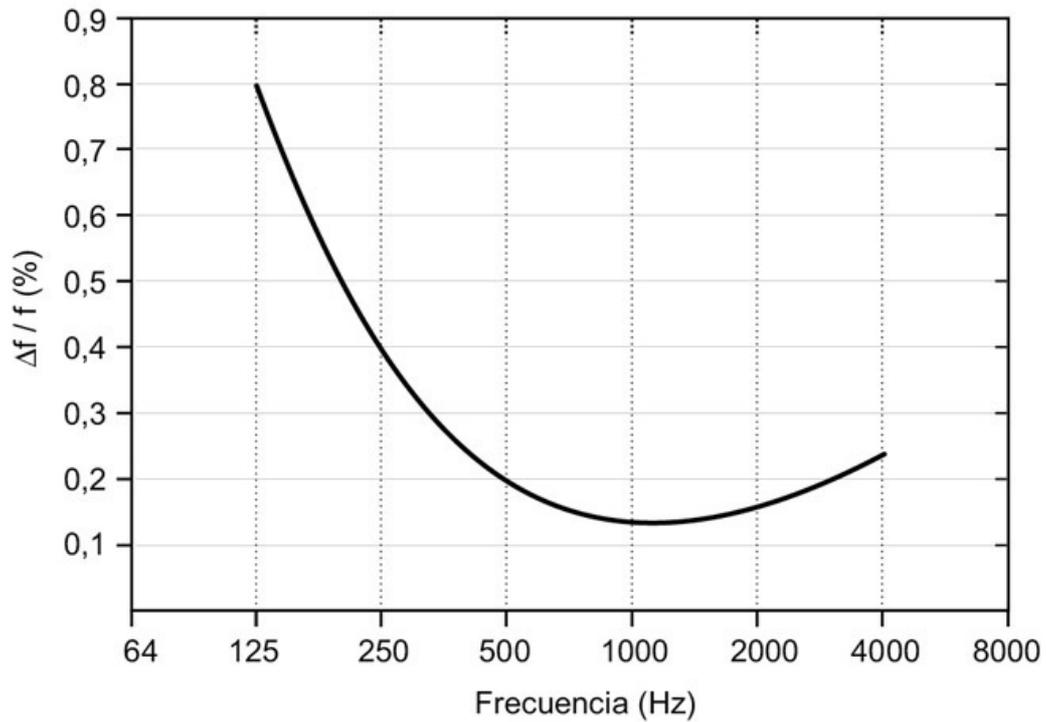
Una de las preguntas elementales relacionada con el tema es la que inquiere sobre nuestra habilidad para discriminar entre dos señales cuyas frecuencias son ligeramente diferentes. La respuesta, obtenida a partir de un procedimiento psicofísico clásico, introduce el concepto de *diferencia apenas perceptible* (DAP o JND)[\[1\]](#) en frecuencia, o *limen* de frecuencia. Para medirla se realiza un experi-

mento en el cual los sujetos oyen alternativamente dos sinusoides de igual sonoridad; mientras una de ellas mantiene constante su frecuencia la otra la modifica levemente. El objeto es determinar la mínima diferencia en frecuencia necesaria para distinguir entre ambas señales. En las figuras 1 y 2 se pueden ver los valores de la DAP en frecuencia en función de la frecuencia de base.

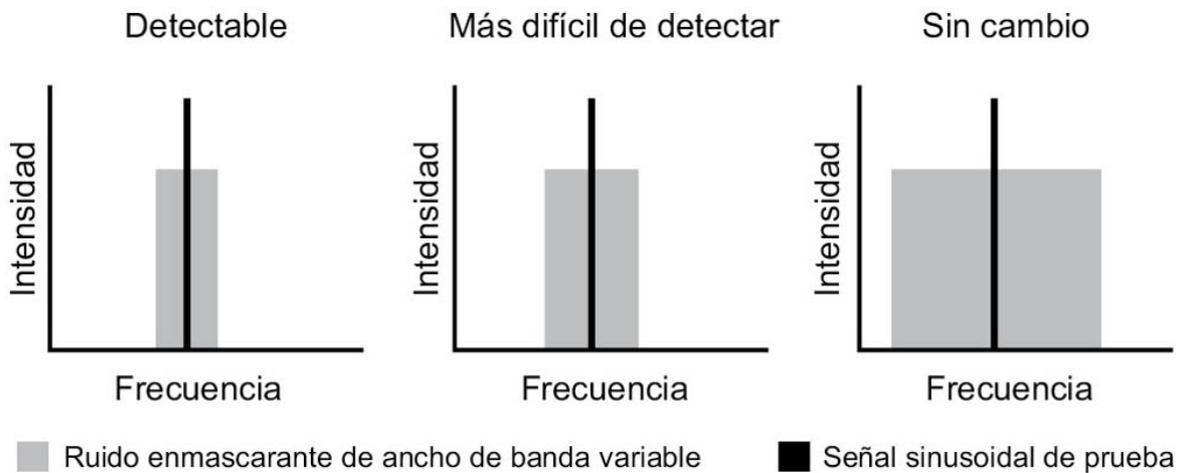
El resultado puede sintetizarse diciendo que la DAP en frecuencia se mantiene cercana a 1 Hz para frecuencias por debajo de 1 kHz y que crece paulatinamente a partir de esta frecuencia. Por encima de 5 kHz la DAP aumenta más rápidamente, y más allá de los 10 kHz, nuestra habilidad para discriminar entre diferentes frecuencias prácticamente desaparece.



**Figura 1.** DAP absoluta en frecuencia; a) obtenida a partir de señales sinusoidales de intensidades medias, entre 60 y 80 dB, b) obtenida a partir de señales sinusoidales de intensidades cercanas al umbral de audibilidad (Wier, Jesteadt y Green, 1977a)



**Figura 2.** DAP relativa en frecuencia



**Figura 3.** Esquema del experimento de Fletcher

La cuestión cambia si en lugar de sinusoides empleamos señales periódicas complejas. Tal como veremos en el capítulo V, en estos casos la DAP puede llegar a ser tan pequeña como 0,1 Hz.

## 2. Bandas críticas

Ya a mediados del siglo XIX algunos investigadores suponían que el sistema auditivo no funciona en bloque y que divide las señales para su tratamiento detallado. Una hipótesis temprana planteaba la presencia de alguna clase de análisis de frecuencia a nivel del sistema auditivo periférico. Este proceso de análisis en frecuencia se ha estudiado indirectamente a partir del fenómeno de enmascaramiento, que ocurre cuando un sonido enmascara o “tapa” a otro de diferente frecuencia. El fenómeno de enmascaramiento se analizará con detalle en el punto 4 de este capítulo.

## Experimento de Fletcher

El experimento fundacional en este campo lo realizó Harvey Fletcher en 1940.[\[2\]](#) Fletcher midió el aumento del umbral de detección para una señal sinusoidal en función de un ruido enmascarante con densidad de potencia constante y banda pasante variable. El ruido estaba centrado a la frecuencia de la senoide y su potencia acústica se incrementaba a medida que crecía el ancho de banda.

El experimento de Fletcher, que fue repetido gran cantidad de veces desde entonces, reveló aspectos desconocidos de nuestro sistema auditivo. En la figura 4 se aprecia un resumen de dicho experimento, realizado por Moore y su equipo en 1993 (Moore, B., 1997). En el gráfico se observa que el umbral de detección para una señal sinusoidal de 2.000 Hz aumenta al principio a medida que crece el ancho de banda del ruido, pero a partir de un valor determinado –400 Hz en la figura– se produce un quiebre marcado de la curva de umbral, que deviene prácticamente constante.

Para dar cuenta de esta singular característica, Fletcher sugirió, siguiendo a Helmholtz, que el sistema auditivo periférico se comporta como si estuviera compuesto por un banco de filtros pasabanda en paralelo (llamados *filtros auditivos*), cuyas bandas se traslapan. Cuando tratamos de detectar una señal en un ambiente

ruidoso, usamos el filtro cuya *frecuencia central* (FC) se acerca más a la frecuencia del estímulo: la señal pasa pero atenúamos gran parte del ruido de fondo, correspondiente a los otros filtros. Sólo la porción de ruido que atraviesa el filtro centrado en la señal la afecta significativamente. Como es de esperar, si se incrementa el ancho de banda del ruido aumenta la potencia de ruido que pasa por ese filtro en particular. Pero una vez que el ancho de banda del ruido supera al del filtro auditivo su capacidad enmascarante deja de crecer. La parte del ruido que cae fuera del filtro sintonizado con la señal no afecta nuestra habilidad para detectarla.

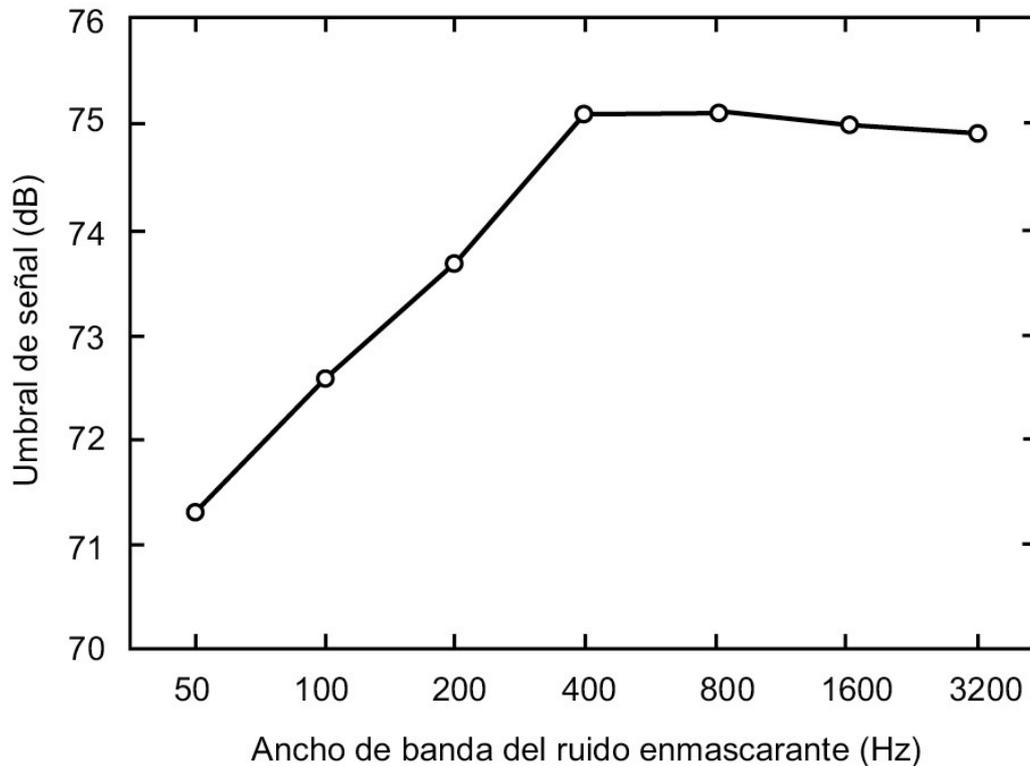
Se asume que el umbral de detección para una señal sinusoidal está determinado por la cantidad de ruido que atraviesa el correspondiente filtro auditivo. Más específicamente, el umbral estaría vinculado a una cierta relación señal/ruido a la salida del filtro. Este modelo es conocido como el *modelo del espectro de potencia*, en el cual los estímulos están representados por su espectro de largo tiempo, sin fluctuaciones transitorias que involucren cambios en las fases relativas.

El término *banda crítica* (BC), acuñado por Fletcher, generalmente se refiere a un hipotético filtro rectangular que simula el comportamiento de los filtros auditivos reales. El *ancho de banda rectangular equivalente* (ERB) es el ancho de banda de un filtro rectangular ideal que posee igual máximo de transmisión que el filtro real y por el que pasa la misma potencia total ante una entrada de ruido blanco. Fletcher estableció las siguientes hipótesis, que definen su *modelo del espectro de potencia*:

- El sistema auditivo periférico contiene un banco de filtros pasabanda lineales.
- Cuando un oyente intenta detectar una señal en un entorno de ruido emplea un filtro auditivo centrado a la frecuencia de la señal. En general es el filtro que entrega una mayor relación señal/ruido de salida.

- Solamente una banda angosta de frecuencias que rodea al estímulo, aquéllas que caen dentro de la bc, contribuye a enmascararlo.
- El umbral de detección de la señal está determinado por la cantidad de ruido que pasa por el filtro auditivo correspondiente. La potencia del estímulo  $P$  dividida por la potencia del ruido dentro de la banda crítica es constante ( $k$ ).

$k = P / (W \times N_0)$	$W =$ ancho de la banda crítica, en Hz.
$W = P / (k \times N_0)$	$N_0 =$ densidad de potencia del ruido, en watts/Hz
	(la potencia en una banda de 1 Hz de ancho).



**Figura 4.** Umbral para una senoide de 2.000 Hz en función del ancho de banda del ruido enmascarante

Midiendo  $P$  y  $N_0$  es posible estimar  $W$ . Según Fletcher, la constante  $k$  vale aproximadamente 1, aunque mediciones recientes la sitúan en 0,4. La diferencia tiene que ver con el contorno espectral del ERB equivalente, que no fue tomado en cuenta por Fletcher.

Sabemos que ninguna de las hipótesis anteriores es estrictamente correcta: los filtros no son lineales, los oyentes combinan la información de varios filtros para aumentar la capacidad de detección y el ruido que cae fuera de la banda pasante del filtro auditivo afecta al umbral de detección. Pero estas imperfecciones no implican que la hipótesis central del modelo esté equivocada. De hecho, gran parte de los avances en el campo se han basado en ella.

La forma rectangular del filtro auditivo propuesta por Fletcher merece un comentario aparte. Si el filtro fuese perfectamente rectangular el umbral de detección debería crecer a una tasa de 3 dB por cada duplicación de la banda de ruido. En la figura 4 se puede ver que la curva de crecimiento tiene una pendiente mucho menor. Este desvío del valor teórico se debe probablemente a dos factores: el filtro no es rectangular, y las pequeñas fluctuaciones del ruido afectan la detección de una manera apreciable.

## **Refinamiento del modelo de Fletcher**

El modelo del espectro de potencia se emplea como punto de partida para el desarrollo de modelos más precisos. Si el ruido enmascarante se representa por su espectro de potencia estacionario  $N(f)$  y la forma espectral del filtro auditivo por  $W(f)$ , la potencia de la señal en su umbral  $P$  se puede calcular con la siguiente expresión:

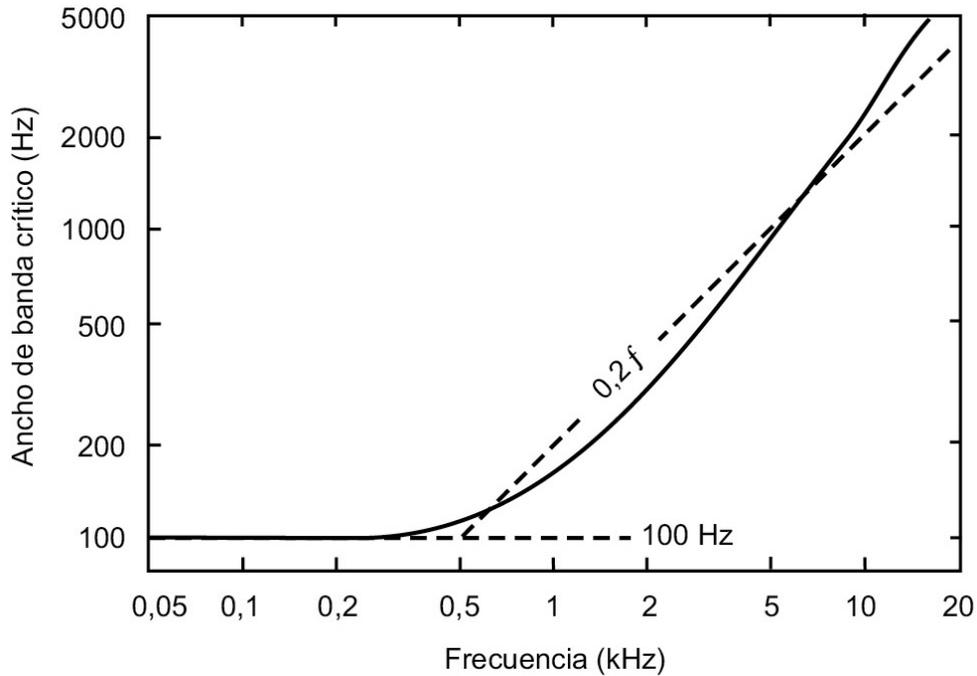
$$P = k \int_0^\infty W(f) N(f) df$$

Midiendo los cambios en el umbral de detección de  $P$  en función del ancho de banda del ruido, es posible evaluar la forma del filtro auditivo  $W(f)$ .

En general se eligen ruidos de banda angosta en vez de sinusoides como enmascarantes a fin de evitar los batidos que se generarían con la señal enmascarada. Para ruidos de banda muy angosta, sin embargo, aparecen fluctuaciones que pueden ser oídas. Otro motivo por el cual se eligen ruidos de banda como enmascarantes es que muchas veces el filtro auditivo que genera la mayor respuesta es uno centrado a una frecuencia diferente de la frecuencia de la señal a detectar. Esta importante característica, en la que el oído explora el campo audible “moviendo” su ventana espectral para buscar la mejor respuesta, se denomina *audición fuera de frecuencia*.<sup>[3]</sup> Para desactivar este efecto se coloca ruido a ambos lados de la zona de exploración con el *método de supresión de bandas por cobertura* de Patterson, que se describe en el apartado siguiente.

Otro método muy empleado para estudiar la selectividad en frecuencia del oído consiste en medir las *curvas de afinación psicoacústica*, en las que se emplea ruido de banda angosta para evitar la interferencia de los batidos.

En la figura 5 podemos apreciar el ancho de banda crítico en función de la frecuencia central de banda. Para frecuencias medias, la banda crítica tiene una extensión de alrededor del 15% de la frecuencia central (corresponde a un intervalo musical a medio camino entre un tono –segunda mayor– y un tono y medio –tercera menor).

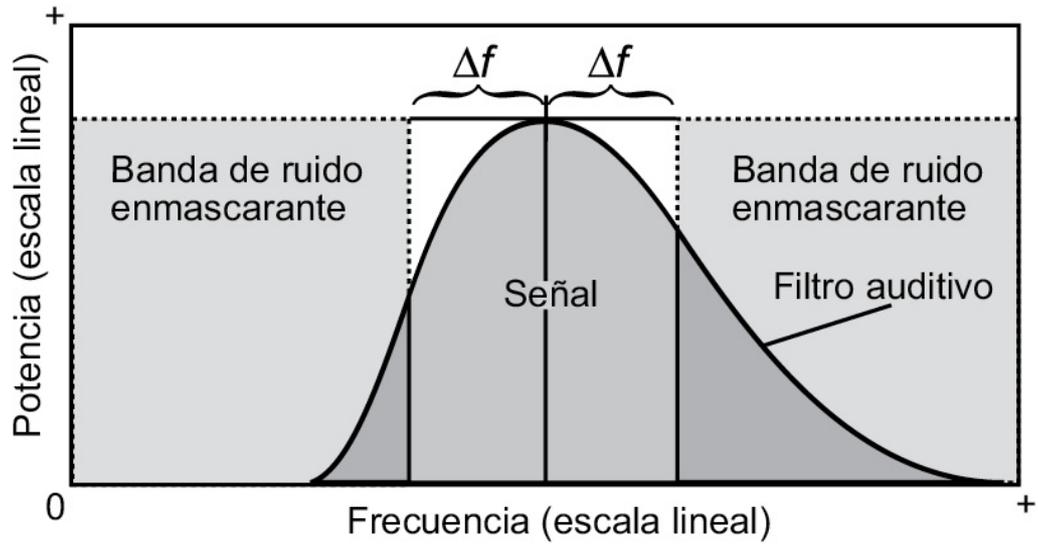


**Figura 5.** Ancho de banda crítico en función de la frecuencia central. Las líneas de puntos indican aproximaciones lineales para baja y alta frecuencia (Zwicker, Flottorp y Stevens, 1957).

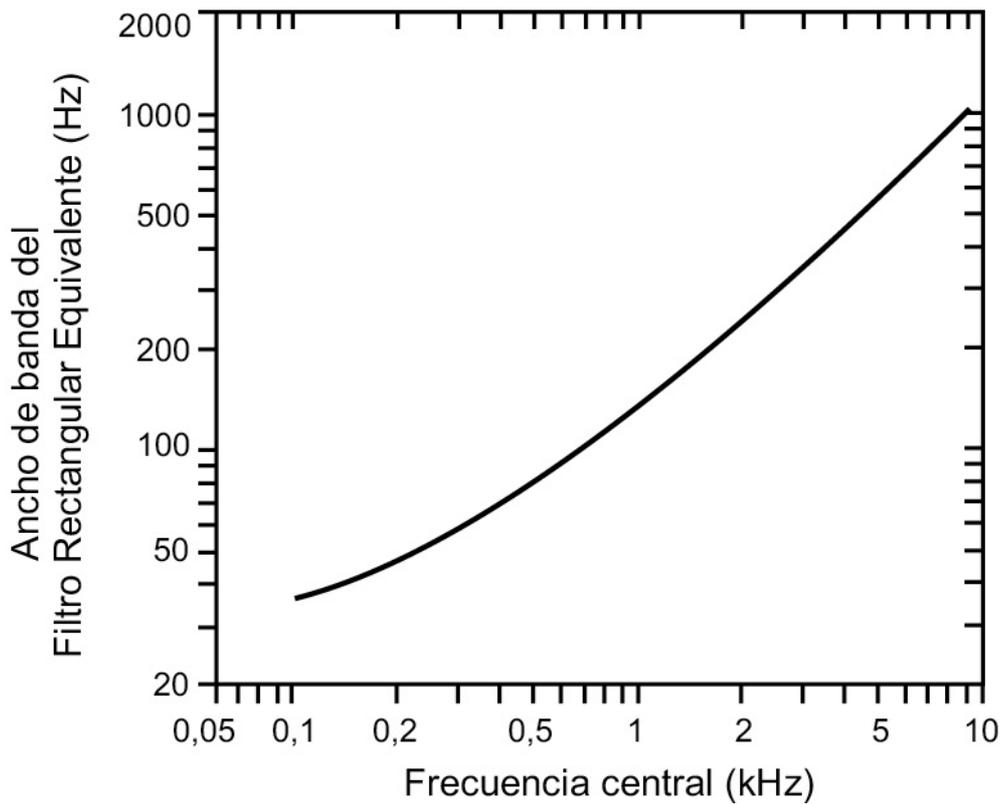
### *Método de supresión de bandas por cobertura*

Para satisfacer las condiciones impuestas por el modelo de Fletcher es necesario, entonces, usar un ruido enmascarante que evite la audición fuera de frecuencia, y que además limite el rango alrededor del cual la relación señal/ruido mantiene valores significativamente altos. El ruido enmascarante diseñado por Patterson, cuyo espectro se puede apreciar en la figura 6, cumple con ambos requisitos (Patterson, R. D., 1976). La forma del filtro auditivo puede estudiarse midiendo el umbral de potencia  $P$  para la señal en función del ancho de la muesca en el espectro del ruido.

El empleo de ruidos enmascarantes de esta clase es muy común en el estudio de señales transitorias, pues las bandas laterales “cubren” los puentes espectrales provocados por las transiciones rápidas de la señal, que aparecen como consecuencia del principio de indeterminación acústico (Basso, G., 1999, 2001a).



**Figura 6.** Esquema del ruido enmascarante usado por Patterson para estimar la forma del filtro auditivo



**Figura 7.** Estimación del ancho de banda del filtro rectangular equivalente en función de la frecuencia central

## *Forma del filtro auditivo*

Experimentos como los anteriores, y otros similares, han llevado a los investigadores a formular que el filtro rectangular equivalente sigue aproximadamente los valores de la figura 7.

Greenwood propuso en 1961 que el ancho del filtro rectangular equivalente representa una distancia constante a lo largo de la membrana basilar. La relación de Greenwood, corregida por Glasberg y Moore (1990), es la siguiente:

$$\text{ERB} = 24,7 (4,37 f + 1)$$

$f$  es la frecuencia central de la banda en Hz. Cada ERB corresponde a una distancia de alrededor de 0,89 mm en la membrana basilar. La conjetura de Greenwood se basa en las mediciones realizadas algunos años antes por Von Békésy.

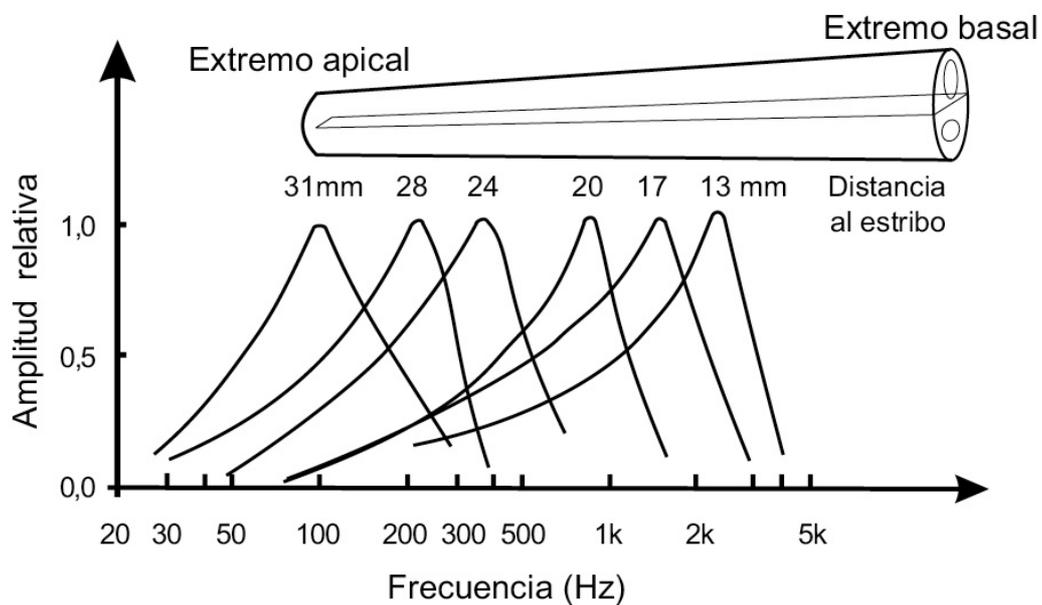
Si el filtro auditivo fuese lineal su forma no cambiaría con la intensidad del estímulo. Pero no es así: a bajos niveles de excitación es aproximadamente simétrico, pero deja de serlo a medida que aumenta la potencia de la señal. En el gráfico de la figura 9 se puede ver la forma del filtro auditivo en función de la intensidad. En particular, la extensión de baja frecuencia posee un ancho de banda mayor que la de alta frecuencia.

Está aceptado que las bandas críticas son continuas pues no hay evidencia experimental de discontinuidad entre ellas. Podemos hablar de la existencia de una banda crítica alrededor de cualquier frecuencia audible.

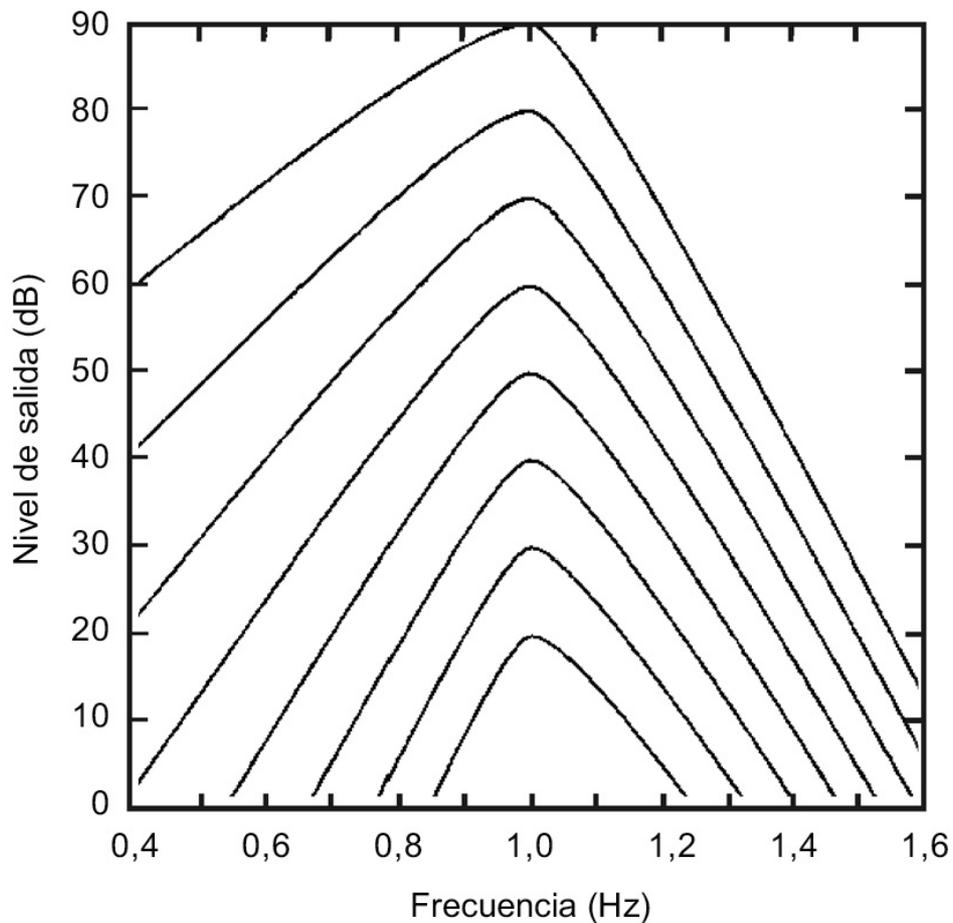
A veces, es útil especificar la escala de frecuencias de un experimento tomando el ancho del filtro rectangular equivalente como unidad:

$$\text{Número de ERB} = 21,4 \log_{10} (4,37 f + 1)$$

En la actualidad se puede predecir a partir de qué nivel una señal compleja puede ser detectada en un ambiente de ruido calculando los umbrales de sus parciales más prominentes. Para frecuencias medias el criterio dice que con una relación señal/ruido de -4 dB la señal comienza a oírse. Es decir, que el estímulo debe superar el nivel a -4 dB del ruido de salida del filtro auditivo que le corresponde. Un modelo que aplica consideraciones similares fue usado con éxito en el diseño de sistemas de comunicación en aeropuertos, a mediados de la década de 1990.



**Figura 8.** Posición del máximo de resonancia en la membrana basilar al ser estimulada por una señal sinusoidal de frecuencia variable (Békésy, G. Von, 1960)



**Figura 9.** Forma del filtro auditivo centrado a 1.000 Hz en función de la intensidad del estímulo, de 20 a 90 dB en pasos de 10 dB

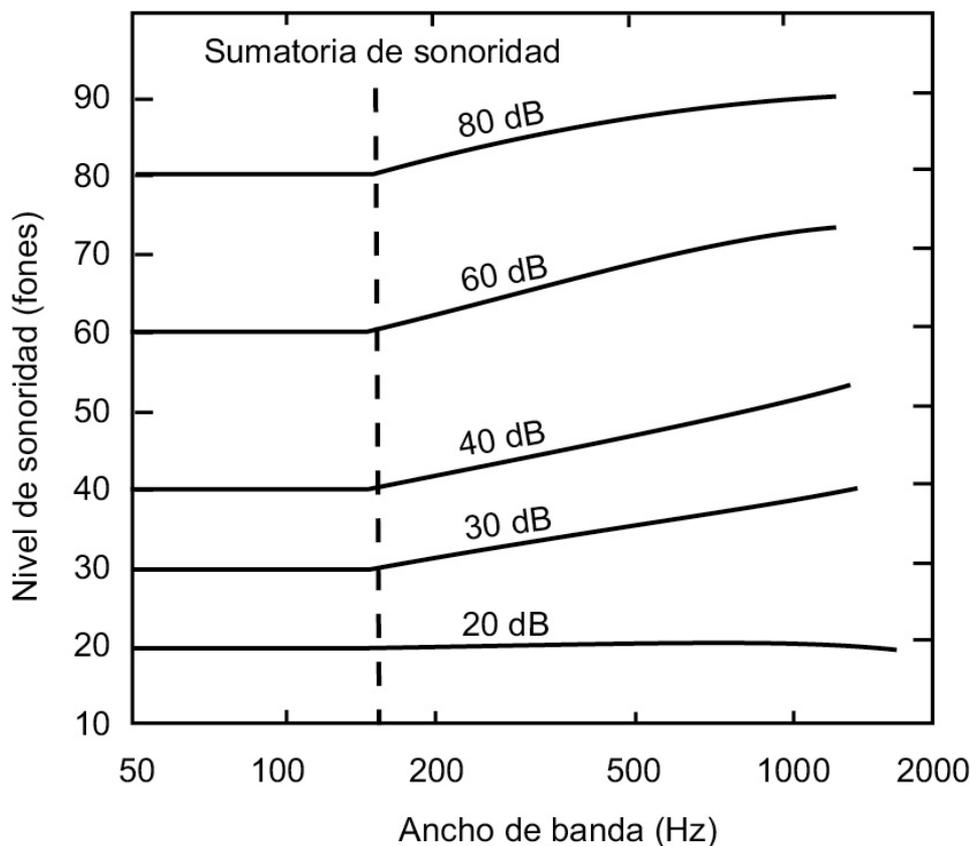
### 3. Sonoridad ante espectros complejos

En el apartado 4 del capítulo II analizamos la influencia de los factores espectrales que intervienen en la percepción de la sonoridad. Podemos ahora examinar el tema desde otra perspectiva.

Consideremos una señal acústica de intensidad fija y ancho de banda variable  $W$ . Mientras  $W$  se mantenga menor que la banda crítica  $BC$  la sonoridad será más o menos independiente de  $W$ , e igual a la de un tono puro de igual intensidad situado en la fre-

cuencia central de la banda. Cuando  $W$  supera la BC, la sonoridad comienza a incrementarse. Por lo tanto, para una determinada cantidad de energía la sonoridad es mayor si la energía se reparte sobre un número de bandas críticas, en vez de estar contenida en una sola de ellas. En la figura 10 se aprecia el aumento de sonoridad aludido. [4]

Este efecto es muy importante en música ya que el aumento de sonoridad depende de la suma de las sonoridades específicas en las diferentes bandas. Por ejemplo, doblar un bajo a la octava es mucho más efectivo, desde el punto de vista de la sonoridad, que doblarlo al unísono aunque la energía física puesta en juego sea la misma.



**Figura 10.** Nivel de sonoridad para un ruido de banda centrado en 1.000 Hz en función del ancho de banda

## Umbral para estímulos complejos

Cuando dos estímulos sinusoidales poco separados en frecuencia se presentan juntos, puede oírse sonido aun cuando cada estímulo por sí mismo esté ubicado por debajo del umbral. Gässler propuso que, dentro de una banda crítica, las energías de los componentes de un estímulo complejo se suman algebraicamente.[5] Spiegel, por el contrario, sugiere que el oído es capaz de integrar la energía sobre anchos de banda mayores que una BC (Spiegel, M. F., 1981). Algunos experimentos realizados con sinusoides muy separadas confirman esto último.

## Sensibilidad a la fase

En un célebre experimento realizado por Zwicker y repetido casi medio siglo más tarde por Sek se emplearon dos señales estacionarias, una modulada en amplitud con un índice de modulación  $m$  y otra modulada en frecuencia con un índice de modulación  $b$  (Sek, A., 1994). Si se igualan los índices de modulación ( $m = b$ ) y  $b$  es pequeño, el espectro de potencia de las señales es idéntico. Las componentes de las señales sólo difieren en sus fases relativas. El resultado de este experimento muestra que somos sensibles a las fases relativas de una señal estacionaria cuando las componentes caen dentro de una misma banda crítica. Por el contrario, cuando las componentes se encuentran separadas en frecuencia por una distancia que supere el valor de una banda crítica, las fases relativas dejan de percibirse. Sólo en este caso parece válida la *ley acústica de Ohm*, que postula la insensibilidad a la fase del sistema auditivo.[6]

## Discriminación de parciales en sonidos complejos

La ley acústica de Ohm también establece que el oído es capaz de percibir los componentes sinusoidales en un sonido complejo. Plomp encontró que esta habilidad llega, según la frecuencia de la

fundamental, hasta el quinto u octavo parcial (Plomp, R., 1976). Con su método –diferente al de Fletcher– estimó un valor para la banda crítica algo menor a la medida por el método de enmascaramiento. Plomp introdujo la hipótesis adicional de que un parcial puede ser oído en el entorno de otros parciales si se encuentra separado de estos últimos por lo menos en 1,25 veces el valor del ERB del filtro auditivo que le corresponde.

Los experimentos de Soderquist parecen probar que los músicos poseen bandas críticas menores a las de los no músicos (Soderquist, D. R., 1970). Sin embargo, Fine y Moore mostraron que no hay tal diferencia (Fine y Moore, 1993). La interpretación más aceptada es la que presume la existencia de algún mecanismo, más entrenado en los músicos, que actúa a niveles no periféricos en la detección de parciales.

## **Selectividad en frecuencia en oídos dañados**

Hay considerable evidencia de que la selectividad en frecuencia disminuye por daño en la cóclea. Una consecuencia es la gran sensibilidad al enmascaramiento que poseen las personas con lesiones cocleares, con la consiguiente dificultad para oír en ambientes ruidosos. Un impedimento adicional aparece ante el análisis perceptual de sonidos complejos como la palabra y la música, por ejemplo en la disminución de la habilidad para distinguir entre diferentes timbres. Con los audífonos convencionales no mejoran estas dificultades pues, aunque conviertan en audibles las señales, no se recupera la selectividad en frecuencia perdida.

## **4. Enmascaramiento**

El *enmascaramiento acústico* es un fenómeno habitual en nuestra vida cotidiana. Cuando elevamos la voz para ser oídos en una reunión, cuando perdemos parte de una conversación que es “tapada” por el paso de una motocicleta con el escape libre, o cuando el rui-

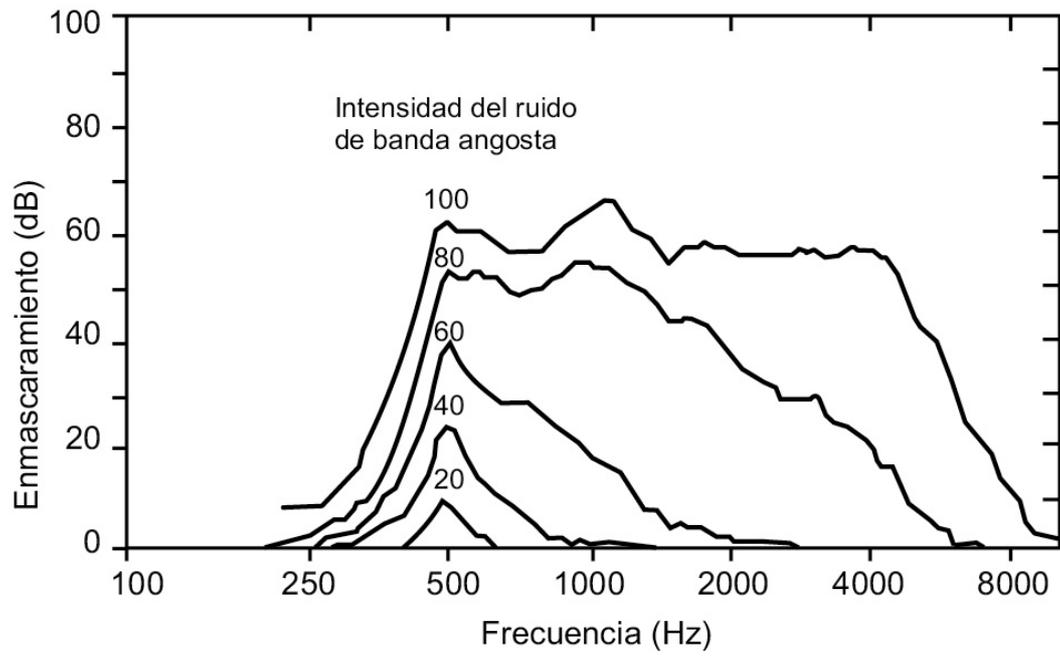
do de una aspiradora nos impide seguir un diálogo en el televisor, lo estamos experimentando. La cuestión puede plantearse de la siguiente manera: si alguien está oyendo un sonido intenso a una cierta frecuencia, ¿qué intensidad deberá tener otro sonido, a una frecuencia diferente, para que lo podamos percibir? El enmascaramiento se define como 1) el proceso que provoca un aumento del umbral de audibilidad para la señal enmascarada o 2) la cantidad de aumento del umbral expresada en dB.

Wegel y Lane publicaron el primer trabajo sistemático de investigación sobre el enmascaramiento de un tono por otro, pero sus resultados se complicaron por la aparición de batidos entre las señales (Wegel y Lane, 1924). Para superar el problema en los experimentos recientes se emplea ruido de banda angosta tanto para las señales enmascarante como enmascarada.

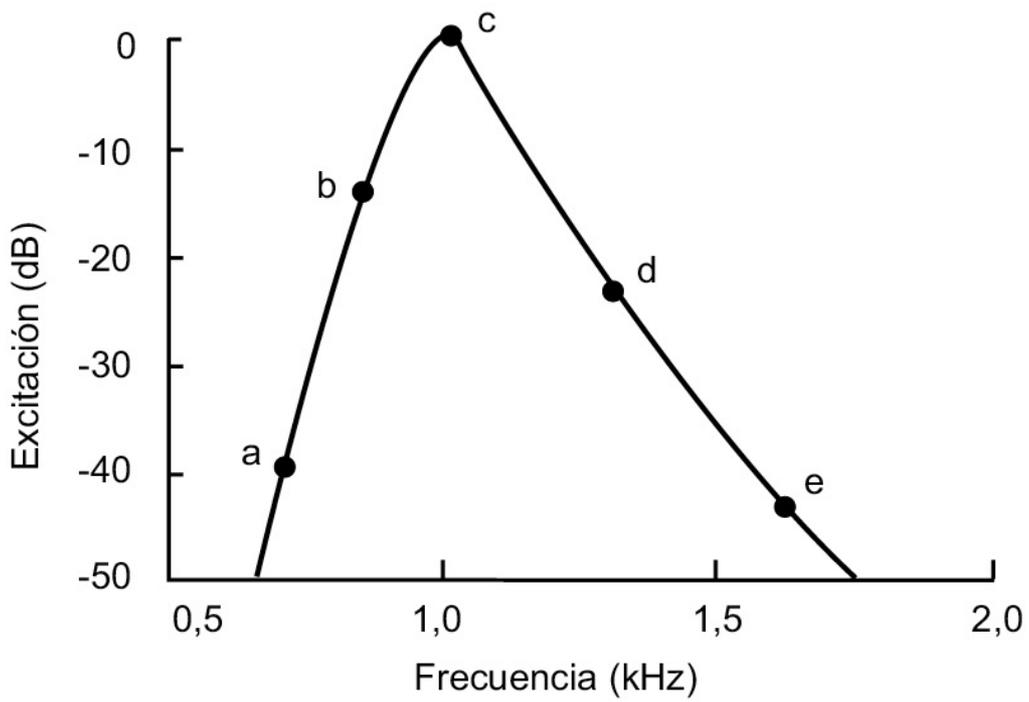
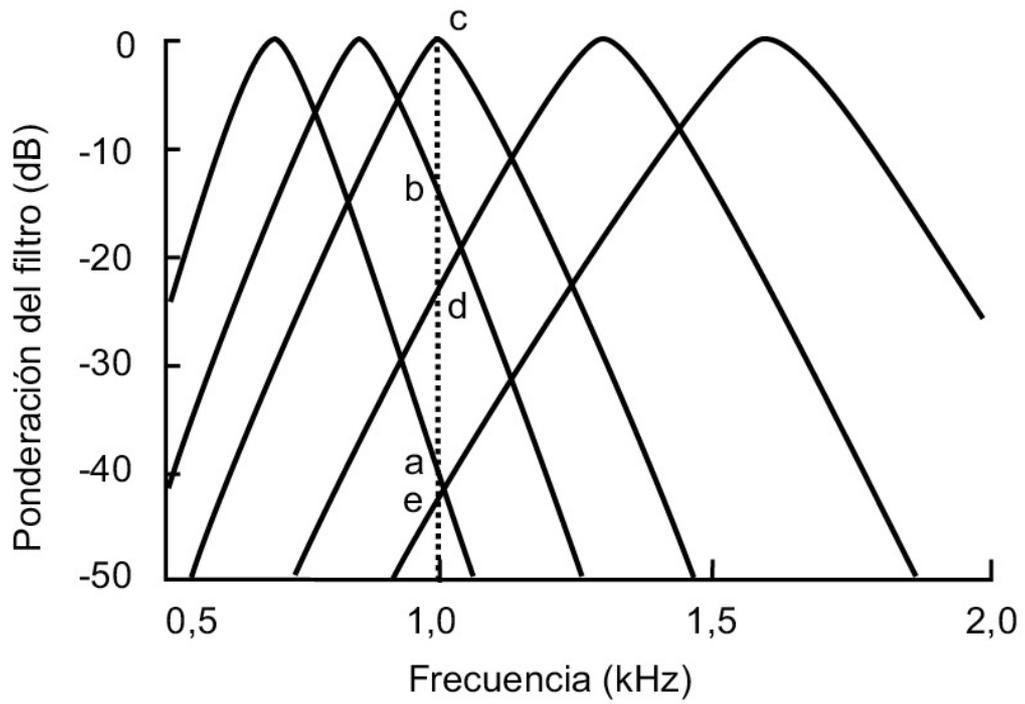
Los patrones de enmascaramiento obtenidos experimentalmente muestran una pendiente más abrupta en la zona de baja frecuencia que en la de alta frecuencia, tal como se aprecia en la figura 11.

Los gráficos anteriores ponen en evidencia el uso simultáneo de más de un filtro auditivo. Una explicación posible es interpretar esta forma como un indicador del modelo de excitación del ruido enmascarante: el patrón de enmascaramiento sería paralelo al patrón de la excitación, pero desplazado verticalmente una pequeña cantidad de dB. En la práctica la cuestión no es tan sencilla, pues también intervienen los factores propios del mecanismo de “oír una frecuencia diferente” y la detección de los productos de combinación entre la señal y el ruido enmascarante.

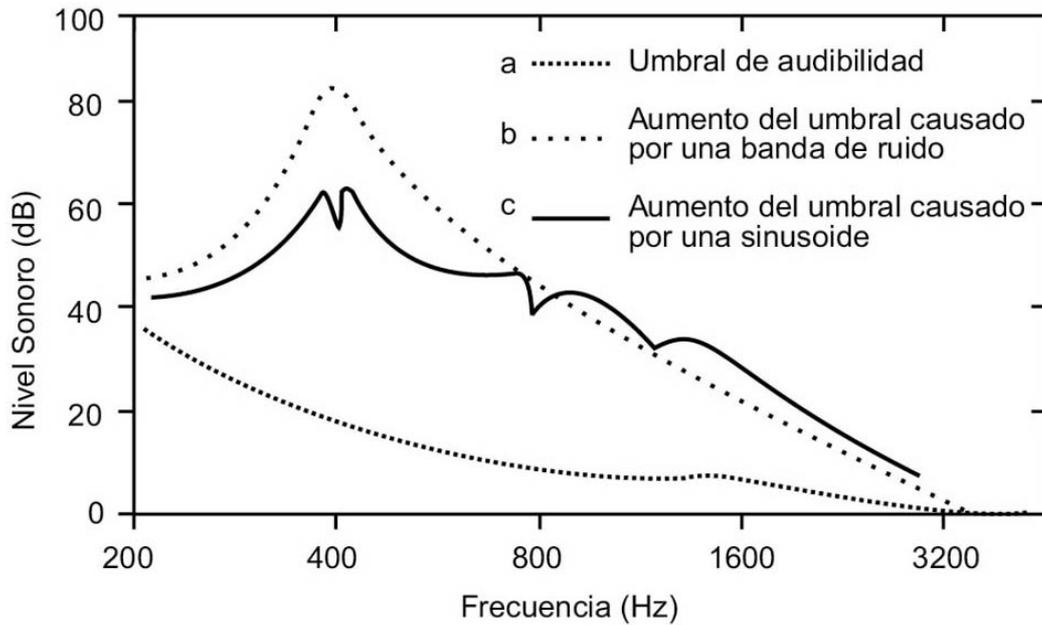
Moore y Glasberg describieron una forma de derivar las curvas de enmascaramiento usando el concepto de filtro auditivo (figura 12), en la que suponen que el ancho de banda de los filtros aumenta al aumentar la frecuencia central  $f_c$  (Moore y Glasberg, 1983).



**Figura 11.** Patrones de enmascaramiento para ruido de banda angosta centrado en 500 Hz



**Figura 12.** Derivación del patrón de enmascaramiento según Moore y Glasberg



**Figura 13.** Curvas de enmascaramiento: a) umbral de audibilidad, b) aumento del umbral causado por una banda de ruido y c) aumento del umbral causado por una sinusoide

A la descripción anterior se le debe agregar la acción de las ondas progresivas en la membrana basilar según el esquema de Von Békésy. Una frecuencia elevada produce un máximo de vibración en la membrana basilar cerca de la ventana oval, mientras que una de baja frecuencia se esparce prácticamente por toda la membrana. Esta característica de “enmascarar hacia las frecuencias agudas” se puede ver en el gráfico de la figura 13.

Como se observa en la figura 13, la segunda señal se percibe con facilidad si su frecuencia está alejada de la frecuencia de la señal enmascarante, pero la dificultad crece apreciablemente a medida que ambas señales se acercan. Esto no debería sorprendernos pues sabemos que, si comparten una misma banda crítica, dos señales se perturban mutuamente y la más débil no se percibe a menos que posea un porcentaje apreciable de la energía de la más intensa, algo así como 6 u 8 dB por debajo de la enmascarante. Cuando la segunda señal estimula una banda crítica diferente, sin embar-

go, el estímulo llega al cerebro por un canal separado y todavía no ocupado: la señal enmascarada puede percibirse estando aun 30 o 40 dB por debajo del ruido.

## **Relación entre las bandas críticas y el mecanismo de enmascaramiento**

Hay muchas semejanzas entre las selectividades en frecuencia medidas tanto en la membrana basilar como a partir de experimentos psicofisiológicos (Moore, B., 1995): cada ancho de banda rectangular equivalente (ERB) corresponde a alrededor de 0,9 mm en la membrana basilar, sin que importe el centro de frecuencia considerado. Sin embargo, a pesar de la buena correspondencia entre el comportamiento de la cóclea y la medida de los ERB, es posible que exista un mecanismo de “afinación” del ancho de banda en etapas superiores de procesamiento. Esto podría lograrse a través de mecanismos de supresión o inhibición lateral. Si la banda crítica depende de un proceso neural inhibitorio, se puede esperar que el mismo demore algunos milisegundos en actuar: la banda crítica para señales cortas debería ser mayor que para señales largas. Los datos experimentales en este punto se contradicen.

## **Modelos de cobertura y supresión**

Hay dos modelos que explican el proceso de enmascaramiento. El modelo de *cobertura* involucra la saturación de la actividad neural evocada por la señal. Si el ruido enmascarante produce una cantidad significativa de actividad en la banda crítica que normalmente responde a la señal, entonces la actividad agregada por la señal resultará cubierta y no se la podrá detectar. Considérese, por ejemplo, el caso de un tono junto a un ruido de banda ancha. El umbral del tono está a -4 dB por debajo del nivel del ruido en su banda crítica. La excitación combinada del tono y el ruido es 1,5 dB mayor que el ruido solo: 1,5 dB representa el incremento mínimo

necesario para la detección del tono (por ejemplo, si el tono estuviera a -20 dB, el nivel total habría aumentado 0,05 dB y no sería audible).

El segundo mecanismo posible, desarrollado por el modelo de *supresión*, se basa en la hipótesis de que el ruido enmascarante suprime la actividad neural que la señal produciría de estar sola.

Por el momento no hay evidencia fuerte que permita descartar alguno de los dos mecanismos, aunque quizá se combinen ambos. Sin embargo, el éxito de los modelos lineales de combinación de filtros auditivos en la explicación del patrón de enmascaramiento inclina la balanza hacia los mecanismos de cobertura, pues la supresión es un proceso altamente alineal.

## Modelos temporales de enmascaramiento

Existen datos que apoyan la hipótesis del uso de los *patrones temporales* de disparo neuronal en el análisis de estímulos complejos (una senoide produce disparos neurales cuyo intervalo temporal es un múltiplo entero de su período). Si la regularidad temporal es menor que cierto valor –como ocurre con el ruido aleatorio– la senoide será enmascarada. El mecanismo temporal parece efectivo únicamente para frecuencias por debajo de los 4 o 5 kHz y cabe suponer la acción de un dispositivo sensible a la fase.

Otro elemento que posiblemente se emplee en la detección de una señal sinusoidal en un fondo de ruido es el cambio en la envolvente de la onda combinada, que fluctúa menos en presencia de la señal sinusoidal.

## 5. Análisis espectro-temporal combinado

El modelo del espectro de potencia asume que, cuando tratamos de detectar una señal sinusoidal en presencia de un ruido enmascarante, hacemos uso de la salida de un filtro auditivo centrado a la frecuencia de la señal. En general, se supone que la prestación se

basa en la salida del filtro que provee la mayor relación señal/ruido y que los umbrales corresponden a una relación señal/ruido constante. Las fases relativas de las componentes y las fluctuaciones de corto tiempo del ruido enmascarante son ignoradas. Este modelo funciona bien en muchas situaciones, pero no en todas. En particular, hay evidencia de que los oyentes a veces realizan comparaciones entre diferentes filtros auditivos. Más aún, ciertas fluctuaciones temporales del ruido enmascarante pueden tener efectos importantes.

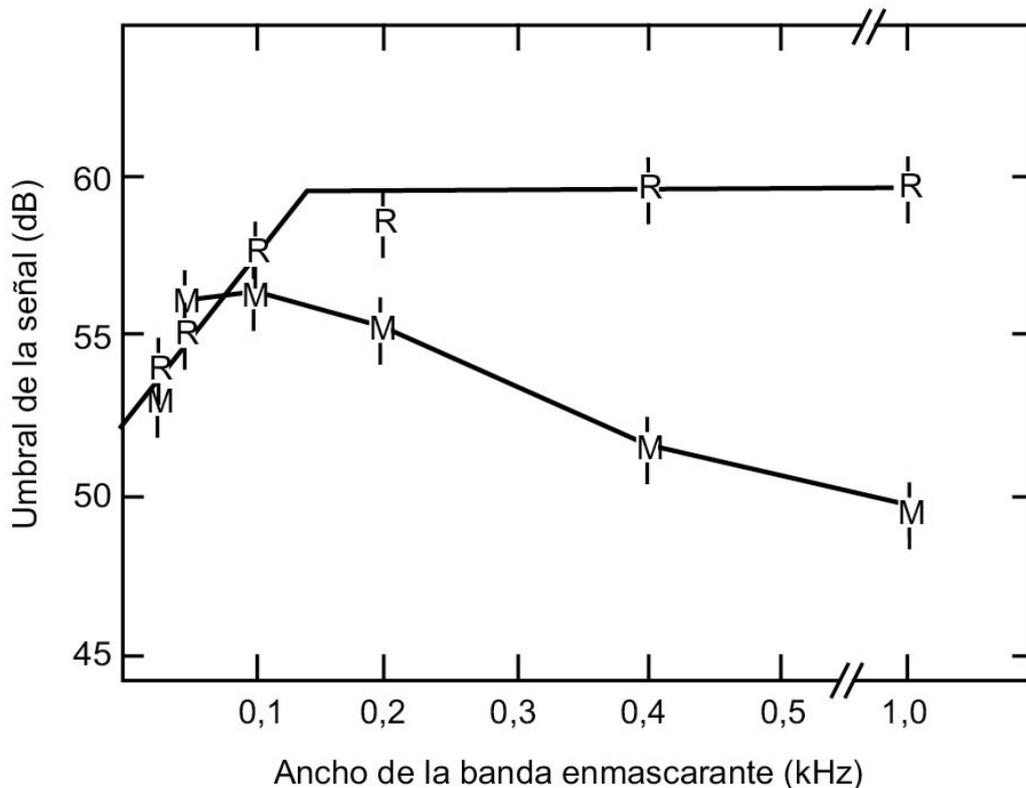
## **Reducción del enmascaramiento por comodulación**

Hall demostró que las comparaciones entre filtros auditivos pueden aumentar la detección de señales sinusoidales en presencia de ruido (Hall, D. *et al.*, 1984). Para que ocurra este efecto debe existir una fluctuación coherente o correlacionada entre diferentes bandas del ruido enmascarante, fenómeno al que Hall llamó *comodulación*.[\[7\]](#) En la figura 14 se aprecia el resultado de su experimento.

Hall usó ruido de una banda de ancho variable. También se pueden emplear dos o más bandas de ruido, incluso una en cada oído, sin que varíe el resultado.

Se asume usualmente que la reducción del enmascaramiento por comodulación depende de la comparación entre las salidas de diferentes filtros auditivos. Sin embargo, también se produce un efecto similar cuando la fluctuación cae dentro de una sola banda crítica, aunque en ese caso la disminución del umbral resulte menor. La reducción del enmascaramiento por comodulación es mayor cuando la variación del ruido enmascarante es lenta y cuando éste cubre un gran ancho de banda. Para señales de hasta 100 ms de duración la reducción del enmascaramiento por comodulación puede llegar a 11 dB. Esta disminución alcanza los 14 dB en el caso de ruido enmascarante de dos bandas cuando la banda agregada se ubica cerca de la frecuencia de la primera banda. También aumen-

ta hasta llegar a un máximo de 16 dB cuando hay más de dos bandas en juego. La magnitud de la reducción del enmascaramiento por comodulación no cambia con la frecuencia central del estímulo.



**Figura 14.** Resultado del experimento de Hall. Los puntos marcados “M” corresponden a los umbrales obtenidos con ruido enmascarante modulado en amplitud con una señal irregular que varía lentamente

La comparación entre las diferentes envolventes temporales de los filtros auditivos parece ser un hecho común en el análisis de los patrones auditivos. Aparentemente juega un papel muy importante en la extracción de señales de ambientes ruidosos y en la discriminación entre varias fuentes acústicas. También da cuenta, al menos en parte, del “efecto de cafetería” (“cocktail party effect”).

Hay dos clases de modelos que intentan explicar la reducción del enmascaramiento por comodulación. Los que integran el primer grupo asumen que el sistema auditivo compara los patrones de modulación de las envolventes de las salidas de los diferentes filtros auditivos: la presencia de una señal es revelada por la disparidad en la modulación entre los filtros.

Un segundo grupo de modelos sugiere que la fluctuación de las envolventes a la salida de los filtros auditivos indican los momentos óptimos en los que se debe escuchar la señal. Es decir, durante los mínimos de la envolvente del enmascarante.

Estos modelos no son mutuamente excluyentes y no parece que la reducción del enmascaramiento por comodulación dependa de un único mecanismo simple.

## **Enmascaramiento no simultáneo**

Los efectos del tiempo sobre el enmascaramiento han sido estudiados en detalle por numerosos investigadores. El enmascaramiento “hacia atrás”, con la señal precediendo al ruido enmascarante, parece provenir de una confusión entre las señales que afecta a los sujetos con poca práctica. De hecho, oyentes entrenados –músicos y técnicos de audio entre otros– obtienen mejores porcentajes de detección que los sujetos promedio. Por otro lado, el enmascaramiento “hacia delante”, con el ruido enmascarante precediendo a la señal, es uno de los tres procesos que alteran el umbral de una señal que tiene lugar después de otro sonido. Los otros dos son la fatiga y la adaptación. Se distingue de ellos por estar limitado a señales que aparecen dentro de los 200 ms después del momento en que cesa el ruido enmascarante.

El enmascaramiento hacia delante tiene las siguientes particularidades: 1) es mayor cuanto más cerca en el tiempo se suceden ambas señales; 2) los incrementos en el nivel del ruido enmascarante no corresponden a incrementos iguales en el umbral de la señal (a

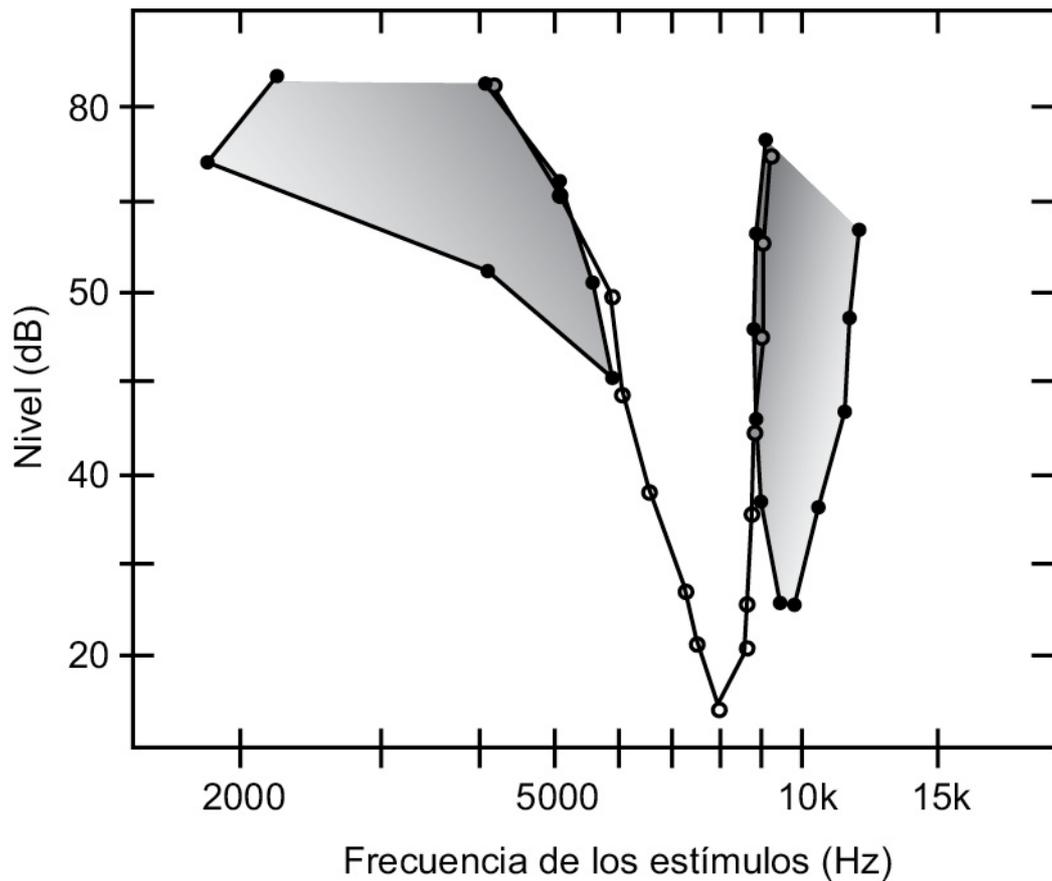
diferencia del enmascaramiento simultáneo, que responde a una relación señal/ruido constante), y 3) depende de la relación de frecuencias entre ambas señales (igual que en el enmascaramiento simultáneo). Los procesos que dan cuenta del enmascaramiento no simultáneo no se conocen con exactitud, aunque probablemente sean consecuencia de la combinación de varios a la vez. La superposición temporal de patrones de vibración en la membrana basilar –una suerte de cola de resonancia o “ringing”– parece ser importante, así como la adaptación de corto término tanto en el nervio auditivo como en los centros neurales superiores.

### ***Supresión lateral en el enmascaramiento no simultáneo***

Los experimentos sobre enmascaramiento simultáneo pueden ser explicados a partir de la superposición de filtros auditivos aproximadamente lineales. Sin embargo, las respuestas neurales son marcadamente alineales. En particular, es importante destacar el hecho de que la respuesta a una señal puede ser suprimida por otra de diferente frecuencia en el fenómeno de *supresión entre dos tonos*, que describe el hecho de que la respuesta ante un estímulo sinusoidal puede anularse por la presencia de un segundo estímulo. En la figura 15 se aprecia este importante efecto.

Los efectos de la supresión comienzan y cesan muy rápidamente, dentro de unos pocos milisegundos luego del comienzo o fin del segundo estímulo.

Para señales más complejas, el fenómeno toma el nombre de *supresión lateral*, en la que una fuerte actividad a una frecuencia central puede suprimir una actividad menor a una frecuencia adyacente. Las crestas del patrón de excitación son resaltadas por los valles que las rodean.



**Figura 15.** Áreas de supresión de un estímulo por otro

La pregunta es ¿por qué la supresión lateral no se revela en el enmascaramiento no simultáneo? Houtgast estudió extensamente el caso a partir del *umbral de pulsación* (Houtgast, T., 1972). En el enmascaramiento simultáneo la supresión se manifiesta porque cualquier reducción en la actividad del ruido enmascarante en la región de frecuencias de la señal se acompaña con una reducción similar en la actividad de la señal enmascarada. En otras palabras, la razón “señal enmascarada / ruido enmascarante” no es afectada por la supresión. En el enmascaramiento no simultáneo la supresión afecta al ruido enmascarante pero no a la señal.

Otra diferencia importante es que la selectividad en frecuencia del enmascaramiento no simultáneo es mayor que en el simultáneo (medido por ejemplo a partir de la curva de afinación psicofísica o PTS).

## ***Examen del perfil espectral***

Green y sus colegas demostraron en 1988 que los oyentes comparan la salida de diferentes filtros auditivos para mejorar la detección de una señal específica, aún en los casos en los que el ruido y la señal fluctúan sincrónicamente. Sugirieron que se percibe un cambio entre el perfil espectral del ruido en presencia de la señal y el perfil en ausencia de la misma. El análisis es más efectivo cuando: 1) el ruido de fondo posee un gran ancho de banda; 2) hay varias componentes de la señal dentro del rango espectral del ruido y 3) el ruido de fondo se compone de bandas de similar intensidad.

En un sentido el examen del perfil espectral no sorprende: la forma del espectro es uno de los factores relevantes a la hora de identificar una vocal o un timbre musical.

## **Enmascaramiento binaural**

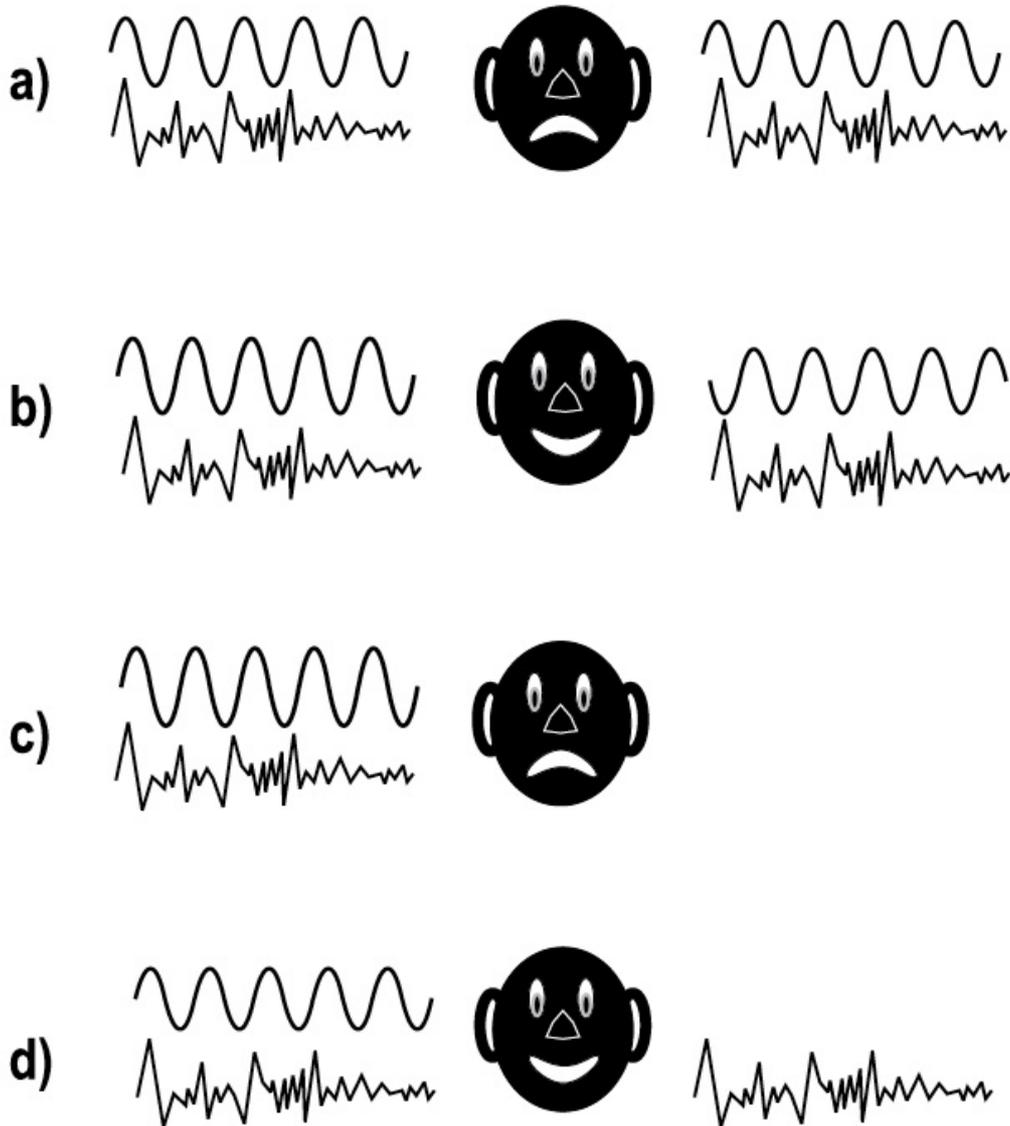
El umbral de enmascaramiento puede ser marcadamente menor cuando la señal se oye con ambos oídos. La diferencia de nivel entre la audición binaural y la monoaural se conoce como *nivel de enmascaramiento diferencial* (MLD). El MLD llega a ser de 15 dB para bajas frecuencias y se reduce a sólo 2-3 dB para señales de 1.500 Hz. No es casual que el límite de 1.500 Hz sea el mismo que el de nuestra capacidad para comparar las fases de las señales entre los dos oídos durante la localización espacial. En la figura 16 se aprecia cualitativamente un experimento clásico en este campo.

Para que el efecto tenga lugar se debe agregar el mismo ruido a los dos oídos (o al menos deben estar correlacionados o provenir de la misma fuente). Se ha observado enmascaramiento binaural en señales sinusoidales, en clics y en el habla. En situaciones reales aparece cuando la señal y el ruido provienen de diferentes posiciones en el espacio y se relaciona fuertemente con el efecto “cocktail party”. La banda del ruido enmascarante que actúa en las condiciones de enmascaramiento diferencial es la que corresponde a la

frecuencia de la señal. Hall sugirió que los filtros auditivos que intervienen en la detección monoaural y binaural son los mismos, pero que la detección binaural depende de la salida de más de uno de ellos en paralelo.

La inteligibilidad del habla en condiciones de ruido ambiental mejora con señales antifásicas, en especial a niveles cercanos al umbral. Las condiciones antifásicas no sólo mejoran la capacidad de detección de señales, sino que también incrementan la habilidad para discriminar señales.[\[8\]](#) Los procesos de enmascaramiento y de localización espacial, sin embargo, parecen procesarse de manera separada.

Los modelos más recientes, como los de Colburn, Stern y Trahiotis, asumen que las tasas de disparo de los diferentes filtros auditivos provenientes de los dos oídos son comparadas por detectores de coincidencia que proveen una especie de salida bidimensional, con una dimensión caracterizada por las frecuencias centrales de los filtros auditivos y la otra por el retardo interaural (Colburn, H. S., 1995). Un complejo modelo, el de Yoichi Ando, da cuenta con bastante detalle del proceso de evaluación interaural y fue utilizado en el diseño de salas para música no amplificada (Ando, Y. *et al.*, 1997a).



**Figura 16.** Enmascaramiento binaural

## 6. Distorsión alinear y sonidos de combinación

El violinista italiano Giuseppe Tartini observó en 1714 que al ejecutar dos sonidos muy intensos se generaba un tercer sonido, más grave, cuya frecuencia correspondía a la diferencia entre las frecuencias de los dos sonidos originales. La aparición de dicho sonido diferencial no debería sorprendernos ya que sabemos que el oído

presenta comportamientos alineales en prácticamente todas las etapas de transmisión y procesamiento de señal.<sup>9</sup>[\[9\]](#) El tímpano, la cadena de huesecillos y el oído interno contribuyen de manera variable a la alinealidad total, aunque es muy difícil hallar el origen exacto de la distorsión en cada caso.

Un sistema de transmisión alineal produce armónicos cuando se lo excita con una senoide pura (Basso, G., 2001a). Los productos de esta distorsión armónica son difíciles de oír pues se confunden con los armónicos propios del estímulo físico y se integran a la periodicidad definida por la fundamental de Fourier de la señal. Pero si el oído es excitado con dos señales sinusoidales, la distorsión alineal produce una gran cantidad de artefactos que se perciben como *sonidos resultantes o de combinación*. En teoría, existen infinidad de sonidos resultantes, pero en la práctica son audibles solamente los de menor frecuencia. Se los puede clasificar de acuerdo con la cantidad de veces que aparecen las frecuencias de los estímulos  $f_1$  y  $f_2$  en la fórmula correspondiente. En la tabla de la figura 17 se indican los sonidos de combinación de menor orden.

En la tercera columna están los armónicos de cada uno de los componentes primarios. Por lo general, los sonidos de combinación de las dos primeras columnas no siguen una estructura armónica.  
[\[10\]](#)

## Sonidos de combinación de segundo orden

El espectro de los sonidos de combinación de segundo orden o *cua-dráticos* se puede apreciar en la figura 18.

Además de los armónicos de los componentes primarios ( $2f_1$  y  $2f_2$ ), se genera un *sonido diferencial* ( $f_2 - f_1$ ) y uno *adicional* ( $f_1 + f_2$ ) de gran amplitud. El sonido adicional es difícil de oír porque cae en una región espectral en la que existen armónicos, si las señales primarias son complejas. También es enmascarado con facilidad por los componentes primarios  $f_1$  y  $f_2$ .

En cambio, el sonido diferencial se percibe con facilidad, sobre todo cuando su frecuencia es mucho menor que  $f_1$ . Aunque la frecuencia resultante ( $f_2 - f_1$ ) es la misma, no hay que confundir el sonido diferencial –producto de la distorsión alineal del oído– con el batido físico, que se origina en la interferencia entre señales sinusoidales –un proceso lineal.

Para un sistema regular con distorsión cuadrática, el nivel del sonido diferencial sigue la siguiente ecuación:

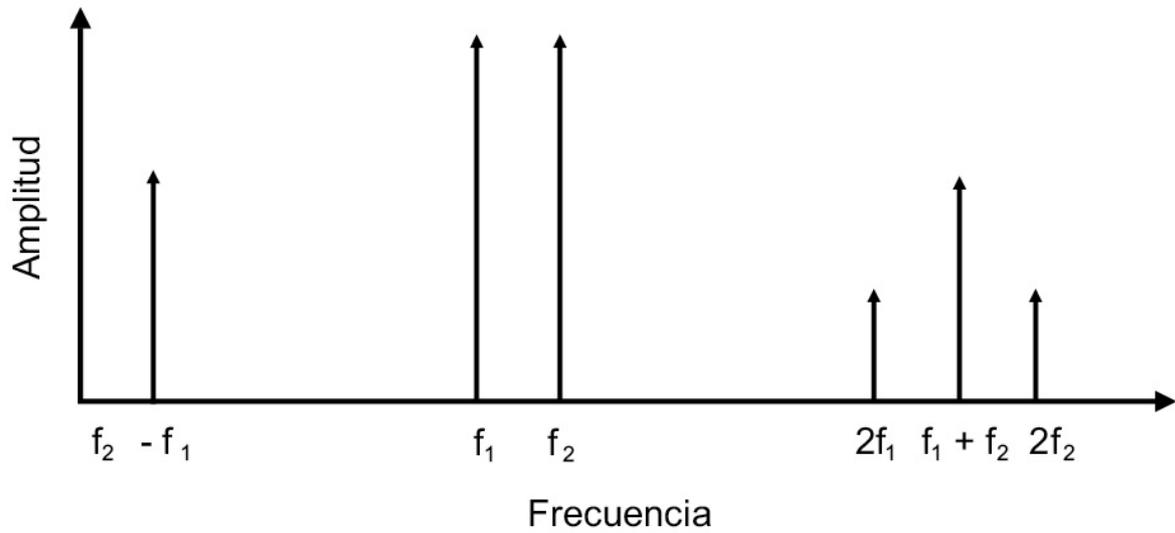
$$NI (f_2 - f_1) = NI (f_1) + NI (f_2) - C_2$$

$C_2$  es una constante que depende de la profundidad de la distorsión cuadrática. Los experimentos psicoacústicos sobre el tema la ubican en un valor aproximado de 130 dB. Es decir, con  $NI (f_1) = NI (f_2) = 70$  dB, el nivel de intensidad del diferencial alcanza 10 dB y no se lo percibe en todas las ocasiones. Pero con  $NI (f_1) = NI (f_2) = 90$  dB el nivel de intensidad del diferencial llega a 50 dB y se lo percibe con relativa facilidad.

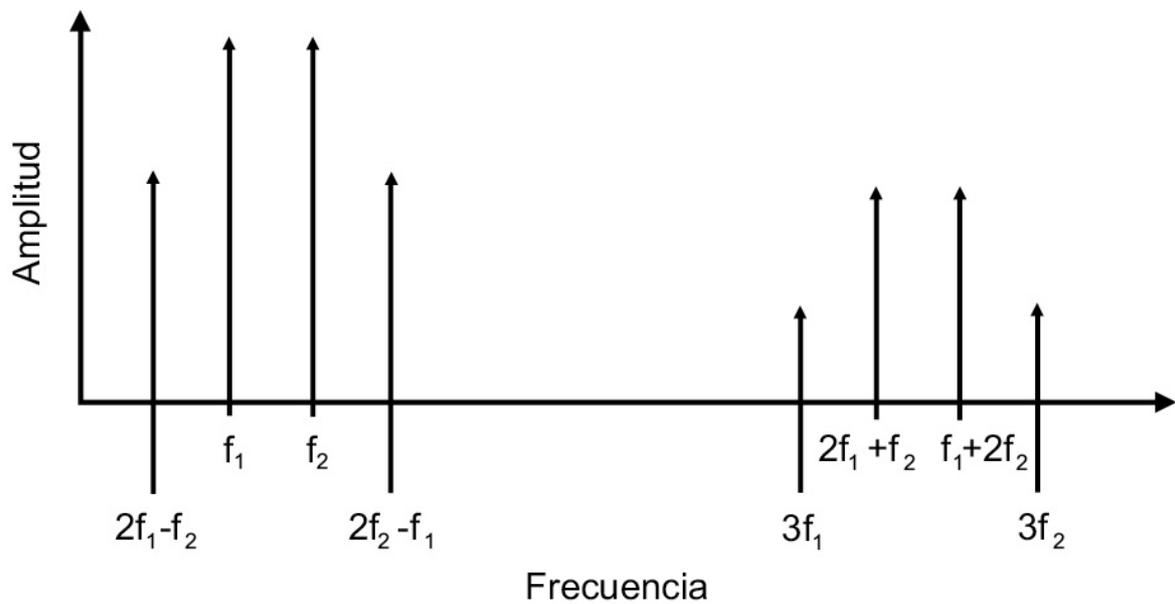
Se asume que la alinealidad causante de los sonidos de combinación de segundo orden se origina en su mayor parte en el oído medio, con una participación menor del tímpano y del oído interno.

Primarias	$f_1$	$f_2$	
Secundarias	$f_2 - f_1$	$f_2 + f_1$	$2f_1, 2f_2$
Tercer orden	$2f_1 - f_2, 2f_2 - f_1$	$2f_1 + f_2, 2f_2 + f_1$	$3f_1, 3f_2$
Cuarto orden	$3f_1 - f_2, 3f_2 - f_1,$ $2f_1 - 2f_2$	$3f_1 + f_2, 3f_2 + f_1,$ $2f_1 + 2f_2$	$4f_1, 4f_2$
...	...	...	...

**Figura 17.** Frecuencias correspondientes a los sonidos de combinación de menor orden. Se presupone  $f_2 > f_1$



**Figura 18.** Distorsión cuadrática pura a partir de dos componentes primarios de frecuencia  $f_1$  y  $f_2$



**Figura 19.** Distorsión cúbica pura a partir de dos componentes primarios de frecuencia  $f_1$  y  $f_2$

### Sonidos de combinación de tercer orden

El espectro de los sonidos de combinación de tercer orden o *cúbicos*, generados por un sistema regular con distorsión cúbica, se

puede apreciar en la figura 19.

Los sonidos adicionales cúbicos de alta frecuencia ( $2f_1 + f_2$ ) y ( $2f_2 + f_1$ ) son extremadamente difíciles de oír. Los sonidos diferenciales cúbicos ( $2f_1 - f_2$ ) y ( $2f_2 - f_1$ ) poseen el triple de amplitud que los terceros armónicos ( $3f_1$ ) y ( $3f_2$ ) de los componentes primarios. En ciertas circunstancias el más bajo de ellos, ( $2f_1 - f_2$ ), es tan audible como el diferencial cuadrático ( $f_2 - f_1$ ).

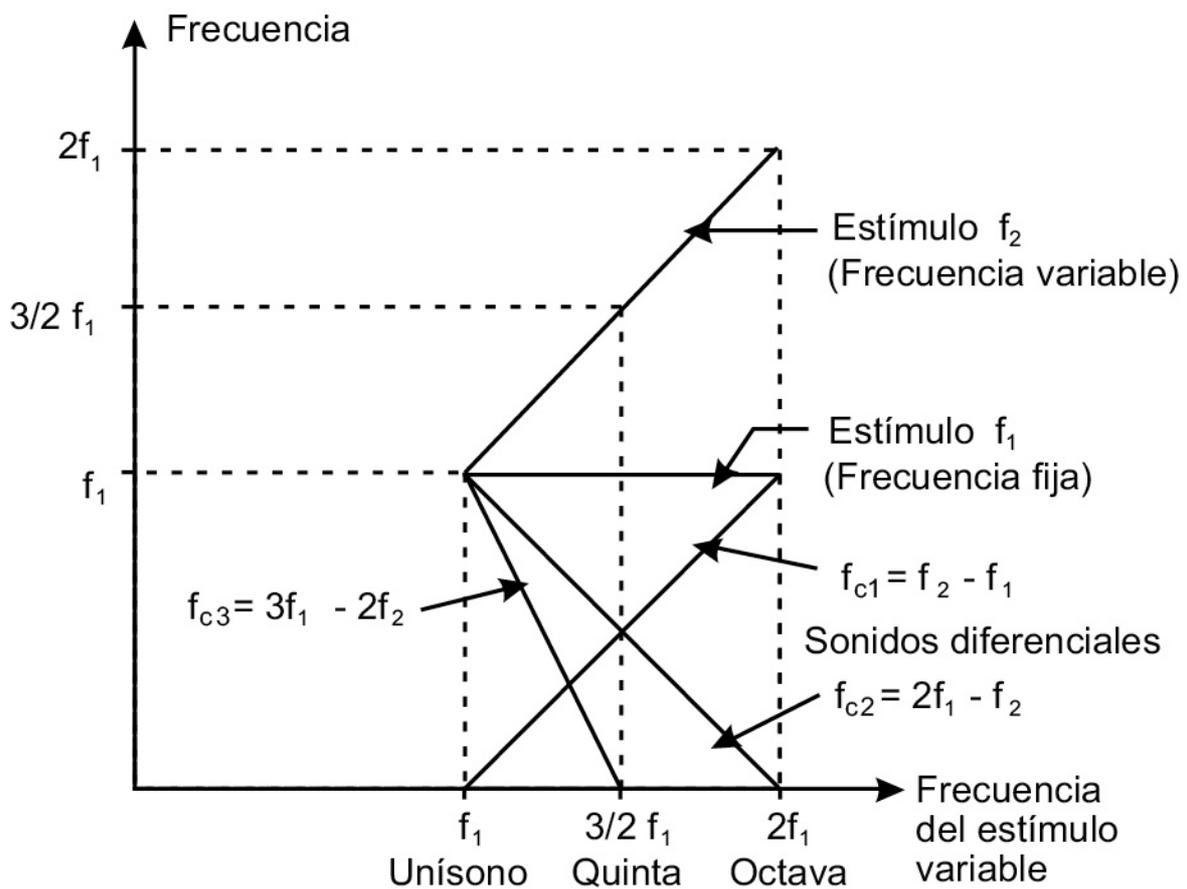
Sin embargo, en este caso la relación entre niveles de entrada y sonoridad percibida no sigue una ley tan simple como en el caso de los sonidos de combinación de segundo orden. El diferencial cúbico ( $2f_1 - f_2$ ) persiste a niveles de estímulo muy bajos y su audibilidad depende en gran medida de la frecuencia de los componentes primarios. Por ejemplo, decrece rápidamente en sonoridad cuando la diferencia entre  $f_2$  y  $f_1$  supera un valor cercano al 30%. Estos hechos revelan la existencia de una *alinealidad esencial*, en la que el sistema no se puede aproximar a un sistema lineal para pequeñas amplitudes de señal. Se sabe que el origen de esta alinealidad esencial aparece más allá del oído medio. Se supone que surge de la actividad de las células motoras en la cóclea o en algún centro de procesamiento neural superior.

En la figura 20 se aprecia el comportamiento de los sonidos diferenciales más audibles en función de la diferencia de frecuencia entre los componentes primarios y en la figura 21 se grafican en notación musical aproximada los diferenciales cuadráticos y cúbicos presentes en algunos intervalos sencillos.

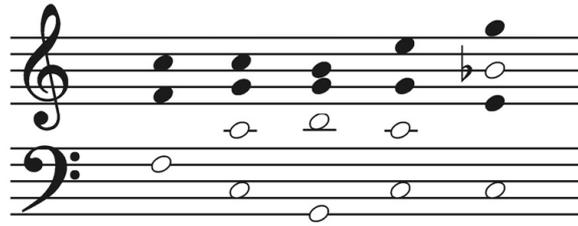
## 7. Algoritmos psicoacústicos de compresión

El fenómeno de enmascaramiento se emplea desde la década de 1980 en el diseño de sistemas para compresión de archivos de audio de gran eficacia. Se los denomina genéricamente *algoritmos psicoacústicos de compresión* y, de todos ellos, en la actualidad el

más difundido es el MP3. El objetivo primario de estos sistemas es el de reducir el tamaño de los archivos digitales de audio que –en codificación PCM estándar– ocupan alrededor de 10 Mb por minuto de registro y tienen una tasa de transmisión de 1,4 Mbps. La compresión permite un ahorro de espacio de almacenamiento y una mayor velocidad de transmisión a través de sistemas de red. A continuación describiremos brevemente las características principales del código MP3, que en principio no difiere sustancialmente de otros formatos de compresión de audio menos difundidos.



**Figura 20.** Diferenciales cuadráticos y cúbicos en función de la diferencia de frecuencia entre los componentes primarios (tomado de Roederer, 1997)



**Figura 21.** Notación musical aproximada de los sonidos diferenciales más destacados

La sigla MP3 es una abreviatura de MPEG-1 Layer 3, un formato definido por el Motion Picture Experts Group (MPEG). Es parte de un extenso programa destinado a establecer estándares para la industria del cine, la televisión, la música y de sistemas multimedia. [11] El algoritmo de compresión/descompresión (Codec) del formato MP3 comprende dos etapas principales. En la primera etapa se remueve la parte inaudible de la señal y en la segunda se eliminan las redundancias del material remanente.

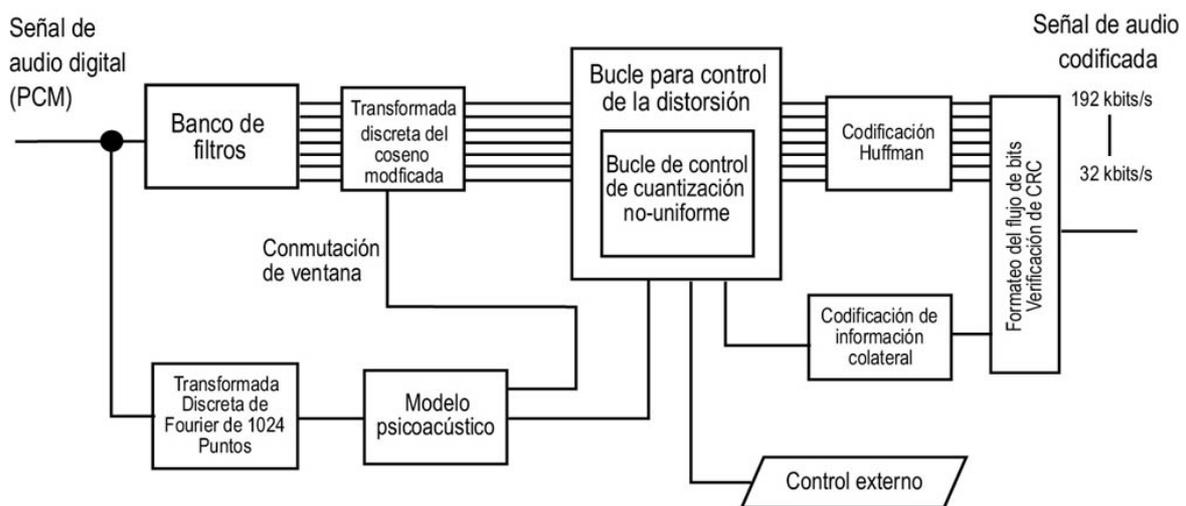
En primer lugar, los archivos de audio, generalmente en dos canales PCM de 44,1 kHz de muestreo y 16 bits de resolución, son transformados en frecuencia y separados en 18 sub-bandas por canal. Cada sub-banda reproduce aproximadamente el tamaño de un ancho de banda crítico. La tasa de transmisión se fija a un valor menor al de un CD de audio para limitar intencionalmente la cantidad de bits disponibles y forzar la compresión. A partir de un modelo psicoacústico se examina cada sub-banda y se decide a cuáles se destina la mayor cantidad de bits. Por ejemplo, las bandas que son enmascaradas por otras más potentes reciben menos bits porque en ellas no se va a percibir el ruido de cuantización, producto inevitable de la menor resolución. En otros términos, se remueve la parte de la señal que no vamos a poder oír al quedar enmascarada.

A la salida del codificador psicoacústico se agrega un codificador por redundancia, por ejemplo uno de Huffman, que permite llevar la tasa de compresión a un valor final de hasta 12:1 sin que se de-

grade significativamente la calidad de audio. En la figura 22 se puede apreciar un diagrama en bloques del proceso completo.

La profundidad de compresión depende de la calidad de audio requerida y del tipo de señal que se va a codificar. Por lo general, una tasa de transmisión de 128 kbps se considera suficiente para alcanzar la calidad de audio de un CD, al menos si se emplea un sistema de monitoreo de calidad media y la música no posee gran rango dinámico.

Es muy importante tomar en cuenta que la codificación MP3 y todas las que emplean esta clase de procesamiento psicoacústico son técnicas a pérdida de información. La señal codificada no puede volver a su estado original una vez finalizado el proceso.



**Figura 22.** Diagrama en bloques de un codificador MP3

## Notas

- [1](#) En inglés, *just noticeable difference*.
- [2](#) Harvey Fletcher (1884-1981), físico e ingeniero estadounidense. Fue uno de los más destacados investigadores en psicoacústica durante la primera mitad del siglo XX.
- [3](#) *Off-frequency listening* en inglés.

- [4](#) Algunos autores lo denominan “suma de sonoridad” (*loudness summation*).
- [5](#) Gässler (1954), en Moore, B. (1995).
- [6](#) Ohm (1843), en Moore, B. (1995).
- [7](#) Hoy se lo denomina reducción del enmascaramiento por comodulación o CMR (*commodulation masking release*).
- [8](#) Un parámetro utilizado en el estudio de salas para música y palabra, el *coeficiente de correlación cruzada interaural* (IACC), cuantifica la correlación existente entre las señales que llegan a los dos oídos.
- [9](#) Aunque aquí adoptamos la denominación tradicional de *sonidos diferenciales* y de *sonidos de combinación*, no los debemos confundir con las señales físicas que los causan ni con las frecuencias de dichas señales físicas.
- [10](#) Una excepción se da cuando una frecuencia primaria es igual o múltiplo de la otra. Por ejemplo,  $f_2 = N f_1$  (Basso, G., 2001a).
- [11](#) La parte de audio de MPEG-1 (norma ISO/IEC IS 11172-3) fue establecida en 1991 y consiste en tres modos de operación de complejidad creciente: Layer-1, Layer-2 y Layer-3. En el estándar MPEG-2, sucesor del MPEG-1, se define el Advanced Audio Coding (AAC) que está reemplazando al MP3 en muchas aplicaciones.

## Capítulo IV

### Procesamiento temporal de señales

Decir que los sonidos del ambiente, de la música y del habla dependen, necesitan y, en última instancia, están hechos de tiempo, es enunciar una verdad de Perogrullo. La naturaleza del sonido es el devenir y el estudio de sus rasgos esenciales ha planteado, y plantea, arduos problemas filosóficos. De las muchas experiencias que nombramos con el término “tiempo”, algunas involucran nociones relacionadas con la sucesión, el presente, el pasado y el futuro. Otras con las ideas de instantaneidad, simultaneidad y duración. También están las que tienen que ver con la flecha temporal y la irreversibilidad del tránsito del pasado al futuro. De la noción de tiempo se derivan de forma más o menos directa otras como movimiento, velocidad, pulso o ritmo. La dificultad del llamado problema del tiempo radica en articular semejante variedad de fenómenos y conceptos dentro un sistema coherente.

Son abundantes y variadas las soluciones propuestas hasta ahora para resolverlo. Pero es posible reconocer en la historia del pensamiento dos grandes tendencias. La primera –que cuenta en sus filas a Aristóteles, Newton y Einstein– considera al tiempo como un hecho físico al que se adapta la mente humana. La segunda supone que el tiempo es un fenómeno que pertenece a la mente humana y que el de los físicos es un constructo proveniente de la idealización de los contenidos mentales. Se pueden situar en esta última línea a pensadores como San Agustín, Brentano o Husserl. Para Paul Ricœur estas opciones no se solapan por entero: a la perspectiva física le falta la capacidad de concebir un presente a partir del instante, y a la perspectiva psicológica le falta sustento a la hora de explicar de forma satisfactoria el durar.

En la actualidad, la mayoría de los investigadores conjeturan que el tiempo no responde a un modelo de tipo unívoco, sino que

constituye una entidad que se debe pensar desde varias perspectivas diferentes. Según Fraser (1992), existen al menos seis categorías temporales:

1. Atemporalidad: en un universo en el que solamente hay radiación electromagnética el tiempo no existe. La emisión y la absorción de un fotón ocurren, en lo que se refiere al fotón, simultáneamente.

2. Prototemporalidad: en un universo de partículas elementales el tiempo es fragmentario. No fluye y carece de dirección. Los sucesos se pueden situar de forma estadística porque la singularización de un instante concreto carece de significado.

3. Eotemporalidad: en el universo macroscópico de la materia dotada de masa el tiempo es continuo, pero inmóvil y sin dirección. No se le aplican los conceptos irreversibles de pasado, presente y futuro. Es el tiempo de las leyes de la mecánica clásica.

4. Biotemporalidad: es el tiempo de los animales. Tiene dirección y existe un pasado, un presente y un futuro, pero sus horizontes son limitados. Los bordes del presente son estables y probablemente característicos de cada especie.

5. Nootemporalidad: es el tiempo de la mente humana. Existe una clara noción de pasado, presente y futuro y sus horizontes son ilimitados. El presente tiene bordes difusos, que varían según la atención.

6. Sociotemporalidad: es el tiempo colectivo que regula el ritmo social. Lo organizan los relojes y los calendarios.

El esquema de Fraser, aunque simple y parcial, sirve para entender que el problema del tiempo no puede abarcarse por completo desde una sola perspectiva.

Al margen de las especulaciones conceptuales, el tiempo es un parámetro acústico importante pues casi todos los sonidos fluctúan durante su desarrollo. Más aún, para sonidos que contienen información, como la palabra y la música, ésta aparece en los cambios

antes que en las partes estables. En estos casos es conveniente distinguir algunas escalas temporales diferentes: 1) la microescala, que va desde algunos milisegundos a varios cientos de milisegundos y que porta la estructura fina del sonido; 2) una escala que contempla los cambios más lentos, caracterizados por las envolventes temporales de los parámetros acústicos, que va desde unas décimas de segundo hasta algunos segundos; 3) otra que contenga las estructuras compuestas que toman varios segundos en desarrollarse, como palabras y motivos musicales, y 4) la macroescala temporal, que puede abarcar minutos enteros y en la que se despliegan oraciones y frases musicales. Esta última escala depende directamente de la memoria de corto término. En los grandes desarrollos musicales, como el de la forma sonata, también juega un rol decisivo la memoria de largo término.

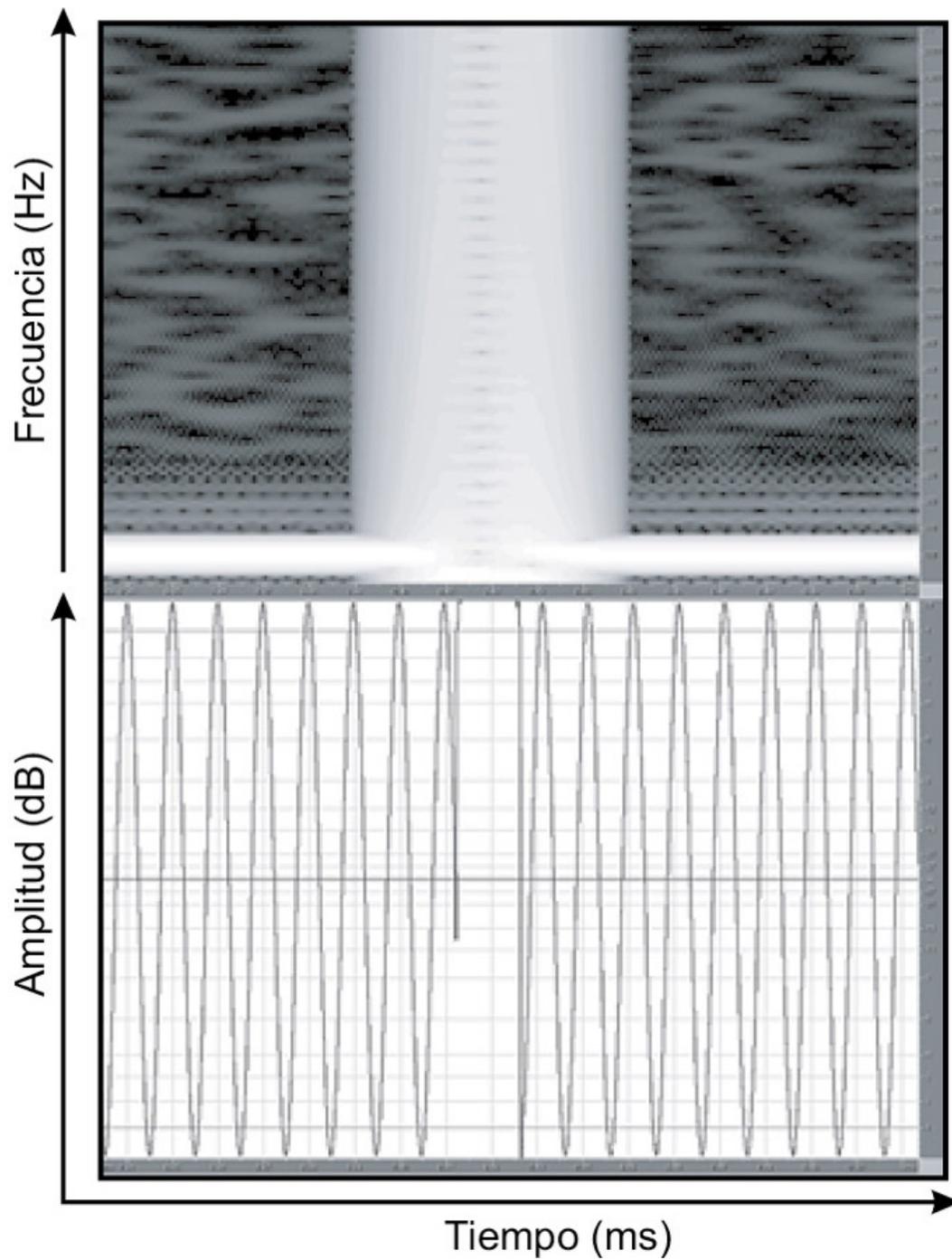
## 1. Resolución e integración temporal

Muchas veces se confunden dos nociones relacionadas con el procesamiento temporal de señales. Una de ellas, la *resolución temporal*, es la habilidad para detectar cambios en el estímulo a medida que transcurre el tiempo. La otra, la *integración temporal*, es la capacidad de agregar información a medida que transcurre el tiempo y permite aumentar la capacidad de discriminación ante un estímulo de larga duración. La mayoría de los estudios psicoacústicos se orienta hacia temáticas relacionadas con la resolución temporal. Sin embargo, la integración temporal juega un rol destacado en ciertos géneros y estilos musicales que emplean la aliteración de sonidos en los que algún parámetro deriva lentamente.<sup>[1]</sup> La mayoría de los autores concuerdan en que la resolución temporal depende principalmente de dos procesos afines. Por un lado se realiza un análisis del patrón temporal dentro de cada canal del filtro auditivo y, por el otro, se comparan los patrones temporales correspondientes a canales diferentes.

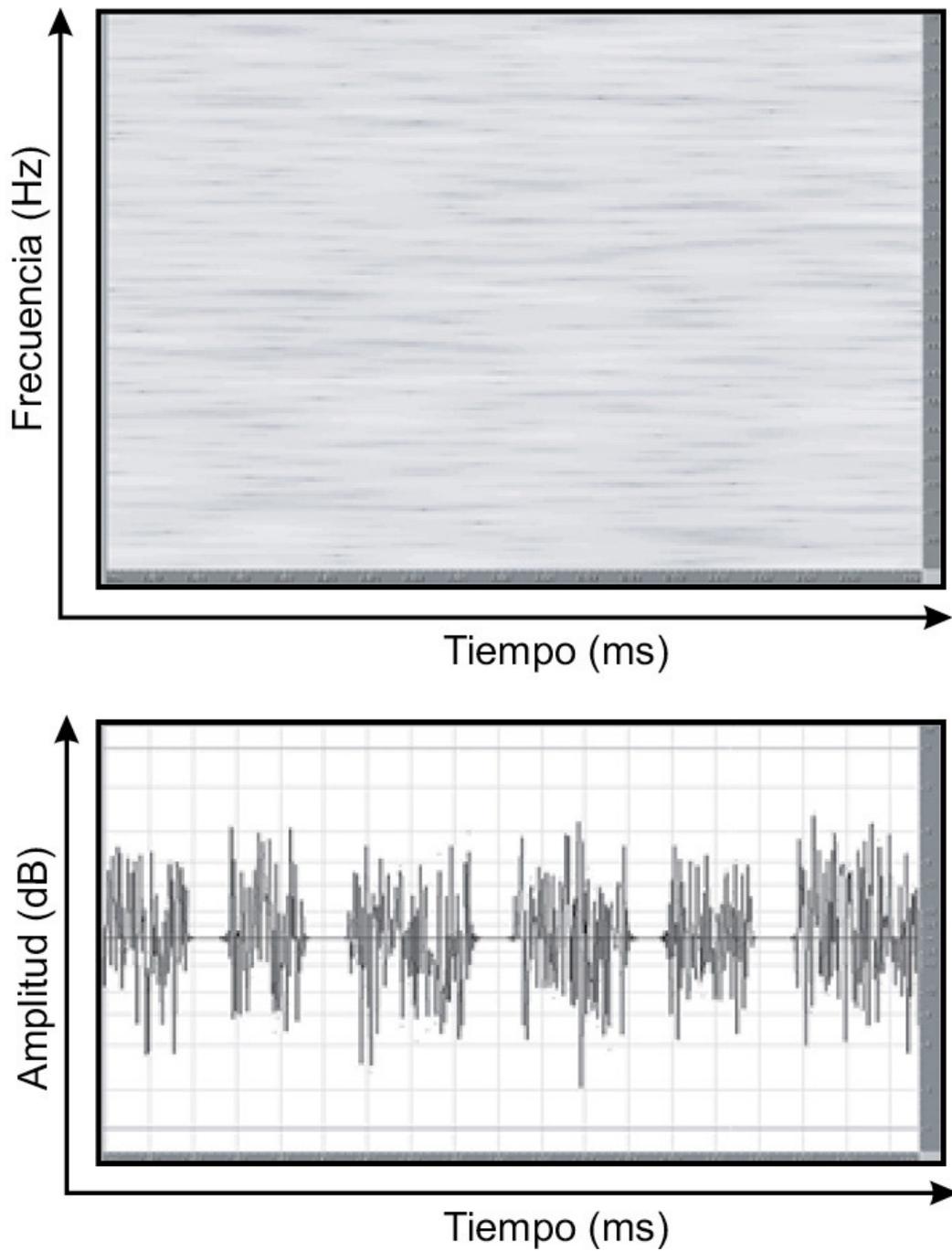
La mayor dificultad al intentar medir la resolución temporal es consecuencia de la inevitable relación entre tiempo y frecuencia propia de todo fenómeno ondulatorio, que depende directamente del principio de indeterminación acústico. Los cambios en el patrón temporal de una señal están acompañados generalmente de una modificación en su espectro. Por ejemplo, al interrumpir brevemente una senoide aparece un espectro de banda ancha que se oye como un clic muy sonoro. En la figura 1 se aprecian los gráficos temporales y sonogramas correspondientes.

Puede ocurrir que una aparente detección temporal sea en realidad una detección espectral –por ejemplo, el espectro de potencia de un par de clics muy cercanos en el tiempo es diferente del espectro de potencia de uno solo de ellos. Para evitar esta clase de confusiones, en las pruebas experimentales se adoptan dos estrategias diferentes. La primera consiste en emplear señales de espectro invariante ante cambios en su patrón temporal. El ruido blanco, por ejemplo, conserva el espectro plano aunque se lo seccione repentinamente, tal como se aprecia en la figura 2.

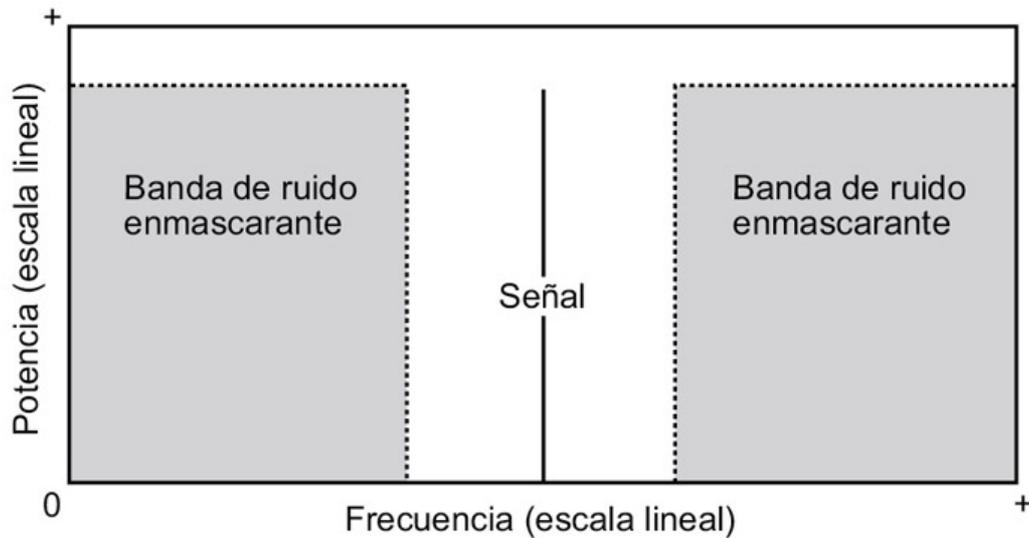
La segunda estrategia emplea estímulos cuyos espectros se modifican al alterarse los patrones temporales –como ocurre con las señales sinusoidales. Para evitar la confusión entre detección temporal y espectral, a la señal de base se le agrega ruido de fondo a fin de enmascarar las bandas de frecuencia que surgen en cada transición. El ruido de Patterson, ya visto en el punto 2 del capítulo III, es uno de los más eficaces ya que no altera a la señal en su banda central (Patterson, R. D., 1976). En la figura 3 se puede apreciar la técnica de Patterson para enmascarar las bandas laterales que aparecen en las transiciones en una señal sinusoidal.



**Figura 1.** Gráfico temporal y espectral de una discontinuidad en una senoide



**Figura 2.** Gráfico temporal y espectral de una discontinuidad en una señal de ruido blanco



**Figura 3.** Técnica de Patterson aplicada a una sinusoide

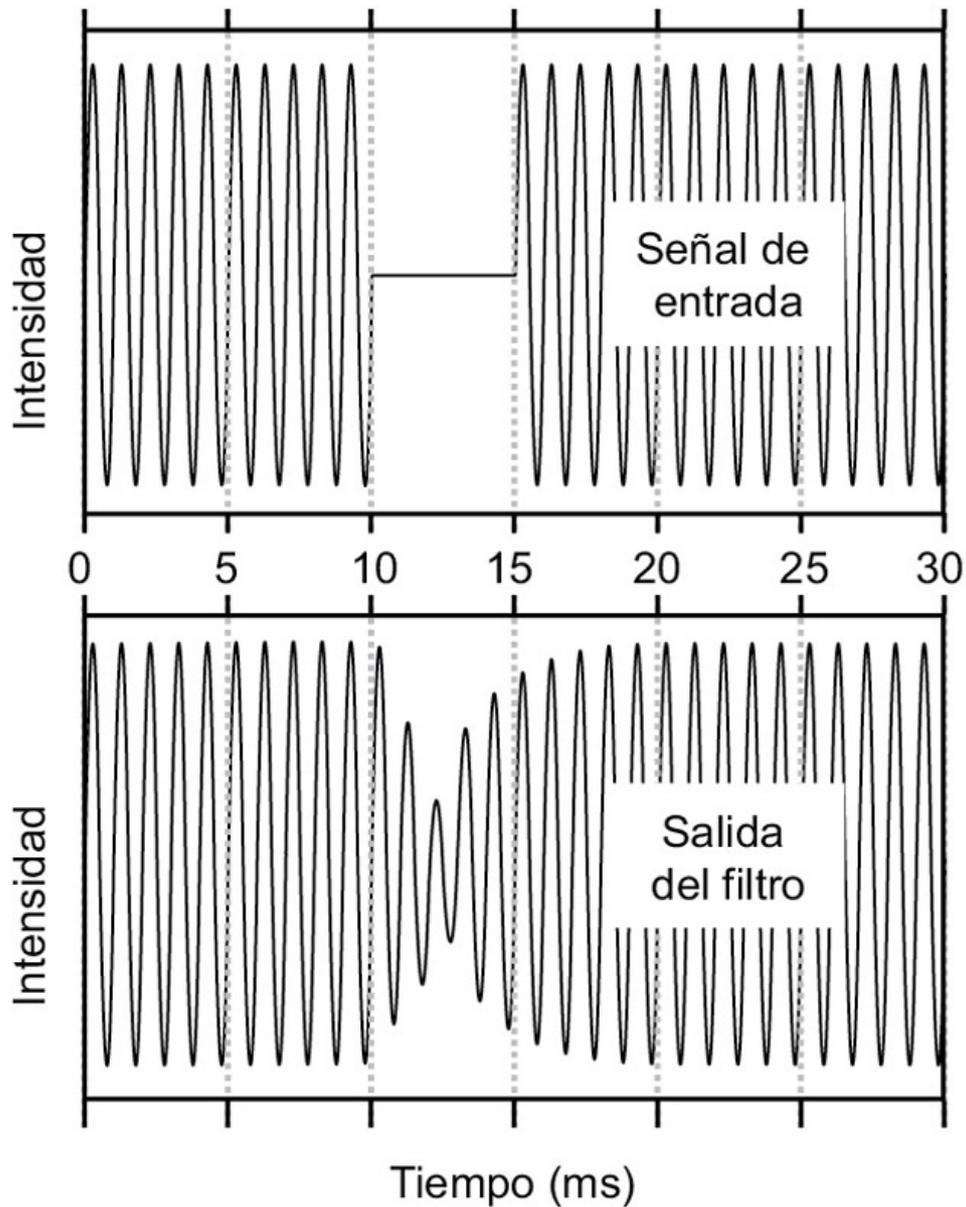
### Efectos de la frecuencia central en la resolución temporal

Se ha sugerido que la resolución temporal debería ser mejor a alta que a baja frecuencia. Una discontinuidad en una sinusoide causa una cola de resonancia (“ringing” en inglés) en cada filtro auditivo, que se mantiene durante más tiempo a medida que decrece el ancho de banda del filtro auditivo correspondiente –por aplicación del principio de indeterminación acústico. En la figura 4 se puede observar la modificación temporal ejercida por el filtro auditivo centrado en 1.000 Hz sobre una señal de entrada de 1.000 Hz.

Como las bandas más angostas se encuentran a baja frecuencia, es lógico suponer que en dicha región del espectro la resolución temporal resulte inferior que en la zona de alta frecuencia. Datos experimentales sugieren que, por encima de los 2.000 Hz, el tiempo de respuesta de los filtros auditivos es muy pequeño y no afecta significativamente la resolución temporal. Por debajo de esa frecuencia la resolución disminuye apreciablemente.

## 2. Resolución temporal en señales de banda ancha

En general, los experimentos en los que se intenta determinar el umbral de detección de discontinuidades (*gaps*) se basan en señales de ruido de banda ancha. El método más utilizado es el de elección forzada entre dos alternativas (2AFC), en el que el sujeto debe decidir cuál de los dos pulsos sucesivos de ruido contiene el *gap*. El umbral de discontinuidad medido en esta clase de experimentos, por ejemplo el realizado por Plomp, es de 2 a 3 ms (Plomp, R., 1976).



**Figura 4.** Respuesta simulada de un filtro auditivo de  $FC = 1.000$  Hz ante una senoide seccionada de  $1.000$  Hz

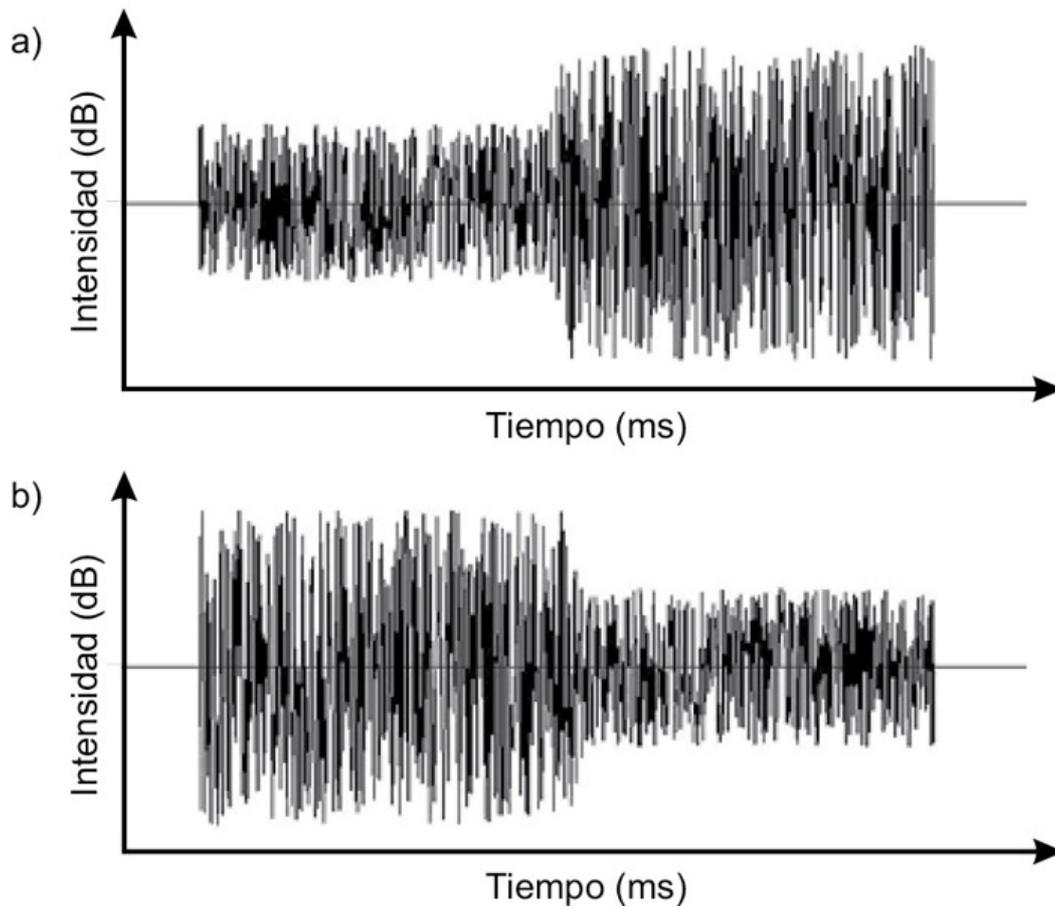
Otro grupo de experimentos emplea señales de banda ancha invertidas en el tiempo. Sabemos que el espectro de potencia de largo tiempo de una señal de corta duración no cambia si ésta es invertida temporalmente. Por lo tanto, si la segunda señal es muy breve y se la distingue auditivamente de la original debe ser por la diferencia en su patrón temporal. El umbral de detección entre dos se-

ñales de banda ancha que cambian de amplitud a la mitad de su desarrollo (de  $A$  a  $2A$ ), pasadas al “derecho” e invertidas, es de 2 a 3 ms, valor que coincide con el hallado por Plomp. Los sujetos no se dan cuenta del cambio de amplitud, pero perciben un cambio tímbrico (oyen “tick” y “tock”). En el gráfico de la figura 5 se pueden ver las señales utilizadas en el experimento citado.

Algunos autores hallaron valores significativamente menores para el umbral de detección, entre 0,25 y 1,8 ms, pero no se conoce la causa de la discrepancia entre ambos grupos de mediciones.

### **3. Resolución temporal en señales de banda angosta**

Para señales de banda angosta, la introducción de una discontinuidad implica “desparramar” el espectro por efecto del principio de indeterminación acústico. De todos modos, la situación es diferente si se trata de señales de banda angosta moderada o con espectro de línea, como es el caso de las sinusoides.



**Figura 5.** Señales de banda ancha empleadas para determinar el umbral de detección de discontinuidades

A continuación se analizan ambas posibilidades.

### **Detección de discontinuidades en ruido de banda angosta**

Para señales de banda angosta moderada el parámetro de comparación es el ancho de banda del filtro auditivo que le corresponde. Tal como hemos visto, la detección de discontinuidades debería ser peor a baja frecuencia porque en esa región espectral el ancho de banda de los filtros auditivos es menor. Los umbrales de detección deberían disminuir a medida que aumenta la frecuencia y por lo tanto el ancho de banda del ruido, hasta llegar al valor de la banda crítica correspondiente. A partir de este punto, un aumento en

el ancho de banda no debería mejorar la detección de la discontinuidad. Sin embargo, los datos experimentales contradicen esta afirmación: cuando se introduce una discontinuidad su gran desarrollo espectral afecta sincrónicamente a una cantidad de filtros auditivos —cosa que no ocurre con las variaciones de amplitud producto de la aleatoriedad propia del ruido. La excitación en paralelo de varios filtros auditivos permite que el umbral de detección resulte menor que el que se obtendría si sólo un filtro fuese puesto en juego. Las mediciones experimentales confirman esta hipótesis, pues para ruido de banda angosta, cuya frecuencia central varía entre 30 y 1.000 Hz, los umbrales de detección de discontinuidades disminuyen desde un valor de 40 ms hasta caer por debajo de 10 ms. Este efecto es de gran importancia para explicar la *paradoja acústica temporal* que analizaremos más adelante.

## **Detección de discontinuidades en sinusoides**

Para evitar la confusión entre detección temporal y espectral, las sinusoides se presentan junto a un ruido continuo con una muesca espectral sintonizada a la frecuencia de la senoide (figura 3). De este modo se consigue enmascarar la dispersión del espectro ocasionada en la discontinuidad.

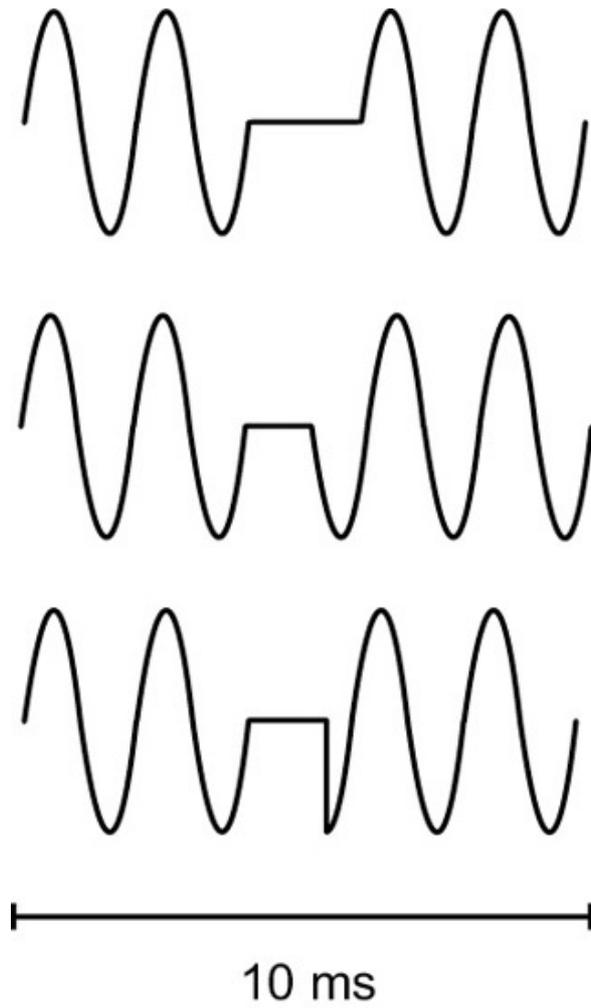
En un célebre experimento, Shailer y Moore estudiaron la detección de discontinuidades en sinusoides enfocando la atención en la fase de reinicio de la señal (Shailer y Moore, 1987). Explicaron sus resultados, que contradecían a la literatura sobre el tema, sobre la base de la cola de resonancia (“ringing”) del filtro auditivo: cuando la senoide cesa, el filtro auditivo sintonizado a su frecuencia continúa respondiendo por un tiempo. Si el reinicio de la señal coincide en fase con esta respuesta del filtro, la detección de la discontinuidad es más difícil. En las figuras 6 y 7 se muestran las señales empleadas y las respuestas del modelo ante discontinuidades de diferente duración.

En los casos en los que la señal “engancha” en fase con la respuesta del filtro auditivo, Shailer y Moore midieron umbrales de 4,5 ms, considerablemente mayores a los 2 o 3 ms esperados.

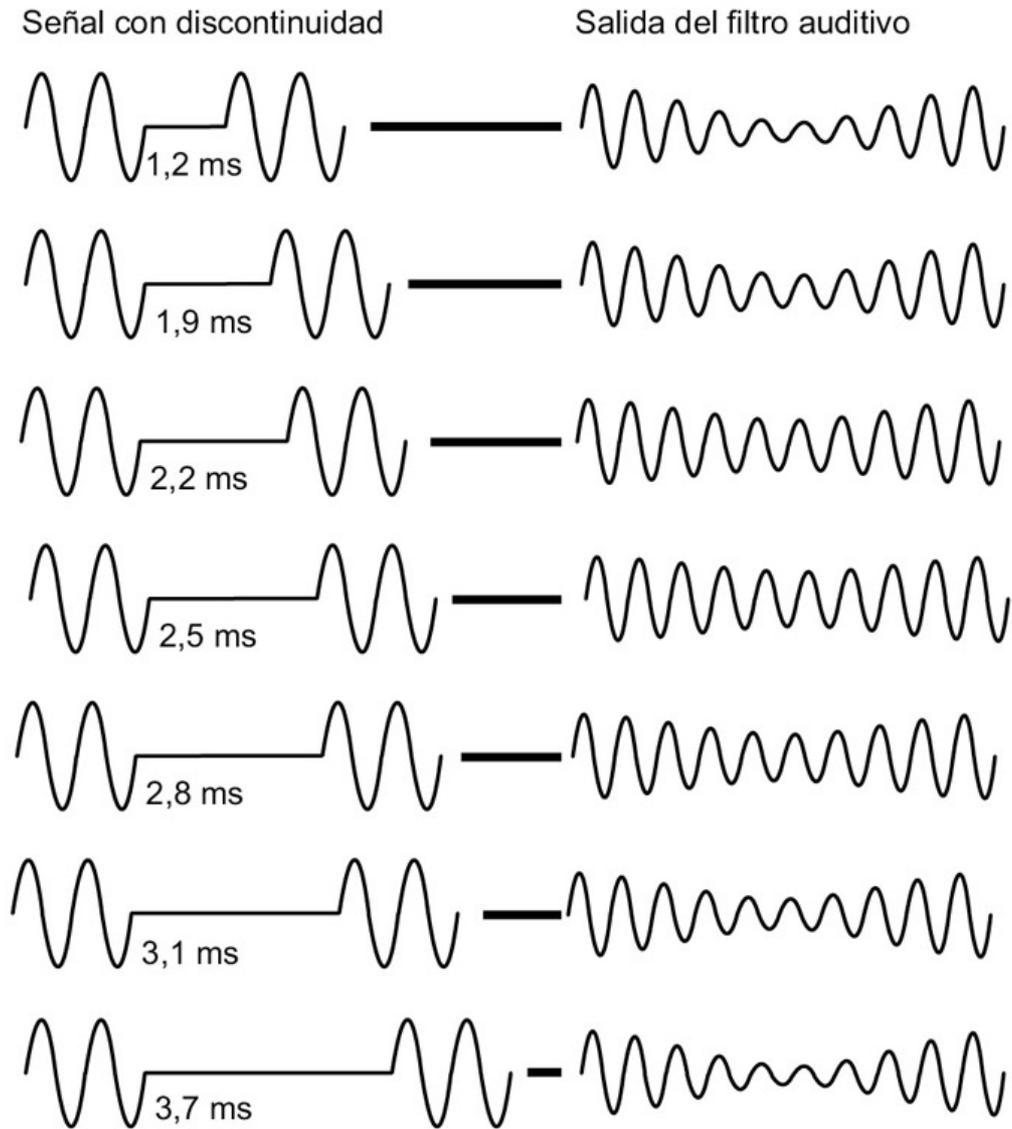
## 4. Modelos de resolución temporal

La respuesta del filtro auditivo a nivel coclear es muy rápida para altas frecuencias y, por lo tanto, se supone que debe existir un limitador a otro nivel de procesamiento neural. En general, se asume que los cambios rápidos son suavizados en alguna de las etapas superiores del proceso. Casi todos los modelos elaborados para dar cuenta de la resolución temporal constan de cuatro etapas, similares a las que se aprecian en la figura 8.

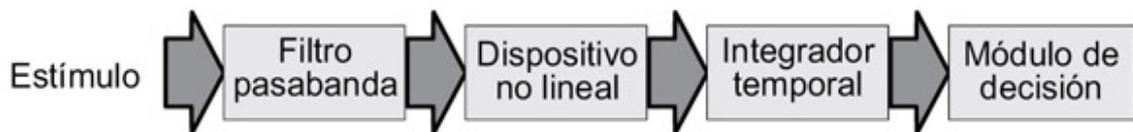
- En primer lugar aparece un filtro pasabanda que corresponde a cada uno de los filtros auditivos conectados en paralelo.
- A continuación, la señal es procesada por algún dispositivo no lineal.
- Luego pasa por un integrador temporal, que puede caracterizarse como un filtro pasabajos.
- Por último, el producto de los primeros tres bloques entra a un módulo de decisión que determina qué es lo que se va a percibir ante la discontinuidad.



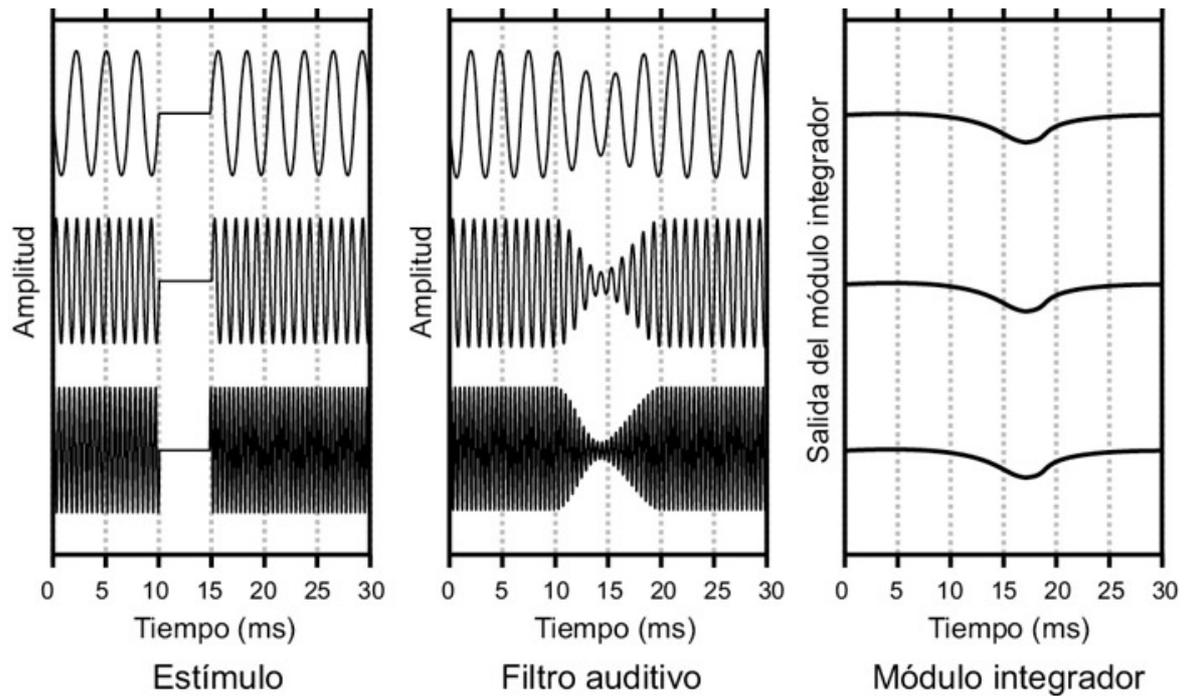
**Figura 6.** Tres condiciones de fase diferentes usadas en el experimento de Shailer y Moore



**Figura 7.** Salida de un filtro auditivo simulado con  $FC = 400$  Hz en respuesta a una senoide, con discontinuidades de 1,2 a 3,7 ms en la condición de fase estándar (experimento de Shailer y Moore)



**Figura 8.** Diagrama en bloques de un modelo de resolución temporal



**Figura 9.** Operación del módulo de decisión en el modelo de Shailer y Moore para estímulos de diferente frecuencia

Estos bloques corresponden a diferentes procesos que transforman la señal del estímulo. De todos ellos, el módulo de decisión es quizá el más complejo y su configuración varía según la tarea que se le asigne. En uno de los modelos más desarrollados, el de Shailer y Moore, la señal es procesada por un módulo de decisión con las características que se aprecian en la figura 9. En este modelo el integrador temporal es una ventana temporal no rectangular.

Se observa claramente que la detección de la discontinuidad resulta independiente de la frecuencia central del filtro auditivo involucrado.

### **Hipótesis del banco de filtros de modulación**

Algunos investigadores han sugerido que el análisis de los sonidos que son modulados en amplitud depende de una parte especializada del cerebro que contiene grupos de neuronas afinadas cada una a una tasa de modulación dada. Cada neurona puede ser conside-

rada como un filtro en el dominio de la modulación y el conjunto de neuronas se conoce como *banco de filtros de modulación*. Este modelo es todavía controversial, pero podría explicar la eficacia de ciertos parámetros de predicción de la inteligibilidad tales como el STI o el RASTI, que trataremos en el capítulo IX.

## 5. Otros procesos temporales

### Discriminación en secuencias de Huffman

Green estudió la discriminación temporal en señales que tienen el mismo espectro de largo tiempo –a gran escala–, pero distintos espectros de corto tiempo –a pequeña escala– (Green, D. M., 1973). Son sonidos breves de banda ancha –parecidos a clics– excepto que la energía de ciertas regiones es retrasada con relación a la de otras regiones. La frecuencia central, la cantidad del retardo y el ancho de banda de la región con retardo pueden variarse independientemente. Estas señales se denominan *sonidos de Huffman*. El experimento de Green postulaba que si los sujetos notaban diferencia entre dos sonidos de Huffman diferentes, quedaba demostrada su sensibilidad al patrón temporal. El resultado confirmó la hipótesis: el umbral de detección del retardo de la región bajo estudio fue de 2 ms, resultado independiente de la frecuencia central de las señales utilizadas. Los sujetos percibieron las diferencias en el patrón temporal como cambios en la calidad del sonido.

### Detección de asincronías en el ataque y la extinción de señales complejas

La asincronía de componentes de una señal compleja es detectada con mayor facilidad en el ataque que en la extinción de la señal. Para señales armónicas se perciben diferencias en el ataque de los armónicos de menos de 1 ms (con umbrales algo mayores para componentes de alta frecuencia), mientras que el umbral en la ex-

tinción va desde valores comprendidos entre 3 y 10 ms –cuando el componente asincrónico cesa después que el resto de la señal compleja– hasta valores entre 10 y 30 ms cuando éste cesa antes que la señal de base. El resultado de estos experimentos es consistente con la importancia perceptual del ataque de un sonido. También está estrechamente relacionado con la adaptación auditiva y con la percepción del timbre instrumental.

Los umbrales de detección de asincronías para señales no armónicas son de 2 a 50 veces mayores que para señales armónicas. Este hecho es explicable en términos de agrupamiento perceptual: es más difícil establecer la sincronía de ataques de señales que parecen provenir de fuentes diferentes.

## **Discriminación de la duración**

Varios estudios revelan que, para valores de duración del estímulo  $T$  superiores a los 10 ms, el umbral diferencial de duración  $\Delta T$  se incrementa con  $T$  y es aproximadamente independiente de las características espectrales de la señal. Esta regla se cumple tanto para la discriminación de la duración en sonidos como en silencios entre sonidos. Valores típicos de  $\Delta T$  son de 1,7 ms para  $T = 25$  ms y de 15 ms para  $T = 320$  ms (Divenyi y Danner, 1977). La fracción de Weber,  $\Delta T/T$ , decrece levemente al aumentar  $T$ . Como con casi todos los parámetros que cuantifican la discriminación auditiva, se encontró que  $\Delta T$  aumenta a intensidades próximas al umbral de audibilidad.

Estos valores se correlacionan con los de integración temporal de la sonoridad, en particular con el *postulado de Henning* que establece que, al evaluar la sonoridad en estímulos de corta duración, prevalece la cantidad total de energía en lugar de la intensidad de la señal, tema que tratamos en el capítulo II, punto 4 (Henning, G. B., 1970).

## 6. Juicios sobre el orden temporal

La habilidad para juzgar el orden temporal en una secuencia de sonidos depende en gran medida de la tarea asignada al sujeto: una cosa es la *identificación* del orden que ocupa cada sonido dentro de la secuencia y otra la *discriminación* entre secuencias con distinto orden, sin que se pueda precisar la ubicación de cada sonido en el conjunto.

En el caso de la identificación del orden temporal, Hirsch encontró que sujetos entrenados necesitaban al menos 20 ms por sonido para determinar el orden en una secuencia (Hirsch, I. J., 1959). En el caso de la discriminación entre secuencias la resolución es más fina, y se necesitan entre 2 y 7 ms por sonido. Estos valores son de gran importancia para comprender el desplazamiento acústico temporal, que presentamos a continuación.

### El desplazamiento acústico temporal

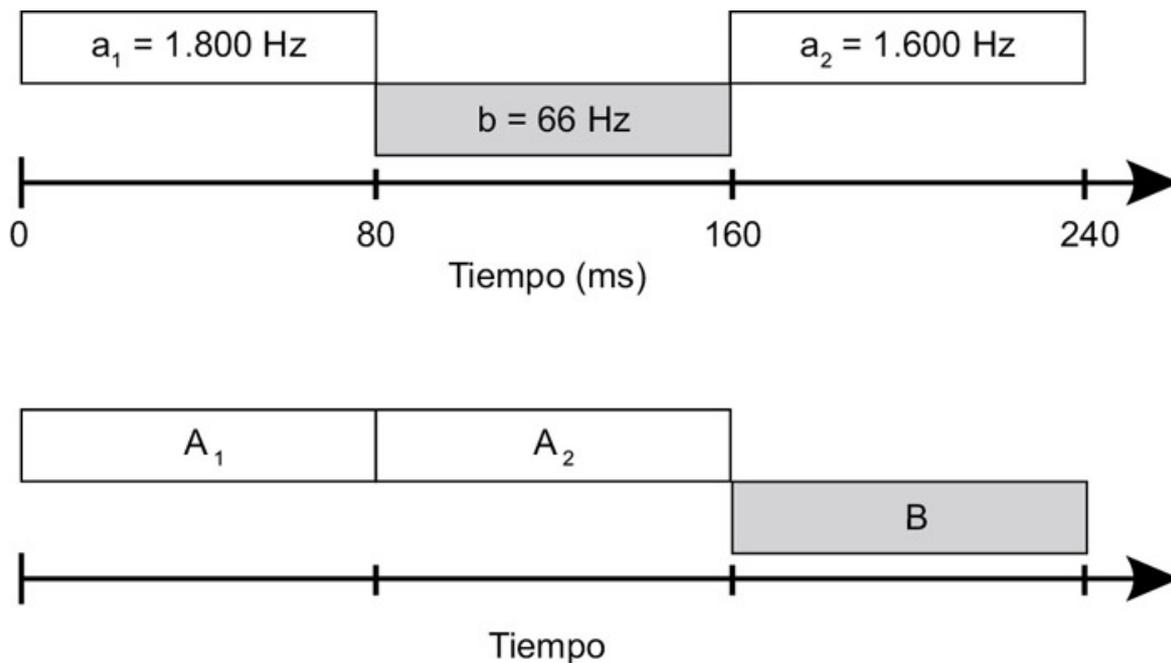
Las pruebas experimentales referidas al desplazamiento entre el tiempo físico y el fenoménico fueron iniciadas por Whilhem Wundt y continuadas luego por Edgard Rubien, quienes lograron transformar un fenómeno de simultaneidad física en una sucesión psicológica. Varias hipótesis fueron propuestas por entonces para dar cuenta del fenómeno, y entre ellas prevaleció la que suponía que la heterogeneidad de los estímulos constituye el factor responsable de los desplazamientos temporales.

En la década de 1970 el psicólogo italiano Giovanni Bruno Vicario desarrolló una serie de trabajos sobre el fenómeno del desplazamiento temporal en la audición, en los que plantea la imposibilidad de establecer una correspondencia biunívoca entre el tiempo físico y el psicológico. Una de las evidencias que esgrime para defender dicha hipótesis es la del desplazamiento temporal de estímulos acústicos extremos: tomando la secuencia [a1 - b - a2] en el *tiempo físico*  $t$ , si el estímulo central  $b$  tiene una frecuencia que di-

fiere de forma notable –por ejemplo cuatro octavas– de los estímulos laterales y escogemos estímulos muy breves –del orden de los 80 milisegundos–, lo que se oye en el *tiempo fenoménico* T es una sucesión distinta [A1 - A2 - B]. Los sonidos más agudos se oyen seguidos y el sonido intermedio, más grave, se ve relegado al final de la sucesión (Vicario, G. B., 1997). En la figura 10 se puede ver un esquema del experimento. La diferencia entre las frecuencias de a1 y a2 corresponde a un intervalo de segunda mayor descendente.

Vicario comprobó que en el caso de que los estímulos a1 y a2 sean notablemente más extensos que b el fenómeno se hace más evidente, incluso desaparece la noción de “localización temporal” de b, que parece flotar sobre los dos sonidos agudos. Llega a la conclusión de que las secuencias de estímulos no homogéneos son más propicias para la producción de desplazamientos temporales.

El desplazamiento temporal no se puede atribuir a ningún proceso fisiológico y no parece sencillo explicar de qué modo relaciones específicas no temporales –como semejanza o disimilaridad– puedan interaccionar con relaciones temporales tales como *antes/después*. La explicación más difundida supone que la posición de los sonidos en el tiempo fenoménico se decidiría luego de haber oído todos los estímulos. La operación, preconsciente, tendría lugar en el lapso de aproximadamente 200 ms que va desde la llegada de las señales al tímpano y la percepción consciente de las mismas (Fraisse, P., 1967).



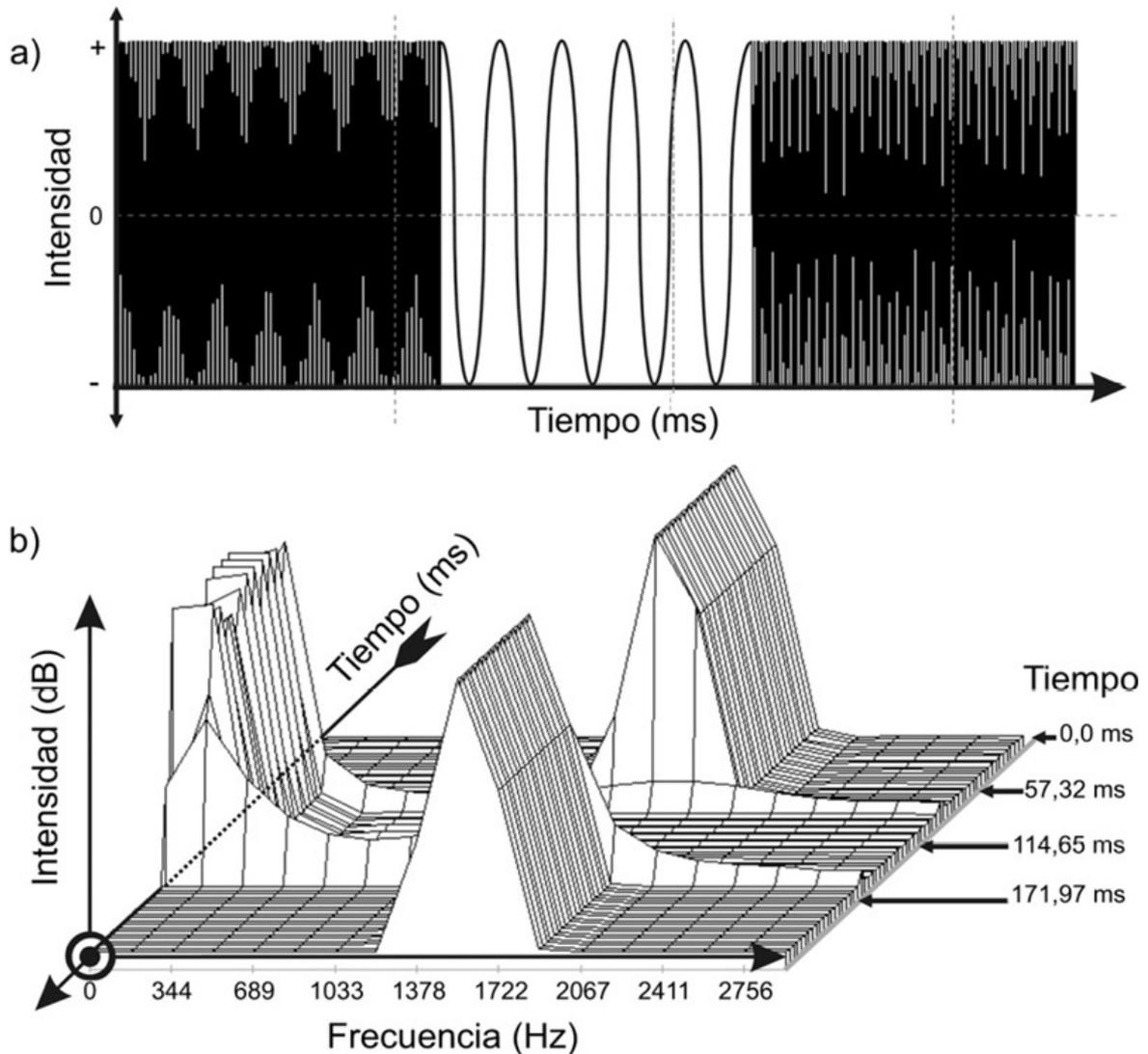
**Figura 10.** Esquema del experimento de Vicario

### *Límites de validez del fenómeno*

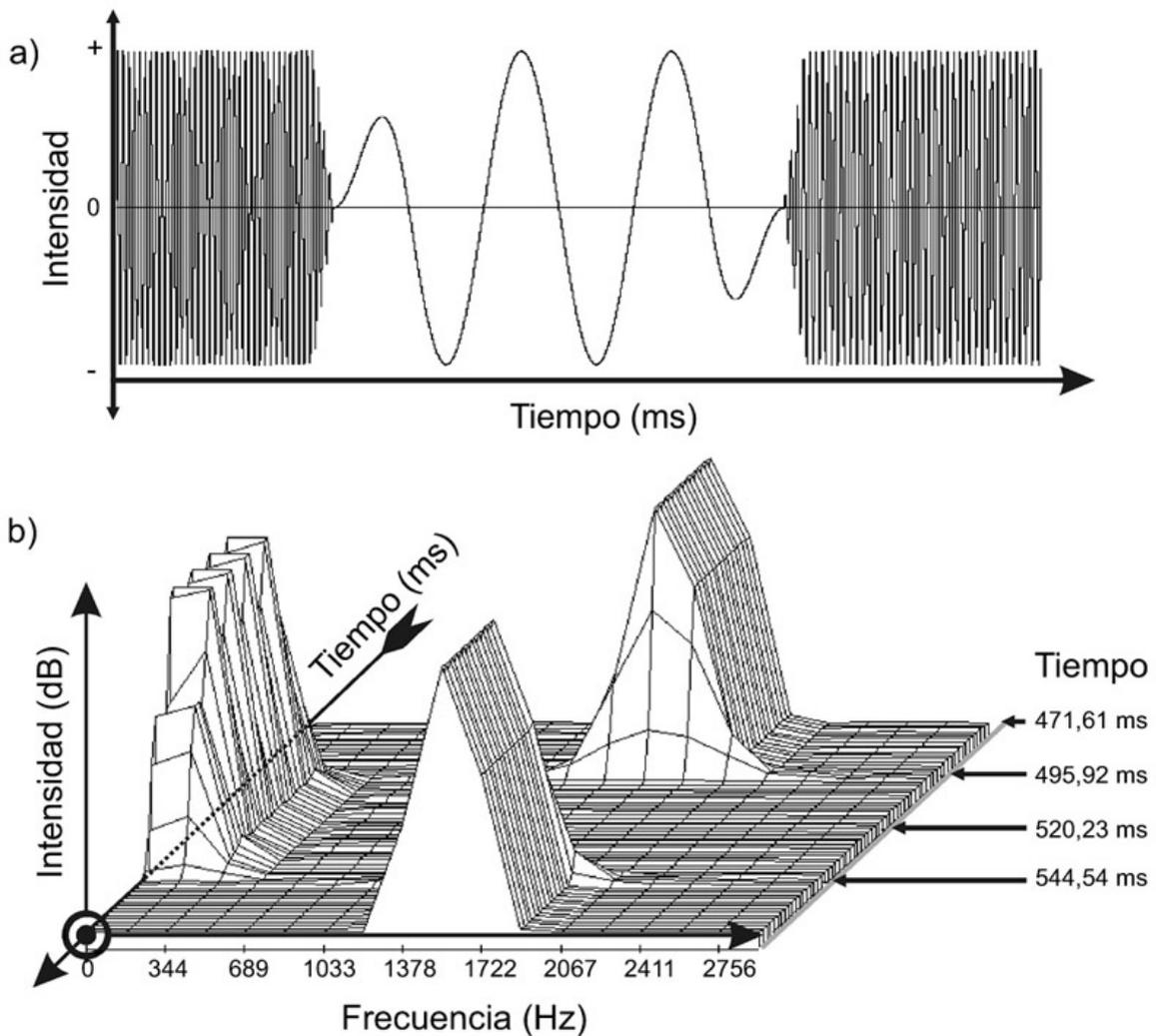
Una condición necesaria para la ocurrencia del desplazamiento temporal es que las distintas señales no deben conectarse espectralmente en las transiciones. De hecho, la reproducción literal del experimento de Vicario no da el resultado esperado. El motivo hay que buscarlo en el perfil dinámico de las señales: si las transiciones son bruscas (figura 11a) aparece un gran ancho de banda que destruye la separación espectral estipulada en la hipótesis original. La aparente anomalía se debe simplemente a que no puede ser quebrantado el límite impuesto por el principio de indeterminación acústico. El efecto se aprecia claramente en el diagrama tridimensional de la figura 11b.

Si se suavizan las transiciones entre las señales, por ejemplo con ventanas temporales gaussianas, se eliminan los puentes espectrales que conectan la región aguda con la grave y el fenómeno de desplazamiento temporal se percibe con claridad. En la figura 12 se observa la señal de prueba modificada y el efecto espectral deri-

vado. Se ve claramente la mayor independencia (separación espectral) entre las señales a1 y a2 con respecto a la señal b, que corresponde a la condición de no homogeneidad de los estímulos postulada en los trabajos iniciales (Basso, Guillén y Liut, 2000).



**Figura 11.** Gráfico temporal y tiempo-energía-frecuencia (TEF) de la señales de la secuencia original

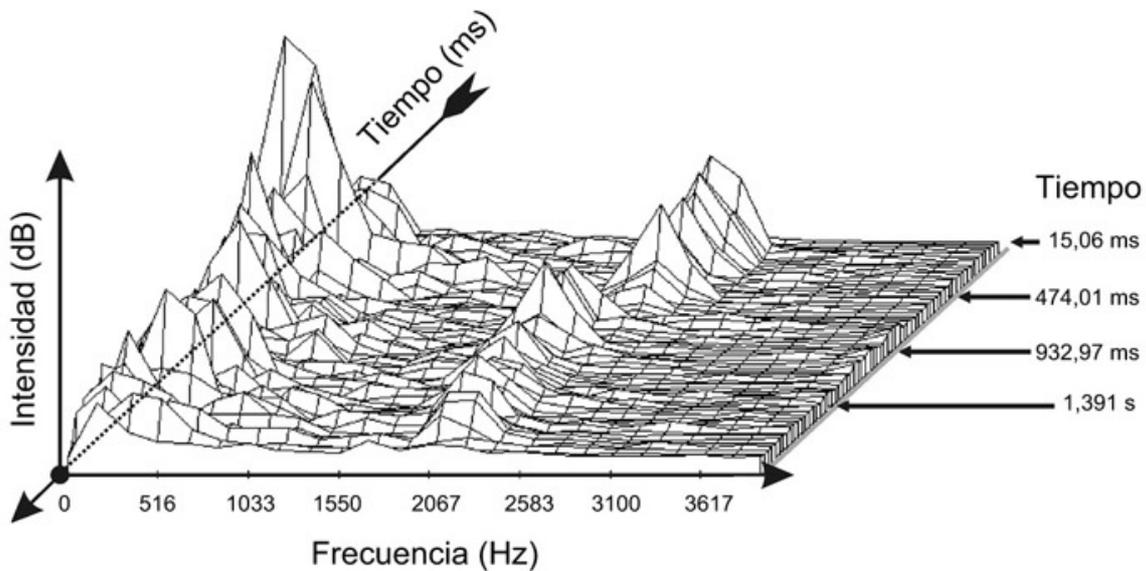


**Figura 12.** a) Detalle de la secuencia temporal con transiciones gaussianas largas. b) Gráfico TEF de la misma secuencia

### *Aplicación musical*

Estas experiencias controladas –como reconocen casi todos los investigadores que participan de ellas–, por su simplicidad distan bastante de lo que sucede en la vida cotidiana. Es posible, sin embargo, encontrar en muchas obras de música el uso –intencional o no– de alguna de estas peculiaridades de la percepción auditiva. Por tratarse de situaciones musicales reales (esto es, dentro del contexto de una “puesta en obra”) no tienen la intensidad y la claridad que se puede lograr en el marco ascético de un laboratorio.

De acuerdo con sus objetivos y con los presupuestos estilísticos involucrados, un compositor puede decidir el modo de aplicación del desplazamiento acústico temporal. Si su intención es independizar y desvincular diferentes elementos de un conjunto, la estrategia planteada permite desdoblarse la estructura en dos tramas temporales separadas y autónomas. Un ejemplo que ilustra esta técnica se encuentra en la pieza para orquesta reducida *Homenaje a García Lorca* (1936) de Silvestre Revueltas, en la que la misma melodía se interpreta en paralelo con una tuba y una flauta *piccolo* a distancias de cuatro y cinco octavas. Aunque el paralelismo absoluto de alturas y duraciones debería ayudar a fundir perceptualmente los instrumentos, la enorme separación de los registros y la ausencia de puentes espectrales rompen la relación temporal entre ellos. Al oír el pasaje se percibe una ambigua sensación de “desajuste”, provocado por el desfase entre tiempo físico y psicológico propio del fenómeno bajo análisis. En la figura 13 se puede ver un gráfico TEF de dicho fragmento.



**Figura 13.** Gráfico TEF del fragmento citado de *Homenaje a García Lorca* de Silvestre Revueltas

Por el contrario, si el propósito del autor se dirige a la fusión de estructuras, el fenómeno de desplazamiento temporal debe ser evitado. Además de reducir la distancia espectral que separa a los estímulos, es necesario agregar datos espectrales que sincronicen las secuencias en las zonas de transición. Ciertas articulaciones y ataques destruyen la paradoja temporal, sincronizando el tiempo físico con el psicológico. Aquí una cuidadosa elección de los medios instrumentales, que incluye tanto los instrumentos mismos como las demás variables involucradas –registro, matiz, articulación, etc.– resulta decisiva.

## Notas

- [1](#) Hay numerosos ejemplos en la música del sudeste asiático. En occidente, la integración temporal constituye un proceso central en el minimalismo.

## Capítulo V

# Percepción de la altura tonal

La *altura tonal* se define técnicamente como “aquel atributo de la sensación auditiva en términos del cual los sonidos pueden ser ordenados en una escala musical” (American Standards Association, 1960).[\[1\]](#) En otras palabras, una secuencia de sonidos que varían de altura tonal tiene el carácter de una melodía. Un ejemplo quizá ayude a delimitar este concepto.

Si se altera la ecualización en un sistema de audio cambia el brillo del sonido y, en cierto sentido, algún aspecto relacionado con su altura. Pero hay otros elementos que no se modifican, pues el cambio en la ecualización no hace que el Do de una flauta se sustituya por un Re, ni que el fragmento module a otra tonalidad. Estas características –que permanecen invariantes frente a los cambios en la distribución de energía del espectro– definen la altura tonal presente en los sonidos llamados *tónicos* (Schaeffer, P., 1966). La tonicidad de un sonido es el rasgo perceptual que permite que se lo identifique con uno de los grados de la escala musical. Lo que cambia al modificarse la posición de los cursores de un ecualizador es la distribución de energía en el espectro de la señal. Esta distribución de energía se asocia con la percepción de la *altura espectral* del sonido: dos guitarras que emiten la misma nota, una de sonido metálico y la otra de sonoridad oscura, producen la misma altura tonal pero distintas alturas espectrales (es habitual decir que la primera es “más aguda” que la segunda).

El grado de tonicidad varía de un sonido a otro: una flauta genera sonidos muy tónicos, una marimba sonidos de tonicidad intermedia y un tambor genera señales de tonicidad nula. En una primera aproximación podemos establecer que el grado de tonicidad guarda relación con el ancho de banda de la señal física: a mayor ancho de banda menor tonicidad, y viceversa. El principio de

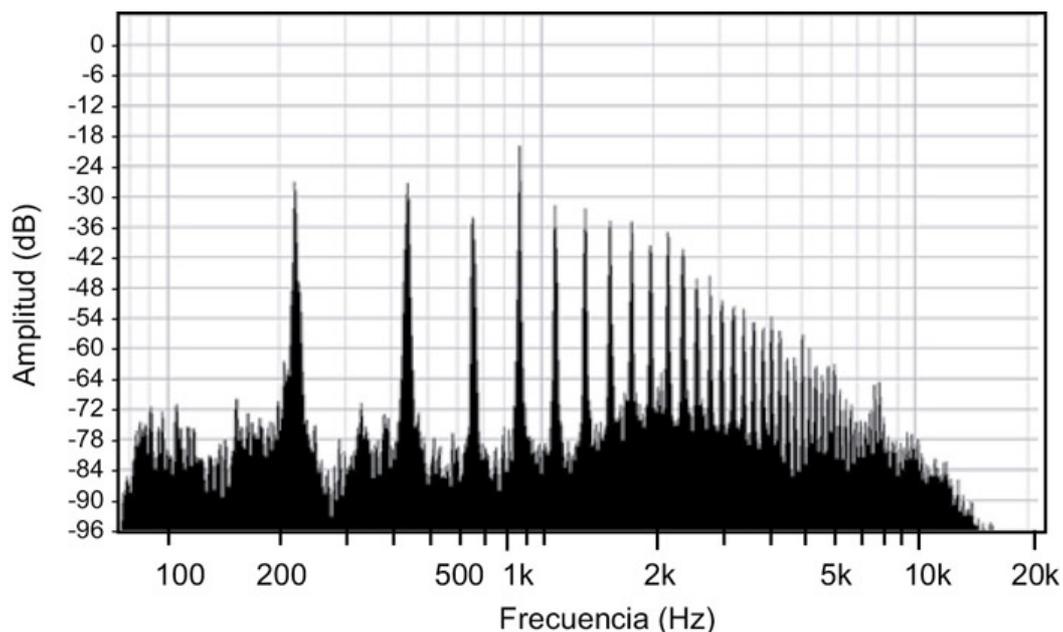
indeterminación acústico gobierna la relación frecuencia/duración en los casos anteriores. La altura tonal se considera completa o “fuerte” cuando permite la formación de acordes y funciones tonales, e incompleta o “débil” cuando no lo hace. Por ejemplo, con tres tom-toms de banda angosta se puede generar una secuencia melódica Do-Mi-Sol, pero si se los ejecuta simultáneamente no se forma el acorde de do mayor: los tom-toms poseen altura tonal incompleta (Basso, G., 2001a).

En general, las señales periódicas son percibidas como sonidos tónicos, a los que se puede asignar un valor de altura tonal definido que se correlaciona directamente con la fundamental de Fourier de la señal. Se presenta aquí una de las propiedades más notables del sentido de audición: percibimos con una altura tonal única los sonidos causados por estímulos periódicos, al margen de la complejidad armónica que posean. En la figura 1 se puede ver el espectro de un sonido de violoncello.

La cantidad de armónicos es aquí muy grande. Sin embargo, se oye claramente una sola altura tonal. En estos casos el sistema auditivo presta más atención a la salida de un hipotético centro de procesamiento de la altura (que señala una única sensación de altura tonal) que a las características de la altura espectral o primaria del sonido. Si se desea oír en detalle cada componente aislado es necesario inhibir el mecanismo anterior, operación que demanda un tiempo considerable y que explica la dificultad de percibir por separado los armónicos superiores en los sonidos de corta duración. Se denomina *audición analítica* a esta última forma de oír y *audición sintética* a la primera. [2]

Algunos autores, entre ellos Terhardt, sugieren que la capacidad para sintetizar gran cantidad de armónicos en una única altura tonal se aprende durante la primera infancia a partir de la reiterada exposición a sonidos periódicos, en especial la voz humana (Terhardt, E. *et al.*, 1982).

A lo largo de este capítulo nos encontraremos constantemente con dos teorías que tratan de explicar la percepción de la altura tonal. Una de ellas es la teoría del lugar (véase cap. I, apartado 4), que parte de dos postulados distintos. El primero establece que el oído efectúa un análisis espectral en el interior de la cóclea: diferentes frecuencias excitan diferentes lugares de la membrana basilar, en una especie de organización tonotópica. El segundo postulado afirma que la altura tonal se relaciona con el patrón espacial de excitación producido por el estímulo. La primera hipótesis es aceptada por la casi totalidad de los investigadores, mientras que la segunda todavía es materia de discusión. Una teoría alternativa a la teoría del lugar es la *teoría temporal*, que sugiere que la altura se relaciona con el patrón temporal de los impulsos neurales originados en el estímulo. Como los disparos de las células nerviosas tienden a ocurrir en una particular fase de la forma de onda, interviene necesariamente la sensibilidad a la fase del sistema auditivo. Como esta sensibilidad desaparece para frecuencias por encima de los 5 kHz, la teoría temporal no explica la percepción de la altura tonal a altas frecuencias.



**Figura 1.** Espectro de un sonido de violoncello

# 1. Percepción de la altura tonal en sinusoides

A comienzos del siglo XX el psicólogo Hermann Ebbinghaus postuló que se debe remover toda la información superflua en el estudio de los estímulos y sus respuestas psíquicas. A partir de esta regla, que mezcla principios tomados de las ciencias físicas y una mala lectura del teorema de Fourier, se impuso una tradición en la que se emplean casi exclusivamente señales sinusoidales para los experimentos de psicoacústica. En consecuencia, también se utilizaron estímulos sinusoidales para estudiar la percepción de la altura tonal. Sabemos, sin embargo, que las sinusoides son señales de prueba inadecuadas porque la altura tonal se asocia en gran medida a la periodicidad de los estímulos armónicos complejos, quizá por nuestra capacidad para agrupar los diferentes componentes espectrales presentes en la voz humana. La altura tonal originada en una senoide resulta imprecisa e inestable, como puede comprobar cualquier oyente ante una melodía realizada con tonos puros —habituales en las señales de espera de los contestadores telefónicos.

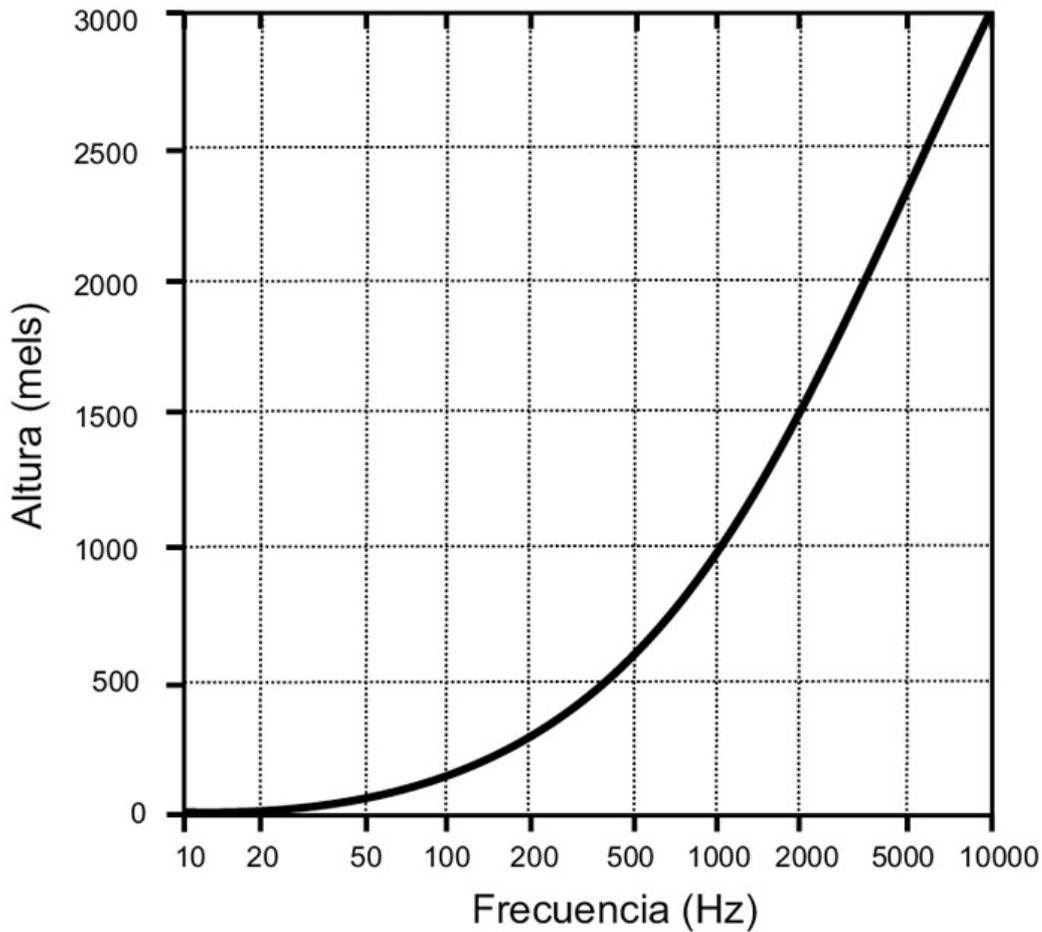
Además, cuando los estímulos son simples y poseen solamente una sola frecuencia destacada, la altura tonal se confunde en parte con la altura espectral. En cierto sentido, una senoide contiene su propio timbre, ya que al realizar un barrido ascendente en frecuencia del único armónico presente (la senoide misma) se excitan sucesivamente zonas adyacentes en la membrana basilar. El resultado perceptual es un glissando acompañado de un cambio continuo del color tonal. Un ejemplo tradicional de esta confusión entre altura tonal y espectral lo brinda la escala de altura, expresada en mels, que se muestra en la figura 2.

Los valores de altura en mels, elaborados a partir de estímulos sinusoidales, difieren notablemente de los obtenidos a partir de señales armónicas complejas. Para algunos autores, la escala en mels describe simplemente el brillo tímbrico de los sonidos simples y no

es aplicable a la altura tonal tal como se la entiende en la práctica musical. A principios de la década de 1970, Fred Attneave y Richard Olson postularon que para estudiar las escalas musicales se deben emplear las señales complejas que se usan en la práctica musical concreta (Attneave y Olson, 1971). Los experimentos que realizaron a partir de este postulado dieron resultados mejor correlacionados con un vínculo logarítmico entre la frecuencia del estímulo y altura tonal que la escala de mels, al menos dentro del rango de las frecuencias fundamentales que se emplean en música.

## **Discriminación en frecuencia**

Como ya señalamos, es necesario distinguir entre *selectividad en frecuencia*, que es la habilidad para distinguir componentes en un sonido complejo, y *discriminación en frecuencia*, la capacidad para detectar cambios de frecuencia a lo largo del tiempo. Para la teoría del lugar ambas habilidades están relacionadas, pues dependen del filtrado en frecuencia de la cóclea. Para las teorías temporales la relación no es tan marcada.



**Figura 2.** Relación entre altura (medida en mels) y frecuencia

El limen –mínima variación detectable en frecuencia– permite evaluar la discriminación en frecuencia y crece monótonamente con el valor de la frecuencia del estímulo sinusoidal. [3]. En las figuras 3 y 4 se pueden observar sus valores absolutos y relativos. El valor relativo del limen, expresado en relación con frecuencia central  $f_c$ , tiende a ser menor a frecuencias centrales y aumenta a medida que el estímulo se desplaza hacia los extremos del espectro audible. Por ejemplo, el limen absoluto es aproximadamente  $\Delta f = 2$  Hz para una frecuencia central de 1.000 Hz y corresponde a un limen relativo de 0,2%.

Es habitual realizar las experiencias de medición del limen a partir de dos métodos: la variación en frecuencia de sinusoides ais-

ladas y la variación en frecuencia de la señal moduladora en un dispositivo de frecuencia modulada (FMDL). [4] Los resultados de las experiencias de FMDL se utilizan, por ejemplo, en el cálculo de los índices de inteligibilidad de la palabra en sistemas de comunicación. En el capítulo IX desarrollaremos este tópico.

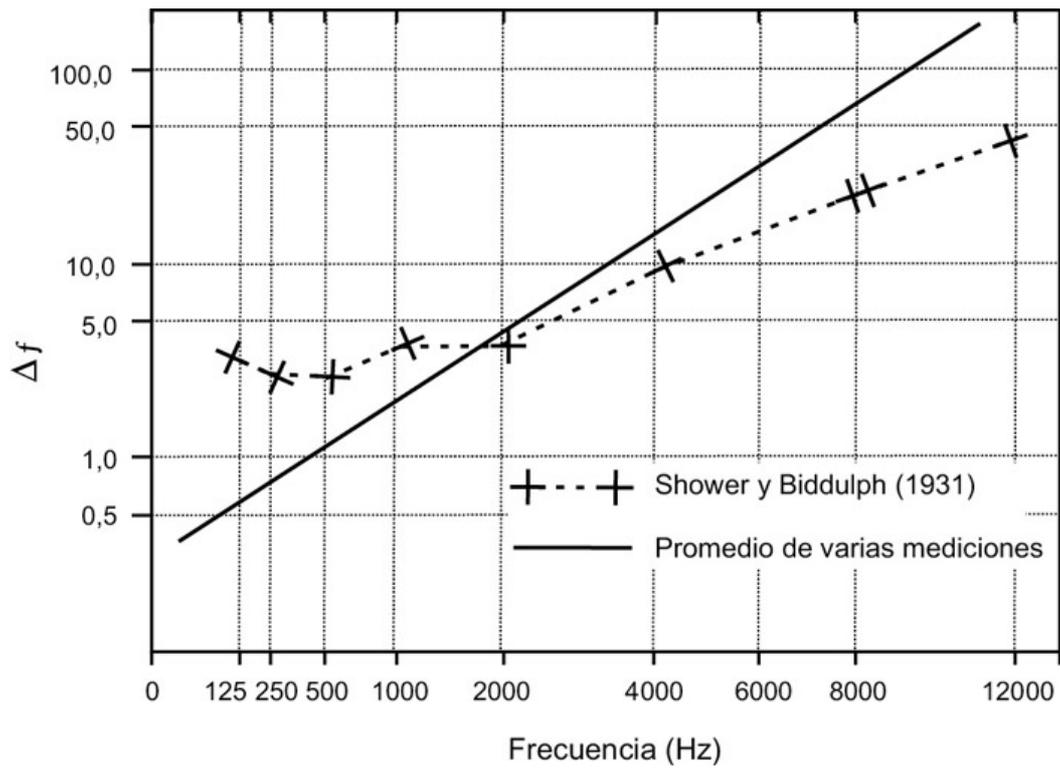
Contrariamente a lo predicho por la teoría del lugar, el DLF no varía con la frecuencia del mismo modo que la banda crítica y disminuye al aumentar la sonoridad del estímulo. A partir de estos datos la mayoría de los investigadores se inclinan por las teorías temporales al proponer sus modelos, aunque los resultados experimentales sugieren que existe un mecanismo temporal para las bajas frecuencias y un mecanismo tonotópico para altas frecuencias.

Zwicker y Henning crearon un modelo mixto que produjo resultados consistentes con los valores medidos de DLF y de FMDL (Zwicker y Fastl, 1999). Este modelo establece que el limen está determinado por la información temporal (sensibilidad a la fase) para frecuencias de hasta 5.000 Hz –la sensibilidad a la fase decrece al llegar a 2 kHz y desaparece a 5 kHz. Por encima de 5.000 Hz tanto el DLF y como el FMDL parecen responder a una suerte de teoría del lugar.

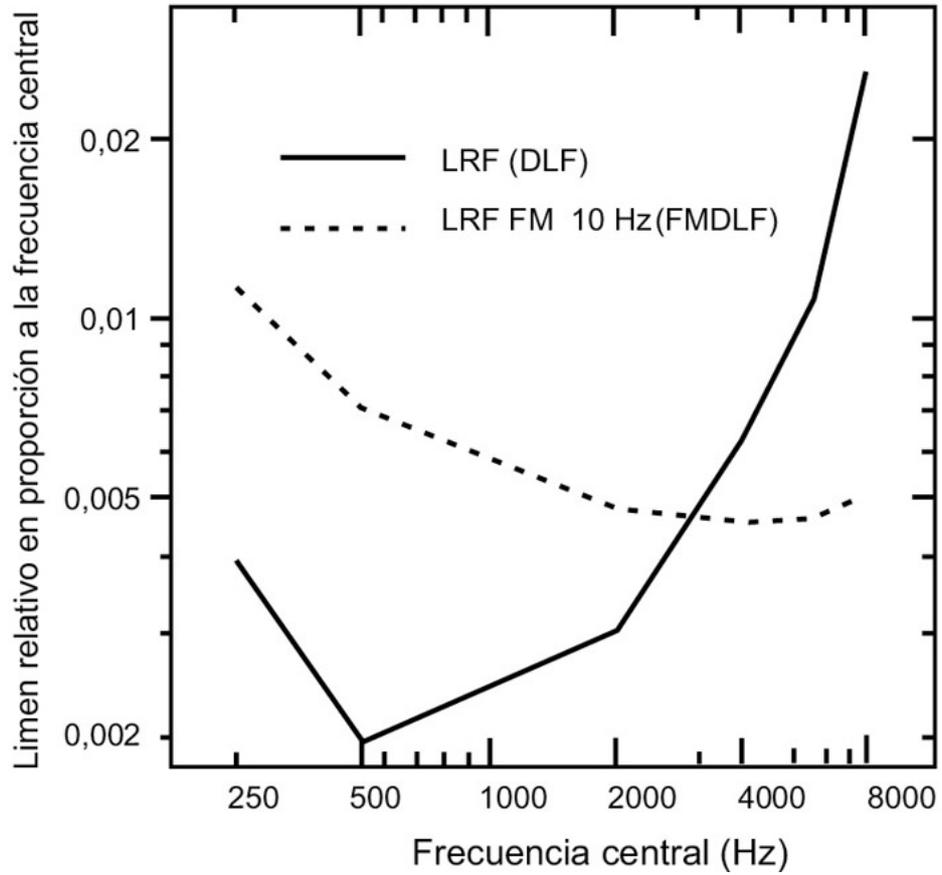
## **Variación de la altura tonal con la sonoridad**

Stevens constató en 1935, a partir de estímulos sinusoidales, que la altura tonal varía con la sonoridad del estímulo (Stevens, S. S., 1935). En la figura 5 se muestran los resultados del experimento original.

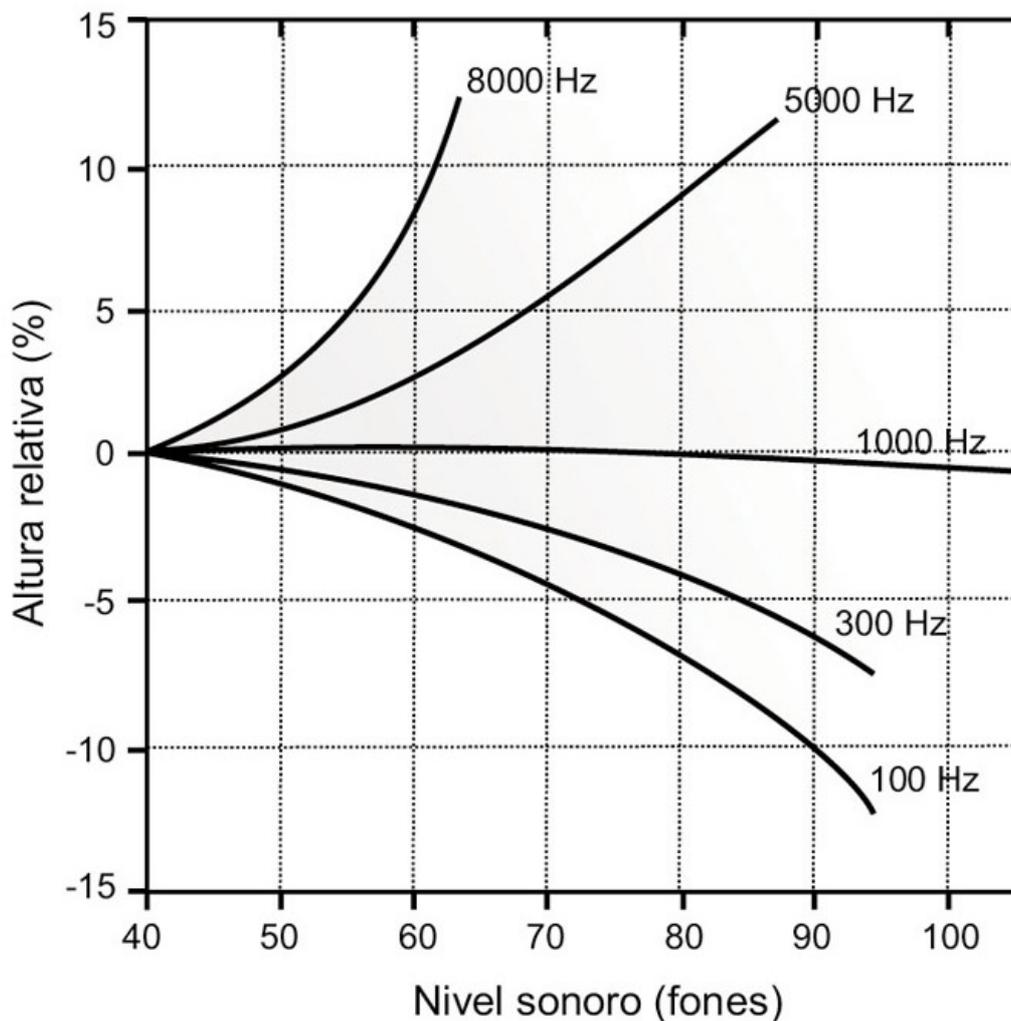
Aunque los datos actuales son menos profundos que los medidos por Stevens, el fenómeno ocurre y, para señales alejadas del centro del espectro audible, la variación en altura tonal puede superar el 5%. Este fenómeno no tiene aún explicación teórica.



**Figura 3.** Limen absoluto en frecuencia para estímulos sinusoidales. En la medición de Shower y Biddulph se emplearon FMDL



**Figura 4.** Limen relativo en frecuencia para estímulos sinusoidales. El FMDL se obtiene con una modulación en frecuencia



**Figura 5.** Variación de la altura con el nivel sonoro

La variación de la altura tonal con la sonoridad desaparece cuando el estímulo está constituido por señales armónicas complejas y no es, por lo tanto, apreciable en la mayoría de los sonidos que se usan en música.

## 2. Percepción de la altura tonal en estímulos armónicos complejos

A pesar de la gran cantidad de estudios realizados a partir de estímulos sinusoidales, la altura tonal es un rasgo propio de los sonidos tónicos derivados de señales periódicas complejas, cuyo espectro es aproximadamente armónico.

El nacimiento del concepto de altura tonal está asociado al desarrollo histórico de los modelos explicativos de la relación frecuencia-altura y del perfeccionamiento de los métodos de medición en psicoacústica. En el siglo XIX, Fechner llegó a la idea de que existe una transformación logarítmica del estímulo físico a partir del uso de métodos diferenciales para sus experimentos. A mediados del siglo XX, Stevens comparó las alturas tonales provenientes de estímulos sinusoidales y propuso la escala de mels. Sus resultados difieren de los de Fechner, en particular en los extremos del registro. Más tarde, Sheppard sostuvo que ambos métodos resultan incorrectos a la hora de evaluar y dar respuesta a la cuestión de la altura tonal, y que se debe centrar el estudio en la tasa de repetición de Fourier de la señal acústica.

Sin duda, hay al menos dos diferentes clases de altura tonal: la que deriva de estímulos diacrónicos o sucesivos y la que proviene de estímulos sincrónicos o simultáneos. Aunque la escala de mels no es una escala apropiada para usos musicales, captura sin embargo algún aspecto importante relacionado con la altura tonal de los sonidos diacrónicos. Algunos datos estadísticos, como la mayor cantidad de intervalos cerrados en el rango central de frecuencias, están relacionados con la escala de mels. También tiene algo que ver con el brillo de los sonidos, tal como comentáramos antes.

La teoría del lugar clásica tiene dificultades a la hora de describir la percepción de señales complejas ya que, por lo general, la altura tonal no corresponde a la posición del máximo de excitación en la membrana basilar. Una demostración concluyente la provee el fenómeno de la fundamental reconstruida o fantasma, descrito por Schouten en 1940 como el “residuo” de batido de un grupo de armónicos consecutivos.[\[5\]](#) La percepción de este residuo no requiere actividad en el punto de la MB que responde a una senoide de altura similar. Tiempo después, Licklider (1956) demostró que la altura del residuo persiste aún cuando los canales en la MB correspondientes a la fundamental de la serie armónica se enmascaran

con ruido de baja frecuencia: los sonidos de baja frecuencia que se perciben, asociados al residuo de batido, aparecen a partir de los canales excitados por los componentes de media y alta frecuencia de la señal.

Varios modelos fueron propuestos para explicar la incidencia de la fundamental reconstruida en la percepción de la altura tonal. Una primera clase de modelos, los llamados *modelos de reconocimiento de patrones*, asume que la altura de un sonido complejo se deriva de las señales neurales correspondientes a los estímulos primarios en la membrana basilar, por ejemplo las frecuencias de los parciales individuales. La segunda clase de modelos sostiene que la percepción de la altura tonal se asienta en la evaluación de intervalos temporales. Veamos cada una con detalle.

## **Modelos de reconocimiento de patrones**

Los modelos de reconocimiento de patrones dividen el proceso de percepción de la altura tonal en dos etapas. La primera consta de un análisis en frecuencia, similar al análisis clásico de Fourier, para determinar las componentes sinusoidales de la señal compleja. La segunda etapa consiste en el reconocimiento de los esquemas o patrones que determinan la altura tonal conforme a la salida de la primera etapa.

En el año 1962 Schouten realizó una experiencia que provocó la aparición de gran cantidad de modelos teóricos que intentaron dar cuenta de los resultados (Schouten, J. F. *et al.*, 1962). A partir de sinusoides moduladas en amplitud generó, en primer término, señales con tres componentes de 1.800, 2.000 y 2.200 Hz. La altura tonal percibida correspondía, como era de esperar, a la de una senoide de 200 Hz, fundamental de Fourier del trío de armónicos anteriores. Luego generó otra señal con componentes de 1.840, 2.040 y 2.240 Hz (que son los armónicos 46, 51 y 56 de una fundamental de 40 Hz). Los oyentes, sin embargo, percibían un sonido compuesto cuya altura tonal predominante era similar a la de una

sinusoide de 204 Hz, junto a dos alturas secundarias que se podían asociar con sinusoides de 185 y 227 Hz. Ninguno de los sujetos que participaron de la experiencia percibió la altura correspondiente a la fundamental de Fourier de 40 Hz.

El modelo propuesto por Walliser, en 1969, para explicar estos resultados consta de dos pasos: 1) se calcula la altura determinada por la diferencia de frecuencia entre parciales vecinos (la tasa de repetición de la envolvente) y 2) se busca el subarmónico del componente de menor frecuencia que mejor se ajuste a la frecuencia encontrada en el paso primero.<sup>6</sup>[6] Este modelo aplicado al experimento de Schouten da 200 Hz en la primera operación [1] y 204, 4 Hz ( $1.840 \text{ Hz} / 9$ ) en la segunda operación [2]. Esta última es la frecuencia que determina la altura tonal.

En 1972 Terdhardt sugirió, inspirado en el modelo de Walliser, que la altura residual es un subarmónico de un componente dominante y discriminable, sea o no el de menor frecuencia. Cuando los parciales están muy próximos en frecuencia no pueden discriminarse y resulta imposible oír la altura residual. Estimó que los parciales dominantes caen en la región espectral comprendida entre 500 y 1.500 Hz.[7]

Unos años después, Terdhardt sugirió que la capacidad necesaria para hacer uso del mecanismo propuesto en su modelo aparece en una fase temprana del aprendizaje, en la que estamos expuestos sobre todo a los sonidos del habla (Terdhardt, E., 1974). Afirmó que aprendemos a asociar la frecuencia de un componente del estímulo con los subarmónicos del mismo: la estimulación de un solo armónico produce alturas probables a las frecuencias de los subarmónicos. Cuando un complejo armónico se presenta, la mayor cantidad de coincidencias a la frecuencia de la fundamental determina la altura tonal del sonido. Para sonidos no armónicos (como el de la experiencia de Schouten) no hay coincidencias exactas y la sensación de altura no resulta tan precisa. Queda claro que este mo-

delo requiere del análisis de más de un parcial de la señal de entrada.

Más o menos en la misma época, Goldstein propuso un modelo en el que la altura tonal de un sonido complejo es derivada por un procesador central que recibe información sólo de la frecuencia de cada componente, soslayando la información aportada por la amplitud y la fase (Goldstein, J. L., 1973). El procesador presume que todos los estímulos son periódicos y que todos sus espectros comprenden armónicos sucesivos. Los errores de estimación de la fundamental ocurren al estimar el número de armónico apropiado. Aplicado al experimento de Schouten, los armónicos de 204 Hz explican razonablemente bien los componentes de 1.836, 2.040 y 2.244, pero no dan cuenta con igual exactitud las frecuencias de 185,5 Hz y de 226,7 Hz –cercanas a las de 185 y 227 Hz percibidas en el experimento original.

Los modelos de reconocimiento de patrones dependen de la resolución espectral de la membrana basilar, pero no implican necesariamente el uso de información tonotópica *per se* como una medida de la frecuencia.

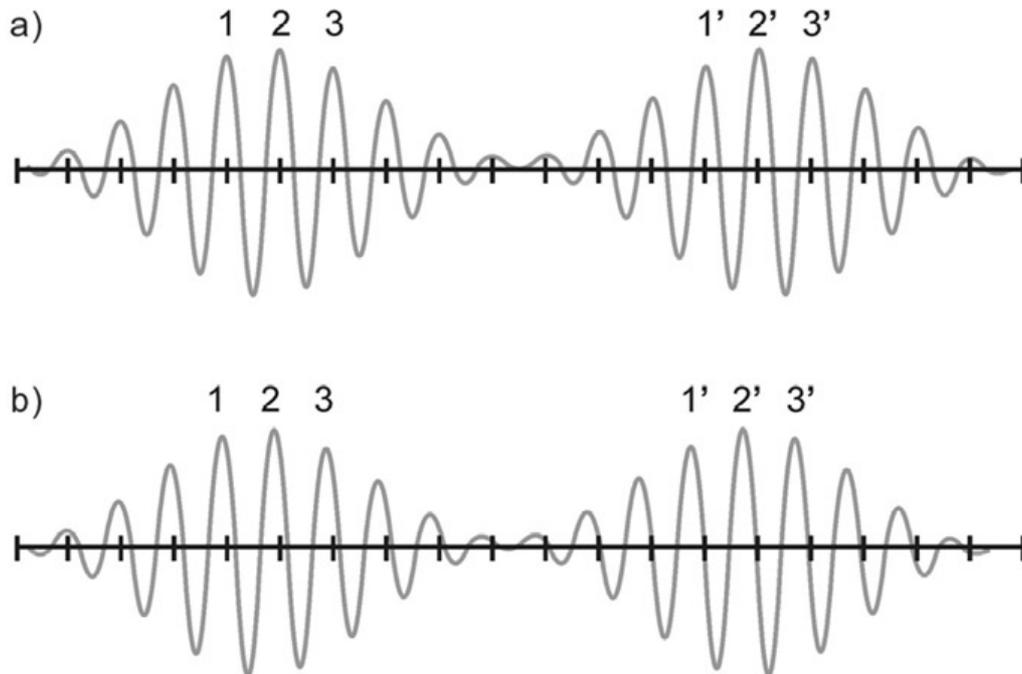
## **Modelos temporales**

Para dar cuenta de su célebre experimento, Schouten propuso un modelo basado en el patrón temporal que desarrolla la onda del estímulo en la membrana basilar. Según su modelo, la altura tonal es determinada por la periodicidad de los parciales involucrados en la determinación del residuo. El peso relativo de cada parcial depende de su capacidad para atraer la atención del oyente, ya sea por sonoridad o por contraste con estímulos previos.

La gran diferencia entre los modelos temporales y los modelos de reconocimiento de patrones es que estos últimos requieren que más de un parcial sea discriminable del complejo para que se perciba la altura del residuo, mientras que los primeros sólo piden que

los parciales interactúen, no que sean discriminables (no es necesario que sean percibidos separados).

En la figura 6 se muestran los resultados temporales del experimento de Schouten.



**Figura 6.** Formas de onda de dos señales moduladas en amplitud. En el primer caso, la portadora (2.000 Hz) es un múltiplo exacto de la moduladora (200 Hz) y el período entre picos correspondientes (1-1', 2-2') es de 5 ms. En el segundo la frecuencia de la portadora aumenta a 2.040 Hz y el intervalo entre picos es algo menor a 5 ms

La altura residual se determina por el intervalo temporal entre máximos en la estructura fina de la onda en la membrana basilar (Schouten registró la señal en el punto de máxima oscilación en la MB). Si hay más de un intervalo temporal posible, la altura tonal corresponde a la cresta más prominente. Si se considera la representación de la señal a nivel neural adquiere significación el intervalo temporal entre los disparos nerviosos similares. Si el sistema

“contase” simplemente la cantidad de disparos por segundo, el modelo de Schouten no podría funcionar.

## **Evidencia experimental relevante a cada modelo**

### ***La región de existencia de la altura residual***

De acuerdo con la teoría de reconocimiento de patrones, la altura residual sólo se puede percibir cuando al menos uno de los componentes sinusoidales del estímulo puede ser oído en forma separada. Es decir, que necesita ser discriminable. Si el estímulo posee únicamente armónicos muy altos no resolubles la altura residual no debería percibirse.

Ritsma encontró que el carácter tonal de la altura residual existe únicamente dentro de una región de frecuencias limitada, que llamó la *región de existencia* (Ritsma, R., 1962). Cuando el número del armónico es muy grande (por encima de 20 o más) se oye un zumbido de altura indefinible. Tiempo después, Plomp halló que la altura residual puede percibirse para tasas de repetición de hasta 1.400 Hz, dato compatible –aunque no idéntico– con la propuesta de Ritsma (Plomp, R., 1967).

A contrapelo de las hipótesis anteriores, Moore estableció que para armónicos de orden elevado hay un rango determinado en el cual se puede oír una altura residual aun cuando ninguno de los armónicos pueda oírse por separado (Moore, B., 1995). Los modelos de reconocimiento de patrones no pueden dar cuenta de este hecho. De todos modos hay que apuntar que los sonidos generados únicamente a partir de armónicos elevados no definen tan claramente la altura residual como aquellos que contienen armónicos más bajos.

Aunque la altura residual se percibe mejor en presencia de gran cantidad de componentes armónicos, Houtgast reportó que si el contexto lleva al sujeto a centrar la atención sobre alturas de un rango particular, aun un solo armónico es capaz de evocar la per-

cepción de una altura residual grave (Houtgast, T., 1976). Cuando la información sensorial se vuelve ambigua, como ocurre en presencia de ruido, esas alturas potenciales se perciben con más facilidad.

En un experimento secuencial de Hall y Peters, en el que tres armónicos sucesivos de 40 ms de duración se presentaban separados por 10 ms de silencio, los sonidos se oían dissociados en condiciones de laboratorio (Hall y Peters, 1981). Pero en presencia de ruido la altura residual asociada, correspondiente a los armónicos en presentación sincrónica, se volvía perceptible. Parece ser que el ruido habilita la aparición de un mecanismo sintético de percepción de la altura.

### ***El principio de dominación***

De acuerdo con la teoría de reconocimiento de patrones, los armónicos discriminables de menor frecuencia son los más importantes para determinar la altura tonal. En cambio, para las teorías temporales los más importantes son los de mayor frecuencia. La altura residual se ha observado únicamente cuando los armónicos están presentes a frecuencias por debajo de 5.000 Hz, región que coincide con la efectividad del procesamiento de la información temporal.

Ritsma sugirió que los armónicos 3, 4 y 5 dominan la sensación de altura tonal en la región de 100 a 400 Hz (Ritsma, R., 1967). Plomp coincidió en parte con este postulado, pero consideró que la región no se define en cuanto al número de armónico, sino que es función de la frecuencia absoluta (Plomp, R., 1967). Para alturas tonales agudas son importantes los primeros armónicos, mientras que para alturas tonales graves cobran importancia los armónicos por encima del quinto.

### ***Evidencia del rol de mecanismos centrales de procesamiento***

Houtsma y Goldstein encontraron que es posible percibir alturas residuales cuando los armónicos sucesivos se presentan alternados entre los dos oídos, por ejemplo los armónicos pares al oído derecho y los impares al izquierdo (estímulos dicóticos): este hecho no puede explicarse conforme a los modelos que requieren la interacción de los componentes en la membrana basilar como el de Schouten (Houtsma y Goldstein, 1972). Implican necesariamente la existencia de un proceso neural central. Los modelos temporales pueden, sin embargo, explicar el reconocimiento de la altura residual a partir de secuencias de disparos neuronales en lugares de procesamiento superiores, más allá de la membrana basilar, en los que se encuentra la información proveniente de los dos oídos.

### ***Alturas tonales basadas únicamente en información temporal***

Miller y Taylor investigaron la sensación de altura tonal provocada por señales de ruido blanco que se conectan y desconectan periódicamente, cuyo espectro asociado no posee máximos ni mínimos destacados (Miller y Taylor, 1948). Encontraron que para una frecuencia de conmutación comprendida entre 100 y 2.000 Hz la mayoría de los sujetos percibieron una cierta altura tonal. Llegaron incluso a construir melodías reconocibles.

Estas experiencias demuestran la existencia de mecanismos temporales en la percepción de la altura tonal, pues los modelos de reconocimiento de patrones no pueden dar cuenta de sus resultados.

### **Modelo de Brian Moore del mecanismo para la percepción de la altura tonal en sonidos complejos**

La evidencia experimental revela que la percepción de una altura tonal particular no depende de la actividad localizada en un lugar específico de la membrana basilar, o de la actividad de un único grupo de neuronas periféricas. Se sabe que la altura tonal inducida

por estímulos complejos armónicos es obtenida, en general, a partir de la interacción de los armónicos más allá de la fundamental de la serie. Por lo tanto, es posible percibir alturas tonales similares provenientes de patrones espectrales diferentes, que comprometen distintos esquemas en la actividad basilar o neural. Para estímulos periódicos que contienen un gran rango espectral los primeros armónicos, hasta el quinto, tienden a dominar el percepto auditivo. Dichos armónicos caen en una zona en la que es posible discriminarlos –resolverlos– por separado y el proceso puede explicarse sobre la base de los modelos de reconocimiento de patrones. Sin embargo, se puede percibir una altura residual aun cuando los componentes son demasiado altos como para ser discriminados por separado. Este hecho no puede ser explicado por los modelos de reconocimiento de patrones, pero sí por los modelos temporales.

Parece que ninguna de las teorías expuestas hasta aquí puede dar cuenta por sí sola del conjunto de datos experimentales. Brian Moore propuso un modelo cualitativo que combina elementos tanto de la teoría de reconocimiento de patrones como de las teorías temporales (Moore, B., 1997). Otros modelos, como los de Van Noordern (1982), Patterson (1987) y Meddis y Hewitt (1991) también combinan aspectos de las dos teorías, aunque los algoritmos matemáticos difieran significativamente en cada caso. El modelo de Moore (esquemático en la figura 7) consta de cinco etapas.

La primera etapa está definida por los filtros auditivos en la membrana basilar. Responden a los armónicos que son aproximadamente sinusoidales en su forma y que pueden resolverse. La respuesta de los filtros a los armónicos más altos combina varios de ellos por banda crítica, y no pueden discriminarse. Aunque la forma de onda de salida de la etapa es compleja, la tasa de repetición corresponde a la que se encuentra en la entrada. A modo de ejemplo, en la figura 8 se aprecia la salida de diferentes filtros auditivos ante un tren de impulsos de 200 pulsos por segundo.

Entrada  
acústica



banco de  
filtros  
pasabanda



Transducción  
nerviosa



Análisis de los  
intervalos  
entre picos



Combinación de  
intervalos con  
frecuencias centrales



Selección de los  
intervalos más  
prominentes



Altura  
tonal

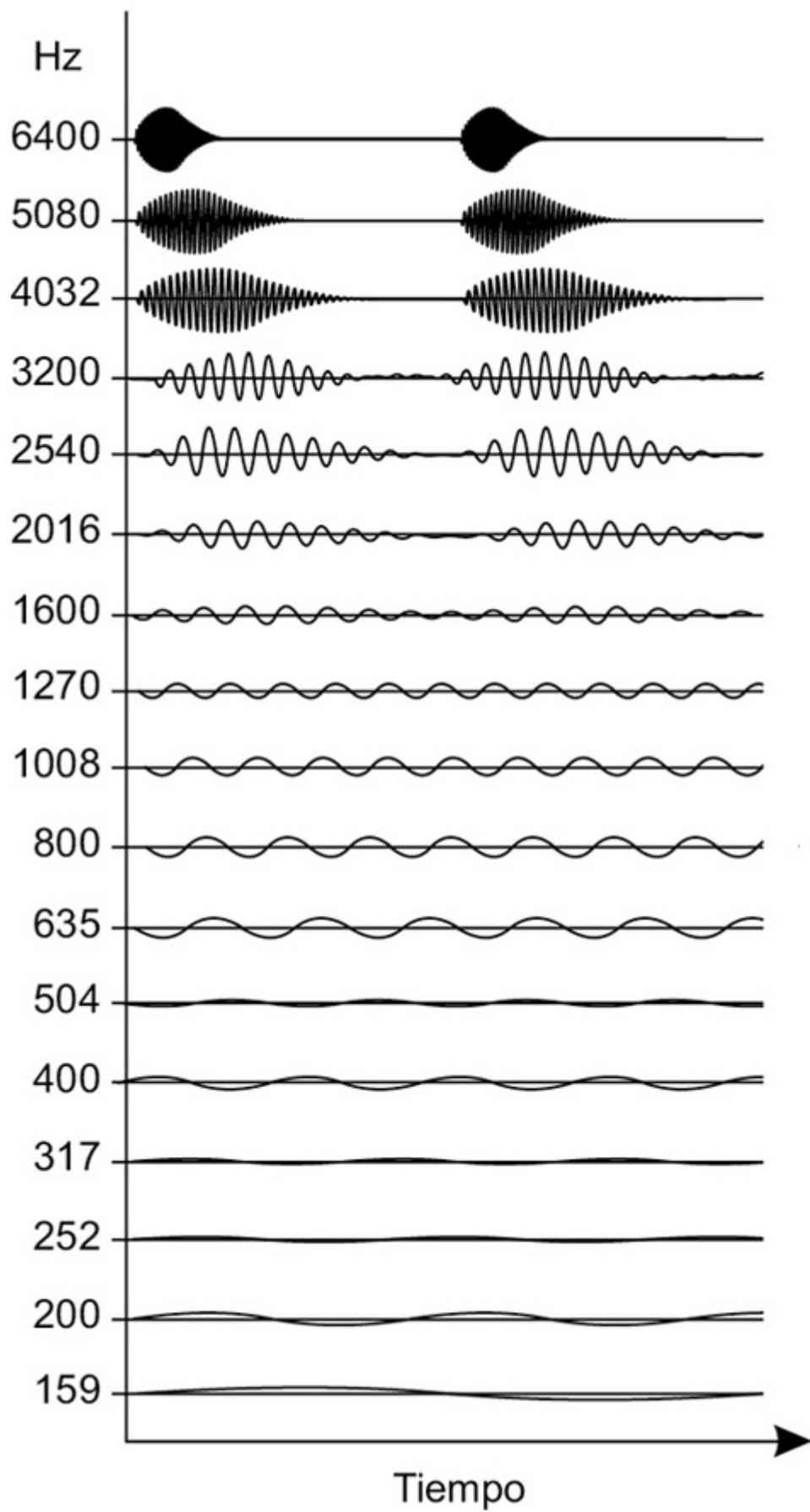
**Figura 7.** Modelo esquemático para la percepción de la altura tonal de Brian Moore

La segunda etapa consiste en la transducción de la salida de los filtros anteriores en impulsos nerviosos.

La tercera etapa define un dispositivo capaz de analizar los intervalos entre máximos a cada frecuencia central  $FC$  (véase la figura 8).

En la cuarta etapa se comparan los intervalos temporales presentes en los diferentes canales y se buscan los patrones comunes a varios de ellos. Este circuito también debe integrar la información a través del tiempo. En general, el intervalo temporal más frecuente corresponde a la componente fundamental.

En la quinta etapa los valores de los intervalos temporales más prominentes llegan a un mecanismo de decisión que selecciona uno de ellos. Este circuito incorpora la memoria y la atención, y puede ser influido tanto por los estímulos inmediatamente precedentes como por el contexto y las condiciones generales de presentación. La altura tonal final (señal de salida del modelo) corresponde al intervalo seleccionado.



**Figura 8.** Simulación de las respuestas de la MB ante un tren de impulsos de 200 pps. Los números de la izquierda indican cada una de las frecuencias centrales FC, que corresponden a las formas de onda de salida de cada filtro

El modelo de Moore explica satisfactoriamente gran cantidad de evidencia experimental. A continuación mencionaremos alguna de las situaciones que logra resolver.

- Cuando la altura residual se obtiene a partir de unos pocos armónicos elevados (por ejemplo, el 12, 13 y 14) que no pueden resolverse por separado, resulta débil y ambigua. Este hecho explica por qué los primeros armónicos tienden a dominar en la percepción de la altura. La altura asociada a los armónicos superiores se hace más clara si aumenta el número de armónicos presentes.
- Un pequeño grupo de armónicos muy altos no genera la sensación de altura tonal, hecho que se puede explicar en términos del limitado rango de intervalos temporales que pueden ser analizados por el grupo de neuronas de cada frecuencia central. Para armónicos por encima del 15, el intervalo temporal correspondiente a la fundamental cae fuera del rango que puede analizar el canal que responde a esos armónicos (este dato es muy importante en el modelo de Moore).
- Cuando los armónicos superan los 5 kHz la estructura fina de la forma de onda a la salida del filtro de la primera etapa no se preserva en el patrón temporal de los impulsos nerviosos.
- La altura de estímulos que no presentan picos espectrales, como el ruido blanco interrumpido periódicamente, se percibe en el modelo a través de los intervalos temporales presentes en los canales de altas frecuencias centrales  $f_c$ , que poseen mayores anchos de banda y por lo tanto mayor defini-

ción temporal (preservan la estructura temporal fina de la forma de onda). La altura tonal de estos estímulos es débil.

- Si se incorpora un mecanismo comparador central de la información proveniente de ambos oídos luego de la tercera etapa, se pueden interpretar los fenómenos dicóticos.

El modelo se parece al de Terhardt en el tratamiento de los primeros armónicos, pero este último no puede explicar la altura producida por los armónicos superiores no discriminables. En el modelo de Moore los sonidos de combinación juegan un rol en la determinación de la altura. El más importante ( $2f_1 - f_2$ ) se percibe aun a bajas sonoridades y puede actuar como un armónico inferior más fácil de discriminar que algunos componentes del espectro de entrada.

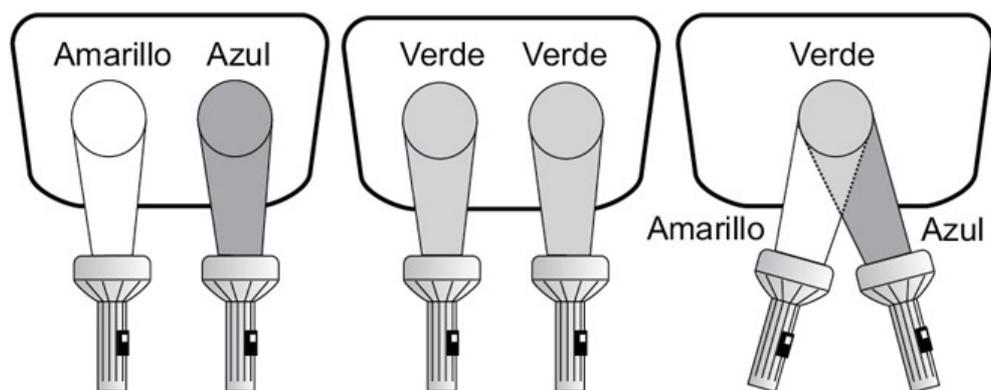
El modelo de Moore se utiliza actualmente en el diseño de traductores de altura tonal para sistemas codificados a número único, por ejemplo para diseñar conversores de altura tonal desde una guitarra al sistema MIDI.

### **3. La altura tonal como un medio morfofórico**

A principios de la década de 1970 Attneave y Olson establecieron un nuevo criterio que permite clasificar de manera diferente los parámetros del sonido. Aunque existen otros atributos para la organización de los sonidos –como el timbre, la sonoridad y la localización espacial– sólo la altura tonal y el tiempo constituyen medios capaces de portar formas, o *medios morfofóricos* [8] (Attneave y Olson, 1971). De acuerdo con los autores, el espacio visual es un medio morfofórico: un triángulo se puede trasladar –o rotar– y sigue siendo el mismo triángulo, su identidad se mantiene bajo ciertas operaciones de transformación. De manera similar las estructuras generadas a partir de la altura tonal, tales como melodías o

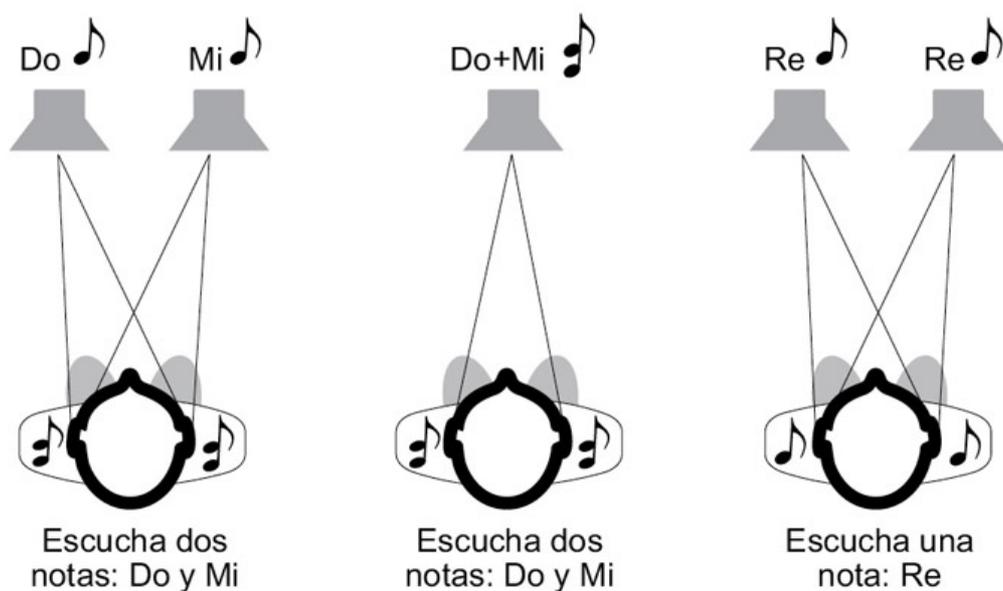
funciones armónicas, pueden ser transportadas hacia arriba o hacia abajo y aún ser reconocidas como las mismas: la información estructural que poseen resulta invariante frente a los transportes en altura tonal.[\[9\]](#) El tiempo también es un medio morfofórico poderoso. Pero otras dimensiones utilizadas en música (como la sonoridad, el timbre y la localización espacial) no lo son. Por ejemplo, no se logra el mismo grado de preservación de la información estructural actuando sobre la sonoridad. En el siglo XX varias corrientes musicales intentaron componer piezas a partir de escalas de sonoridad, de ataques o de otros parámetros sonoros, pero fracasaron en el intento de equiparar su capacidad estructural a la de la altura tonal o el tiempo.[\[10\]](#)

En la misma línea, el psicólogo cognitivista Michael Kubovy distinguió entre *atributos dispensables* e *indispensables* (Kubovy, M., 1981). En visión la posición espacial es un atributo indispensable, mientras que el color es dispensable, hecho relacionado con su incapacidad analítica frente a los colores: al mezclar azul y amarillo vemos un nuevo color, el verde, pero no sus componentes iniciales. En audición la altura tonal es un atributo indispensable, mientras que la localización espacial no lo es. La “indispensabilidad” del espacio en visión y de la altura tonal en audición es paralela al hecho que ambos son atributos correspondientes a medios morfofóricos. Entonces el parámetro análogo del espacio visual no es el espacio auditivo, sino la altura tonal. En la figura 9 se grafican las diferencias entre atributos dispensables e indispensables en visión y en la figura 10 en audición.



En la percepción visual el color es dispensable, la posición no

**Figura 9.** Atributos dispensables e indispensables en visión



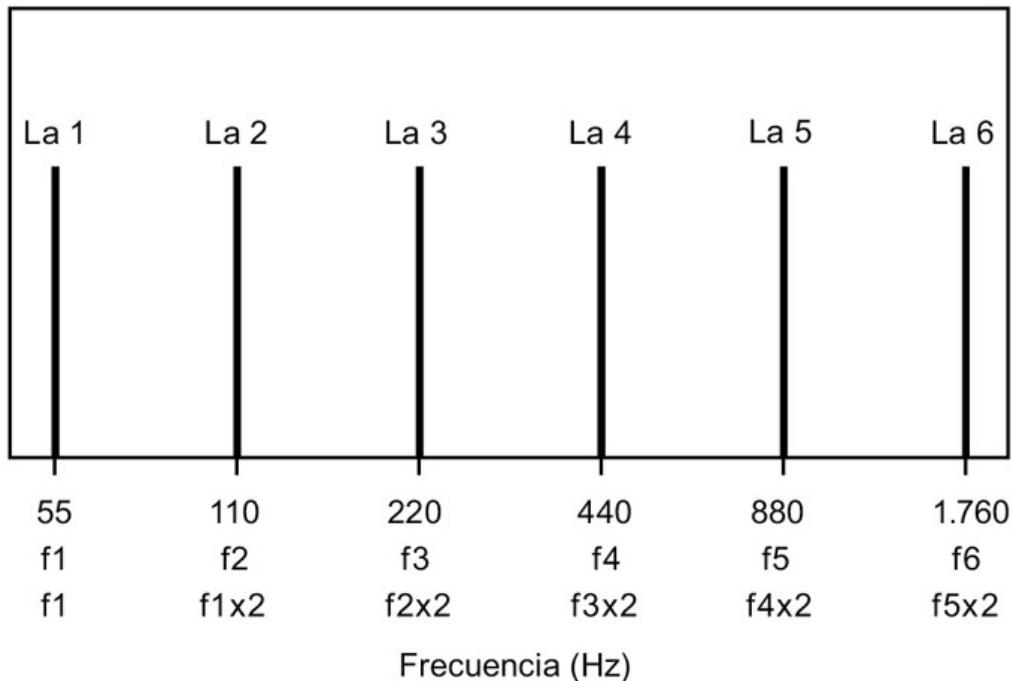
En la percepción auditiva la altura es indispensable,  
la posición es dispensable

**Figura 10.** Atributos dispensables e indispensables en audición

## 4. Representación gráfica de la altura tonal

La representación gráfica tradicional de la altura tonal emplea una escala logarítmica unidimensional de frecuencias que preserva los

intervalos musicales bajo diferentes transformaciones, pero no alcanza a describir las complejas relaciones que operan entre los intervalos musicales. En la figura 11 se puede verificar que las octavas ocupan segmentos de la misma longitud a lo largo del espectro audible. Lo mismo ocurre con el resto de los intervalos.[\[11\]](#)



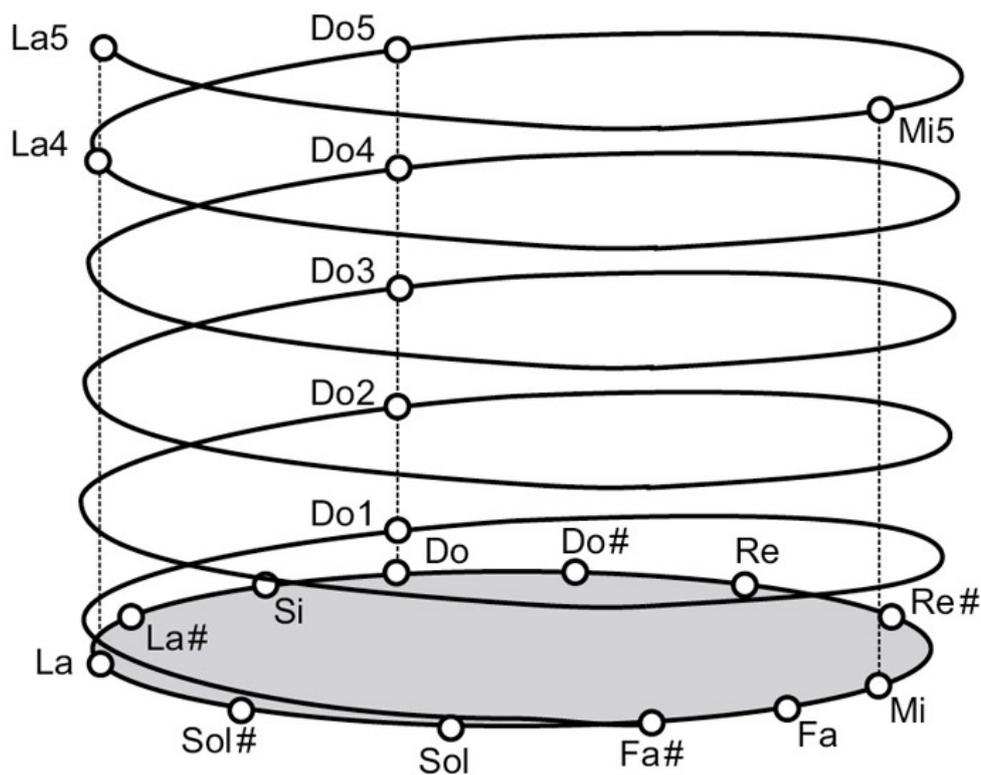
**Figura 11.** Representación logarítmica unidimensional de la altura tonal

Sin embargo, una escala logarítmica unidimensional no refleja el hecho de que ciertos intervalos, como las octavas y las quintas justas, poseen características especiales. Ciertos estudios etnomusicológicos hasta pretenden demostrar que dichos intervalos son culturalmente universales, pero el debate sobre el punto sigue lejos de alcanzar una conclusión definitiva.[\[12\]](#)

A mediados del siglo XIX Drobisch concibió una representación tridimensional en forma de hélice que distingue entre la altura unidimensional (a la que llamó *pitch height*) y la altura tonal (*pitch croma* según Drobisch).<sup>13</sup>[\[13\]](#) En la figura 12 se puede apreciar la versión de Shepard de la hélice de alturas tonales.

En el esquema los sonidos de cada familia tonal (por ejemplo todos los la a distancia de octava entre sí) se hallan encolumnados: el La5 queda más cerca del La4 que del Mi5, disposición que corresponde a la proximidad según la consonancia musical, al menos en la música tonal tradicional.

Si se desea suprimir la altura tonal (*croma*) sin eliminar la altura unidimensional (*height*) lo más sencillo es reemplazar las señales armónicas originales por ruido de banda angosta. En otras palabras, se remueven los elementos que aportan tonicidad y se dejan solamente los que corresponden a la altura espectral del sonido. En ese caso la hélice se reduce a un solo eje vertical, el que define la altura unidimensional.



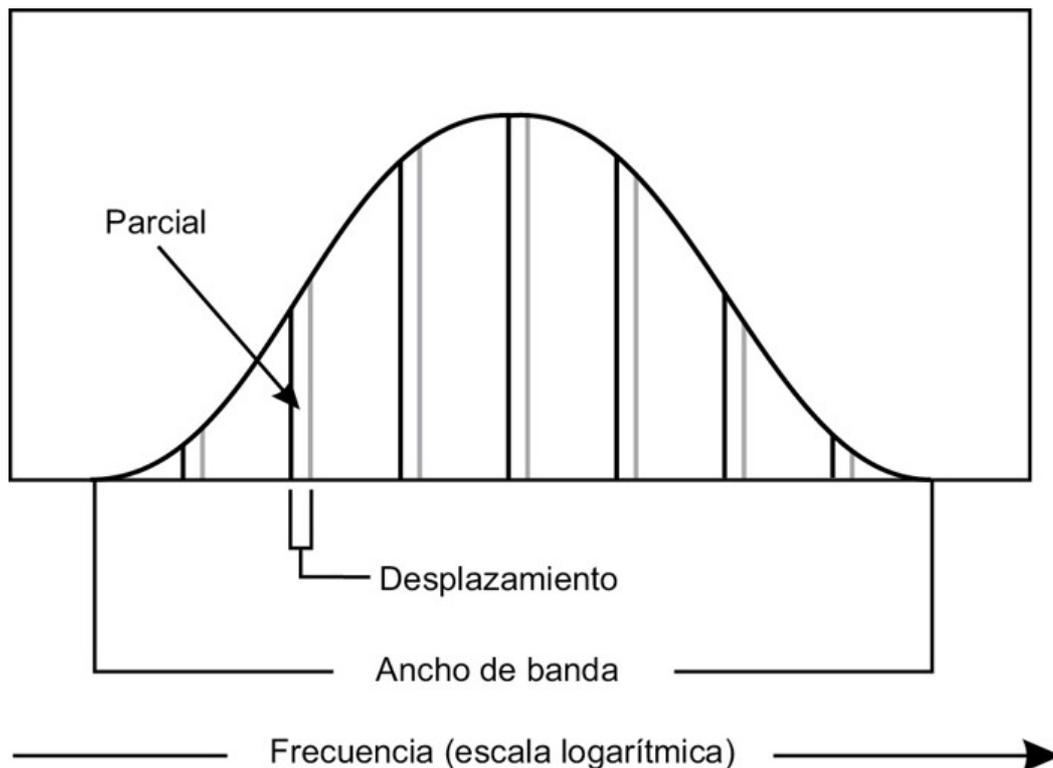
**Figura 12.** Hélice de alturas tonales

En cambio, si se quiere suprimir la altura unidimensional (*height*) y dejar sólo la altura tonal (*croma*), es necesario generar una señal que obligue a la hélice a colapsar en un círculo como el dibujado

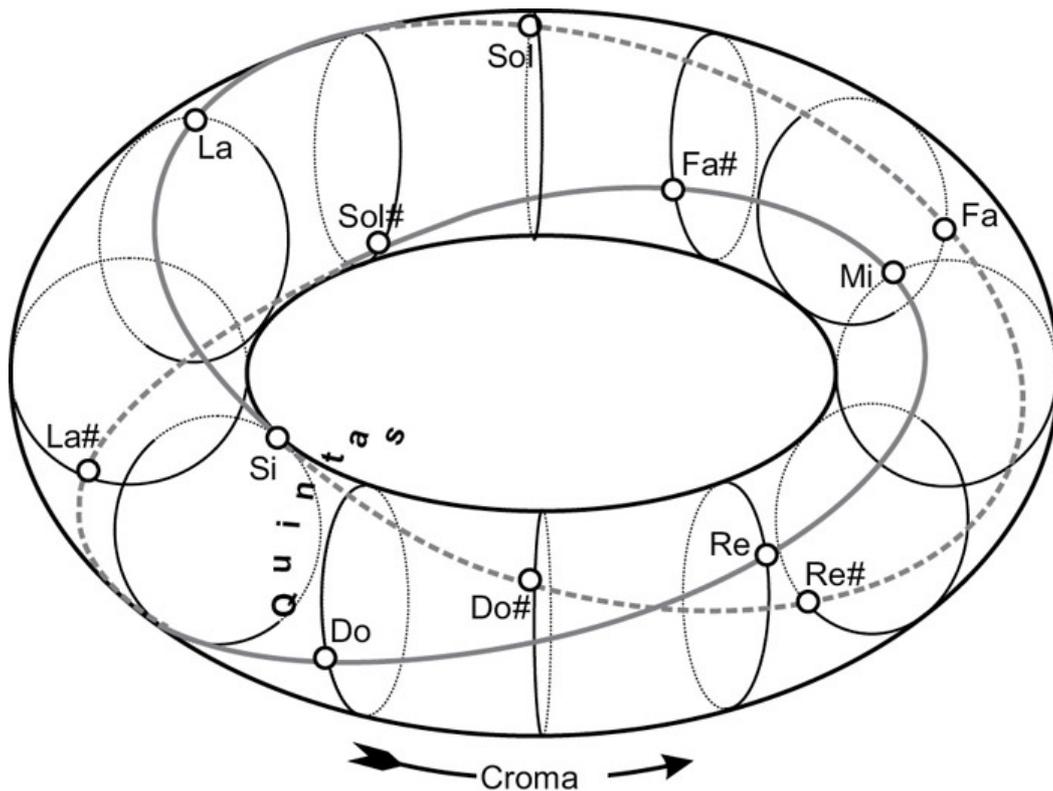
en la base de la figura 12. Roger Sheppard publicó en 1964 un trabajo en el que describe la construcción de señales capaces de suspender la percepción de la altura unidimensional, en las que la altura tonal parece ir hacia el agudo (o el grave) indefinidamente (Sheppard, R. 1964).[\[14\]](#)

## Sonidos de Sheppard

Los sonidos de Sheppard consisten en una gran cantidad de componentes sinusoidales a distancia de octava que abarcan por completo el rango de frecuencias audibles, modulados en amplitud por una envolvente espectral en forma de campana que tiende a cero para altas y bajas frecuencias (Risset, J. C., 1978b). La figura 13 muestra el espectro de una señal de Sheppard.



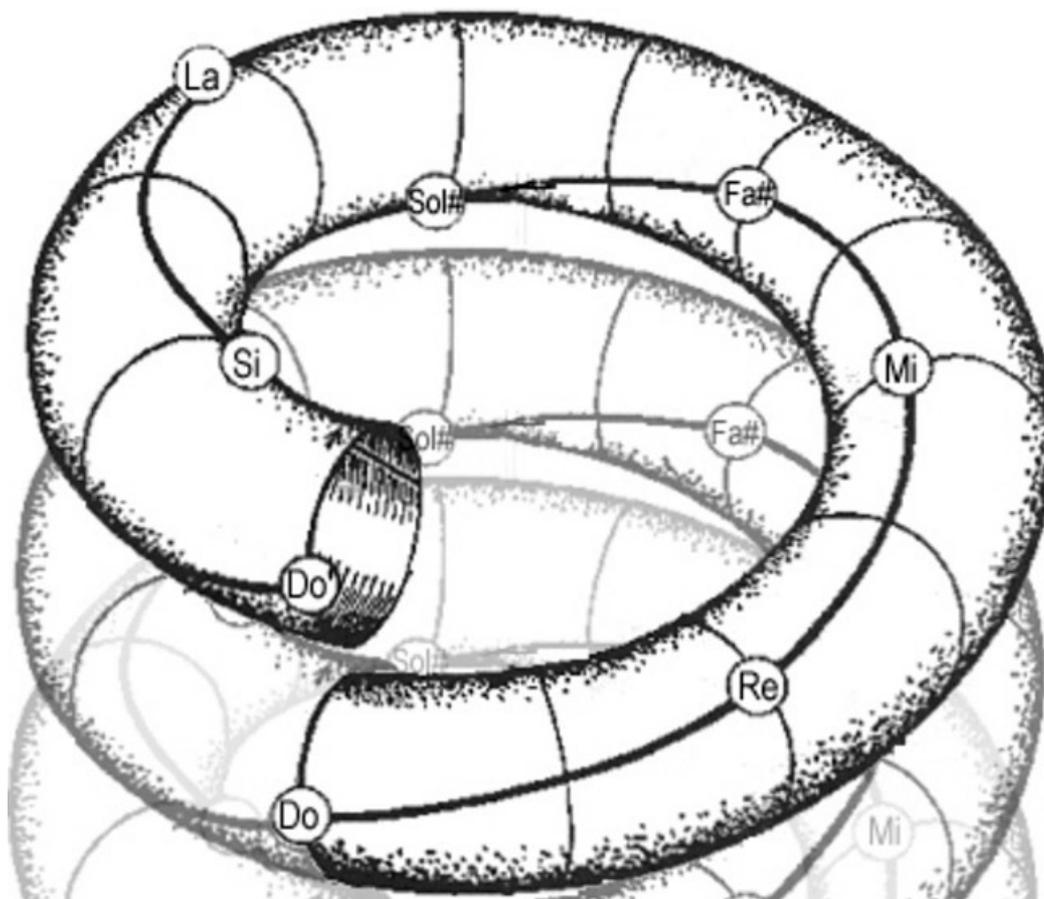
**Figura 13..** Señal de Sheppard



**Figura 14.** Toroide que permite representar las relaciones de octava y de quinta

A medida que transcurre el tiempo, la frecuencia de cada parcial se va corriendo hacia arriba, en una suerte de glissando en octavas paralelas, mientras que la amplitud es controlada por la envolvente en forma de campana. Nuevos parciales ingresan en la parte grave del espectro a medida que superan la barrera del umbral de audibilidad, a la vez que los más agudos se reducen hasta desaparecer. La cantidad de parciales y la energía de la señal son constantes, pero se percibe un incremento continuo de la altura tonal. Los sonidos que descienden indefinidamente se obtienen de manera similar, desplazando hacia abajo la frecuencia de los componentes.

Una ilusión equivalente a la de los sonidos de Sheppard se puede obtener con el otro medio morfofórico conocido, el tiempo. Ligeti y Risset emplearon procesos de este tipo en varias de sus composiciones.



**Figura 15.** Hélice toroidal que permite representar la altura tonal, la altura (*height*) y las relaciones de octava y de quinta

Aunque la hélice de Sheppard de la figura 12 preserva la relación de octava, no registra la importancia que poseen los intervalos de quinta en la música tonal tradicional. El toroide de la figura 14 sí lo hace, pero pierde la capacidad de graficar la altura unidimensional (*height*).

Para representar al mismo tiempo la altura unidimensional (*height*), la altura tonal (*croma*) y las relaciones especiales de octava y quinta es necesario un espacio de cinco dimensiones. La hélice toroidal resultante sería similar a la representada en la figura 15.

Si además deseamos representar los intervalos de tercera mayor y tercera menor, imprescindibles a la hora de definir la modalidad de un acorde, la figura se complica hasta el punto de resultar im-

posible de dibujar. Solamente se pueden plantear las ecuaciones algebraicas que correspondan a cada modelo.

## 5. Altura tonal e intervalos musicales

Un desarrollo que intente exponer la complejidad de la percepción de los intervalos y de la generación de escalas musicales requiere por lo menos un libro completo.[\[15\]](#) Aquí intentaremos presentar los tópicos más relevantes sin adentrarnos en el estudio completo de los pormenores propios de la práctica musical.

### Percepción de intervalos musicales

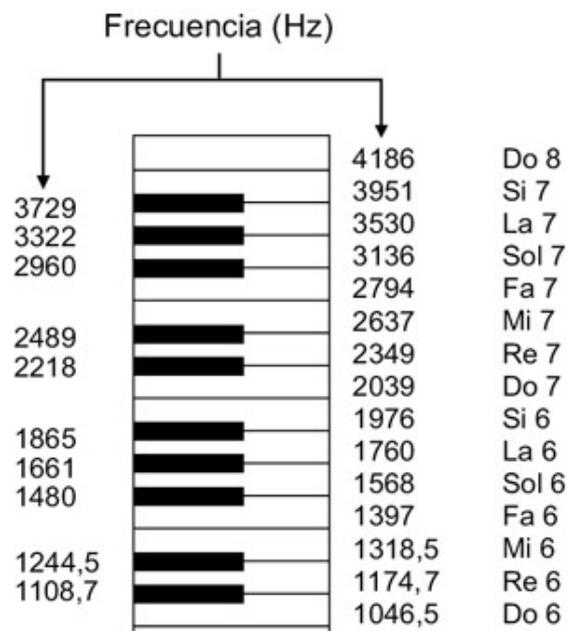
Noam Chomsky distinguió entre *competence* (habilidad para reconocer las reglas que gobiernan la lengua) y *performance* (habilidad para realizar un acto de habla) en su teoría de la gramática y la sintaxis. Aunque tengamos almacenadas internamente las reglas de la lengua, debido a nuestras limitaciones para procesar y emitir sentencias complejas, muchas veces no las aplicamos correctamente al oír o producir frases completas. Del mismo modo, las reglas que gobiernan las relaciones interválicas en música deben ser conocidas e internalizadas (es decir, “naturalizadas”) por los oyentes. Veamos algunos casos propios de la música occidental tradicional, en la que dichas relaciones ya han sido largamente naturalizadas.

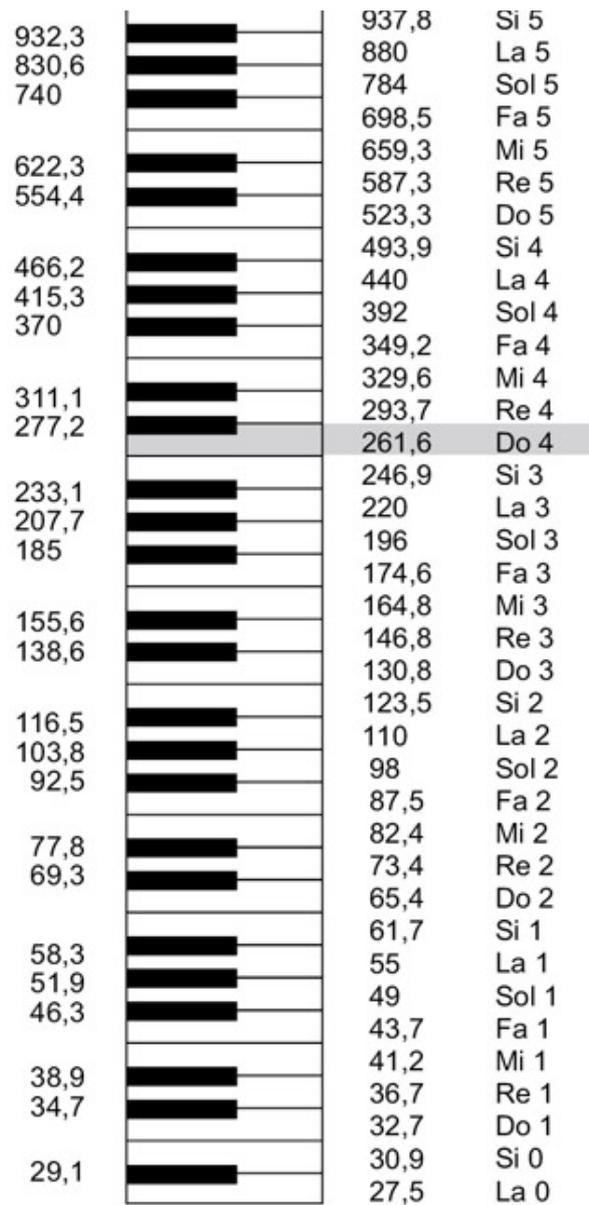
Dos sonidos cuyos estímulos están separados por un intervalo de octava suenan, en algún sentido, de manera similar. En muchas culturas se les denomina con el mismo nombre dentro de la escala musical (son, por ejemplo, los sonidos de la familia del Do). Desde el punto de vista del estímulo, la frecuencia de la segunda señal se ajusta a un valor cercano al doble de la frecuencia de la frecuencia de la primera señal. En otros términos, si la relación de frecuencias fundamentales es aproximadamente  $f_2 \approx 2 f_1$  se percibe un intervalo de octava. Otros intervalos pueden definirse también a partir de relaciones similares. Así, una quinta justa corresponde a la rela-

ción  $f_2 \approx (3/2) f_1$  y una cuarta justa a  $f_2 \approx (4/3) f_1$ . Las relaciones son solamente aproximadas pues la definición de los intervalos depende de gran cantidad de factores, la mayoría culturales, alguno de los cuales veremos más adelante.

Las relaciones anteriores, sin embargo, sólo se perciben claramente cuando las dos frecuencias se ubican por debajo de 5.000 Hz. Existe gran cantidad de evidencia experimental que indica que los estímulos cuyas frecuencias fundamentales (las que deciden la periodicidad de la señal compleja) superan el valor de 5.000 Hz no definen la altura tonal. Por ejemplo, una secuencia de sonidos no produce la sensación de melodía por encima de los 5 kHz. Y los sujetos con oído absoluto fracasan repetidamente a altas frecuencias. Estos datos son consistentes con el hecho ya visto de que a alrededor de los 5 kHz cambian los mecanismos de percepción de la altura: un mecanismo temporal actúa a bajas frecuencias y un mecanismo tonotópico lo hace a altas frecuencias. Tanto el sentido de la altura tonal como la sensibilidad a la fase se pierden más allá de los 5.000 Hz.

Puede ser interesante repasar el rango de frecuencias fundamentales que abarca el teclado de un piano moderno: la frecuencia más alta no llega a 4.200 Hz (véase la figura 16).





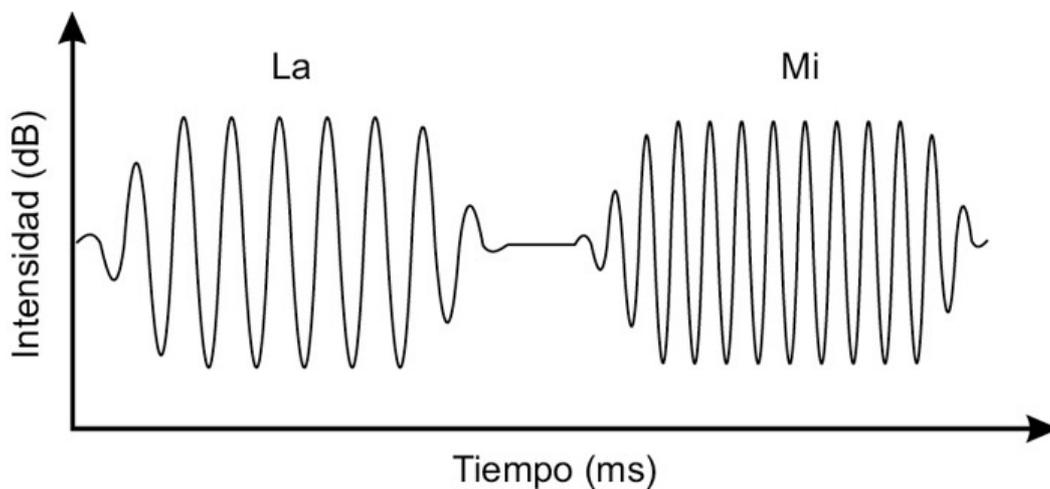
**Figura 16.** Rango de frecuencias fundamentales de un piano moderno

### Intervalos sincrónicos y diacrónicos

Los dos mecanismos que actúan en la percepción primaria de la altura, el tonotópico y el temporal, se manifiestan de manera diferente si se presentan intervalos sucesivos o *diacrónicos* o intervalos simultáneos o *sincrónicos*. Cuando los estímulos son diacrónicos, la sucesión de intervalos posee carácter melódico. La evaluación de la

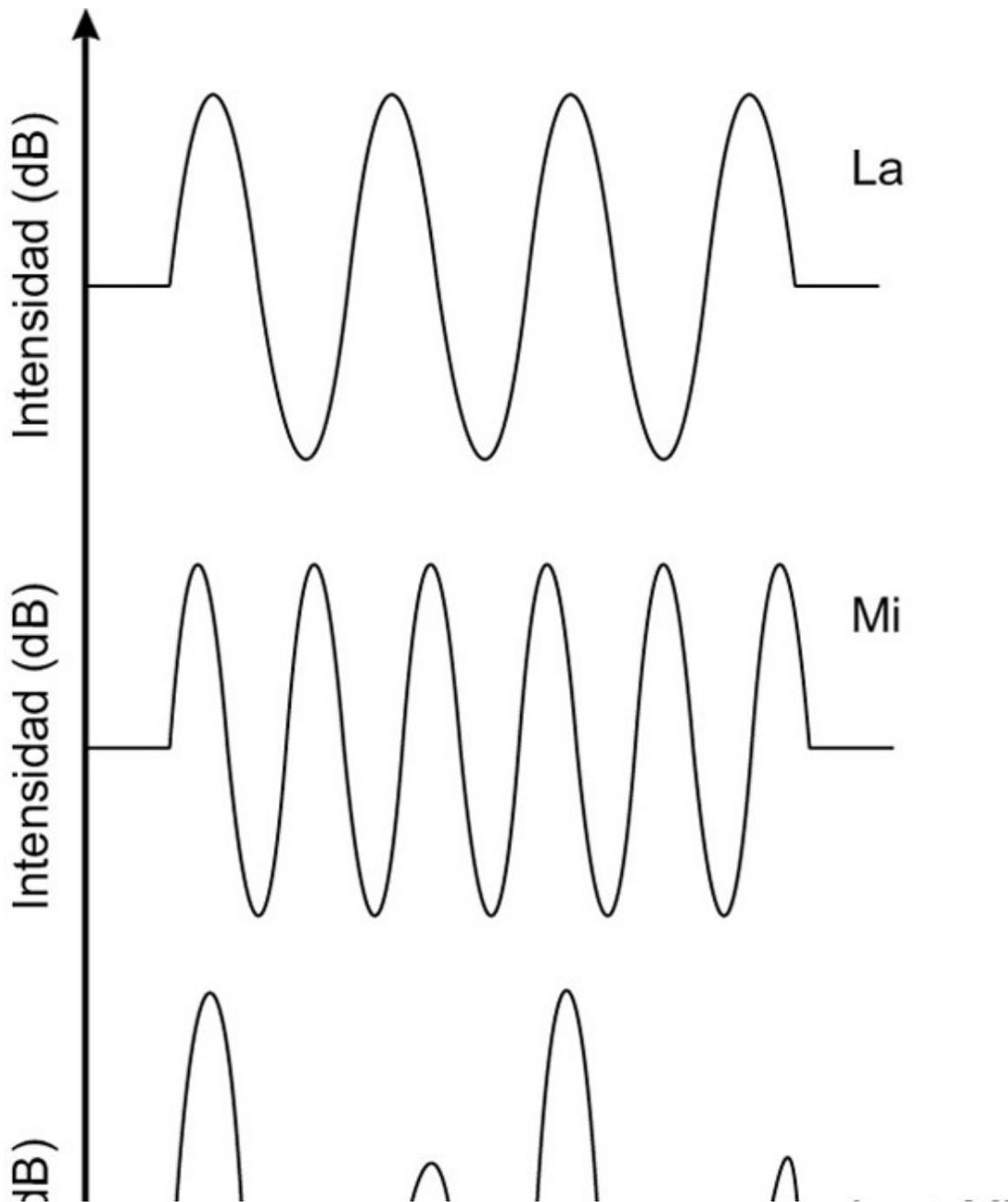
diferencia de altura entre estímulos –el intervalo diacrónico– puede derivar únicamente de la evaluación de la distancia que existe entre puntos de vibración en la membrana basilar. Si los estímulos son periódicos, con espectro armónico, se evalúa la distancia entre los dos patrones de excitación. En estos casos el mecanismo tonotópico es el único que puede activarse porque, al no actuar los dos estímulos al mismo tiempo, no es posible el análisis temporal simultáneo entre ambos. En la figura 17 se aprecia un gráfico temporal de las señales que inducen un intervalo diacrónico de quinta justa.

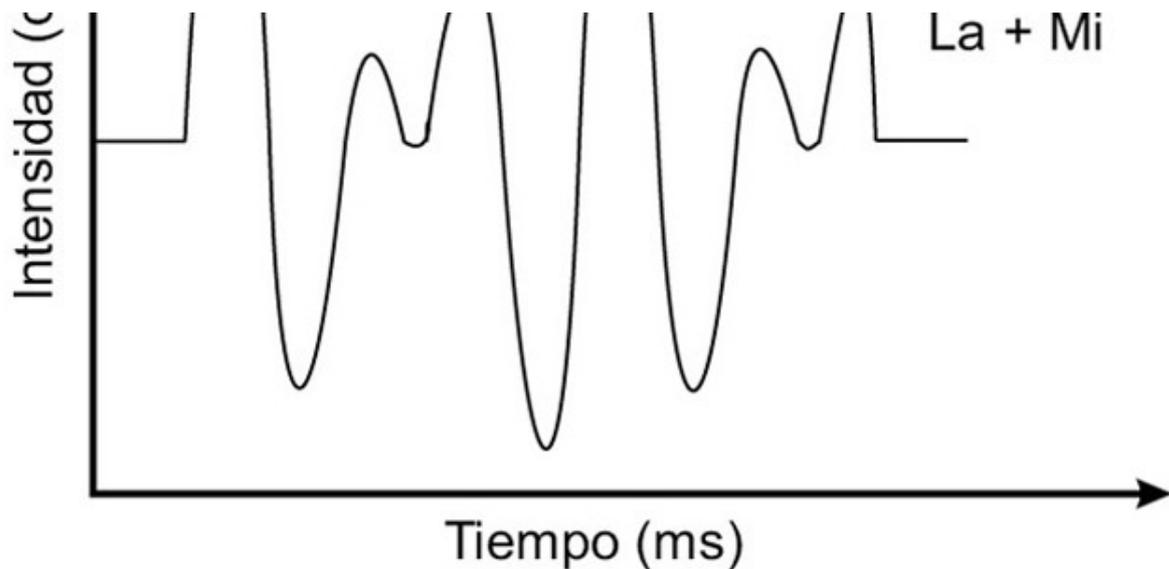
Juzgamos los intervalos sucesivos casi de manera táctil, comparando el punto de máxima excitación del segundo estímulo con el recuerdo del punto de máxima excitación de primero. En este caso, el rol de la memoria de corto término resulta determinante. Si además actúa la memoria de largo término, hecho que ocurre en alrededor del 1% de la población, se dice que el oyente posee oído absoluto. El *oído absoluto* es la habilidad de reconocer la altura tonal de un sonido sin compararla con una referencia cercana en el tiempo.



**Figura 17.** Señales asociadas a un intervalo diacrónico de quinta justa

En cambio, cuando los estímulos se presentan de manera sincrónica, la manifestación simultánea de dos o más sonidos posee carácter armónico. [16]. La evaluación de la diferencia de altura entre estímulos –el intervalo sincrónico– puede derivar en este caso tanto del mecanismo tonotópico anterior como del análisis temporal simultáneo de las dos señales. En este último caso la aparición de batidos y sonidos diferenciales prevalece como criterio de consonancia y correcta afinación, al menos en la música tonal occidental.



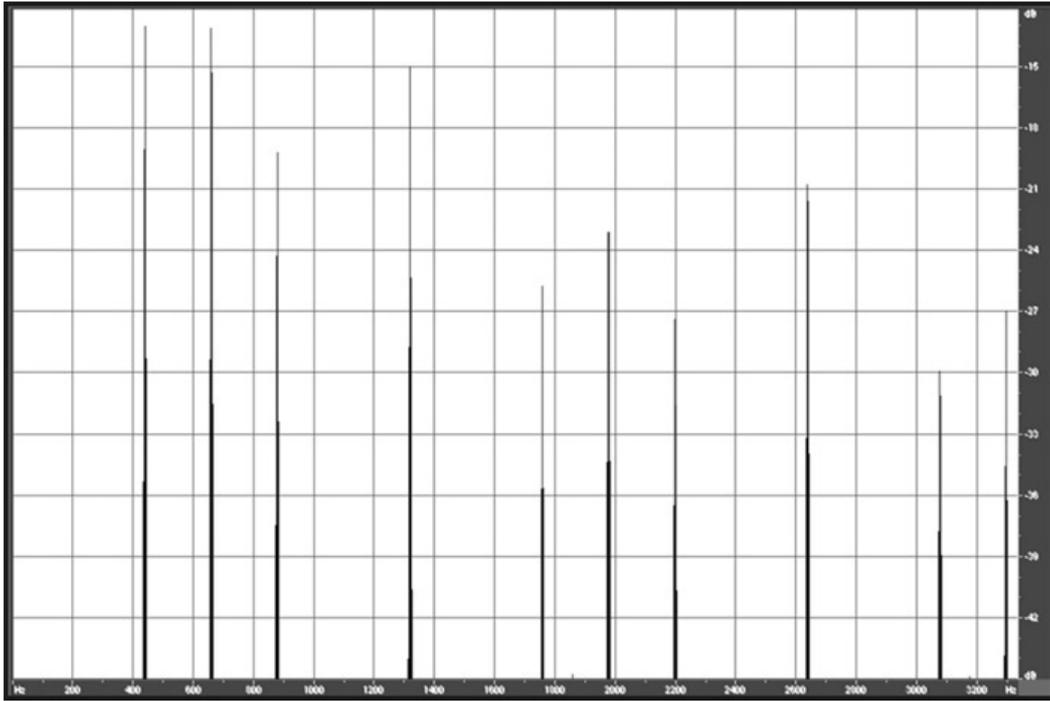


**Figura 18.** Señales asociadas a un intervalo sincrónico de quinta justa

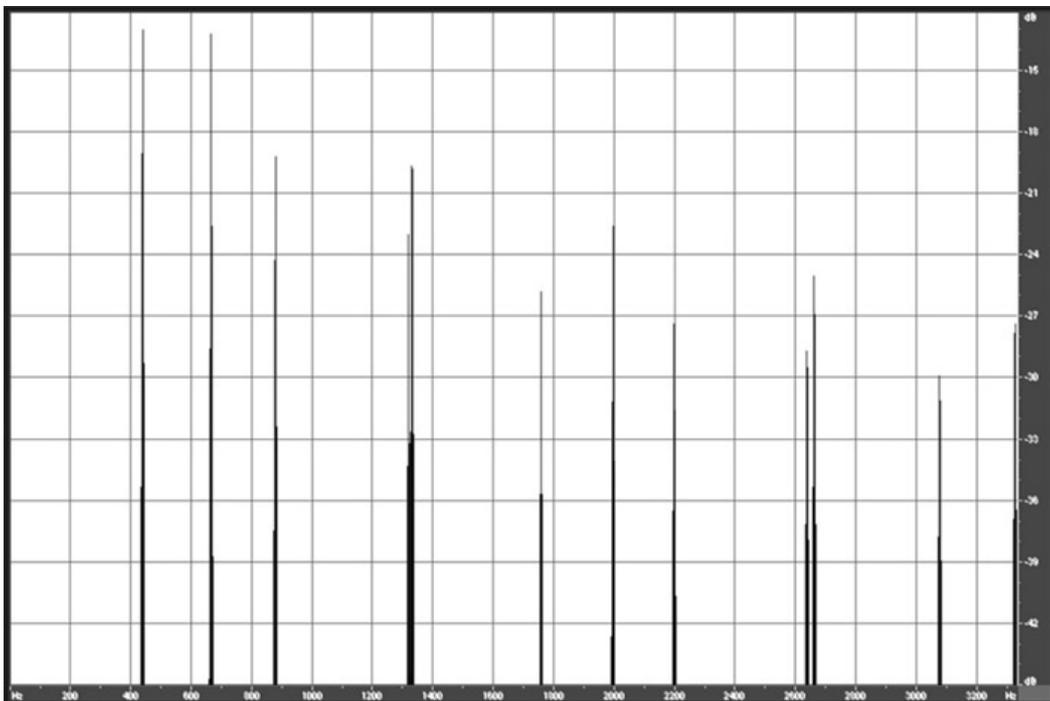
Tomemos como ejemplo un intervalo sincrónico de quinta justa. Cada vez que la señal más grave, un La<sub>4</sub> de  $f = 440$  Hz completa dos ciclos la más aguda, un Mi<sub>5</sub> de  $f' = 660$  Hz completa 3, tal como se aprecia en la figura 18.

La exacta coincidencia de fase cada dos ciclos del La<sub>4</sub> –o tres del Mi<sub>5</sub>– impide la aparición de batidos y diferenciales. El gráfico espectral del intervalo anterior es el que se muestra en la figura 19.

¿Qué ocurre si desafinamos un poco esta quinta justa, por ejemplo, aumentando la frecuencia del Mi<sub>5</sub> apenas 2 Hz para llevarlo a 662 Hz? Aparecen necesariamente batidos de primer orden que crecen en frecuencia a medida que aumenta el orden de los armónicos involucrados.



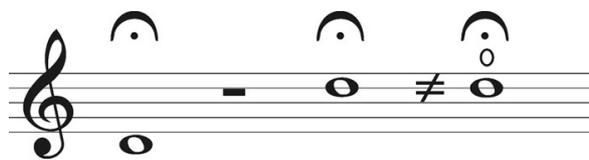
**Figura 19.** Espectro de Fourier asociado a un intervalo sincrónico de quinta justa



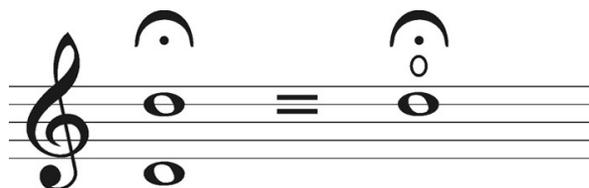
**Figura 20.** Espectro de Fourier asociado a un intervalo sincrónico de quinta justa desafinado

El batido más grave se da entre el tercer armónico de La4 (un Mi6 de  $f \cdot 3 = 1.320$  Hz) y el segundo armónico de Mi5 (otro Mi, pero esta vez de  $f' \cdot 2 = 2 \times 662$  Hz = 1.324 Hz) con un período de batido de  $fb1 = 4$  Hz. El siguiente ocurre entre el sexto armónico de La4 ( $f6 = 2.640$  Hz) y el cuarto armónico de Mi5 ( $f'4 = 2.648$  Hz). Ahora el período de batido es  $fb2 = 8$  Hz. La secuencia continúa hasta, teóricamente, alcanzar el límite del rango audible.

En la música occidental tradicional, consonancia armónica y correcta afinación de acordes significan mínimo batido. Por este motivo, para evaluar los intervalos sincrónicos prevalecen los criterios temporales –que aseguran el control de los posibles batidos– sobre los criterios tonotópicos. Hay que aclarar que en la música de muchas culturas el batido entre sonidos no es considerado un elemento negativo o disonante, sino que se lo integra como parte de la textura del sonido.



**Figura 21.** Octava diacrónica ejecutada con un violín



**Figura 22.** Octava sincrónica ejecutada con un violín

Un ejemplo práctico para demostrar la diferencia entre una octava diacrónica y otra sincrónica se puede realizar con un instrumento de arco, por ejemplo un violín. Primero se toca la tercera cuerda al aire (Re4), se hace un pequeño silencio y se toca en segunda cuerda la octava (Re5) del sonido anterior, buscando la mejor afina-

ción posible. A continuación se compara esta última nota con el segundo armónico de la tercera cuerda (Re5): el Re5 en segunda cuerda es casi siempre más agudo que el armónico.

¿Qué ocurre? La octava diacrónica fue afinada de acuerdo con la memoria auditiva y no pudo controlarse por batido. Por lo general se la afina un poco más grande que la que corresponde a la relación 2:1. A su vez, la frecuencia del segundo armónico, si la cuerda es de buena calidad, es casi exactamente el doble de la frecuencia de la cuerda a aire.

En cambio, si la octava anterior se interpreta sincrónicamente (un bicorde) es posible llevar el batido entre ambos sonidos a cero. En este caso la altura del Re5 en segunda cuerda y el armónico en tercera son de la misma altura tonal.

## **Consonancia y disonancia**

Los conceptos de consonancia y disonancia son, obviamente, culturales y dependen del contexto histórico, del género, del estilo y hasta de las características particulares del compositor de cada pieza de música. En muchos textos se leen frases del tipo: “cuando los sonidos musicales relacionados por una razón simple suenan simultáneamente el sonido es agradable” o “la disonancia puede ser explicada en parte por el batido entre componentes en la membrana basilar”, que pretenden establecer causas perceptuales primarias e independientes del marco cultural de referencia. Dichos intentos de universalizar el concepto de consonancia son rebatidos al confrontarlos con la gran cantidad de lenguajes musicales no occidentales en los cuales no se cumplen. Incluso dentro del área de la música occidental misma se crearon, a partir de fines del siglo XIX y principios del XX, sistemas que redefinen completamente la noción tradicional de consonancia.[\[17\]](#) Un estudio exhaustivo del tema necesariamente debería contemplar el desarrollo contextual e histórico de cada lenguaje musical particular. Dicho estudio no po-

drá soslayar el análisis de las combinaciones de alturas tonales, de sus campos de acción y de las jerarquías relativas que ocupan en los diferentes sistemas musicales. En otras palabras, el estudio que imaginamos tendría que encarar ampliamente un análisis escalístico completo de cada sistema musical.

## Notas

- [1](#) En la literatura en inglés se la denomina *pitch chroma*.
- [2](#) Por ejemplo, Gérard Grisey tuvo que ralentar las masas sonoras en *Partiels*, compuesta en 1975, para que se pudiera oír la evolución de cada componente (parcial) aislado.
- [3](#) Se la conoce como DLF por sus iniciales en inglés (*difference limen in frequency*) y, como tal, pertenece al conjunto de parámetros de discriminación agrupados como “mínimas diferencias perceptibles” (JND, de *just noticeable difference*).
- [4](#) FMDL (*difference limen in modulation frequency*).
- [5](#) La fundamental reconstruida corresponde al primer armónico en el desarrollo de Fourier de una señal periódica. Véase el Apéndice I.
- [6](#) Walliser (1969), en Moore (1995).
- [7](#) Terhardt llama *subarmónicos* al conjunto de las fundamentales de Fourier cuyas series armónicas incluyen el componente del estímulo en cuestión. Sus frecuencias coinciden con los submúltiplos de la frecuencia del componente. No se los debe confundir con el producto de una hipotética resonancia inferior.
- [8](#) *Morphophoric medium* en inglés.
- [9](#) El ejemplo clásico consiste en transportar un fragmento musical un cierto intervalo, por ejemplo una quinta justa ascendente.
- [10](#) Podemos citar aquí la pieza para piano *Modes de valeurs et d'intensités*, compuesta en 1949 por Olivier Messiaen (1908-1992), que emplea por primera vez la técnica serial no solamente para las alturas sino también para las duraciones, los ataques y las intensidades.
- [11](#) Estrictamente, la representación tradicional no describe alturas tonales sino las frecuencias de los estímulos asociados.
- [12](#) Como contraejemplo, Unemoto comenta que no existe una palabra en el idioma japonés para designar el intervalo de octava (Unemoto, T., 1989).
- [13](#) Drobisch (1855), en Cook, P. (1999).

- [14](#) Se puede hallar un paralelo visual en el grabado “Ascendiendo y descendiendo” de M. C. Escher.
- [15](#) Este libro, en este momento simple expresión de deseo, tendría naturalmente un lugar en la colección Música y Ciencia.
- [16](#) El término “armónico” se emplea aquí en un sentido amplio y significa ocurrencia simultánea. De manera similar, en el párrafo anterior se habla de carácter melódico para describir la ocurrencia sucesiva. Se podría trazar un correlato con las estructuras verticales y horizontales en una partitura o en un sonograma.
- [17](#) Que alguno de los sistemas experimentales que aparecieron en el siglo XX hayan sido acusados de “axiomáticos”, y hasta de “tecnocráticos”, no invalida la afirmación anterior.

# Capítulo VI

## Timbre

Aprendimos en nuestra etapa escolar que el sonido tiene por atributos la altura, la intensidad, la duración y el timbre. El timbre se refería a cierta cualidad del sonido que nos permite reconocer los instrumentos de una orquesta o distinguir entre sonidos diferentes. Pero, a diferencia de la altura, de la intensidad y de la duración, el timbre no es un atributo perceptual de descripción sencilla.

Entendemos la percepción como un proceso en el que usamos la información provista por nuestros sentidos para armar representaciones mentales del mundo que nos rodea. Una parte importante en esta tarea es decidir qué partes de la estimulación sensorial corresponden al mismo objeto o evento físico, y cuáles no. La necesidad biológica del timbre también puede ser comprendida desde una perspectiva evolutiva, a partir de la adaptación del sentido del oído para proporcionar información sobre la base de las vibraciones mecánicas del aire.

Bregman opina que el timbre surgió de la necesidad de distinguir entre dos objetos auditivos cuando éstos no pueden identificarse a partir de parámetros simples (Bregman, A., 1994). Si un rasgo auditivo se mantiene dentro de límites unidimensionales, como ocurre con la altura tonal, podemos operar hasta una cantidad máxima de cinco o seis elementos individuales a la vez. Para manejar simultáneamente una mayor cantidad de unidades de información se requieren más dimensiones y el timbre aportaría esas dimensiones adicionales.

Según Shepard, el sistema perceptual de los animales nace con la incorporación sensorial de las regularidades propias del ambiente en el cual han evolucionado.<sup>[1]</sup> Por ejemplo, somos muy sensibles a ciertos aspectos del sonido que no se modifican durante el recorrido entre la fuente y el oyente, como la altura tonal. Por

el contrario, somos prácticamente insensibles a las características que varían de manera aleatoria con el contexto, como la fase y la intensidad relativa de los componentes en una señal compleja. De acuerdo con esta concepción, sostenida por gran cantidad de investigadores, los parámetros “fuertes” que determinan el timbre de los sonidos estarían relacionados con las características de la señal acústica que resisten la distorsión propia del medio que nos rodea.

## 1. Definiciones de timbre

El primer problema que aparece al estudiar el concepto de timbre es la vaguedad y ambigüedad del término. Repasemos las dos definiciones más utilizadas.

1. Una manera clásica de definir timbre es asociándolo a las fuentes acústicas de origen. En música, los nombres de los instrumentos se usan como etiquetas para denotar los timbres, por ejemplo “timbre de flauta” o “timbre de violín”. Si uno desea comunicar la impresión general dada por un sonido, este método de nombrarlo puede ser adecuado. Las etiquetas nominales actúan de modo análogo a los nombres de los colores: el término “rojo” es suficiente para pedir un tomate maduro en una verdulería, pero resulta demasiado amplio y vago para describir los miles de “rojos” que se usan en la industria gráfica. Del mismo modo, hablar de “sonido de flauta” no nos sirve para diferenciar entre los sonidos de dos flautas diferentes, ni para distinguir los sonidos que obtienen dos flautistas del mismo instrumento. Como ejemplo, podemos mencionar la mutación que ocurrió con el concepto de ruido blanco: en principio describe una característica físico-matemática de una señal acústica de espectro plano, pero se convirtió entre los músicos electroacústicos en una nueva unidad cultural, el “ruido blanco” como un timbre más de su repertorio.

Como veremos más adelante, el uso del timbre a manera de etiqueta implica el reconocimiento del sonido sobre la base de un es-

quema previo en el que la memoria auditiva juega un papel central.

2. La definición académica de timbre se puede encontrar, por ejemplo, en el apartado sobre terminología acústica del ASA. Es repetida hasta el cansancio en multitud de normas y textos, y dice menos de lo que no dice: “timbre es el atributo de la sensación auditiva en términos del cual un oyente puede juzgar que dos sonidos similares, con igual sonoridad y altura tonal, son diferentes”.[\[2\]](#) Ésta es, quizá, la definición menos comprometida de toda la historia de la acústica. Hasta se podría dudar de su competencia en tanto definición. Implica, por ejemplo, que los sonidos a los que se le puede asignar timbre poseen altura tonal; luego, los sonidos no tónicos –como los de un platillo o de un tambor–, ¿no poseen timbre! Sólo nos quedan dos posibilidades: desistimos de estudiar más de la mitad de los sonidos que nos interesan, o buscamos otra definición. Como veremos más adelante, este enunciado está profundamente influido por el modelo acústico clásico de timbre.

Además, ¿qué significa “dos sonidos similares”? ¿Alcanza esta semejanza a la ubicación espacial, o a la tasa de repetición si existiera? ¿Difiere el timbre de un registro estereofónico de una trompeta si uno intercambia los canales derecho e izquierdo? ¿Es distinto el timbre de dos notas de una flauta del timbre de tres notas del mismo instrumento? Son demasiadas preguntas sin respuesta, y quizá convenga aquí repetir la opinión de Bregman sobre el punto: “Creo que la definición de timbre de la ASA debería ser la siguiente: ‘No sabemos cómo definir el timbre, pero no es ni la sonoridad ni la altura tonal’” (Bregman, A., 1994).[\[3\]](#)

Una posible explicación del lugar especial que ocupan la sonoridad y la altura tonal en la definición académica es que ambos parámetros son unidimensionales y de fácil control en los experimentos de acústica y psicoacústica. Otro motivo es sin duda el lugar preponderante que ocupó la altura tonal en las investigaciones

acústicas aplicadas tanto a la música occidental como a la no occidental, al menos hasta la década de 1980.

Sin duda, el timbre es más que un simple catálogo de sonidos o que lo que queda de un sonido al quitarle la sonoridad y la altura tonal. Necesitamos, por lo visto, una nueva definición. Sería conveniente que esta nueva definición contemple una descripción sencilla de los elementos que usamos para percibir el timbre y que, en lo posible, se relacione claramente con las señales acústicas que lo causan. Hace años que los investigadores en acústica y los músicos electroacústicos trabajan para desarrollar una colección ordenada de términos tímbricos descriptivos, aunque con resultados inciertos.

Las definiciones tradicionales de timbre presentan otra dificultad. Un evento físico puede comprender más de una ocurrencia: los pasos de una persona, la lluvia sobre una chapa, los sonidos granulados que se usan en electroacústica, varios instrumentos de música tocando una secuencia isorrítmica, etc. Todos estos ejemplos se oyen como unidades acabadas pues forman parte de una representación perceptual coherente, pero no se adaptan con facilidad a la noción limitada de *sonido* que aparece en las definiciones de timbre. El concepto de *objeto auditivo*, que analizaremos en detalle en el capítulo VIII y que comprende la fusión de las diferentes características de la señal de entrada en una única imagen coherente, resulta mucho más adecuado en estos casos.

## **2. Estudios y modelos explicativos sobre el timbre**

El timbre de un sonido es un rasgo perceptual de naturaleza diferente a la de los otros rasgos ya analizados, como la altura tonal o la sonoridad. Por ejemplo, no es posible ordenar los sonidos de menor a mayor en una sola dimensión tímbrica: el timbre es, en consecuencia, un rasgo multidimensional.

Durante los últimos doscientos años se estudió con cierto detalle el timbre de los sonidos. Se realizaron gran cantidad de estudios experimentales y se propusieron una pluralidad de modelos explicativos, tanto desde el punto de vista de la señal acústica como del de las representaciones psíquicas de los sonidos. Casi todas las experiencias se dedicaron al estudio del timbre en los instrumentos de música. En la mayoría de los estudios el timbre no es considerado como un objeto de estudio en sí mismo, sino como una característica perceptiva que permite identificar una categoría particular de fuentes sonoras.

Las señales preferidas en los estudios de laboratorio son mayoritariamente estacionarias en el tiempo –sinusoides o combinaciones de sinusoides–, pues son fáciles de producir, de controlar y de medir. Aunque las señales estacionarias son muy raras en la naturaleza, los cantos y gritos de muchos animales se les pueden aproximar por tramos. Los sonidos generados por algunos instrumentos de música también poseen segmentos que se pueden equiparar a ondas estacionarias. Otros grupos de estudio emplean pulsos de ruido, blanco o rosa, en vez de señales estacionarias deterministas. De nuevo la elección recae sobre un artefacto técnicamente simple. Aunque los pulsos de ruido son más habituales en la naturaleza que las señales estacionarias, ocurren muy raramente en estado puro. Como sabemos, el universo acústico que nos rodea es muy complejo y todo intento de simplificación implica riesgos reconocidos. Muchos de los desvíos y desajustes que aparecen entre los modelos acústicos sobre el timbre y la realidad ambiental –o musical– tienen su origen en una extrapolación incorrecta de los resultados de los experimentos de laboratorio a situaciones acústicas reales.

Al margen de las herramientas metodológicas adoptadas, la mayoría de los modelos y teorías sobre el timbre intentan vincular ciertos parámetros de la señal acústica con las consecuencias perceptuales asociadas, comprender nuestra capacidad para asignarle unidad a los diferentes sonidos que provienen (o parecen provenir)

de la misma fuente acústica y descubrir las propiedades que facilitan la categorización tímbrica de los objetos auditivos.

En lo que sigue presentaremos algunos de los modelos explicativos más destacados sobre el timbre. Al final de esta parte resumiremos las características que, a nuestro juicio, sintetizan el estado actual de la cuestión.

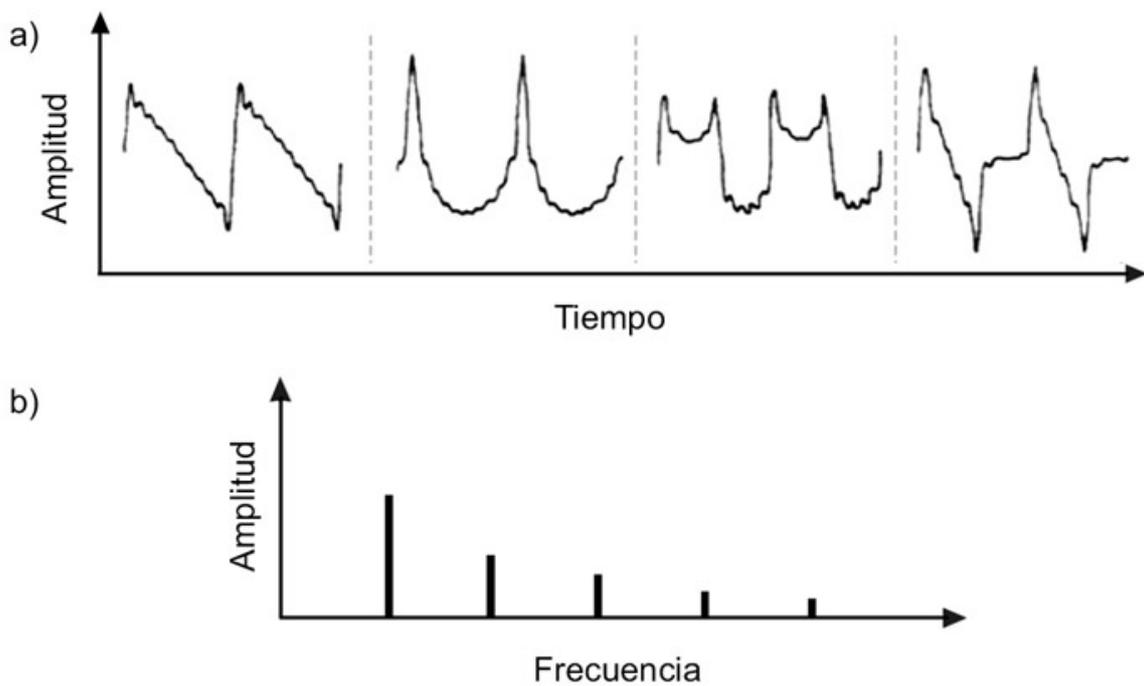
## **El modelo clásico**

El modelo clásico empleado para explicar el timbre –al menos en los instrumentos de música– establecía que, así como la frecuencia fundamental determina la altura percibida y la intensidad física la sonoridad, el timbre tiene origen en la amplitud y distribución del espectro promedio de la señal acústica. El modelo estaba basado en la ley acústica de Ohm, enunciada en 1843, según la cual el oído es capaz de discriminar los armónicos que componen una señal periódica, pero no puede percibir las diferencias de fase entre los componentes. En la figura 1 se pueden ver los oscilogramas de cuatro señales periódicas con igual espectro de potencias y diferentes fases relativas entre armónicos. Si pudiéramos oírlas no encontraríamos diferencias apreciables.

De modo más preciso, la ley acústica de Ohm establece que un oyente no puede distinguir entre dos señales con la misma distribución de frecuencias y amplitudes en los armónicos, pero con una distribución de fases diferente. En este modelo el timbre depende únicamente del espectro de potencias de la onda, que es sólo una parte del análisis completo de Fourier (Basso, G., 2001a).

Según la tradición, el defensor más destacado del modelo clásico fue Herman von Helmholtz. Sin embargo, Helmholtz tuvo la precaución de limitar su alcance a las señales estacionarias: “Ciertas características particulares de los sonidos de varios instrumentos dependen del modo que comienzan y cesan [...]”, de manera que estudió solamente “[...] las peculiaridades de los sonidos musicales

que continúan uniformemente”.[\[4\]](#) Los seguidores del modelo –casi todos los físicos, acústicos y músicos hasta bien entrada la década de 1960–, no tuvieron la prudencia de Helmholtz y proclamaron resuelto el asunto.<sup>5</sup>[\[5\]](#) A lo sumo, reconocían que la teoría estaba limitada a las ondas estacionarias periódicas que, según ellos, componían la mayor parte de las señales de los instrumentos de música.

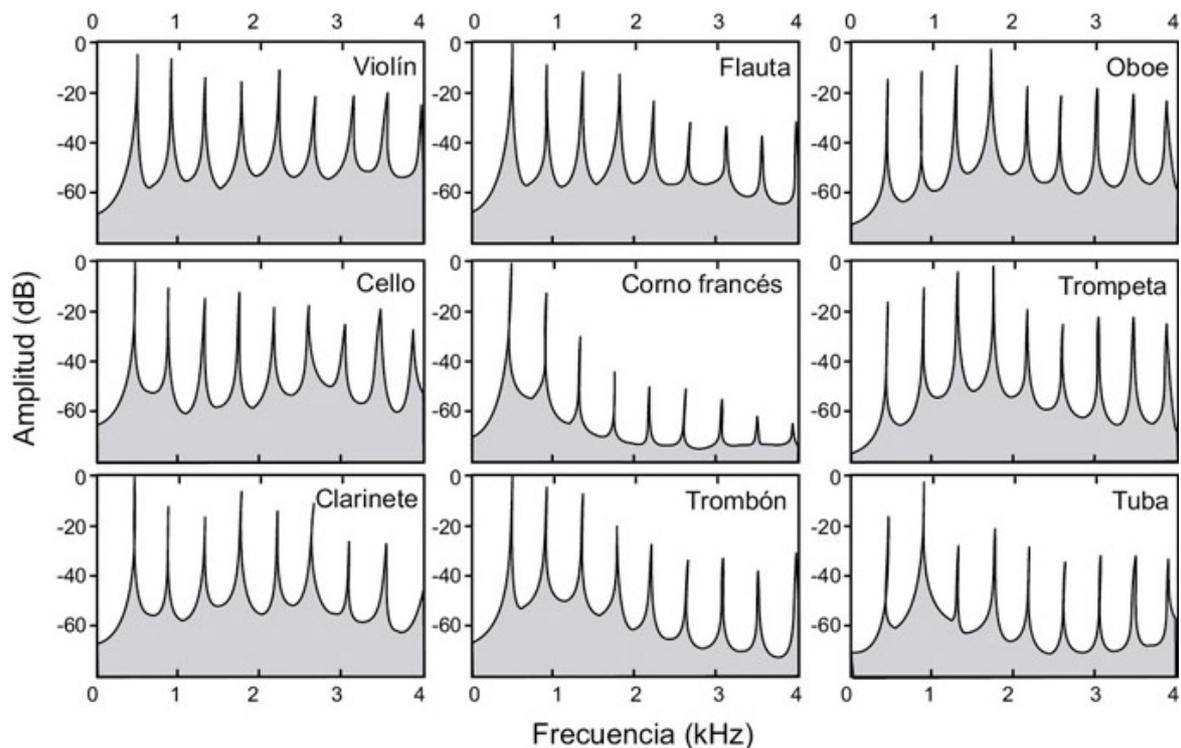


**Figura 1.** a) Oscilogramas de cuatro señales periódicas con igual espectro de potencias y diferentes fases relativas; b) espectro de potencias de las señales anteriores

Llevar esta teoría a la realidad no fue tarea simple. Se sabía desde principios del siglo XX que los espectros de los sonidos varían en el tiempo, así que la pregunta central fue: ¿cuál de todos los espectros sucesivos debía tomarse para determinar el timbre del sonido? La respuesta la dio Hall en 1937 al establecer que se debían promediar temporalmente los espectros, y que el resultado de este promedio era el responsable del timbre.[\[6\]](#) Un ejemplo representa-

tivo de estudio experimental basado en el modelo clásico se puede apreciar en la figura 2.

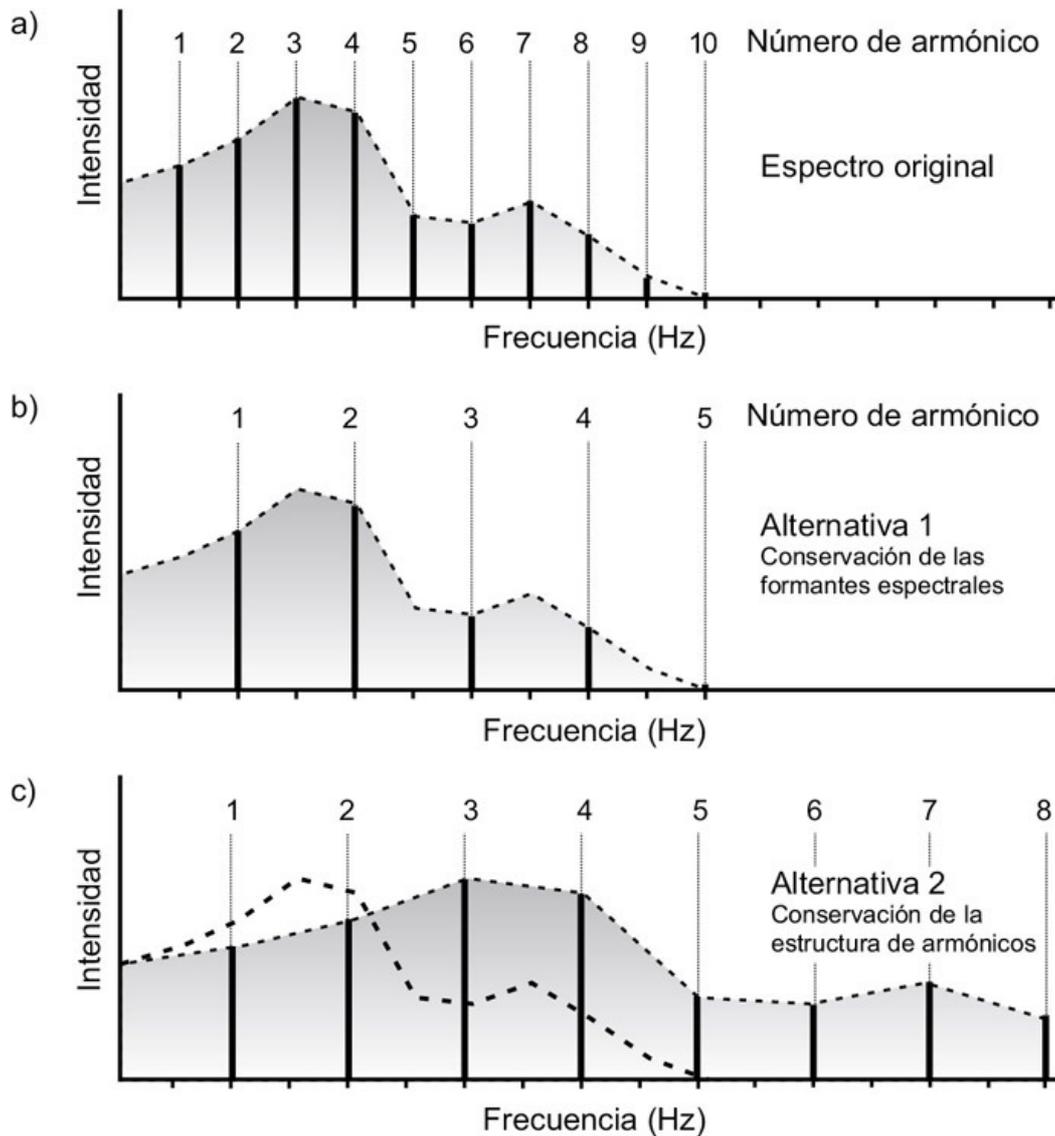
Si el timbre es consecuencia de la distribución de los armónicos o, expuesto de otra manera, de la forma de onda estacionaria, una señal cuya forma se mantiene no debería sufrir modificaciones tímbricas. Pero uno de los experimentos acústicos más simples contradice esta afirmación.[\[7\]](#) La forma de onda de una senoide no cambia al cambiar su frecuencia, pero si se genera un glissando que recorra buena parte del rango audible pocas personas dirán que el timbre permanece inmutable. En realidad, a bajas frecuencias la senoide suena blanda y oscura, mientras que a altas frecuencias parece dura y brillante.[\[8\]](#) ¡La senoide, el “ladrillo” con el que se pretendía construir el edificio del timbre poseía, a su vez, un timbre particular! O peor aún, mutaba de timbre a medida que cambiaba de frecuencia. La conclusión del experimento es categórica: no existe una relación uno a uno entre forma de onda y timbre. Uno de los pilares del modelo clásico se desmoronaba.



**Figura 2.** Espectros de sonidos de diferentes instrumentos según Wedin y Goude (Handel, S., 1995)

Otra cuestión que generó un intenso debate fue la de la invariancia del timbre ante un cambio de frecuencia. Para que el timbre no varíe, ¿deben conservarse las relaciones de amplitud entre armónicos para preservar la forma de onda escalada, o deben conservarse las posiciones absolutas de la envolvente temporal? El modelo clásico sostenía claramente la primera opción. En un célebre trabajo, Slawson demostró que el modelo también fracasaba en ese punto (Slawson, W., 1968). En la figura 3 se pueden ver las dos opciones. Los oyentes juzgaron que la alternativa 1 conserva mejor el timbre que la alternativa 2 –defendida por el modelo clásico.

Durante el año 1926, en pleno auge de la teoría clásica, Metfessel publicó un artículo en el que ponía en duda la simplista proposición “forma de onda estacionaria igual timbre” (Butler, D., 1992). La inversión temporal de una señal de audio –posible desde la aparición del registro magnético durante la segunda guerra mundial–, en la que se mantiene sin cambios el espectro de potencias promedio pero se alteran las fases relativas, demostró que había algo más que distribución espectral en el tema del timbre (cualquiera puede comprobar el cambio dramático que se oye al invertir temporalmente un registro de piano, por ejemplo mediante el uso de software de audio en una computadora).



**Figura 3.** Experimento de Slawson sobre la invariancia del timbre ante un cambio de frecuencia, a) espectro original, b) alternativa 1: frecuencias fijas o conservación de las formantes espectrales; c) alternativa 2: frecuencias relativas o conservación de la estructura de armónicos

¿Por qué continuamos reconociendo la fuente acústica al oír música en una pequeña radio, o cuando oímos un instrumento en una sala con una respuesta en frecuencia irregular? En ambos casos el espectro de potencias original se altera profundamente y también

debería alterarse, de acuerdo con el modelo clásico, el timbre que percibimos.

Hemos examinado varias de las dificultades que encontró la teoría clásica del timbre desde su enunciación a mediados del siglo XIX. Aunque no daba cuenta de numerosos hechos, seguía siendo considerada correcta por la mayoría de los especialistas, quienes suponían que solamente restaba realizar algunos ajustes. Esta situación se mantuvo hasta la década de 1960, cuando la aparición de la síntesis acústica por computadora terminaría definitivamente con el modelo. Como sabemos, un modelo sirve para explicar hechos conocidos y para predecir o crear otros nunca experimentados. Durante el siglo pasado se potenció la necesidad de aplicar las nuevas tecnologías a la construcción y a la simulación de instrumentos musicales. La llegada de la energía eléctrica primero, del control electrónico después y, por último, de la síntesis digital, permitió gradualmente la construcción de sonidos a partir de la descripción física de las señales asociadas a tales sonidos. El mayor esfuerzo tecnológico se apoyó en los tratados de acústica que repetían, obviamente, el modelo clásico. En la mayoría de los casos el producto de la síntesis instrumental se parecía muy poco al instrumento original. Los sonidos resultaban apagados, carentes de identidad y de interés.<sup>[9]</sup> La conclusión fue entonces evidente: la descripción física de la causa del timbre era inadecuada, pues fallaba en el decisivo test de la síntesis acústica. Este fracaso puede describirse como una auténtica “catástrofe del timbre” de la acústica del siglo XX. El modelo clásico debía ser reemplazado por otro.

## **Análisis multidimensional**

La falla principal del modelo clásico fue, sin duda, que en su formulación soslayaba completamente la importancia de los factores temporales. Helmholtz ya había sospechado que el ataque y la extinción de ciertos sonidos contenían información relevante para la

determinación del timbre. Y resultaba evidente que la correspondencia entre los datos físicos y los rasgos percibidos no era simple y lineal como, por ejemplo, en la relación frecuencia-altura.

Para no repetir el error cometido durante el apogeo de la teoría clásica, en el que se intentó imponer un concepto físico para explicar la percepción, durante la década de 1960 se idearon y realizaron gran cantidad de experiencias psicoacústicas. Las motivaban la necesidad de encontrar los datos físicos empleados efectivamente en la percepción del timbre, comprender la estructura de las representaciones psíquicas del timbre de un sonido y hallar una correspondencia entre datos físicos y representaciones mentales.

Aunque parece haber una gran cantidad de maneras de describir la forma en que un sonido se distingue de otro, la psicoacústica necesita encontrar las más sencillas para simplificar el problema. Es interesante aquí comparar el estudio del timbre con el del color. Un color está determinado físicamente por un espectro de luz característico. Existen infinidad de espectros posibles y podemos percibir miles de colores. Sin embargo, todos los colores pueden describirse con sólo tres dimensiones: dos de ellas definen la cromaticidad y la otra el brillo. ¿Es posible describir las diferencias de timbre entre los sonidos a partir de un número pequeño de dimensiones, como ocurre con el color? Para responder a esta pregunta se emplea el *método del análisis multidimensional*, mediante el cual se procesan estadísticamente los juicios de los oyentes sobre el grado de similitud entre dos sonidos elegidos de un conjunto restringido. La hipótesis de trabajo del método afirma que existe un número limitado de dimensiones capaces de describir la totalidad de los timbres posibles. Si dos sonidos tienen un valor cercano en una dimensión, son similares en esa dimensión aunque no lo sean en las otras. Bregman lo explica de la siguiente manera: “Cuando comenzamos a pensar en las similitudes entre cosas, inmediatamente nos damos cuenta que una cosa puede parecerse a otra de maneras diferentes. Una pelota se parece a algunos juguetes en su

redondez, a otros en su color rojo, a otros en su tamaño, [...] el hecho de que algo pueda parecerse a otras cosas de diferentes maneras lleva naturalmente a la idea de la existencia de dimensiones de similitud”.[\[10\]](#)

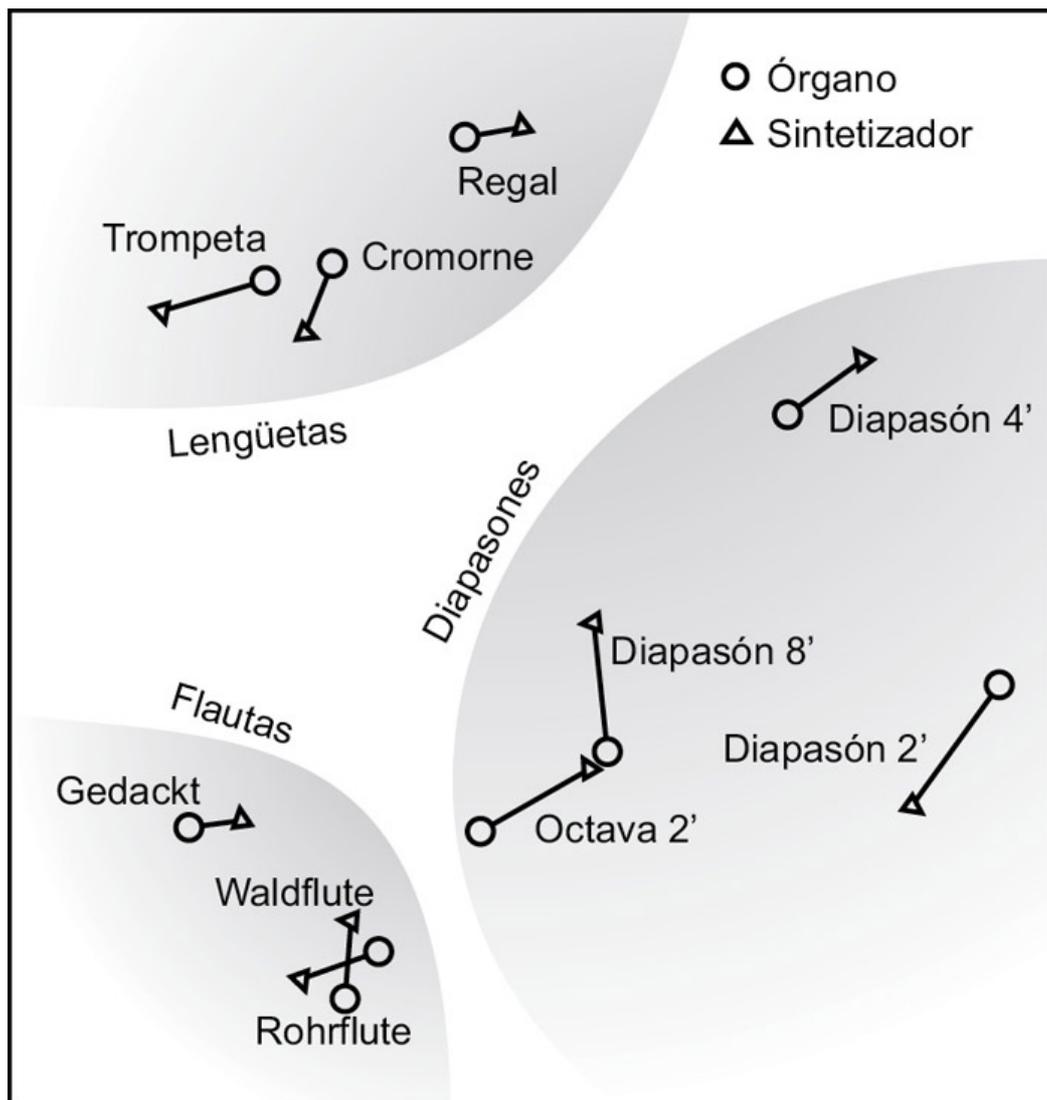
A continuación describiremos tres destacados estudios de análisis multidimensional sobre el timbre.

*Estudios de Plomp.* En 1967 Reinier Plomp descubrió que las diferencias tímbricas se correlacionaban bien con diferencias en el espectro cuando se consideraban los niveles de 18 bandas de tercio de octava, apenas mayores que las bandas críticas. Dicho de otra forma, el timbre se relacionaría con el patrón espectral de excitación del sonido. Como hay 37 bandas críticas cubriendo todo el rango audible habría que considerar otras tantas dimensiones. Para corroborar estas conjeturas decidió realizar un estudio sistemático. Los primeros trabajos de evaluación multidimensional fueron realizados entre 1960 y 1976 (Plomp, R., 1967, 1976). Empleó como datos de partida sonidos de violín, viola, violoncello, oboe, clarinete, fagot, trompeta, corno francés y trombón. Determinó que los juicios sobre diferencias entre los sonidos podían resolverse con sólo tres dimensiones perceptuales.

En una segunda etapa intentó detectar la clase de información acústica usada para determinar el timbre. Analizó físicamente las señales de los instrumentos filtrándolas en 15 bandas de tercio de octava, siguiendo aproximadamente la distribución de las bandas críticas en el oído. A partir de la conjetura de que las 15 bandas no eran totalmente independientes unas de otras, redujo toda la información espectral a cuatro dimensiones básicas. La intención de Plomp era rotar y acomodar estas cuatro dimensiones físicas para ajustarlas a las tres dimensiones perceptuales derivadas del análisis de diferencias en timbre. La hipótesis principal establecía que los oyentes emplean la correlación entre diferencias espectrales para percibir el timbre de los sonidos. Aunque Plomp permite la

aparición de nuevas dimensiones perceptuales en casos particulares (por ejemplo, la “tasa de ruido” es decisiva para evaluar el sonido de un cascabel), sugiere que siempre se reducen en última instancia a no más de tres. Simplemente, cada sonido polariza el campo perceptual de acuerdo a sus características particulares.

Aunque el modelo de Plomp introduce el análisis multidimensional en el estudio de la percepción del timbre, sigue fiel al modelo espectral estático de la teoría clásica. Quizá por eso su mejor resultado se ha obtenido con sonidos cuyo cuerpo es muy estable en el tiempo, como lo son los que generan los tubos de órgano. En la figura 4 se aprecia un diagrama bidimensional en el que se comparan los juicios sobre los sonidos de los tubos reales con sus simulaciones espectrales. La dimensión vertical describe la cantidad de armónicos superiores (de menor a mayor cantidad), mientras que la horizontal es producto de una combinación compleja de propiedades espectrales.



**Figura 4.** Representación bidimensional del timbre de sonidos de órgano (círculos) versus sonidos de órgano sintetizados (triángulos) (Rasch y Plomp, 1999)

Una de las críticas más importantes que se le han hecho a este tipo de análisis es que depende fuertemente del conjunto de sonidos elegidos. En cierta medida, para cada conjunto de datos se obtienen resultados diferentes. Además, las diferencias en el timbre “estático” de Plomp no son suficientes para definir un objeto auditivo, en parte porque la señal acústica puede ser alterada espectralmente por las reflexiones en una sala. En la época de los trabajos citados se sabía que el reconocimiento de los instrumentos musicales,

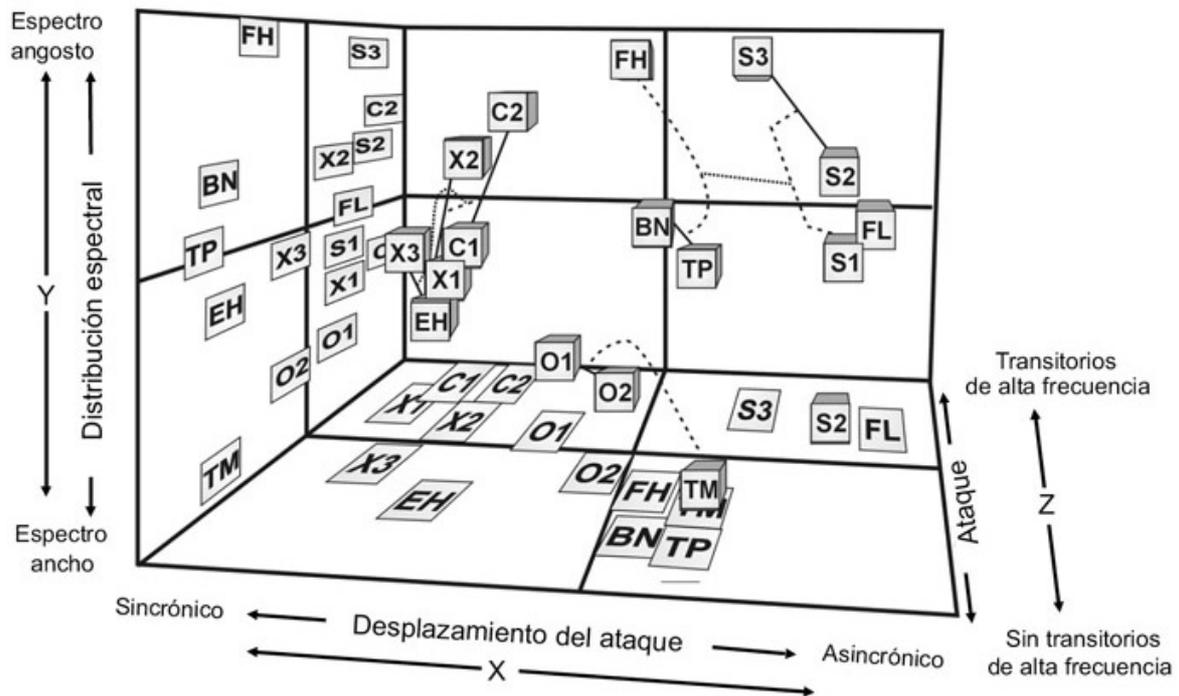
por ejemplo, depende con fuerza del ataque y de la estructura temporal de la envolvente dinámica del sonido.[\[11\]](#) Sin embargo, Plomp no incorporó factores temporales en su modelo.

*Modelo de Grey.* Entre 1975 y 1979 John Grey realizó una serie de experiencias tendientes a construir un espacio tímbrico vectorial, similar al que se emplea para describir los colores. En la primera de ellas empleó dieciséis sonidos de la misma altura, intensidad y duración, provenientes de 12 instrumentos tónicos (es decir, de altura tonal definida). Digitalizó las señales, las simplificó reemplazando las microvariaciones temporales por secciones rectilíneas, y se las presentó de a pares a veinte oyentes. Los sujetos debían cuantificar la similitud entre los sonidos de cada par mediante el uso de una escala graduada de uno a treinta. Grey interpretó que las distancias psicológicas se podían correlacionar negativamente con las similitudes percibidas. En otras palabras, a menor similitud, mayor distancia tímbrica. Los resultados de esta experiencia fueron graficados en el espacio vectorial tridimensional de la figura 5. Cada dimensión, aunque obtenida por análisis multidimensional, representa los atributos físicos que Grey consideró mejor apareados con las respuestas perceptuales obtenidas (Grey, J., 1975).

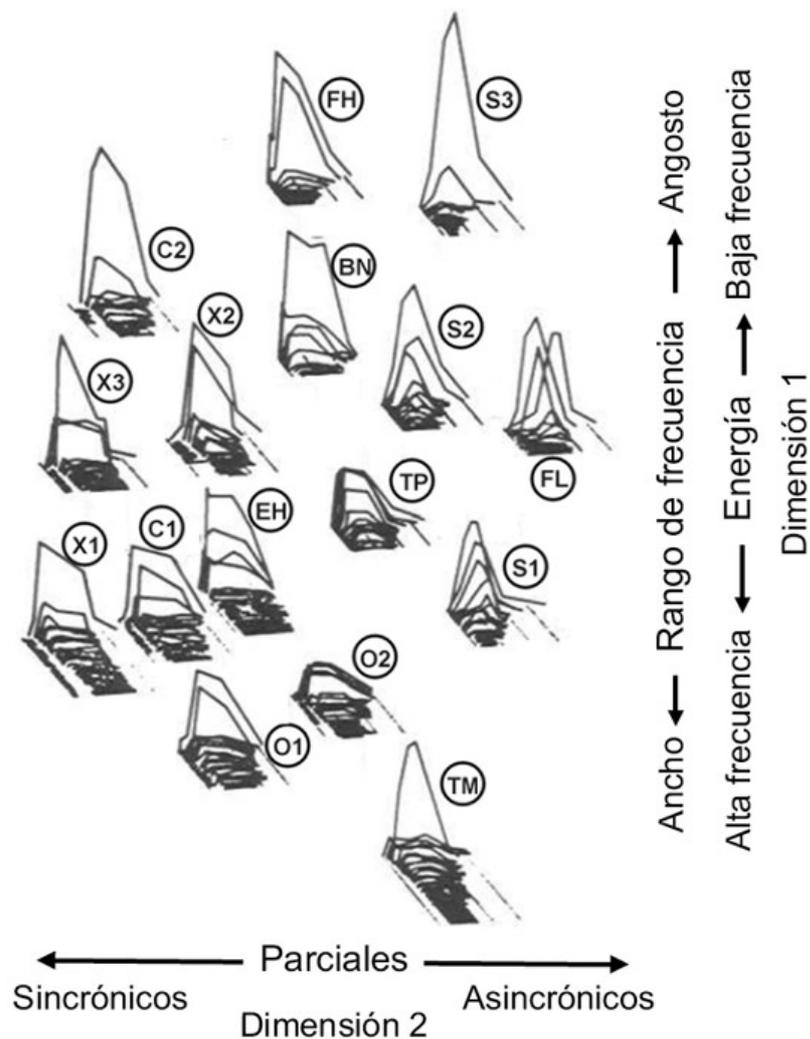
La dimensión vertical (Y) representa la distribución de la energía en los parciales armónicos de la señal. Se trata principalmente de una *dimensión espectral*. Se puede pensar en ella como representando el grado de “brillo” del sonido, y en los análisis tradicionales queda definida por la envolvente espectral. La dimensión horizontal (X) representa en parte el grado de fluctuación de la envolvente espectral durante el desarrollo del sonido y en parte el grado de sincronismo en el ataque de los diferentes armónicos. Es una *dimensión espectro-temporal*, denominada “flujo espectral” por Krumhansl (McAdams y Bigand, 1994). Permite distinguir entre sonidos de evolución estática y sonidos de evolución dinámica. La tercera dimensión (Z) representa la presencia de vibraciones inar-

mónicas previas a la parte armónica principal del sonido. Se puede hablar de una “calidad de ataque” y es esencialmente una *dimensión temporal*.

En una publicación de 1977, Grey ilustró dos de las proyecciones bidimensionales de su espacio tridimensional con los diagramas tiempo-energía-frecuencia (TEF) de las señales de los instrumentos (figura 6).



**Figura 5.** Representación tridimensional de similitudes tímbricas. Oboes (O1, O2); corno inglés (EH); fagot (BN); clarinete alto en Mib (C1); clarinete bajo en Sib (C2); saxofón alto mf (X1); saxofón alto p (X2); saxofón soprano en Sib (X3); flauta (FL); trompeta (TP); corno francés (FH); trombón con sordina (TM), violoncello *sul ponticello* (S1), normal (S2) y *sul tasto* (S3) (Grey, J., 1975)



**Figura 6.** Ejemplo de una proyección bidimensional del espacio de timbres (los instrumentos son los mismos de la figura 5) (Grey, J., 1977)

Las conclusiones que el autor extrajo de este estudio, y de otros similares, fueron muy valiosas. En primer lugar, constató que la representación mental del conjunto de timbres es relativamente estable y más o menos idéntica entre un sujeto y otro. Se comprobó que un cambio en una dimensión física particular (por ejemplo en la envolvente espectral) provocaba un desplazamiento en la dimensión perceptiva correspondiente (“brillo” en este caso). En ciertas

ocasiones se producían también desplazamientos en otras dimensiones. Quizá la conclusión más controvertida esté relacionada con la continuidad de las transiciones tímbricas: “No hay fronteras perceptuales marcadas en la identificación entre dos sonidos interpolados [...] la ausencia de límites abruptos y la sensibilidad al contexto sugiere que, definidas estrictamente, las transiciones no se perciben categorialmente”.[\[12\]](#)

De acuerdo con las experiencias de Grey, es probable que existan distintas estrategias perceptuales asociadas a la discriminación tímbrica, que se activan en función de la característica particular más pronunciada –temporal o espectral– del sonido en cuestión. Quizá estas diferencias en las respuestas estén fundamentadas en la polarización de la atención hacia lo que los oyentes seleccionan, en un proceso preconsciente, como la característica tímbrica distintiva en cada caso.

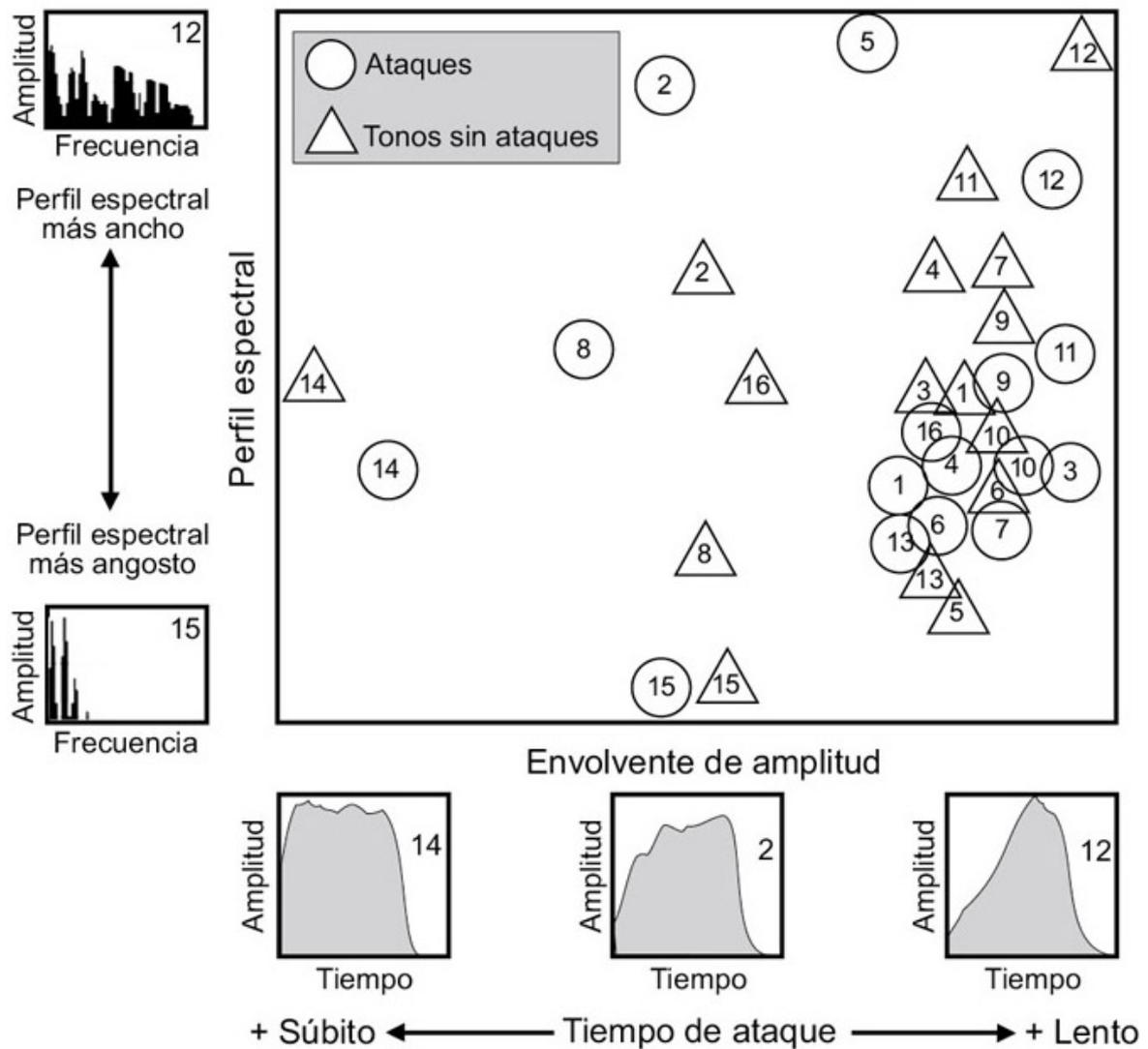
Se puede poner en duda la hipótesis de Grey según la cual sonidos extremadamente complejos no difieren entre sí más que en ciertas dimensiones subyacentes comunes. Es posible que algunos timbres posean características únicas no codificables en dimensiones continuas. Un buen ejemplo se encuentra en la preeminencia de armónicos impares en el clarinete, que contribuye a su identificación más que otros datos acústicos específicos como la envolvente de amplitud. En cambio, en los sonidos en los que la envolvente de amplitud posee características destacadas, como en el trombón o la flauta, ésta prima en la identificación por sobre la estructura armónica.

Con todo, la más profunda de las objeciones a los modelos basados en espacios vectoriales ataca directamente su naturaleza: un espacio vectorial es esencialmente lineal. Sabemos, sin embargo, que en la percepción del timbre ocurren grandes saltos y bifurcaciones marcadamente alineales. Un modelo como el de Grey quizá sirva solamente para describir regiones tímbricas acotadas, con

comportamientos lineales locales, ocupadas por sonidos de características afines.

*Estudios de Iverson y Krumhansl.* Iverson y Krumhansl analizaron la percepción del timbre empleando sonidos de instrumentos musicales tónicos en tres condiciones diferentes: señales completas, sólo los primeros milisegundos de las señales (ataques) y señales completas menos los ataques (es decir, el cuerpo y la extinción) (Iverson y Krumhansl, 1993). Como en el estudio de Grey, los sujetos juzgaron la similitud de cada par de instrumentos tomados de un total de 16. Los juicios de similitud fueron notablemente consistentes en las tres condiciones impuestas. Los autores graficaron parte de los resultados en el espacio bidimensional de la figura 7.

La dimensión perceptual horizontal está relacionada con las diferentes envolventes de amplitud. Los instrumentos de ataque rápido, como los de percusión, se encuentran a la izquierda, mientras que los de ataque gradual, como los de viento, a la derecha. La dimensión vertical está asociada con la envolvente espectral (el brillo aumenta hacia arriba). Iverson y Krumhansl sostienen que los oyentes basan sus juicios en las propiedades acústicas que se encuentran en la totalidad del desarrollo del sonido. El proceso físico de generación vincula fuertemente las diferentes partes entre sí. El ataque, el cuerpo y la extinción son interdependientes y no se contradicen. Este argumento explicaría la causa por la cual con sólo una parte del sonido –el ataque o el cuerpo, según el caso– podemos identificar su timbre. Otra conclusión derivada de este estudio establece que los datos que usamos para percibir el timbre dependen del contexto. La representación psíquica resultante es función tanto de la duración, intensidad y envolvente espectral de la señal acústica, como del banco de sonidos almacenados en la memoria del sujeto, su estado de atención y su experiencia previa. En este punto, los autores declaran la probable inexistencia de invariantes acústicas subyacentes en la percepción del timbre de los sonidos.



**Figura 7.** Representación bidimensional de similitudes tímbricas. Comparación entre ataques aislados (círculos) y sonido sin los ataques (triángulos): fagot [1]; violoncello [2]; clarinete [3]; corno inglés [4]; flauta [5]; corno francés [6]; oboe [7]; piano [8]; saxofón [9]; trombón [10]; trompeta [11]; trompeta con sordina [12]; tuba [13]; campana tubular [14]; vibráfono [15] y violín [16].

*Conclusiones sobre los modelos multidimensionales.* A pesar de la gran cantidad de información que aportaron hasta la fecha, los intentos tendientes a encontrar un número reducido de dimensiones no han llegado a resultados concluyentes. Ninguno de ellos demostró que las dimensiones sugeridas son completamente adecuadas,

aun dentro de un universo restringido de estímulos. Nuestra objeción al modelo de Grey puede extenderse, en menor o mayor medida, al resto de los estudios sobre el timbre basados en el análisis multidimensional.

## **Modelos verbales**

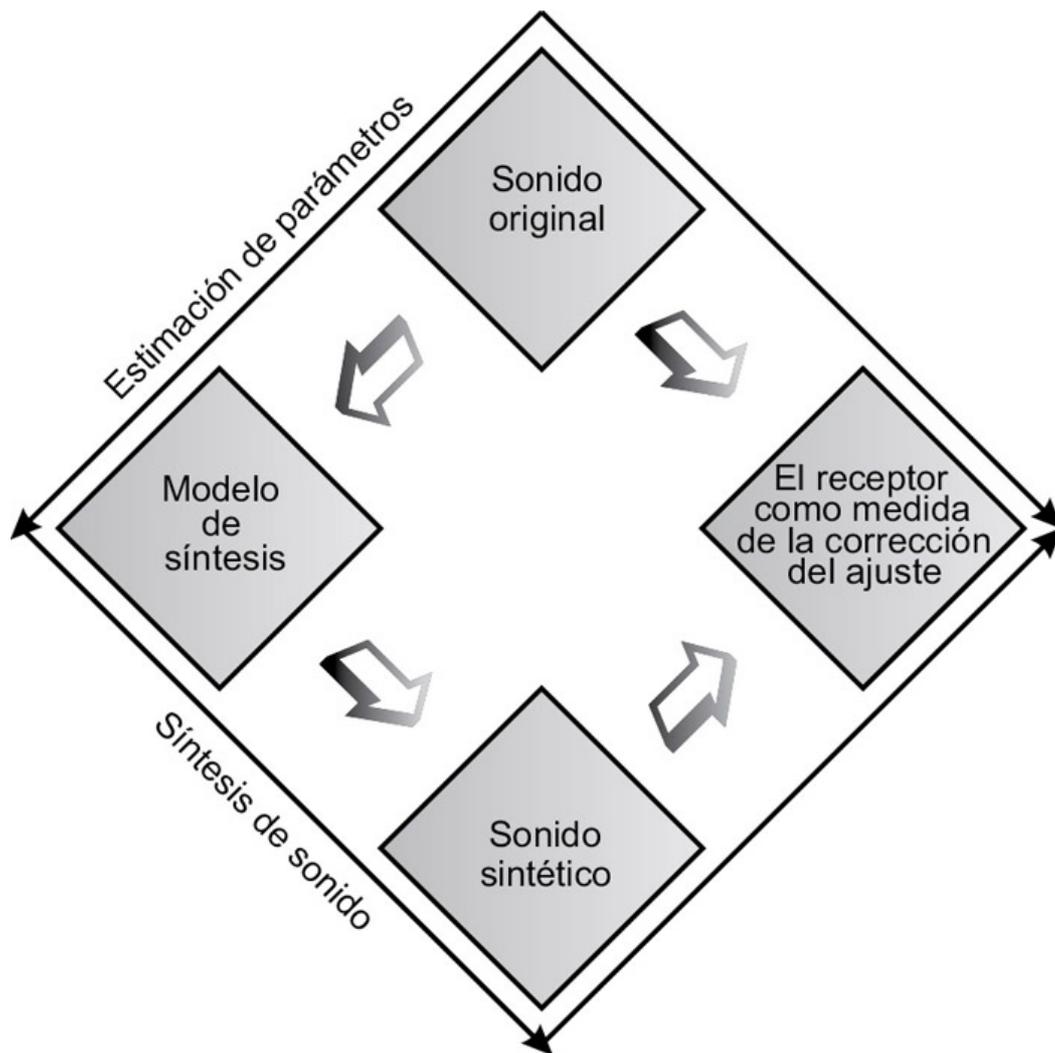
Desde una perspectiva distinta, Von Bismark intentó reducir el número de variables necesarias para determinar el timbre a partir de la verbalización de las percepciones (Bismark, G., 1974). Que es, de hecho, lo que hacemos habitualmente cuando tratamos de describir un sonido en una conversación. Seleccionó 35 sonidos diferentes y le pidió a los oyentes que los califiquen, con un valor de 1 a 7, con relación a 30 cualidades o dimensiones diferentes. Cada dimensión se definió a partir de una pareja de adjetivos contrarios como “duro-blando”, “seco-vivo”, “opaco-brillante” y así hasta llegar a las 30 incluidas en el estudio. Los sonidos provenían de generadores de señales periódicas complejas, con un número variable de armónicos, y de pulsos de ruido rosa. Se hacían pasar a través de filtros que los modificaban de acuerdo con determinadas envolventes espectrales, entre las que figuraban las correspondientes a las vocales en idioma alemán. Aplicando el método de selección de componentes principales, el autor redujo las 30 dimensiones verbales originales a cuatro dimensiones perceptuales. Las dos dimensiones más destacadas representaban los pares “punzante-apagado” y “compacto-disperso”. La dimensión más importante –definida por el par “punzante-apagado”– se relacionó con el centro de gravedad del espectro de la señal, que depende de la relación entre la energía de alta y baja frecuencia, y se parece al “brillo” presente en casi todos los estudios sobre timbre. La segunda dimensión en importancia –“compacto-disperso”– permite separar los sonidos tónicos del ruido. En esta experiencia, como en todas las de su clase, el resultado es muy sensible a la selección inicial de sonidos. Por ejem-

plo, la dimensión “compacto-disperso” no hubiera aparecido si no se incluían pulsos de ruido entre los estímulos de partida.

Otra posibilidad de acercamiento verbal al timbre la dan las “experiencias de campo”, en las que simplemente se recopilan y clasifican los términos que las personas –músicos o no músicos, según el caso–, usan para describir los sonidos. Aquí la dificultad está dada por la multiplicidad de universos lingüísticos existentes, aun dentro de una misma comunidad. Los mejores trabajos son los que restringen las muestras a un universo muy acotado. Dos buenos ejemplos son los trabajos de Kendall y Cartarette, en los que analizan los términos que aparecen en el trabajo de Von Bismark citado y en el tratado de orquestación de Walter Piston (Kendall y Cartarette, 1993). El fin práctico más significativo de esta clase de estudios es obtener un diccionario de términos que, al alimentar un sistema basado en una red neural, permita que los músicos controlen de manera verbal ciertos programas de síntesis por computadora.

## **Modelos de análisis por síntesis**

El timbre de los sonidos también se puede estudiar desde una perspectiva totalmente diferente, que parte de la producción antes que del análisis. La idea central consiste en el desarrollo de modelos de síntesis que permitan generar sonidos sintéticos indistinguibles de los sonidos “naturales” que se desean imitar. Si ambos sonidos, el sintético y el original, se oyen exactamente iguales, entonces la descripción acústica del sonido –adoptada para elaborar el modelo– es acertada. El método se denomina *análisis por síntesis*. Jean Claude Risset ilustra el proceso en un diagrama similar al de la figura 8 (Risset y Wessel, 1999).



**Figura 8.** Diagrama conceptual del proceso de análisis por síntesis

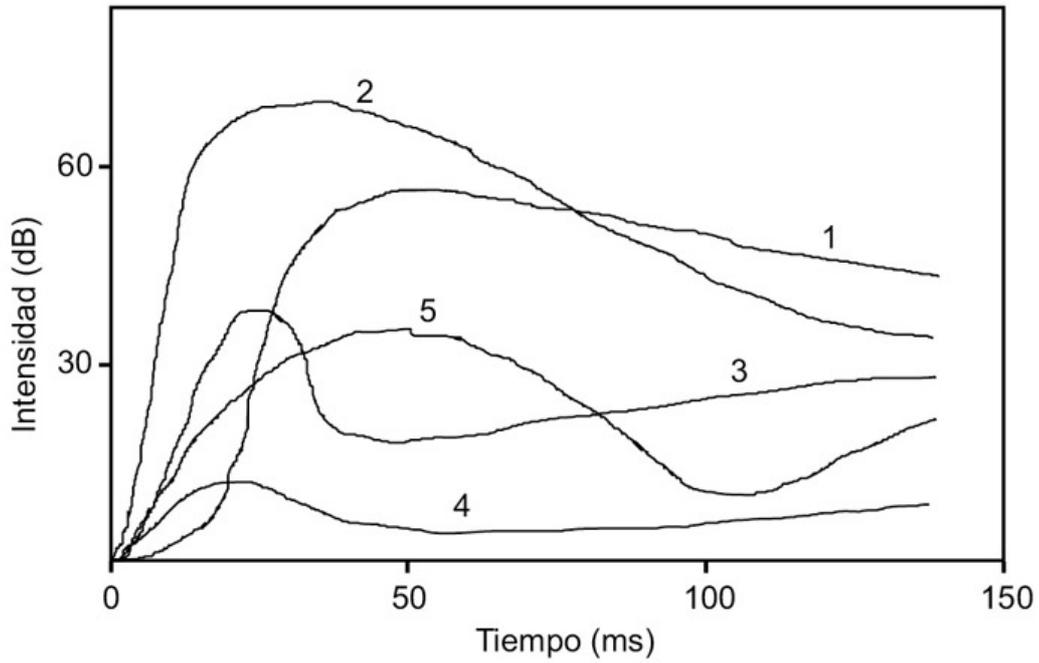
Quizá un ejemplo célebre sirva para aclarar el procedimiento usual en esta clase de investigaciones. En la década de 1960 Risset y Max Mathews intentaron, sin éxito, imitar los sonidos de los instrumentos de bronce. La teoría clásica –con sus espectros estáticos– fallaba completamente, y el intento por dotar a las señales de un comportamiento temporal sencillo –compuesto por un ataque, un cuerpo y una extinción– tampoco mejoraba mucho los resultados (Risset y Mathews, 1969). Para destrabar la cuestión, registraron magnéticamente el sonido de una trompeta real ejecutada por un profesional. La grabación fue realizada mediante el método de

análisis sincrónico de altura en una cámara anecoica –con un tiempo de reverberación muy bajo– para eliminar los efectos acústicos propios de una sala normal. El análisis sincrónico presupone que la señal es cuasi-periódica y describe el desarrollo de cada armónico individual en el tiempo. El resultado de un análisis de esta clase se muestra en la figura 9.

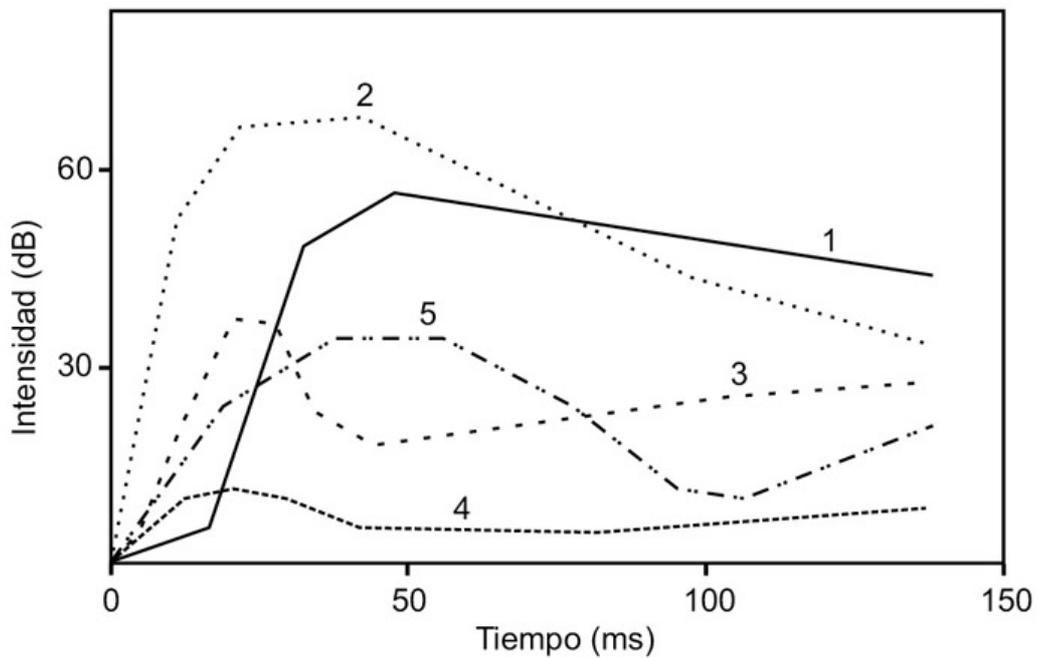
A partir de este análisis del sonido original, los autores simulaban la señal simplificándola de dos maneras: eliminaron las fluctuaciones de alta frecuencia y aproximaron las curvas por tramos lineales. La señal sintetizada quedó como la que se ve en la figura 10.

Las dos señales resultaron indistinguibles cuando se las comparó auditivamente. Las simplificaciones, por lo tanto, no afectaron los datos relevantes presentes en el sonido original.

Aunque el modelo era simple, para la época resultaba muy costoso a nivel computacional (en el estudio original se simulaban las evoluciones temporales de los primeros 13 armónicos). A partir de sucesivas simplificaciones, Risset y Mathews encontraron que: 1) la variación temporal del espectro es crítica para determinar el timbre; 2) las variaciones de alta frecuencia y pequeña amplitud no se perciben; 3) la dinámica del crecimiento de cada armónico define gran parte del timbre de los bronce; 4) existe una formante ubicada entre 1.000 y 1.500 Hz de gran importancia para la discriminación tímbrica; 5) la proporción de armónicos superiores aumenta significativamente con la intensidad. Empleando estos nuevos conocimientos lograron crear un modelo computacionalmente más económico que superó exitosamente la prueba de discriminación auditiva.



**Figura 9.** Resultado del análisis sincrónico de altura del sonido de una trompeta (Basso, G., 2001a)



**Figura 10.** Síntesis de la señal anterior simplificada

El ejemplo anterior, aunque limitado a unos pocos sonidos de trompeta, ilustra las características y ventajas del análisis por síntesis. Para empezar, permite evaluar cada parámetro del sonido de manera individual, uno por vez, evitando la complejísima tarea de lidiar con espacios multidimensionales en los que no se sabe bien qué cosa produce cada efecto perceptual. No menos importante, saca provecho inmediato del avance tecnológico en los campos de la acústica y de la informática, a la vez que impulsa nuevas investigaciones y desarrollos. Los modelos de análisis por síntesis han demostrado ser eficaces para generar conocimiento científico y técnico, reducir la cantidad de información necesaria para expresar una señal acústica, producir variantes y modificaciones nunca experimentadas y controlar la prosodia musical en tiempo real.

Una descripción detallada de las técnicas empleadas contendría la historia de la electroacústica durante los últimos 60 años. Sólo para enumerarlas, podemos citar la síntesis aditiva, la síntesis sustractiva, la síntesis alineal o global –a este grupo pertenece la síntesis por FM de Chowning–, el modelado físico y, actualmente en desarrollo, la síntesis basada en principios perceptuales.

## **Estudios sobre el timbre en la voz humana**

Parece que poseemos capacidades auditivas especiales para comprender la voz humana, que no empleamos para oír el resto de los sonidos. Es algo similar a lo que ocurre en la decodificación visual de los rostros. Se sabe que para el reconocimiento del rostro de una persona se usan recursos diferentes de los que se usan para percibir el resto del campo visual (existen personas que no pueden distinguir los objetos, pero que leen perfectamente los rostros, y viceversa). Quizá estas capacidades tengan que ver con la importancia social de la información y con lo parecidos que son los rostros –y las voces– entre sí; es probable que se necesite cierta capacidad potenciada y específica para distinguir entre una voz y otra.

Por otra parte, existe fuerte evidencia de que la palabra y la música se procesan en lugares diferentes de la corteza cerebral. Anteriormente se creía que cada uno de los hemisferios cerebrales estaba especializado en uno de estos dominios, pero en la actualidad se sabe que tanto la producción como la percepción de música y palabra involucran varios subprocesos, cada uno de ellos lateralizado de modo diferente. Por ejemplo, se descubrió que las tareas de discriminación y de identificación (que implican un acceso al léxico verbal) son independientes y están localizadas en diferentes lugares de la corteza.

El problema de la identificación tímbrica de las voces es mucho más difícil que en el caso de los demás sonidos. No sólo porque podemos identificar miles de voces diferentes (contra un grupo reducido de etiquetas tímbricas generales), sino por el pequeño rango de variación de los rasgos no tímbricos en la voz. Sabemos que las frecuencias fundamentales típicas en hombres, mujeres y niños están comprendidas entre 80-240, 140-450 y 170-600 Hz, respectivamente. Y que el rango de variación en frecuencia para un hablante promedio alcanza una relación de 2 a 1 (en el canto no profesional es de 3 a 1). Pese a esta amplitud de valores espectrales, se comprobó que un cambio de apenas 5% en la frecuencia de alguna de las tres primeras formantes puede reducir drásticamente la capacidad de identificación de una voz conocida.

Aunque aparentemente la frecuencia fundamental, la frecuencia de las formantes —especialmente entre 1.000 y 2.500 Hz— y el ancho de banda de las formantes participan en la identificación de la voz humana, se presume que la información acústica relevante depende de las características particulares de cada locutor (Van Dommelen, W., 1990). En otros términos, no existe un conjunto fijo de datos invariantes que se apliquen a todas las situaciones. Por ejemplo, puede ocurrir que la señal acústica de la voz del hablante A posea cierta cantidad de información específica, de la cual el oyente extrae un subconjunto de datos adecuado para su identifi-

cación. Pero en el caso del hablante B el subconjunto anterior podría resultar inapropiado, y el oyente debe elegir otro diferente.

## Estudios sobre el “color” del sonido

Un grupo particular de estudios fue orientado sobre lo que se llamó el “color” del sonido. El color del sonido no tiene, según sus defensores, propiedades temporales y se aplica a los atributos estacionarios del sonido, no a los transitorios. De acuerdo con Wayne Slawson, “un sonido puede cambiar de color a lo largo del tiempo, pero el cambio en sí mismo no es un color independiente. No más que un cambio en un color visual puede ser un color independiente”.[\[13\]](#) Slawson intentó descubrir el modo de controlar el color del sonido mientras cambia la altura tonal, la sonoridad o algún otro rasgo del sonido. Propuso un sistema de control basado en el modelo físico de fuente y filtro. Por ejemplo, en una guitarra la cuerda es la fuente de oscilación y la caja de resonancia el filtro. A partir del modelo propuso una serie de reglas de control y transformación.

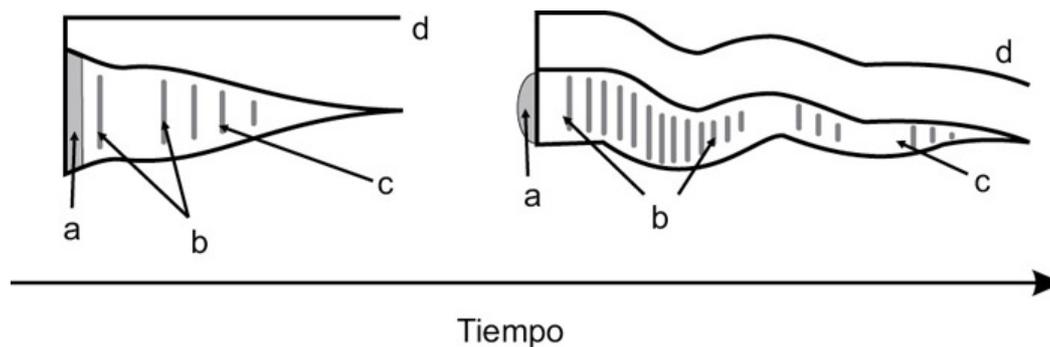
1. El color del sonido está asociado con el filtro, no con la fuente. Para mantener el color constante se deben mantener constantes las características del filtro.
2. Para mantener el color del sonido constante se debe mantener la envolvente espectral constante.[\[14\]](#)
3. El color del sonido posee cuatro dimensiones independientes: apertura, agudeza, lasitud y pequeñez.
4. Para transponer un color respecto a una dimensión, se debe agregar un valor constante en esa dimensión.
5. Para invertir un color respecto a una dimensión, se debe invertir el signo del valor en esa dimensión.[\[15\]](#)

Estas reglas definen claramente un espacio vectorial, similar al espacio de los colores visuales propuesto por Schrödinger. A lo con-

troversial de la conjetura, Slawson le sumó ciertas afirmaciones que se pueden calificar, por lo menos, de arriesgadas: “el color del sonido aparece como un elemento musical viable que puede ser sometido a un control estructural preciso, en todo análogo al que se le aplica a la altura tonal”. Aquí aparece por primera vez el debate sobre la capacidad del timbre para sustentar transformaciones estructurales o, dicho de otro modo, sobre la capacidad morfofórica del timbre (más adelante desarrollaremos este punto). Al margen de la audacia de las afirmaciones de Slawson, en sus conjeturas se perciben claros resabios del modelo clásico del timbre.

En sintonía con Slawson, Yee On Lo acepta que la envolvente espectral determina en gran medida el timbre de los sonidos, pero se diferencia al asignarle también gran importancia a la evolución temporal del espectro. Afirma que percibimos los cambios espectrales a través de “cuadros” definidos como la mínima porción repetitiva de la onda. Estos cuadros actúan como las transiciones entre fonemas en la voz hablada. Así como los cuadros proporcionan la microestructura temporal –del orden de los milisegundos–, la macroestructura –del orden de los segundos– está también construida a partir de la concatenación de cuadros. Como apunta Lo: “El timbre es la evolución de los cuadros”.[\[16\]](#) Existen ciertos cuadros críticos que definen los cambios importantes a lo largo del tiempo. En el modelo de Lo, la percepción del timbre se funda en tres parámetros: la disposición de los cuadros críticos, la envolvente de amplitud y la trayectoria de la altura tonal. En la figura 11 se aprecian los diagramas de Lo para dos sonidos diferentes.

El modelo de Yee On Lo se puede ver como una versión cuantificada a diferentes niveles de la concepción que asocia el timbre con la evolución temporal del espectro de la señal.



**Figura 11.** Diagramas de Lo de dos sonidos diferentes (Lo, Y. O., 1987)

## Eventos acústicos diferentes a la voz y a los sonidos musicales

Aunque la mayoría de los estudios sobre el timbre tratan sobre los sonidos de la voz hablada o de la música, algunos investigadores centraron su atención sobre sonidos complejos propios de la vida cotidiana. Comentaremos brevemente cuatro de ellos.

1. D. J. Freed estudió los sonidos producidos por diferentes baquetas al golpear cuatro cacerolas de distinto tamaño (Freed, D. J., 1990). Aisló cuatro parámetros de “previsión de timbre”: la *energía global*, la *distribución espectral*, la *velocidad de evolución espectral* y la *rapidez de alteración del evento*. Las evaluaciones de los sujetos sobre el tipo (dureza) de la baqueta empleada se mantuvieron totalmente independientes del tipo de cacerola golpeada, a pesar de que los estímulos contenían información tanto sobre el excitador como sobre el resonador. Los parámetros de previsión de timbre de Freed se correlacionaron bien con las evaluaciones de la dureza de la baqueta. Actualmente se continúa investigando con el fin de individualizar los índices invariantes que permiten a los oyentes caracterizar y reconocer al excitador y al resonador de forma separada.

2. B. H. Repp estudió los sonidos que producen las manos al aplaudir. Encontró que los oyentes coincidían al juzgar el sexo del “aplaudidor”, aunque los juicios no siempre correspondían al sexo real de la fuente (Repp, B. H., 1987). Según el autor este estudio refleja la influencia de los estereotipos culturales auditivos.
3. Warren y Verbrugge analizaron las respuestas auditivas frente a objetos de vidrio que rebotaban o se rompían (Warren y Verbrugge, 1984). Analizó la correspondencia entre patrones rítmicos, la intensidad y las envolventes espectrales y los juicios auditivos asociados.
4. N. J. Vanderveer estudió la identificación libre y la clasificación de eventos acústicos complejos como el tintineo de llaves, el crepitar del papel, etc. (Vanderveer, N. J., 1979). Según S. Handel, sería natural indicar los sonidos a partir del evento mecánico que los engendra, sin duda a partir de invariantes transformacionales (Handel, S., 1989). Los grupos estarían parcialmente determinados a partir de las propiedades acústicas de los materiales utilizados (es decir, las invariantes estructurales).

A los estudios anteriores les cabe, potenciada, la crítica que se le hace a casi todos los estudios sobre el timbre: dependen fuertemente del conjunto de sonidos escogidos.

## **Resumen y conclusiones**

Los resultados de los estudios mencionados, y de otros similares, son muchas veces contradictorios. En general, todos aceptan que la percepción del timbre depende de la forma de onda en su conjunto, pero discrepan sobre cuáles de sus características deben considerarse relevantes. A modo de resumen, podemos establecer una serie de elementos, aceptados por la mayoría de los investigadores,

que intervienen en la discriminación de los timbres instrumentales aislados.

- La información presente en el ataque de la señal.
- La información relativa a la evolución de la envolvente espectral, principalmente en el cuerpo de la señal.
- La presencia de pequeñas variaciones aleatorias de frecuencia en los componentes discretos de la señal.
- Otras características se discriminan con mayor dificultad:
- El grado de coherencia de la variación en frecuencia de los componentes discretos.
- El grado de simplificación en las variaciones de amplitud y frecuencia (válido sobre todo para los modelos de análisis por síntesis).

Esta enumeración es válida solamente como referencia ya que deben agregarse numerosas singularidades y excepciones. En la lista que sigue mencionamos algunas de ellas.

- Las variables significativas para la percepción del timbre musical existen en, al menos, dos niveles temporales diferentes: un nivel se asienta en los cambios que ocurren en algunos milisegundos y el otro registra las variaciones más lentas, del orden del segundo.
- Cuando el estudio permite que participen las características de ataque de los sonidos, éstas dominan por sobre los otros factores presentes en la señal acústica. Solamente el brillo, determinado por la distribución espectral, permanece presente como un factor tímbrico de importancia. Un error sistemático cometido por muchos investigadores que pretenden eliminar el ataque de la señal aparece al seccionarlo abruptamente: lejos de eliminar el ataque, se lo reemplaza por otro de banda ancha.[\[17\]](#) Algo similar ocurre en los experimentos de restauración fonémica, en los que ciertas sílabas de una frase son su-

primidas. El reconocimiento es mayor cuando se reemplazan las sílabas ausentes por silencios que cuando se enmascaran las mismas sílabas por un ruido. En el segundo caso se reemplaza una sílaba por otra cosa, en el primero no.

- El brillo del sonido figura como un parámetro significativo en todos los estudios mencionados. La gran discusión gira en torno a la naturaleza de la envolvente espectral. Algunos autores consideran que su importancia guarda estrecha relación con la frecuencia fundamental de la señal, mientras que otros sostienen que las bandas de frecuencia (en algunos casos las formantes) deben ocupar regiones espectrales absolutas. La mayor parte de la evidencia experimental confirma la segunda hipótesis.
- Cuando sonidos de un mismo género se hallan situados en contextos musicales de cierta complejidad, en particular en texturas isorrítmicas, los índices de naturaleza temporal parecen perder importancia desde el punto de vista de la discriminación, en beneficio de los índices ligados a la envolvente espectral. Parece ser que, cuando la complejidad aumenta (complejidad armónica o cantidad de voces) las diferencias temporales finas no son tan bien preservadas como las diferencias espectrales. Un buen ejemplo se puede encontrar en los corales para maderas en algunas sinfonías románticas. En esos fragmentos, los transitorios de las notas aisladas son menos importantes que la información presente en la parte del cuerpo de la señal, ligada a la envolvente espectral.
- Ciertos sonidos parecen poseer características o dimensiones tímbricas exclusivas, no comparables con las usuales en la mayoría de los instrumentos. La preeminencia de armónicos impares en el clarinete –cuya sola aparición sobrepasa en importancia a las demás características de la señal–, el particular transitorio de ataque de la trompeta, la presencia de ruido rosa en el sonido de un platillo, la envolvente de amplitud en el

trombón y en la flauta, son buenos ejemplos de dimensiones tímbricas particulares no extrapolables a la totalidad de los instrumentos de música.

### **3. Propiedades particulares del timbre**

A continuación describiremos algunas de las propiedades particulares de la percepción auditiva que fueron analizadas, y en algunos casos descubiertas, a partir de las investigaciones sobre el timbre de los sonidos.

#### **Constancia del timbre**

Es evidente que existe alguna forma de invariancia o de constancia tímbrica, que se comprueba en el hecho de que una fuente acústica puede reconocerse en una gran variedad de circunstancias. Por ejemplo, un saxo se identifica como tal al margen de la altura tonal o de la dinámica que esté generando. Se lo reconoce sin dificultad aun si lo oímos a través de un medio tan distorsivo como una pequeña radio portátil o a través de una puerta de madera. La voz de una persona conocida se identifica en una habitación silenciosa o en una reunión muy concurrida, en la que se mezcla con otras voces, con la música y con los ruidos de propios de la fiesta. Los espectros que nos llegan son muy diferentes en todos estos casos. Sin embargo, la noción de constancia tímbrica no se comprende bien a partir de los estudios clásicos de discriminación.

¿Existen correlatos físicos que den cuenta de esta constancia tímbrica en diferentes contextos? Para reconocer el timbre único de una voz se deben elegir, aislar y analizar las partes de la señal que aportan la información adecuada. Una mala elección implicará un cambio en el timbre percibido y una identificación equivocada. En este punto resulta interesante destacar que la envolvente espectral de una señal acústica sufre, inevitablemente, modificaciones importantes entre la emisión y la recepción. Ya sea por la acción

de filtrado en el medio de propagación, o por distorsión en las reflexiones y refracciones que puedan ocurrir, el espectro que le llega al oyente depende de la geometría límite del espacio acústico y de la ubicación de la fuente y el receptor en dicho espacio (expresado con mayor precisión, la alteración del espectro depende de la impedancia compleja que encuentra la señal en el trayecto emisor-receptor). [18]. Las variaciones temporales, en especial las de pequeña escala presentes en el ataque, preservan mucho mejor el patrón estructural original luego de transformaciones como las anteriores. Se supone que estos patrones temporales invariantes cumplen una función destacada en la percepción de la constancia tímbrica.

Habitualmente reconocemos sin error el timbre de los sonidos, lo que significa que elegimos los elementos de información adecuados en contextos enteramente diferentes. Como en el caso de las constancias visuales, la constancia del timbre sólo puede ser explicada en función de complicados procesos en las etapas superiores del tratamiento neural. El tratamiento de las señales de entrada al sistema auditivo, o de las representaciones auditivas primarias, no alcanza para dar cuenta del fenómeno.

Como ocurre en el reconocimiento visual de los rostros, se supone que la aparición de ciertos rasgos específicos alcanza para situar un sonido en una categoría tímbrica particular –por ejemplo, para reconocerlo como sonido de saxofón–, al margen de la existencia o no de otras características menos significativas. Lo interesante en el caso del timbre es que el conjunto de rasgos imprescindibles no es el mismo para todos los sonidos, como ya vimos en el apartado anterior.

## ¿Timbres metaméricos?

El sentido de la audición es claramente analítico con relación a las alturas tonales, relacionadas con la frecuencia fundamental de la señal acústica. Es decir, en presencia de una mezcla espectral dis-

tinguimos cada uno de los componentes de la mezcla. Por ejemplo, cuando combinamos un Do y un Mi oímos tanto la tercera mayor (la combinación) como cada una de las notas individuales. En contraste, el sentido de la vista no es analítico con relación a los colores, relacionados con la longitud de las ondas lumínicas: al mezclar azul y amarillo vemos el color verde (la combinación), pero ya no percibimos el azul y el amarillo que lo forman. Esta característica de la visión permite crear colores metaméricos, causados por espectros diferentes pero indistinguibles a la vista. ¿Existen los timbres metaméricos o, lo que es equivalente, dos señales físicas diferentes que se oigan exactamente igual? De acuerdo con los estudios realizados hasta ahora, la respuesta es no (al menos en un sentido estricto, pues se pueden hallar dos señales diferentes que cumplan una misma función tímbrica –por ejemplo en una fusión de objetos auditivos– aunque se los pueda discriminar).

Por otro lado ¿existe un “unísono tímbrico”, más allá del caso trivial de dos sonidos de igual intensidad y altura tonal? Es decir, dos sonidos de diferente altura tonal e idéntico timbre (¿no sería éste el ideal de un constructor de pianos?). Algunos autores creen que sí, aunque sus argumentos no son plenamente convincentes (Van Noorden, L., 1975; Erickson, R., 1975). Otra característica de la audición, relacionada con lo anterior, es la inexistencia de alguna cosa parecida a una “octava tímbrica”: no es posible obtener un ciclo de timbres a partir del cual construir una “escala tímbrica”, a semejanza de la escala de alturas tonales. Por supuesto, lo anterior no es más que otra forma de decir que el timbre es un rasgo multidimensional.

## **El timbre como un medio morfofórico**

En el capítulo V mencionamos la posibilidad de clasificar los rasgos acústicos en morfofóricos, capaces de portar formas invariantes, y no morfofóricos. Comentamos que la altura tonal y el tiem-

po constituyen medios morfofóricos poderosos. Sin embargo, tanto Attneave y Olson, como más adelante Michael Kubovy y Roger Shepard, consideraron el timbre como un medio no morfofórico. Es evidente que a partir del timbre no se puede codificar la misma clase de información estructural que se obtiene operando con la altura tonal. Sin embargo, descartar toda capacidad de organización estructural cimentada en la modulación tímbrica no parece razonable. La codificación fonética del habla se asienta principalmente en las variaciones del timbre en función del tiempo, y la armonía musical puede ser considerada como una particular codificación del timbre de los sonidos (cada función tonal posee un “color” particular).[\[19\]](#)

## **Timbre y atención**

Una característica destacada de la percepción auditiva se relaciona con la atención. Si comparamos las propiedades de los eventos y rasgos acústicos que nos permiten mantener la atención selectiva, veremos que son (casi) las mismas que se emplean para discriminar y preservar el timbre de un sonido ante cambios de contexto. En ambos casos el acento está puesto sobre el esfuerzo para identificar y seguir auditivamente un objeto acústico particular.

## **4. Comentario final**

Lo visto en este capítulo vale solamente para caracterizar el timbre de un único objeto auditivo. Todavía no consideramos las características que debe poseer una señal acústica para que percibamos uno o una multiplicidad de objetos acústicos, tema que será desarrollado en el capítulo VIII.

La conclusión lógica de este capítulo estaría dada por el análisis de los usos musicales del timbre, quizá ordenados de acuerdo con su progresión histórica. Pero tal empresa ocuparía, por lo menos, un libro similar a éste. Así que lo dejamos para más adelante.

# Notas

- [1](#) Shepard (1981), citado en McAdams y Bigand (1994).
- [2](#) American Standards Association (1960), traducción de GB. Esta definición es defendida, entre otros, por Crowder, aunque impone una restricción temporal que se suma a las condiciones originales (Crowder, R. G., 1989). Otros autores (Plomp, R., 1976) restringen la definición a la clase de los sonidos complejos estacionarios en el tiempo.
- [3](#) Michel Chion es más sarcástico: “[...] ‘altura, intensidad, duración, timbre’ equivale a ‘altura, peso, edad, fisonomía general’ [...]” (Chion, M., 1998).
- [4](#) Helmholtz, H. von (1885-1954). Trad. de GB.
- [5](#) Sin intención de agotar la lista, esta posición se puede ver reflejada en los trabajos de Rayleigh (1945), Culver (1947), Olson (1952), Benade (1960), Beranek (1962), Backus (1969), Feynman (1971) y Sundberg (1991).
- [6](#) Hall (1937), en Cook, P. (1999).
- [7](#) experimento fue realizado por primera vez por Köhler en 1915 (Risset y Wessel, 1999).
- [8](#) Que en un texto sobre el timbre estemos usando palabras como “duro”, “oscuro” o “brillante”, tomadas en préstamo de otros sentidos, demuestra lo lejos que estamos de tener resuelto el asunto.
- [9](#) Algunos de estos instrumentos, sin embargo, lograron imponerse como generadores de timbres originales y novedosos. Tal es el caso del órgano Hammond.
- [10](#) Bregman, A. (1994). Trad. de GB.
- [11](#) El ejemplo clásico en esa época consistía en invertir temporalmente un registro de audio: la envolvente espectral promedio no cambia, pero el timbre sí lo hace.
- [12](#) Grey, J. (1977). Trad. de GB.
- [13](#) Slawson (1985), en Butler, D. (1992). Trad. de GB.
- [14](#) Esta regla no es más que una variación de la anterior.
- [15](#) Slawson (1985), en Butler, D. (1992). Trad. de GB.
- [16](#) Lo, Y. O. (1987)]. Trad. de GB.
- [17](#) Este efecto es consecuencia del principio de indeterminación acústico (Basso, G., 2001a).
- [18](#) De aquí surge la fragilidad de la *altura espectral* ante la menor modificación de la señal acústica. Esta propiedad, opuesta a la solidez que presenta la *altura tonal* frente a modificaciones similares o superiores, se desarrolla en Basso, G. (2001a).
- [19](#) En la música contemporánea la distinción entre armonía y timbre resulta muchas veces desdibujada, como puede comprobarse en piezas como *Aventures* de Ligeti y *Modulations* de Grisey.

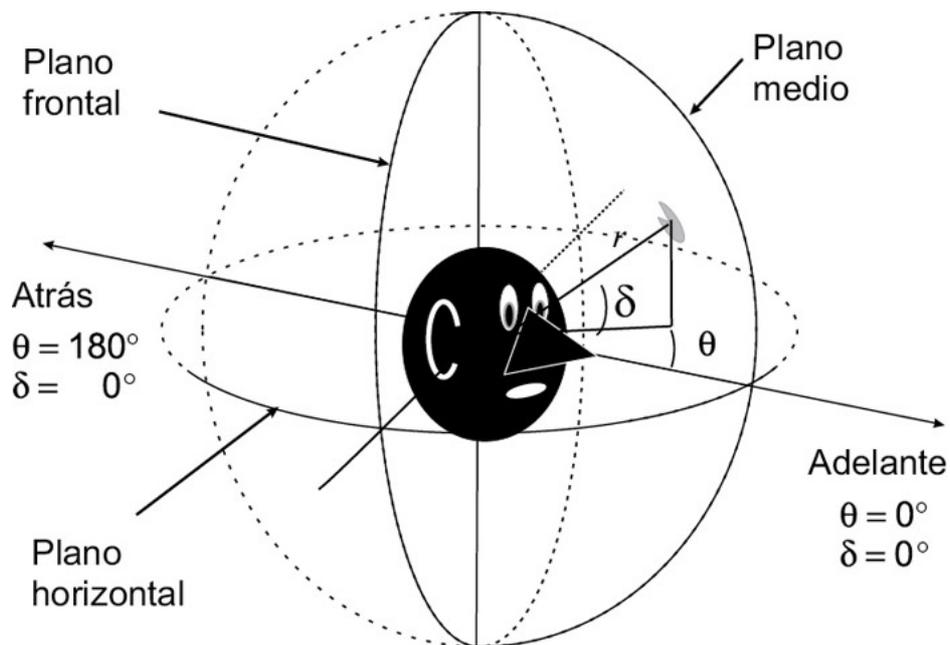
# Capítulo VII

## Percepción auditiva del espacio

Para analizar la percepción auditiva del espacio es usual distinguir entre localización y lateralización. El término *localización* se refiere a los juicios vertidos sobre la dirección y distancia a la que se encuentra una fuente acústica, que pueden o no coincidir con su ubicación física. El término *lateralización* se usa para describir la localización aparente de un sonido “en la cabeza” del oyente. La lateralización puede verse como una versión de laboratorio de la localización y es un componente central en el diseño de campos acústicos virtuales.

Para caracterizar al espacio acústico vamos a emplear coordenadas esféricas cuyos parámetros son el ángulo de azimut  $\vartheta$ , el ángulo de elevación  $\delta$  y la distancia  $r$  a la fuente, tal como se ve en la figura 1.

Comenzaremos nuestro estudio de la percepción espacial con el caso particular de los estímulos sinusoidales estacionarios. Más adelante veremos lo que ocurre con las señales generalizadas, variables en el tiempo y de espectro compuesto.



**Figura 1.** Sistema de coordenadas esféricas:  $\vartheta$ , azimut;  $\delta$ , elevación;  $r$ , distancia

## 1. Localización de sinusoides

En la localización de señales que poseen una sola línea en su espectro –como ocurre con una senoide– los factores espectrales no aportan información espacial relevante. Por este motivo los datos empleados en la localización de sinusoides son principalmente binaurales. De todos ellos se destacan claramente la *diferencia interaural de intensidad* (IID) y la *diferencia interaural de tiempo* (ITD).

### Diferencia interaural de intensidad

Cuando la fuente acústica se ubica fuera del plano medio del oyente, la intensidad de las señales que llegan a cada uno de sus oídos es ligeramente diferente. En la mayoría de los casos, esta diferencia de intensidad no es provocada por la desigualdad de recorrido entre la fuente y cada oído, sino por la sombra acústica que causa la cabeza al interponerse entre la fuente y el oído más alejado.

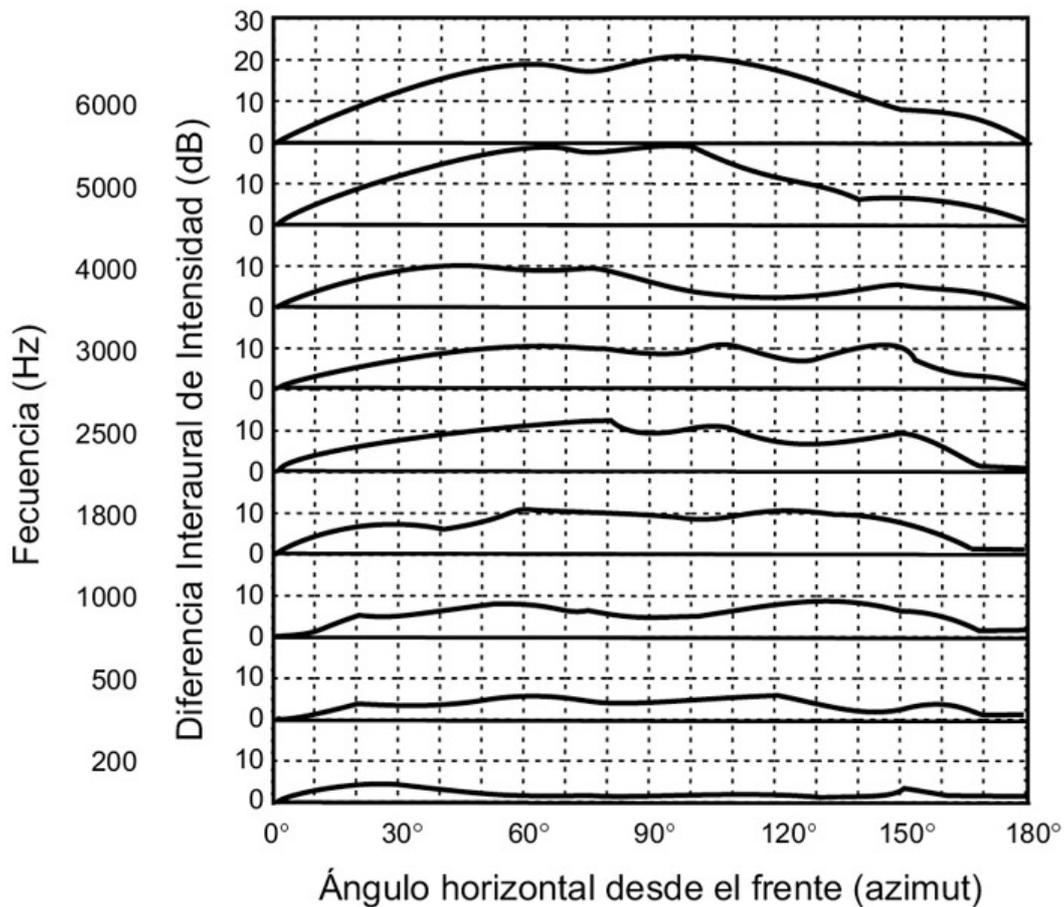
En la figura 2 se puede ver la diferencia interaural de intensidad (IID) en función del azimut  $\vartheta$ . La frecuencia de la señal sinusoidal generada por la fuente se emplea como parámetro de la familia de curvas.

Una cabeza humana promedio se puede aproximar a una esfera de alrededor de 20 cm de diámetro. A causa del fenómeno de difracción la cabeza no va a provocar una sombra acústica nítida por debajo de 1.700 Hz, y va a resultar acústicamente transparente por debajo de 500 Hz. Como puede verse con claridad en la figura 2, la IID es ínfima a frecuencias por debajo de los 500 Hz, pero puede llegar hasta 20 dB a frecuencias elevadas. En otros términos, la difracción de las ondas en la cabeza limita la eficacia del mecanismo de

detección de diferencias interaurales de intensidad a la parte superior del espectro audible.

En cuanto al mínimo cambio detectable en la IID, se ha comprobado que llega a 1 dB para señales frontales siempre que la frecuencia de la señal supere los 1.000 Hz.

Un buen ejemplo relacionado con la IID lo provee la reproducción estereofónica bicanal, en la que casi toda la información espacial está codificada únicamente en términos de diferencias de intensidad (esto no es así, debemos aclarar, en los registros de audio bien cuidados en los que se contempla también la información de fase).



**Figura 2.** Diferencia interaural de intensidad (IID) en función del azimut  $\vartheta$  y de la frecuencia

## Diferencia interaural de tiempo

Para fuentes acústicas ubicadas fuera del plano medio del oyente, la diferencia interaural de tiempo (ITD) aparece como consecuencia de la diferencia en el tiempo que le lleva a la onda alcanzar cada uno de los oídos. La ITD varía entre 0 y 690 ms para señales cuyos ángulos de azimut corresponden a 0° y 90°, respectivamente. Como una onda acústica que se propaga en el aire tarda 30  $\mu$ s en recorrer 1 cm, para ir de un oído al otro necesita aproximadamente 700  $\mu$ s.

Si las señales son estrictamente sinusoidales, una diferencia de tiempo equivale a una diferencia de fase. A bajas frecuencias la información contenida en la diferencia de fase es significativa, pero a altas frecuencias la longitud de onda es menor que la distancia entre oídos y la diferencia de fase provee datos ambiguos. A una longitud de onda de 23 cm, similar a la distancia promedio entre oídos, le corresponde una frecuencia de 1.500 Hz. Si la señal posee 10 kHz entran varios ciclos completos en esa distancia y la fase deja de aportar información espacial unívoca. En la figura 3 se pueden apreciar estas dos situaciones.

Los movimientos de la cabeza reducen en parte la ambigüedad de fase, pero ésta resulta muy grande para frecuencias por encima de los 1.500 Hz y el mecanismo de ITD pierde toda efectividad.

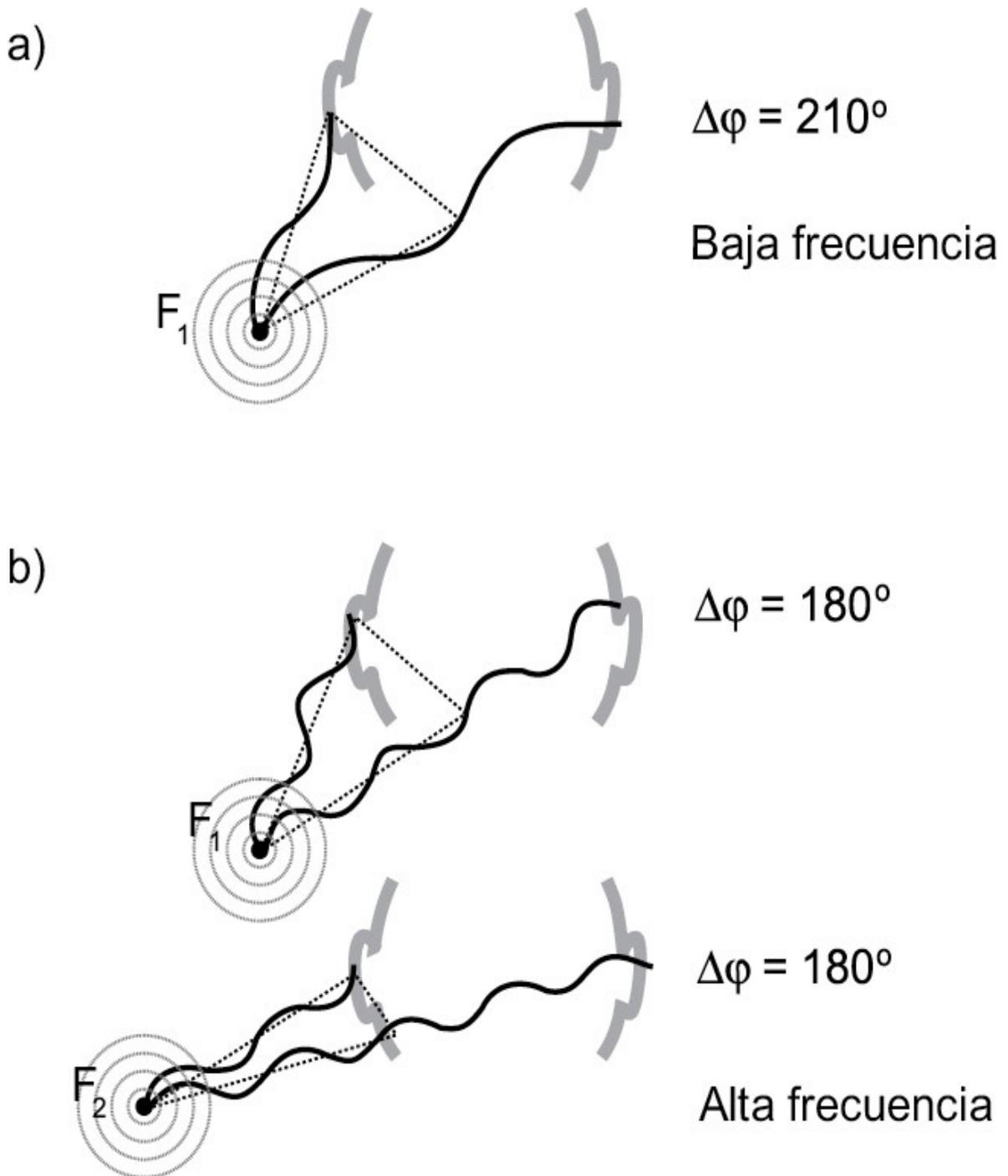
En la figura 4 se aprecia la diferencia de recorrido en función del ángulo de entrada de las señales que llegan a los oídos.

Si denominamos  $r$  al radio de la cabeza, la diferencia de recorrido  $d$  es:  $d = r \vartheta + r \sin \vartheta$

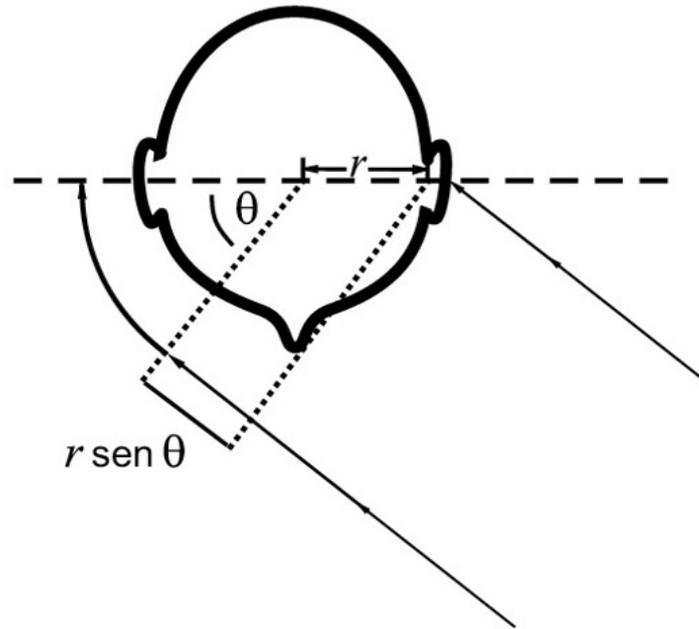
En la figura 5 se puede ver el gráfico de las diferencias temporales de llegada en función del ángulo  $\vartheta$ . La curva se quiebra a 90° por simetría bilateral.

Tan importante como la ITD es el *mínimo cambio de ángulo perceptible* (MAA). Cuando el azimut  $\vartheta$  es cero los umbrales de detección para cambios en el ángulo de llegada son muy pequeños, del orden de 10 a 20 –que corresponden a una diferencia temporal de 10  $\mu$ s. Para otros ángulos horizontales de llegada el MAA varía sensible-

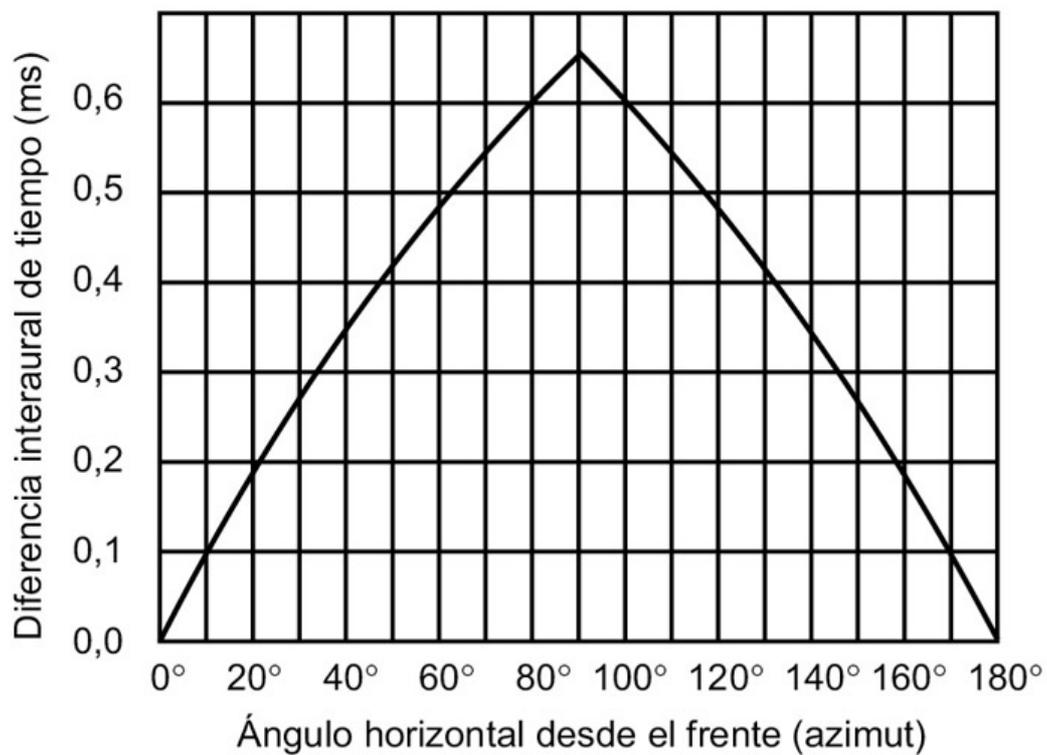
mente con la frecuencia (figura 6). A 1.800 Hz y un azimut de 600 el MAA no puede discernirse.



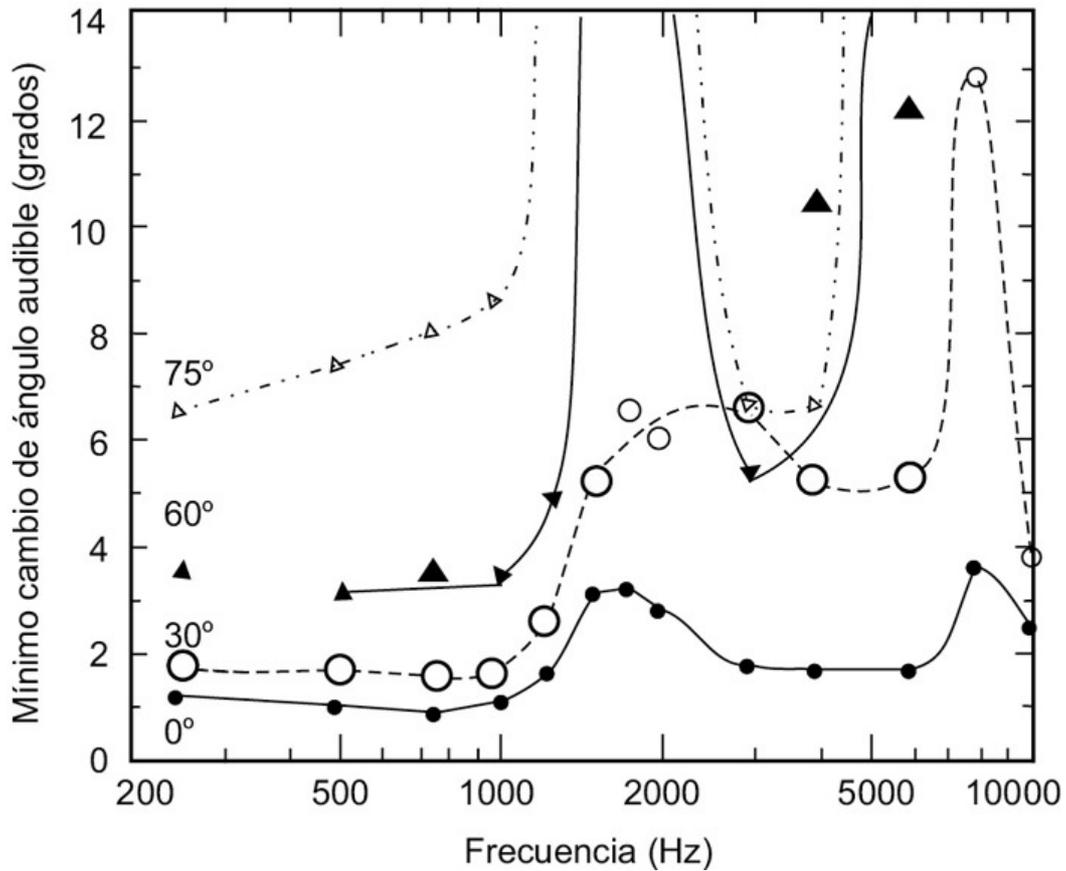
**Figura 3.** Pérdida de efectividad de la ITD a alta frecuencia por confusión de fase. El mecanismo no puede distinguir entre la localización de las fuentes F1 y F2 si emiten a alta frecuencia



**Figura 4.** Cálculo de la diferencia de recorrido entre las señales que llegan a ambos oídos



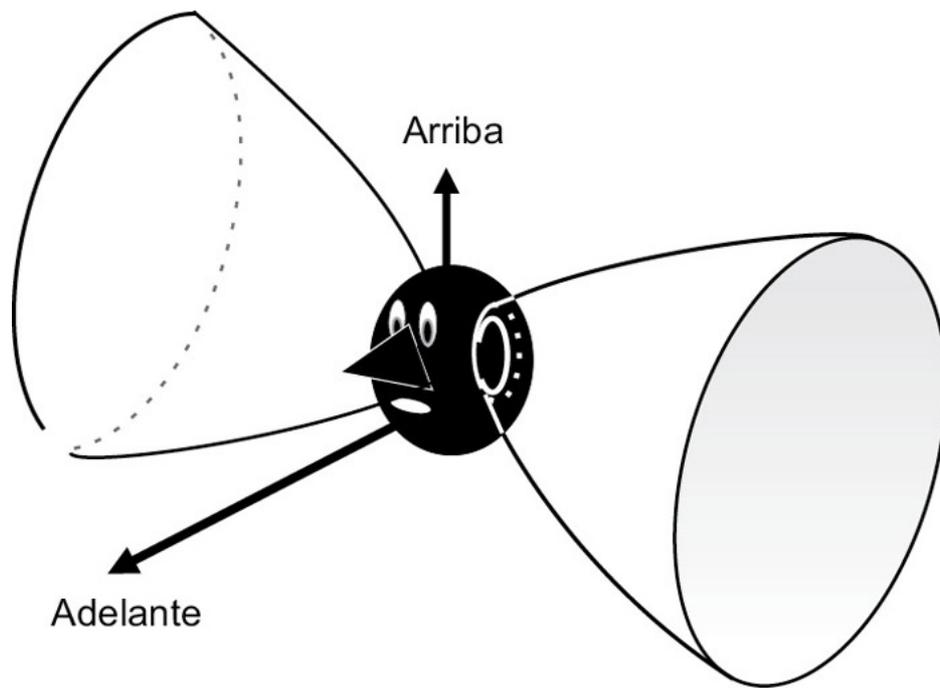
**Figura 5.** Diferencia interaural de tiempo ITD en función del azimut



**Figura 6.** Mínimo cambio de ángulo perceptible (MAA) en función de la frecuencia. El parámetro variable es el azimut  $\vartheta$

### *El cono de confusión y los movimientos de la cabeza*

Como la diferencia interaural de tiempo depende de la diferencia de recorrido de las ondas entre ambos oídos, la línea que va de un tímpano al otro define un eje de simetría axial alrededor del cual se encuentran lugares geométricos con el mismo ITD. Aparecen así los denominados *conos de confusión* a cada lado de la cabeza tal como se aprecia en la figura 7.



**Figura 7.** Cono de confusión lateral

Las ambigüedades relacionadas con este cono de confusión, al igual que las vinculadas con el plano medio, pueden ser resueltas moviendo la cabeza. Si la rotamos 200 a la izquierda y la imagen auditiva se desplaza 200 en su posición lateral aparente, entonces uno tiende a suponer que la fuente se sitúa en el plano horizontal. Si, en cambio, el movimiento de la cabeza no es acompañado por un cambio en la imagen perceptual, la fuente de sonido se localiza directamente arriba o abajo del oyente. La visión y el sentido del equilibrio juegan un papel importante en este proceso, pues ayudan a definir la posición de la cabeza en el espacio. A niveles elevados de procesamiento neural la representación espacial involucra la integración de la información proveniente de diferentes sentidos.

Como resumen –y solamente para señales sinusoidales–, es posible sostener que los datos físicos aportados por la IDD son útiles a altas frecuencias, mientras que los provenientes de la ITD lo son para bajas frecuencias. Este relevo en los mecanismos de percepción espacial fue denominado en 1907 *teoría doble* por Lord Rayleigh, pero no se verifica de manera acabada para señales de espectro complejo.

Los umbrales de cambio en la percepción de las diferencias de tiempo son menores cuando la ITD es cero, con la fuente ubicada en el plano medio. El umbral de variación para la IDD también es menor cuando la IDD es cero. En definitiva, la resolución espacial es mejor para sonidos que llegan de frente o desde algún lugar en el plano medio, con 00 de azimut.

En algunos casos la localización monoaural es tan buena como la binaural, hecho que sugiere el uso de algún mecanismo adicional no contemplado en la IDD ni en la ITD, tal como veremos más adelante.

## **Batidos binaurales**

Una notable característica de nuestra audición, en la que se comprueba la subsistencia de la información de fase más allá del sistema auditivo periférico, se revela con la aparición de batidos binaurales. Éstos se manifiestan al presentar dos señales sinusoidales de frecuencias apenas diferentes ( $\Delta f < 20$  Hz), una en cada oído. La existencia de los batidos binaurales depende de la interacción de la información de las señales que parten de ambos oídos a niveles superiores de procesamiento neural. Proveen una demostración irrefutable de que la descarga de las neuronas preserva la información de fase del estímulo.

Los batidos binaurales son menos nítidos que los batidos monoaurales, con las dos sinusoides del estímulo presentes en un solo oído. Se los percibe sólo por debajo de los 1,000 Hz, a diferencia de los monoaurales que se oyen en todo el rango de frecuencias audibles. Además, estos últimos se perciben mejor cuando las intensidades de las dos sinusoides son similares. Los batidos binaurales, por el contrario, se oyen aun con grandes diferencias de intensidad entre los estímulos (la sensibilidad a la fase existe dentro de un gran rango de intensidades), y un autor hasta sostuvo que una de las señales puede quedar por debajo del umbral de audibilidad. Estudios recientes no han corroborado esta afirmación.

## 2. Localización de señales no sinusoidales

Los sonidos reales, a diferencia de las sinusoides teóricas, tienen comienzo y fin. Como sabemos, todo cambio temporal en una señal acústica va asociado a cambios en su espectro, tanto más profundos cuanto más rápido ocurre la transición temporal. Precisamente, estas irregularidades espectrales aportan nueva información útil para localizar fuentes acústicas. El sistema auditivo las aprovecha de dos maneras diferentes: un primer mecanismo compara el espectro presente a la entrada del sistema auditivo con espectros previos o con espectros almacenados en la memoria de largo término. Como esta operación puede realizarse con un solo oído involucra un proceso monoaural. El segundo mecanismo compara las características espectrales de las señales que llegan a ambos oídos y es, obviamente, binaural.

### Localización monoaural: el rol del pabellón auricular

En 1969 Butler sugirió que el pabellón auricular provee información sobre la localización vertical y sobre la discriminación entre los campos acústicos anterior y posterior. Es decir, permite localizar fuentes ubicadas en el plano medio, imposibles de resolver con la IID y la ITD (Butler, D., 1969).

Hoy está aceptado que el pabellón auricular modifica el espectro del estímulo en función del ángulo de incidencia de la onda en relación con el eje medio de la cabeza. La acción del filtro se mide comparando el espectro del estímulo exterior con el espectro de la señal en la entrada del canal auditivo externo. La razón, expresada en dB, se conoce como la función de transferencia de la cabeza (HRTF) [1] y se puede apreciar en la figura 8.

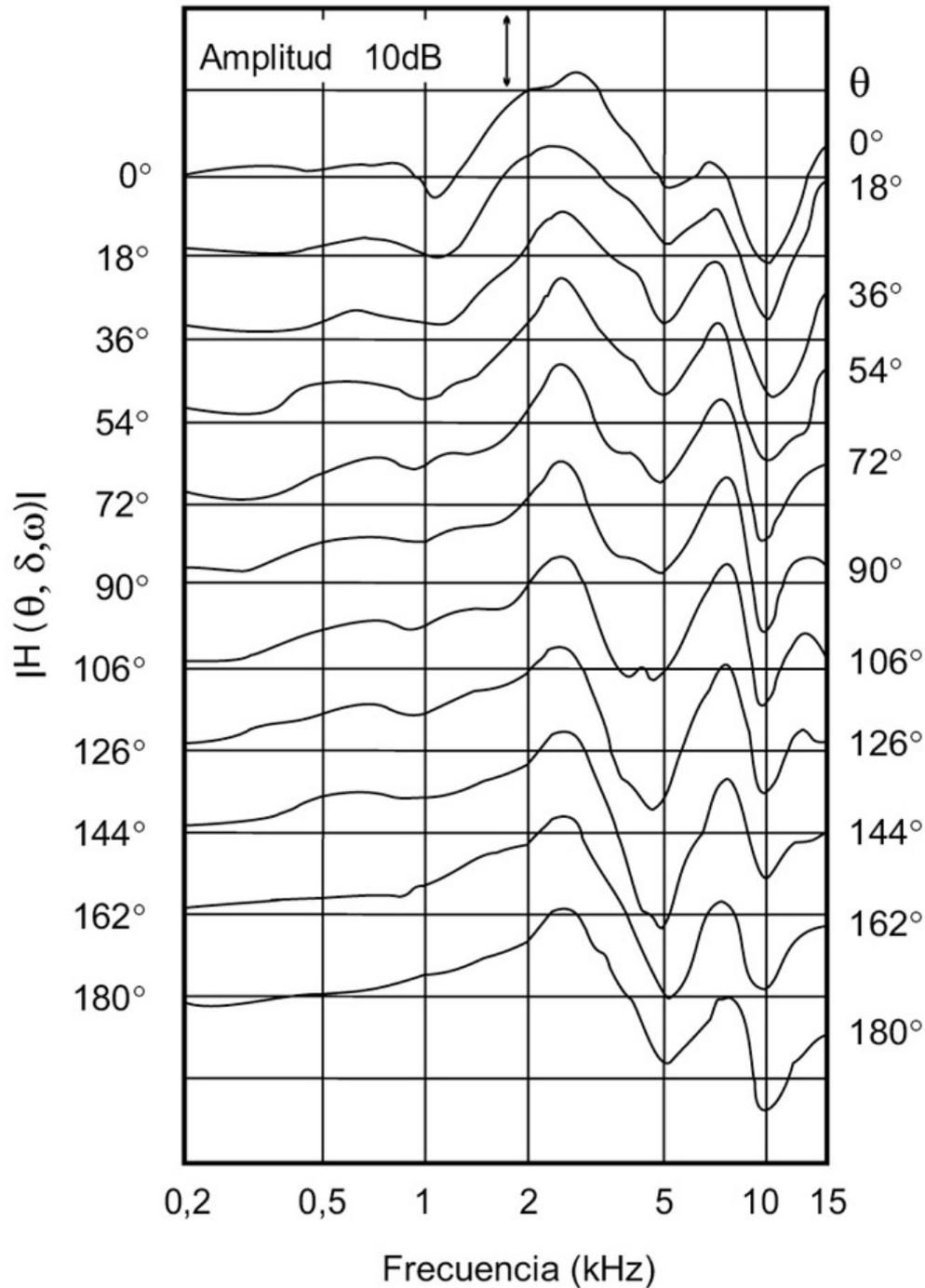
La información provista por el pabellón auricular, dado su tamaño –unos 27 cm<sup>2</sup>–, es efectiva a frecuencias por encima de los 6 kHz. Entre 500 Hz y 6 kHz la cabeza, actuando como pantalla, funciona de manera similar.

En experimentos con ruido de banda angosta, la elevación aparente deriva de la coincidencia entre los picos de la HRTF y la frecuencia central del ruido. Por ejemplo, un ruido diótico centrado en 8 kHz se oye “arriba” al margen de su ubicación real, pues en esa posición la función de transferencia posee un pico de amplitud.[2]

Si el oyente desea hacer un uso eficiente de los datos espectrales asociados con la dirección de la fuente acústica, es necesario que distinga entre los picos y valles relacionados con la dirección (filtros HRTF) y los propios de la fuente y de las reflexiones generadas en las superficies cercanas. Esto significa que el conocimiento previo de la fuente y de las condiciones acústicas del ambiente es relevante. Sin embargo, como cada pabellón auricular provee filtros diferentes y conocidos, se los puede separar de los datos espectrales externos.[3] G. Plenge presentó evidencia de que cuando un sujeto no está familiarizado con las características de la fuente y del ambiente acústico la localización monoaural se degrada (Plenge, G., 1974). Pero en compensación se requieren muy pocos segundos para entrar en la nueva situación. Esta *acomodación espacial* es de gran importancia en la audición y ejecución de música en vivo.

## **Empleo de datos espectrales en la localización binaural**

En señales de espectros de banda las diferencias interaurales de tiempo proveen información para localizar la fuente que no está sujeta a las ambigüedades de fase típicas de las sinusoides. Si el espectro cubre un gran rango de frecuencias, las ambigüedades de fase pueden ser superadas a través de comparaciones entre los canales del filtro auditivo involucrados. D. W. Grantham consideró que la diferencia interaural de tiempo debe definirse a partir de señales de espectro de banda y propuso llamarla *verdadera* ITD (Grantham, G., 1995).



**Figura 8.** Funciones de transferencia desde el campo externo hasta la entrada del canal auditivo externo (HRTF) con el ángulo horizontal  $\vartheta$  como parámetro

A continuación mencionamos algunos experimentos en los que se destaca la aplicación de datos espectrales en la localización binaural.

### ***Localización de señales transitorias***

Klump y Eady midieron los umbrales para la discriminación de la ITD para tres tipos de estímulos cortos. Fueron necesarios 9  $\mu$ s para detectar un cambio en el ITD ante un estímulo de ruido de banda angosta de 1,4 s de duración; 11  $\mu$ s para sinusoides de 1.000 Hz recortadas por una ventana rectangular de 1,4 s; y 28  $\mu$ s para clics de 1 ms de duración (téngase en cuenta que 10  $\mu$ s corresponden a un desplazamiento lateral de 10). Los umbrales de ITD para trenes de impulsos repetidos fueron del orden de los 10  $\mu$ s (Klumpp y Eady, 1956).

### ***Localización en función de la frecuencia***

La discriminación de la posición lateral sobre la base de los retardos temporales entre los dos oídos depende principalmente del contenido de baja frecuencia de señales de banda ancha, por ejemplo clics. Por debajo de 1.500 Hz las fases de las oscilaciones de decaimiento en la onda de los dos oídos pueden compararse, proveyendo información sobre los tiempos de llegada relativos de los clics. Para frecuencias por encima de 1.500 Hz la información de la estructura fina del decaimiento se pierde. Sólo queda la referida a la envolvente de decaimiento.

### ***Importancia de las disparidades en el cuerpo de la señal***

Las disparidades en el cuerpo de una señal compleja (que implican la percepción de la diferencia entre fases) son más importantes que las del ataque, aun para pulsos de 10  $\mu$ s de duración. Sólo para impulsos muy cortos prevalece la disparidad del ataque (que en este caso es la única disponible, pues el cuerpo queda enmascarado bajo el ruido de banda ancha provocado por la intervención del principio de indeterminación acústico).

### ***Correspondencia tiempo-intensidad***

Si se presentan clics idénticos por medio de auriculares, la imagen se percibe “en el centro” de la cabeza. Si a continuación se retarda el clic del oído derecho, por ejemplo 100  $\mu$ s, la imagen se desplaza hacia la izquierda. Pero si la intensidad de la señal en el oído derecho aumenta, es posible colocar la imagen nuevamente en el centro de la cabeza. La cantidad de tiempo necesario para compensar 1 dB de diferencia interaural se denomina *razón de correspondencia*. Este fenómeno apoya la teoría que establece que las diferencias en tiempo y en intensidad son codificadas en el sistema nervioso de manera similar, y es de gran importancia para la implementación de sistemas virtuales de espacialización.

### ***Adaptación binaural***

El sistema auditivo binaural parece procesar sólo el ataque del conjunto en un tren de pulsos. Los clics luego del primero aportan poca información para la localización. E. R. Hafter presentó evidencia de que un disparo adecuado –un pulso de ruido de baja intensidad o un hueco (*gap*) en el tren de clics– puede desactivar la adaptación binaural. En su experimento los clics que aparecen luego de la perturbación son más efectivos que los anteriores –adaptados– para localizar espacialmente la fuente. Hafter y Buell dicen: “[...] El sistema auditivo se vuelve insensible a los datos interaurales luego del ataque, a la vez que monitorea el entorno a la espera de nuevas condiciones. Cuando algo inesperado ocurre, el entorno espacial es nuevamente examinado” (Hafter y Buell, 1990).

### ***Interferencia binaural***

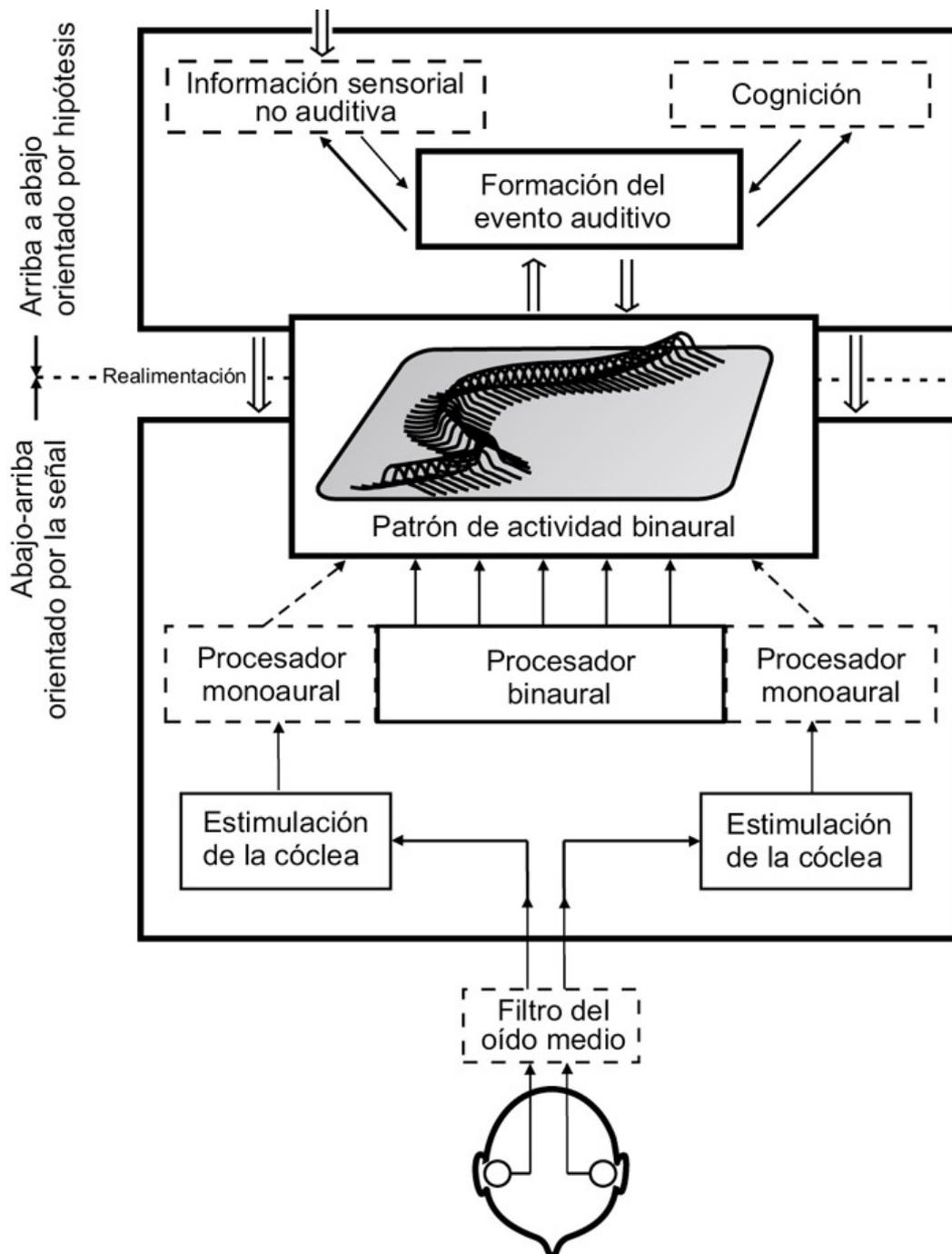
Los ruidos de baja frecuencia degradan fuertemente la lateralización de las señales de alta frecuencia, mientras que la recíproca no es cierta. Una interpretación de la interferencia binaural establece que ésta resulta de la fusión de la señal y del ruido en una imagen auditiva única, no jerarquizada.

### 3. Modelos de procesamiento espacial

Se han propuesto gran cantidad de modelos que intentan dar cuenta de la multiplicidad de mecanismos presentes en la percepción espacial. En la figura 9 se puede apreciar el esquema de Blauert, que incluye todos los bloques que aparecen en los modelos más detallados (Blauert, J., 1997).

En cuanto a los modelos de aplicación, uno de los más interesantes es el propuesto por Yoichi Ando a fines de la década de 1980. Su esquema se puede ver en la figura 33 del capítulo I.

El modelo de Ando postula la existencia de un mecanismo precortical de extracción de la autocorrelación de las señales que provienen de ambos oídos, la posterior correlación cruzada interaural entre las vías derecha e izquierda, y la especialización de los hemisferios cerebrales para el proceso de factores espaciales o temporales. Para ciertas tareas complejas (como la propia del efecto “cocktail party”) es necesaria la interacción de ambos hemisferios: el izquierdo –temporal– tiene a su cargo el procesamiento de la palabra, mientras que el derecho –espacial– procesa la localización de las fuentes. Este modelo fue utilizado con éxito en el diseño de salas para música no amplificadas, entre ellas el célebre auditorio de Kirishima, en Japón.



**Figura 9.** Modelo esquemático de Blauert para la audición espacial

## 4. El efecto de precedencia

Sabemos que el sentido de la audición, a diferencia de la vista, se concentra en las fuentes acústicas antes que en las reflexiones provo-

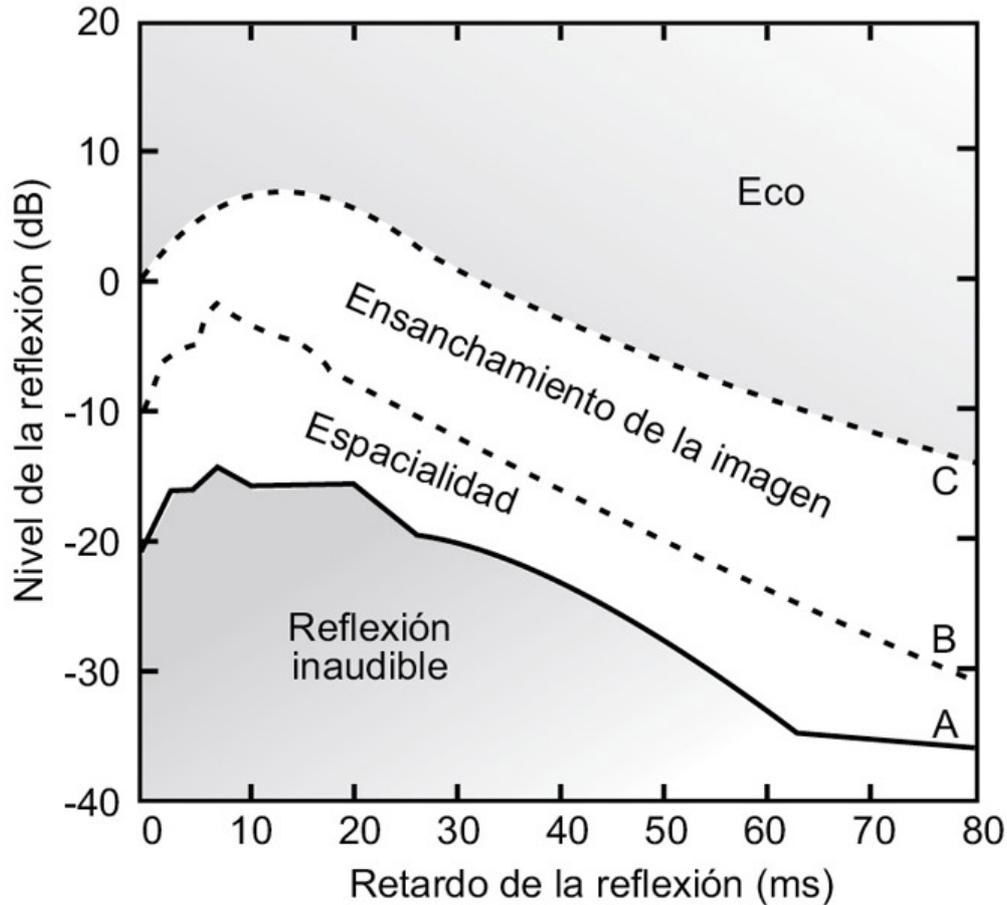
cadadas por los objetos del entorno. Un ejemplo cotidiano ocurre al localizar la ubicación de una fuente acústica en una sala reverberante donde la mayor parte de la energía proviene de reflexiones. Wallach (1949) y Hass (1951) investigaron este fenómeno con dos señales asincrónicas y los resultados pueden sintetizarse en las siguientes reglas.

- Dos sonidos sucesivos se perciben como uno solo si el intervalo temporal entre ellos es lo suficientemente corto. El límite superior es de 5 ms para clics y de 40 ms para sonidos de carácter complejo, como palabra y música.
- Cuando ocurre lo anterior, la locación de la fuente está determinada por la del primer sonido. Este fenómeno se denomina *efecto de precedencia*, o *efecto Hass*.
- El efecto de precedencia aparece sólo para sonidos de carácter discontinuo o transitorio.
- El segundo estímulo puede tener alguna influencia en la localización de la fuente (su máximo aporte es un cambio de 70 en la localización aparente).
- Si el intervalo entre estímulos es de 1 ms o menor, el efecto de precedencia no actúa; se obtiene algún tipo de compromiso entre las ubicaciones espaciales de las dos fuentes. Este hecho se denomina *suma de locaciones* (Blauert, J., 1997).
- Si el estímulo retrasado es lo suficiente intenso, entre 10 y 15 dB por encima del primero, el efecto no ocurre.
- Se asume corrientemente que el efecto trabaja más efectivamente cuando ambos estímulos se parecen. P. Divenyi demostró, sin embargo, que las diferencias espectrales no cuentan si se mantiene la envolvente temporal (Divenyi, P., 1992). Esta conclusión es razonable, pues las reflexiones pueden modificar el espectro de la señal por distorsión armónica.
- El efecto es más notorio en el plano horizontal, aunque también aparece en el plano medio.

- El efecto de precedencia puede tomar algún tiempo en aparecer (muy importante en grandes salas e iglesias).
- Ciertas modificaciones en las condiciones acústicas del ambiente pueden destruir temporariamente el efecto de precedencia.
- La habilidad para detectar cambios en el intervalo temporal del estímulo retrasado puede mejorar notablemente con la práctica.

El efecto de precedencia no involucra una completa supresión de la información de los “ecos”. El oyente puede distinguir cualitativamente cualquier cambio en el patrón temporal de los estímulos. Es más, este patrón aporta información sobre la sala y la posición de los objetos y paredes en ella. El efecto de precedencia seguramente aparece en un nivel relativamente alto del proceso cognitivo. Parece que los posible “ecos” son suprimidos después de haber sido evaluados como consistentes con las expectativas del oyente en relación con la fuente acústica y al ambiente. Cuando ocurre un cambio inesperado la máscara del efecto de precedencia puede retirarse momentáneamente. En el gráfico de la figura 10 se puede ver con claridad la región de validez del efecto de precedencia.

En la región superior las reflexiones superan los límites del efecto de precedencia y se las percibe como ecos aislados, en la parte intermedia las reflexiones aportan información espacial del entorno acústico, y en la región inferior las reflexiones se vuelven inaudibles. Cuando las reflexiones poseen retardos iguales o menores a 1 ms aparece el fenómeno de *filtro en peine*, que se oye como una especie de coloración espectral.[\[4\]](#)



**Figura 10.** Región de validez del efecto de precedencia

## Refuerzo electroacústico en grandes espacios

El efecto de precedencia se emplea para preservar la localización espacial de la fuente física cuando se emplean varios altavoces separados por distancias significativas. Por ejemplo, en el sistema de dos fuentes electroacústicas de la figura 11, el oyente percibiría antes y con mayor intensidad al altavoz 2. El altavoz 1, cercano a la fuente física, se manifestaría como un eco.

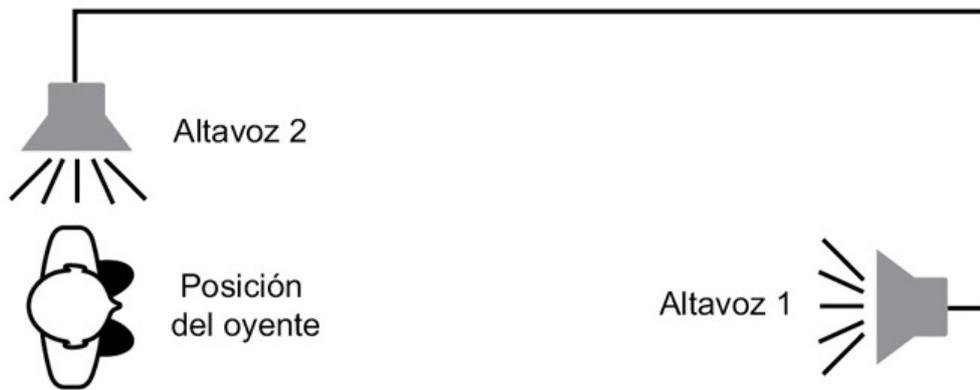
La secuencia temporal que le llega al oyente se aprecia en la figura 12. Es evidente que percibirá como fuente acústica primaria al altavoz 2.

La solución a este problema consiste en retardar la señal del segundo altavoz hasta ubicarla dentro del límite de Hass. El esquema

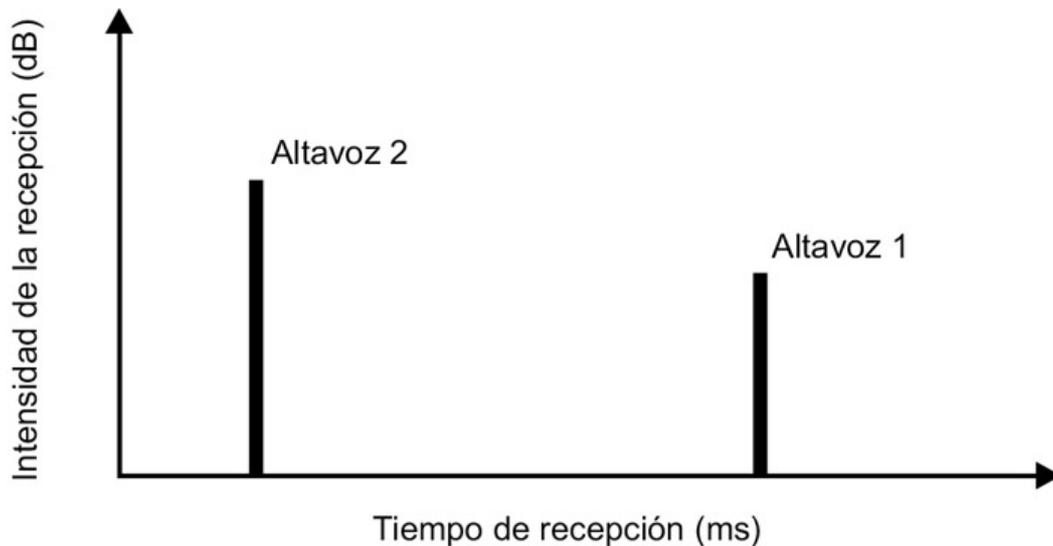
del dispositivo se muestra en la figura 13.

El resultado final se puede evaluar en el gráfico temporal de la figura 14.

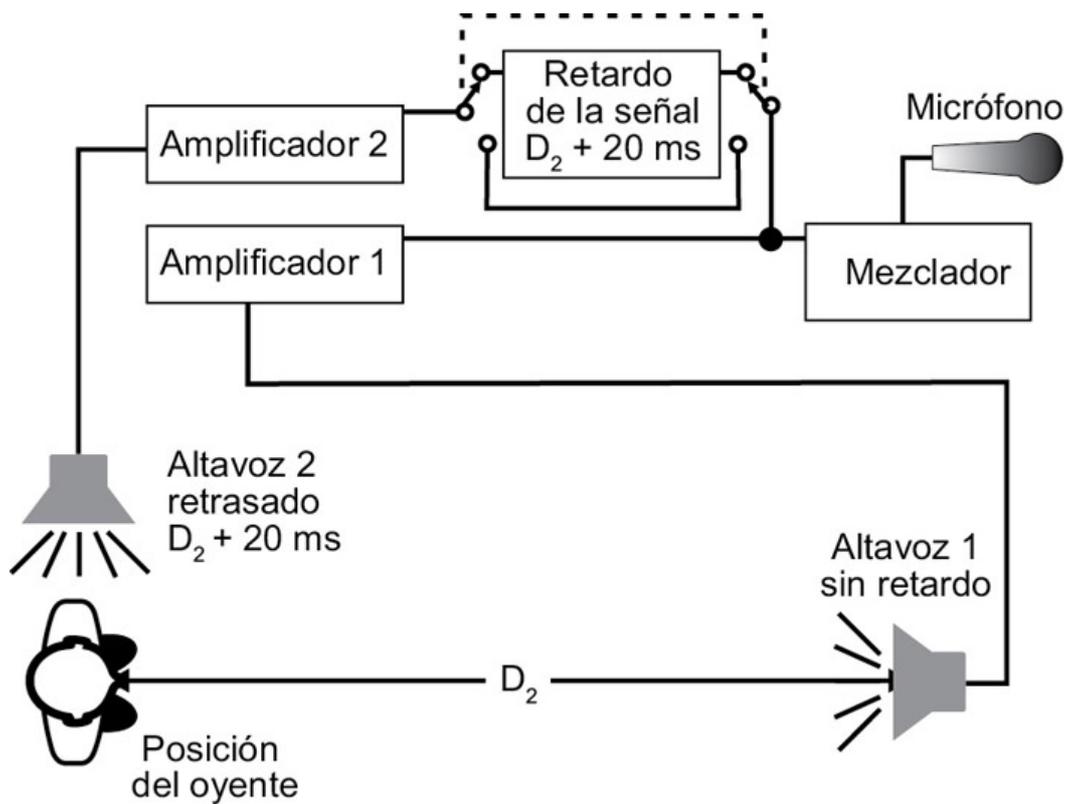
Ahora se recibe primero la señal del altavoz 1. La señal del altavoz 2 ocupa una posición que cae dentro de los límites establecidos por el efecto de precedencia. Para el oyente la locación de la fuente queda establecida por el primer sonido que le llega, que es el que emite el altavoz 1.



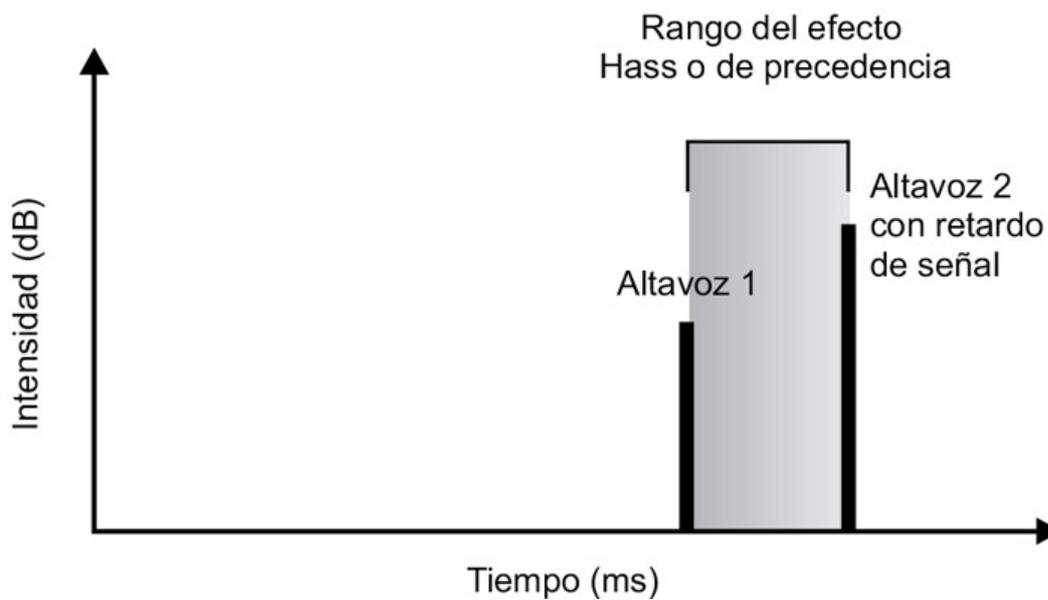
**Figura 11.** Sistema de refuerzo electroacústico con dos altavoces distantes entre sí



**Figura 12.** Diagrama temporal en el punto de recepción a partir del sistema de refuerzo electroacústico con dos fuentes



**Figura 13.** Aplicación del efecto de precedencia a un sistema de refuerzo electroacústico



**Figura 14.** Diagrama temporal en el punto de recepción luego de instalar el dispositivo de retardo

## 5. Percepción de la distancia

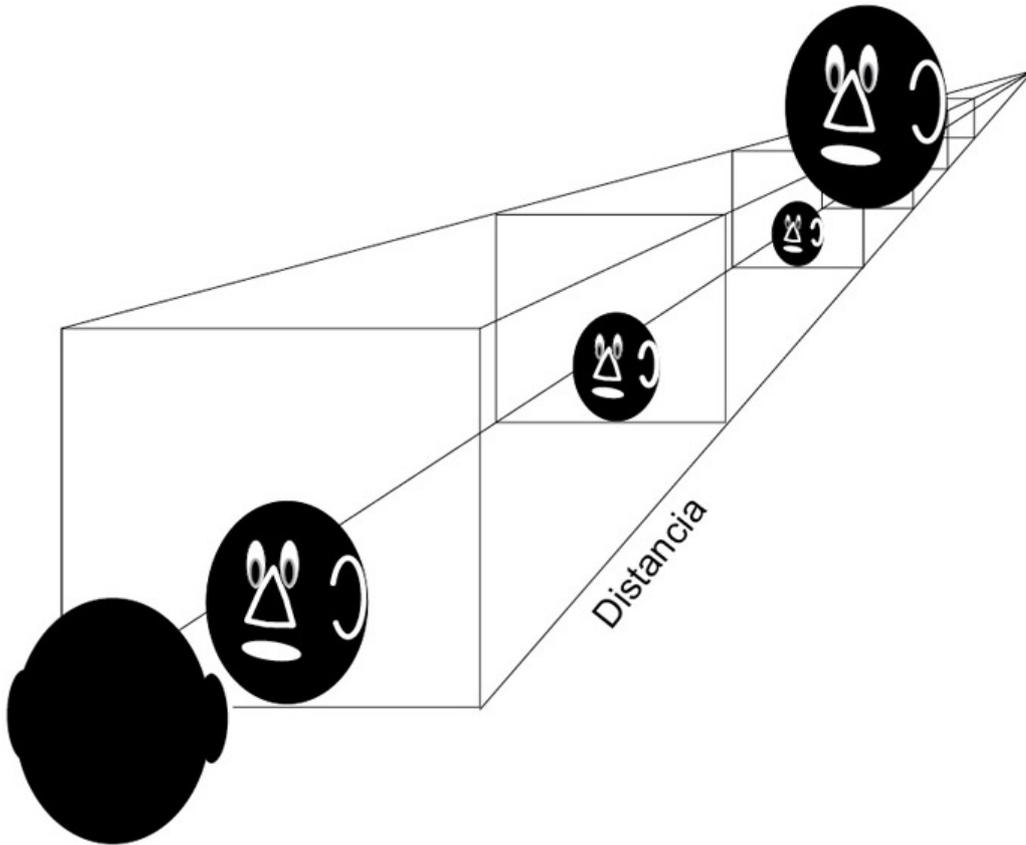
Hasta ahora analizamos los mecanismos empleados para determinar la dirección de llegada de las ondas acústicas desde una fuente externa. Nos queda por ver de qué manera evaluamos la distancia a la que se encuentra dicha fuente. En el sistema de coordenadas esféricas de la figura 1, al comienzo de este capítulo, nos faltaría determinar la distancia  $r$  a la que se encuentra la fuente.

### Perspectiva auditiva

La perspectiva acústica, al igual que su contraparte visual, es una importante fuente de información ambiental. Las imágenes acústicas de gran desarrollo en azimut se reducen a un punto a grandes distancias. Por ejemplo, el *ancho acústico aparente* de una orquesta sinfónica, muy grande en las primeras butacas de la platea, disminuye a unos pocos grados al oírla en un parque a 300 m de distancia. [5] El efecto es equivalente al de las líneas paralelas que convergen en el infinito. Si continuamos desarrollando la analogía con el sentido de la vista, se podría decir que la sonoridad disminuye con la distancia tal como ocurre con el tamaño aparente de los objetos y que la definición tímbrica se atenúa de modo similar al gradiente de color en la visión. [6] Aunque no siempre estudiada en detalle, la distancia aparente de la fuente acústica que define la *perspectiva auditiva* compone una importante dimensión psicoacústica. Los investigadores que desarrollan dispositivos acústicos para el cine y sistemas multimedia han concentrado gran parte de su interés en el perfeccionamiento de modelos virtuales de perspectiva auditiva.

La sonoridad, cuando interviene como un parámetro de entrada para evaluar la distancia de las fuentes acústicas, deja de ser un correlato unidimensional de la intensidad física para convertirse en un rasgo complejo que involucra más de una dimensión. Los objetos en el mundo son, en general, de tamaño constante, pero la imagen de

un objeto en la retina crece y decrece cuando éste se acerca o se aleja de nosotros. Este hecho es conocido como *constancia de tamaño*. En la figura 15 se puede apreciar una ilustración clásica de la constancia visual de tamaño.



**Figura 15.** Constancia visual de tamaño. La cabeza más cercana al observador tiene el mismo tamaño que la “gran cabeza” alejada

En el dominio de la audición la *constancia de la sonoridad* funciona de modo análogo a la constancia de tamaño para el sentido de la vista. Si una fuente acústica de intensidad constante se aleja de nosotros a campo libre, la intensidad física que nos llega disminuye proporcionalmente con el cuadrado de la distancia. No debe sorprendernos entonces que una fuente cuya intensidad decrece parezca alejarse. Es decir, también podemos experimentar la sensación de oír una fuente acústica que disminuye de intensidad sin alejarse –un

instrumento musical en *decrescendo*–, cuyo equivalente en el campo visual podría ser un globo que se desinfla.

Imaginemos un experimento: un oyente observa a dos cantantes, una cercana y la otra alejada. La cantante cercana produce un sonido pp, mientras que la otra emite el mismo sonido pero en ff. Los sonidos poseen la misma altura tonal y duración. Si al oyente se le pide que evalúe la sonoridad de ambas cantantes, dirá que la más alejada supera a la cercana. Suponiendo que la intensidad del pp sea  $1/128$  de la intensidad del ff, y que por estar a 50 m la intensidad de la cantante alejada llega con  $1/2.500$  ( $1/d^2 = 1/502$ ) de la emitida, la onda cercana posee alrededor de 20 veces más energía que la alejada. Sin embargo, el oyente insistirá con su afirmación. En la figura 16 se puede ver este experimento.

El experimento anterior muestra que existen otros datos no tomados en cuenta, diferentes de la sonoridad, que nos aportan información sobre la intensidad física y la distancia de las fuentes acústicas. Hasta aquí no tomamos en cuenta el cambio espectral operado en las señales de las cantantes en función del esfuerzo vocal. En general, la envolvente espectral cambia de forma –para favorecer a los armónicos más agudos– a medida que la intensidad y el esfuerzo vocal aumentan.

La figura 17 sintetiza los trabajos de John Chowning en una cantidad de experimentos similares al expuesto (Chowning, J., 1999). Representa una generalización del cambio de la composición espectral en función de la altura tonal, la dinámica y la distancia. A causa de la gran cantidad de dimensiones necesarias para describir los datos físicos se han incluido espacios bidimensionales (espectros de potencia) dentro de un espacio tridimensional cerrado.

Es posible ahora comprender al oyente del experimento: no juzgaba la distancia y la sonoridad de la fuente únicamente a partir de la intensidad que percibía. En la figura 18 se esquematiza una situación simétrica a la de la figura 16.

Una fuente de información adicional la proporciona la *reverberación* del recinto. La relación entre la intensidad de la reverberación y la onda directa es interpretada por el oyente como un índice de distancia. En un espacio cerrado típico la intensidad del campo reverberante no varía con la distancia a la fuente, mientras que el sonido directo sí lo hace. Sabemos, por experiencias previas, que el nivel del campo reverberante depende de la intensidad de la fuente que lo excita y usamos este conocimiento para evaluar la intensidad y distancia de las fuentes acústicas en recintos cerrados.

En la figura 19 se comparan dos señales inmersas en el mismo campo reverberante. La señal A se ubica a cinco unidades de distancia del oyente y posee una intensidad relativa seis veces menor que la señal B, a una distancia de dos unidades. Ambas llegan al punto de recepción con la misma intensidad directa. Sin embargo, la relación con el campo reverberante aporta la información necesaria para evitar la confusión de distancias sin necesidad de apelar a diferencias espectrales.

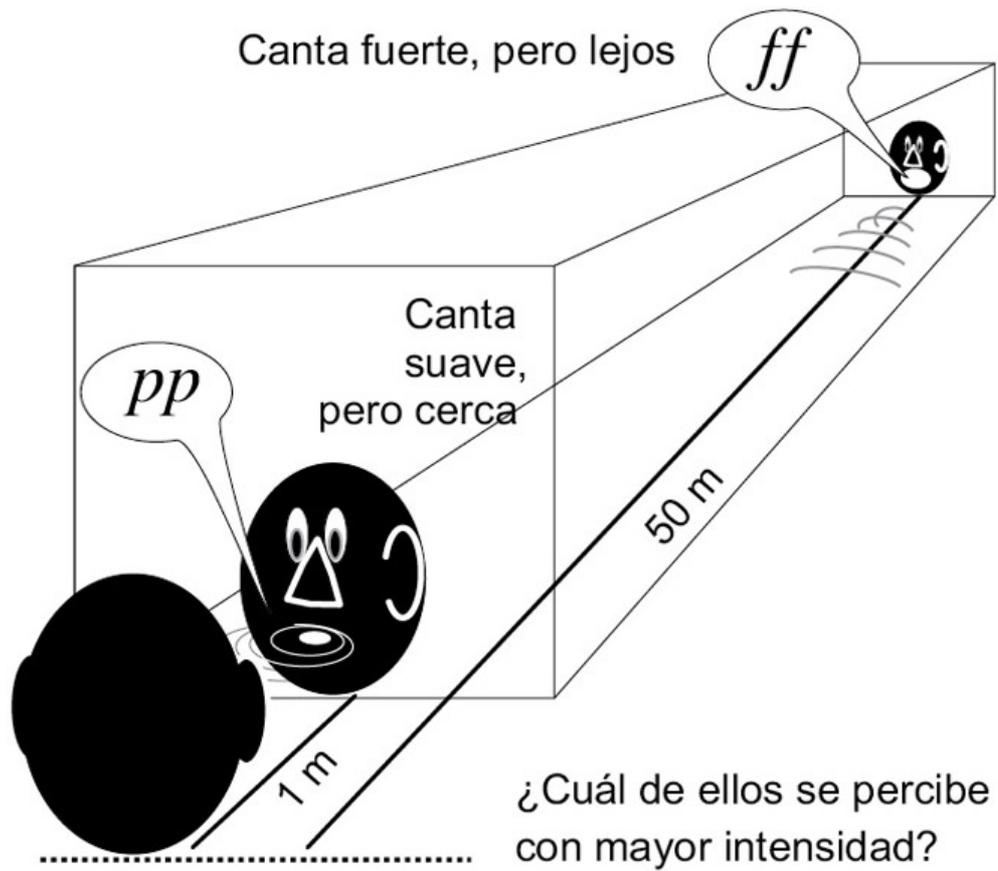
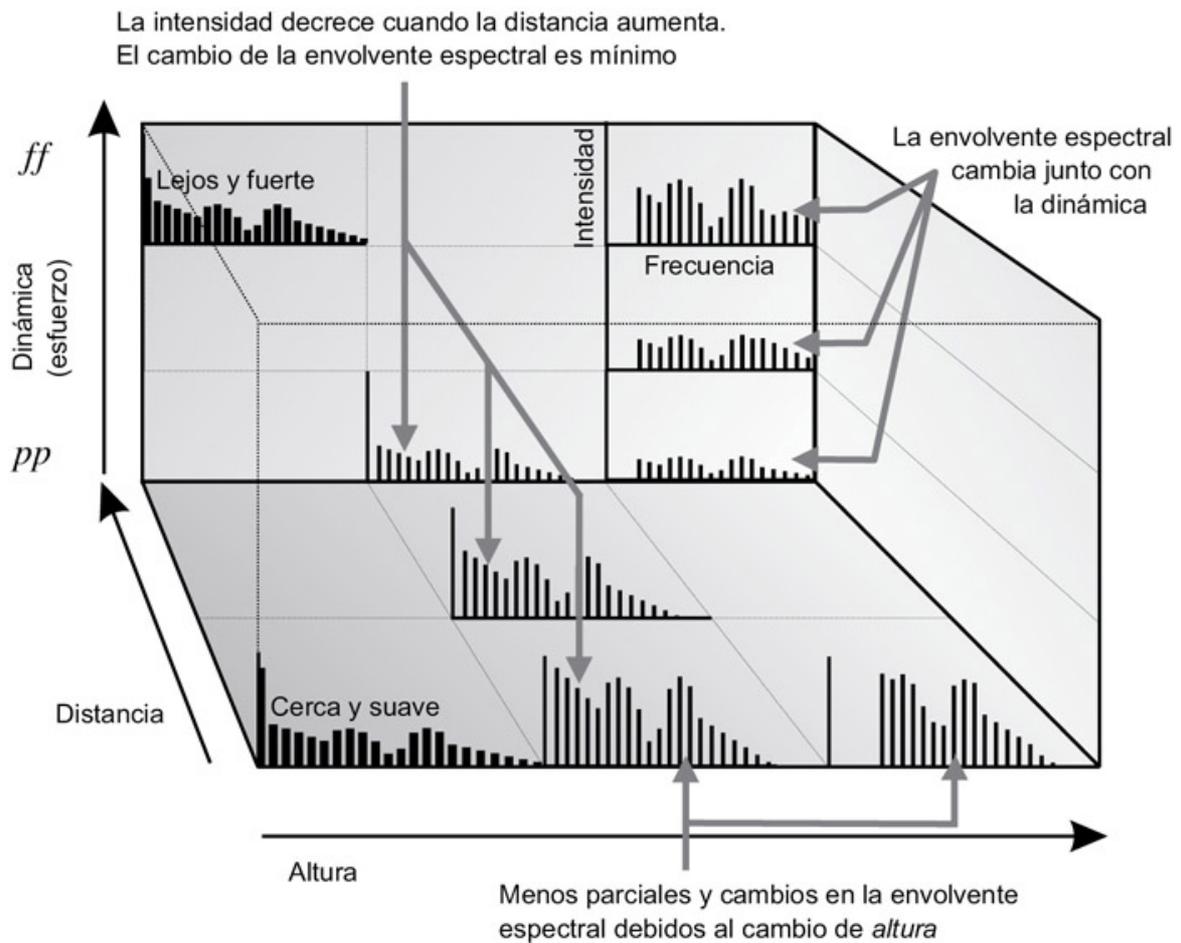
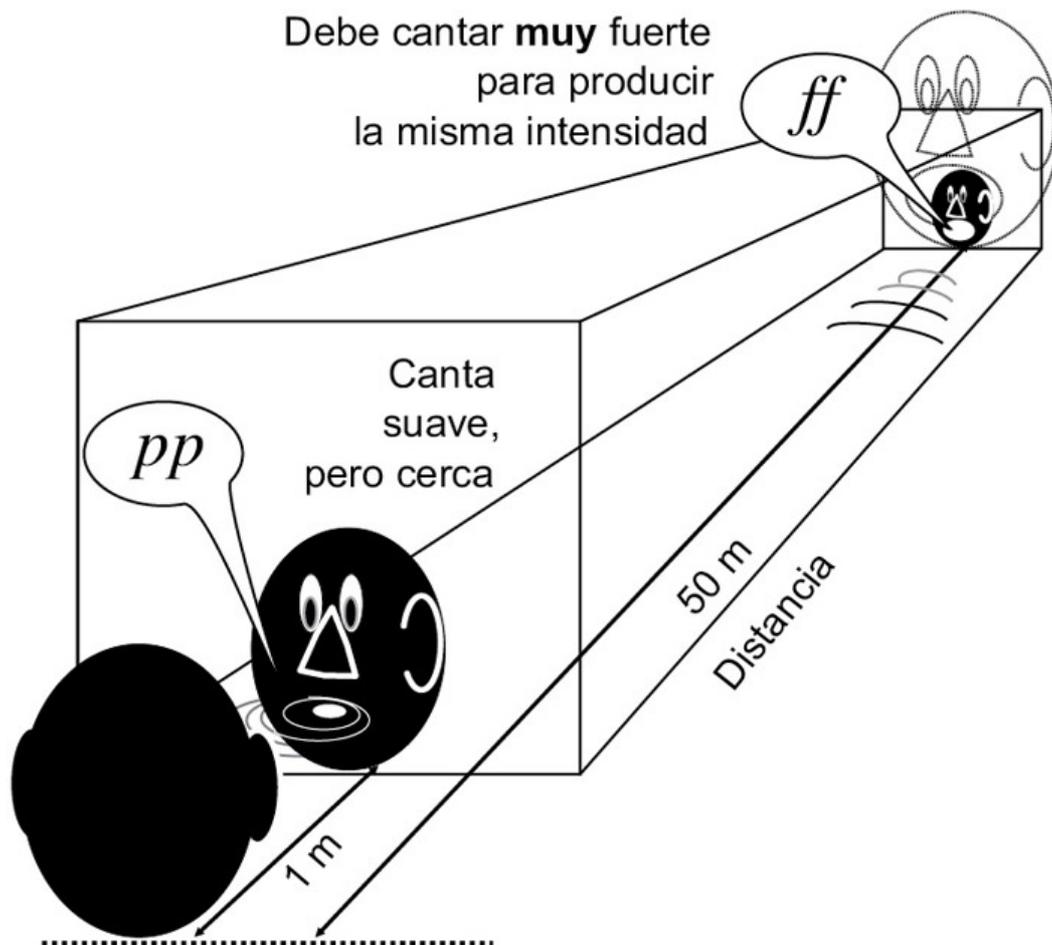


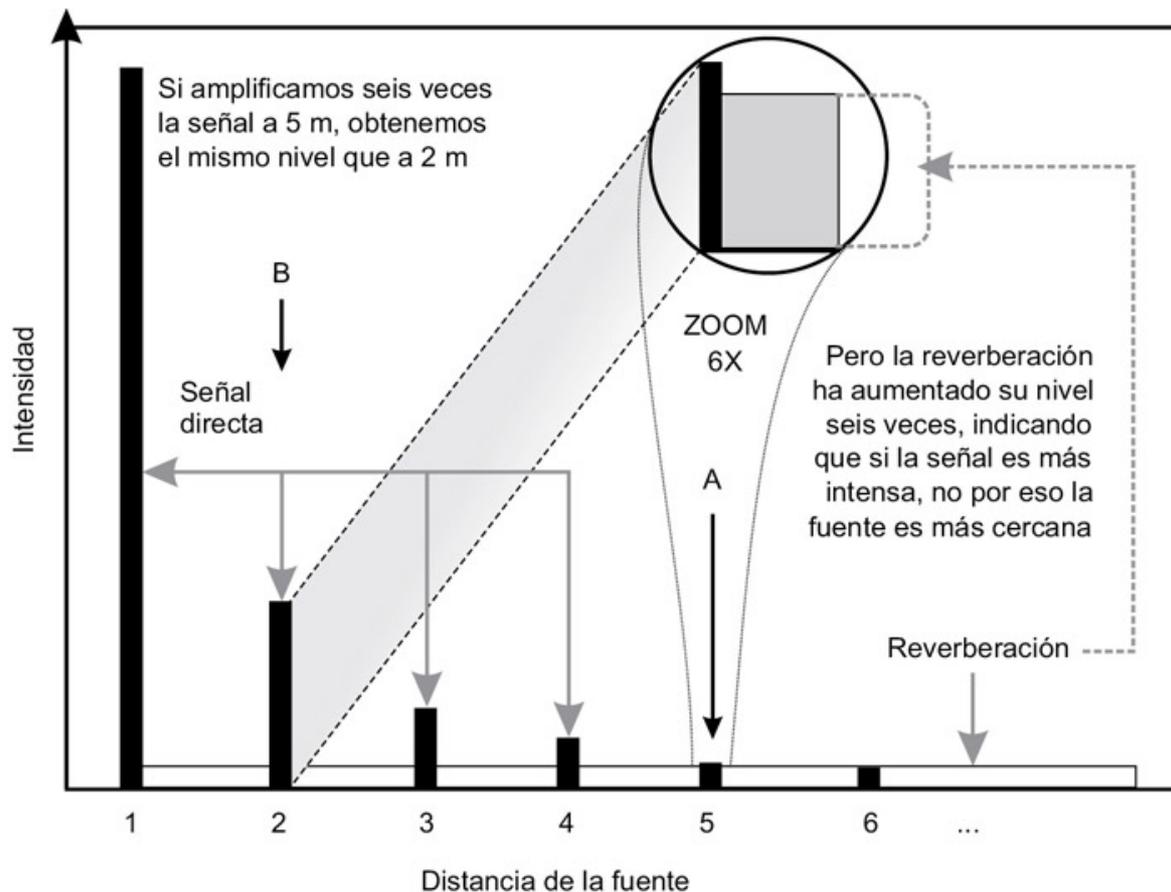
Figura 16. Constancia de sonoridad



**Figura 17.** Espacio distancia-intensidad-frecuencia según Chowning



**Figura 18.** Percepción de la distancia: los sonidos de ambas cantantes poseen para el oyente la misma intensidad, pero la composición espectral le permite al oyente saber que la más alejada canta en *ff*



**Figura 19.** Razón señal directa/campo reverberante en función de la distancia

Aunque la analogía con el sentido de la vista nos ha resultado útil hasta ahora, debe señalarse que la constancia de sonoridad es más compleja que la constancia de tamaño: pocos objetos visuales varían de tamaño ante nuestros ojos (como en el ejemplo de un globo que se desinfla), pero las fuentes acústicas varían normalmente de intensidad en lapsos muy breves, especialmente en música. Por eso quizá el mecanismo de percepción asociado a la constancia de sonoridad necesite información relacionada con índices espectrales y con la interpretación de la relación señal directa/señal reverberante. La evaluación aislada de la intensidad parece no alcanzar.

En general, los juicios auditivos sobre la distancia a la que se encuentra una fuente acústica son relativamente imprecisos, y los erro-

res con relación a la distancia física son del orden del 20% en situaciones normales.

Es interesante describir lo que ocurre cuando los oyentes son privados de algunos de los datos mencionados. Por ejemplo, en el campo libre no actúa el sistema de evaluación de la perspectiva relacionado con la reverberación y la precisión de los juicios con relación a la distancia se reduce significativamente. Estas situaciones se han estudiado en experiencias de laboratorio, con fuentes sinusoidales ubicadas en cámaras anecoicas, en las se preserva solamente la información relacionada con la intensidad física.

## 6. Percepción del movimiento

La habilidad para percibir el movimiento lateral de una fuente acústica se mide empleando el *mínimo ángulo de movimiento audible* (MAMA). Los experimentos han mostrado que el sistema auditivo es poco sensible al movimiento de las fuentes acústicas, aunque se comporta mejor ante movimientos lentos. Para desplazamientos angulares del orden de los 150/s, el MAMA es de 50. Pero para movimientos angulares veloces, del orden de los 900/s, el MAMA crece hasta llegar a 210.

El *efecto Doppler*, responsable del cambio de altura en las sirenas de las ambulancias cuando pasan a nuestro lado, es una importante fuente de información sobre el cambio en el sentido del movimiento relativo entre el emisor acústico y el oyente. Es un fenómeno físico que hace que la frecuencia aparente en el punto de recepción aumente si éste se acerca a la fuente y que disminuya si se aleja de ella. En la figura 20 se ilustra el caso con una fuente acústica en movimiento hacia un receptor en reposo.

Si la fuente se encuentra quieta la onda acústica llenará la distancia  $FR$  que la separa del receptor en un tiempo  $\Delta t = FR/c$ , con  $c$  como la velocidad del sonido en el aire. La longitud de onda en este caso es  $\lambda = FR / f \Delta t$ . Pero si la fuente se desplaza hacia el receptor a velocidad  $vF$  la onda acústica recorrerá en el mismo tiempo

una distancia menor  $F'R$ . La fuente se habrá movido en ese lapso una distancia  $F'R - FR = v_F \Delta t$  y la nueva longitud de onda será  $\lambda' = F'R / f \Delta t$ . Tomando en cuenta la relación general  $v = \lambda f$  y luego de algunas operaciones algebraicas sencillas se tiene que:

$$f' = \frac{c - v_F}{c} f$$

En la ecuación anterior  $f$  es la frecuencia de la señal emitida por la fuente y  $f'$  es la frecuencia aparente en el punto de recepción. Si se considera también la posibilidad de movimiento del receptor a velocidad  $v_R$  la ecuación se modifica levemente:

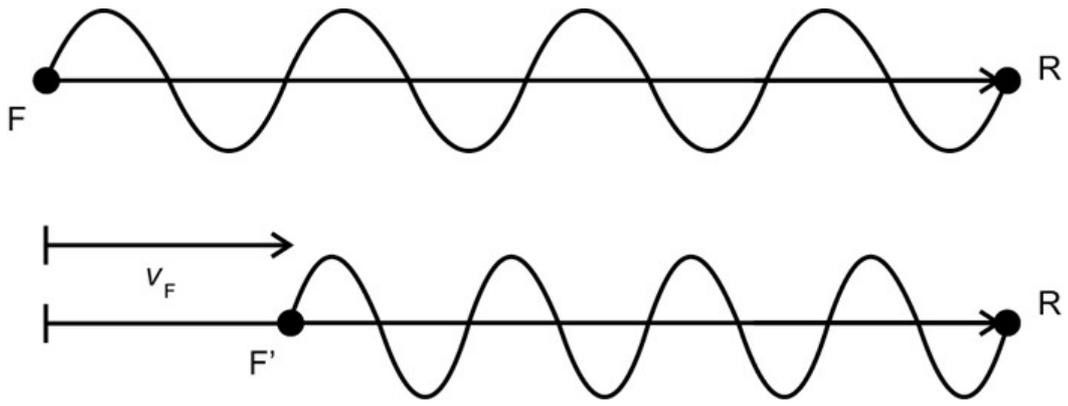
$$f' = \frac{c - v_F}{c - v_R} f$$

El efecto Doppler se emplea desde hace tiempo en música electroacústica a partir de módulos específicos de procesamiento. Casi no es necesario comentar la importancia que tiene una buena simulación Doppler en el caso de los efectos especiales para cine y multimedia.

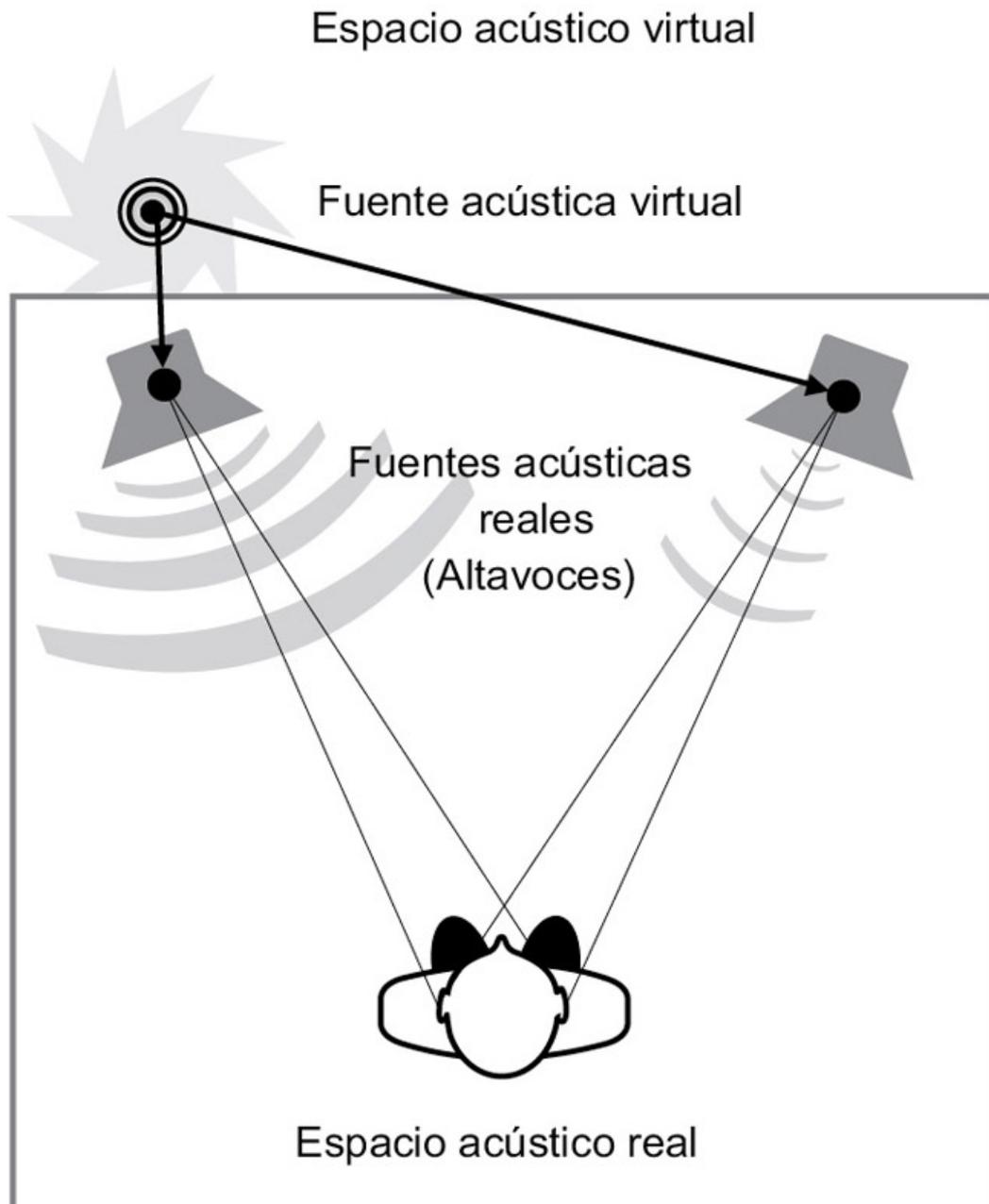
En cuanto a la percepción del efecto, se reduce a la percepción de una variación continua de frecuencia, que ya examinamos en el capítulo III.

## 7. Campos acústicos virtuales

En un campo acústico habitual, con el oyente inmerso en su interior, la información espacial depende de los conocidos datos espectrales, temporales (ITD) y de intensidad (IID) que aportan las señales que le llegan a ambos oídos. En un sistema electroacústico de reproducción, sin embargo, a la información anterior se suman las referencias espaciales codificadas en la señal grabada.



**Figura 20.** Efecto Doppler con una fuente  $F$  en movimiento a velocidad  $v_F$  hacia un receptor  $R$  en reposo



**Figura 21.** Campos acústicos definidos por un sistema estereofónico bicanal

Por ejemplo, el registro de un concierto en un gran auditorio contiene información sobre la reverberación y las características espectro-temporales de la sala. Asimismo, es función de los vínculos espaciales entre cada instrumento y cada uno de los micrófonos utilizados. Cuando esta grabación se reproduce en otro recinto a través de un sistema estereofónico bicanal, el campo acústico que percibe el oyen-

te se desdobra entre: a) un *campo acústico real*, que es en el que está inmerso y que detecta a partir de los sonidos que él mismo produce —como voz hablada, ruidos de pisadas, etc.— y b) un campo acústico virtual generado por la fuente electroacústica a partir de las referencias codificadas en el registro original. En la figura 21 se puede ver un esquema que ilustra dicha situación.<sup>7</sup>[7]

La situación ideal sería que el campo acústico real no interfiriese negativamente con el espacio acústico virtual que define la grabación. Pero esto ocurre por lo general en salas de audición cuidadosamente diseñadas, por ejemplo una sala de control de un estudio de grabación. Lo habitual es que el campo acústico de la sala real degrade la información espacial que contienen las señales electroacústicas. Aunque en el ejemplo nos referimos a un sistema estereofónico bicanal, lo mismo ocurre con los demás sistemas de espacialización electroacústicos que existen en el mercado, tales como las diferentes versiones conocidas como *surround* o las que se derivan de la codificación Blumlein original.<sup>8</sup>[8]

## 8. Resumen

Nuestra habilidad para localizar sonidos es muy buena en el plano horizontal, algo menor en el plano vertical (elevación) y regular para la distancia.

La localización espacial es el resultado del procesamiento combinado de datos. Para fuentes acústicas corrientes los datos descritos en este capítulo son accesibles simultáneamente, al menos en circunstancias normales. Es esta riqueza de información, muchas veces redundante y reforzada por datos visuales, lo que hace la localización más precisa y confiable.

Un capítulo aparte podría dedicarse al diseño de campos acústicos virtuales. Todos los implementados hasta ahora, desde los dispositivos monofónicos hasta los actuales algoritmos de espacialización para cine y multimedia, pasando por la estereofonía bicanal, hacen uso de las citadas capacidades auditivas. En algunos casos, la con-

tradicción entre datos de entrada (por ejemplo, una IID que corresponde a una fuente a la izquierda del oyente con una de ITD relacionada con la misma fuente ubicada a la izquierda) puede degradar e incluso destruir la localización auditiva.

## Notas

- [1](#) Del inglés *Head related transfer functions*.
- [2](#) Cuando las señales que llegan a los dos oídos son idénticas se dice que el estímulo es *diótico*
- [3](#) Un ejercicio interesante consiste en “aplastar” los pabellones contra la cabeza y salir a dar una caminata. Es notable como se altera nuestra percepción acústica del espacio al modificar levemente los filtros HRTF.
- [4](#) *Comb filter* en inglés (Basso, 2001a).
- [5](#) El parámetro subjetivo empleado en acústica de salas se denomina “ancho aparente de la fuente” (AWS), bien correlacionado con el factor objetivo de espacialidad (1-IACC).
- [6](#) La analogía con el sentido de la vista, aunque inexacta, resulta útil para ejemplificar los diferentes aspectos implicados en la percepción de la perspectiva auditiva.
- [7](#) Por supuesto que el término “real” está empleado aquí como una etiqueta práctica, sin connotaciones epistemológicas. En cierto sentido, por supuesto, ambos campos acústicos son “reales” a su manera.
- [8](#) Por ejemplo, Dolby-Surround, Dolby-Digital, THX, Ambisonics y muchos otros.

# Capítulo VIII

## Percepción de objetos auditivos

Desde Aristóteles, la percepción se entiende como el proceso a través del cual utilizamos la información provista por nuestros sentidos para armar representaciones mentales del mundo que nos rodea. En general, desde una perspectiva realista (la de Aristóteles lo era) la tarea de la percepción es tomar los datos de los sentidos y deducir de ellos una representación útil del ambiente exterior. En la vida diaria no percibimos los rasgos del sonido –altura, sonoridad, timbre, duración– de modo aislado, sino que los integramos en la representación de fuentes acústicas u objetos auditivos. Una parte importante en el proceso es decidir qué partes de la estimulación sensorial corresponden al mismo objeto o evento ambiental. La forma final de cada representación mental de un objeto –visual o auditivo– formado a partir de un cierto agrupamiento de datos sensoriales, es una propiedad emergente. Esta propiedad no corresponde a ninguno de los datos tomados de manera individual y surge a partir de nuestra capacidad de organización psíquica de los datos. La cita de Helmholtz reproducida en el capítulo II (“[...] estamos muy bien entrenados para evaluar los objetos que nos rodean, pero fallamos completamente cuando intentamos observar las sensaciones *per se* [...]”) alcanza aquí plena significación.

### 1. Ecología acústica y análisis auditivo del entorno

Albert Bregman llamó *análisis auditivo de escenas* al proceso que permite reunir en una unidad perceptiva el conjunto de datos provenientes de una fuente acústica externa (Bregman, A., 1994). Estrictamente, cada oído no recoge más que los cambios temporales de la presión acústica en el tímpano que le corresponde. Un *agregado auditivo* es, según Bregman, la asociación de las partes del es-

pectrograma neuronal (conjunto de datos sensoriales) que parecen pertenecer a un mismo evento perceptual. Una fuente acústica es una entidad física que genera ondas en el aire. Un agregado auditivo es un percepto causado por un grupo de elementos sonoros que se experimenta como un todo y que parece emanar de una única fuente acústica externa.[\[1\]](#) Normalmente, las propiedades emergentes que definen cada agregado están correlacionadas con las propiedades de los objetos físicos asociados.

En visión se puede describir el problema del análisis óptico del ambiente en términos del correcto agrupamiento de las regiones estimuladas en la imagen retiniana. En nuestro caso ¿qué elementos se deben agrupar para ensamblar un sonido o, mejor aún, un objeto acústico? No es posible comparar las estrategias de agrupamiento de ambos sentidos pues existe una diferencia crucial en el modo en que usamos la energía acústica y la lumínica para obtener información del mundo exterior. En audición, hacemos uso de la energía emitida por los objetos (las fuentes acústicas) y casi no tomamos en cuenta la energía reflejada en otros objetos físicos. Se podría afirmar que existen al menos tantas fuentes acústicas como objetos sonoros en una situación normal. Las reflexiones –que cambian en función del contexto espacial– degradan la información que nos llega de cada una de las fuentes. En consecuencia, necesitamos atenuar la información que llega por reflexión y colocar en primer plano la que nos llega directamente (en una sala de conciertos situamos auditivamente el piano en el escenario, aun cuando gran parte de su energía nos llega por reflexión desde las paredes y el cielorraso de la sala). Por el contrario, la información visual está causada principalmente por la reflexión de la luz en los objetos de interés. Encendemos una lámpara no porque nos interese mirarla, sino para ver por reflexión los objetos de la sala. A diferencia de la gran cantidad de fuentes acústicas que nos acompañan desde siempre, hubo una sola fuente de luz a la vez a lo largo de gran parte de nuestro recorrido evolutivo. En resumen, los da-

tos auditivos y visuales son diferentes y se complementan para brindarnos un panorama verosímil del mundo exterior.

## **2. Información utilizada para separar los objetos auditivos**

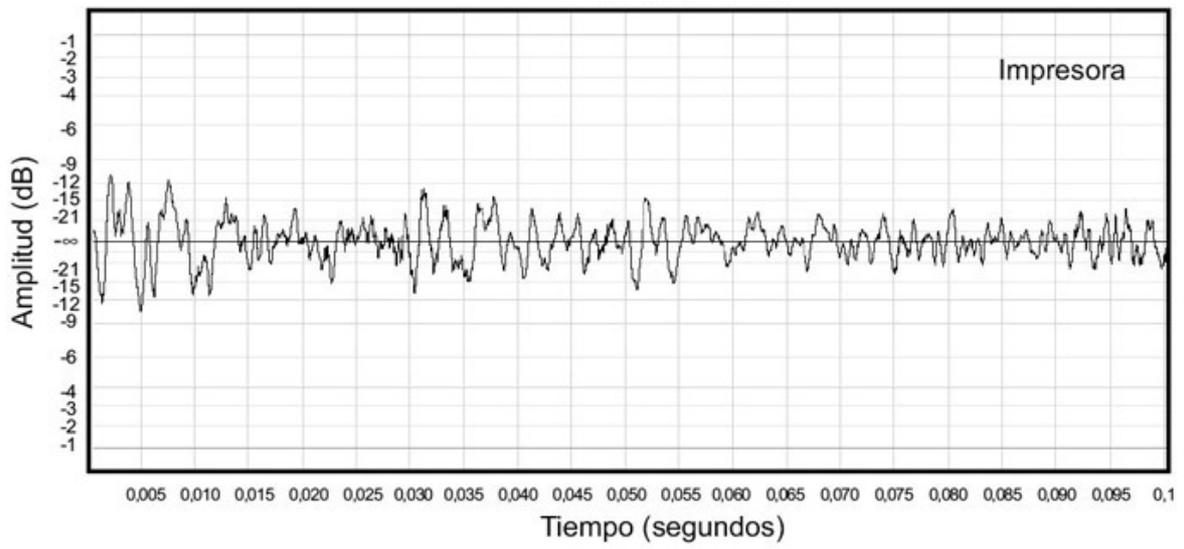
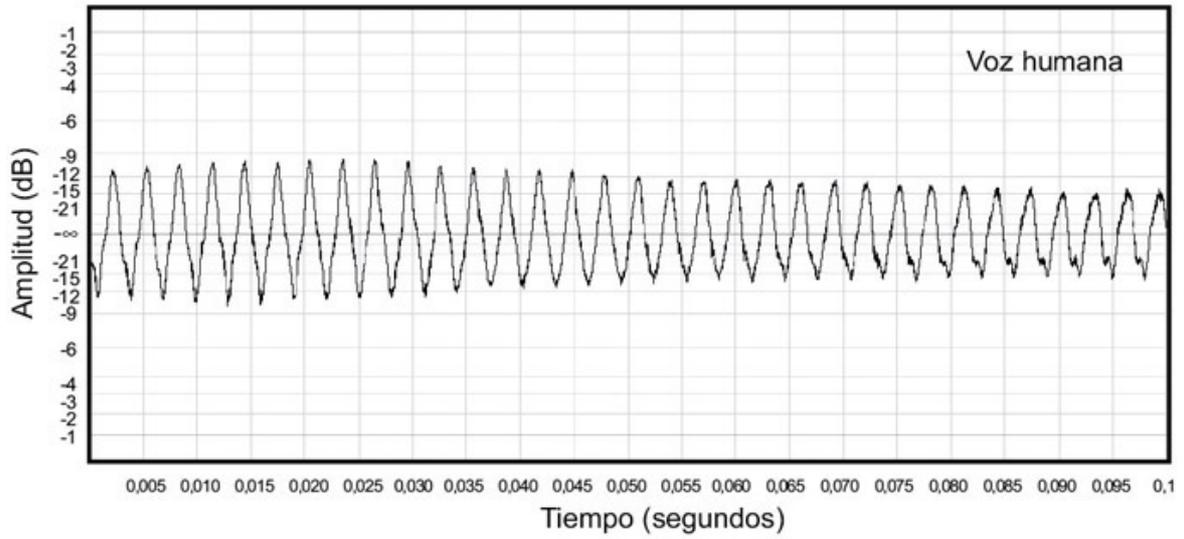
La fusión o fisión simultánea de la información auditiva y la construcción de agregados son ejemplos de organización perceptual. A partir de ellos analizamos el entorno acústico al distribuir los datos sensoriales entre los diferentes objetos auditivos potenciales. Podemos ilustrar la distinción entre los conceptos de fuente acústica y de agregado acústico a partir de las señales de la figura 1. En ella se ven varias ondas acústicas aisladas que se han reunido en el último gráfico. Es muy difícil distinguir cada componente en la mezcla final, pero realizamos esta operación inmediatamente con el oído (si pudiéramos reproducir las señales en un equipo de audio) gracias al proceso de agrupamiento perceptual. Cada una de las fuentes corresponde a un solo agregado acústico.

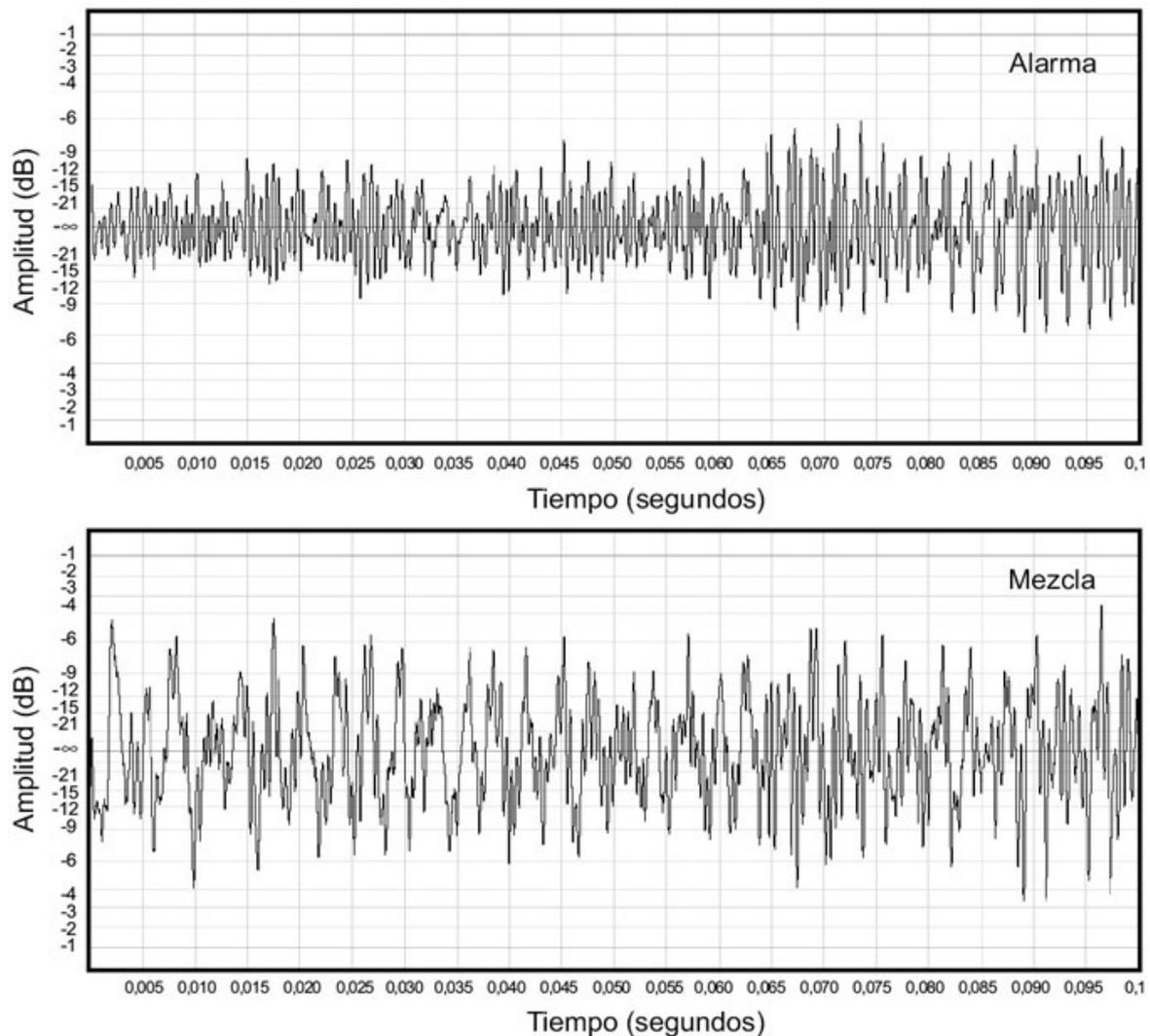
Parece haber diferentes mecanismos de segregación auditiva que responden en función de la tarea a realizar. Cuando los oyentes intentan mantener integrada en un solo agregado una secuencia de sonidos, la segregación es involuntaria y opuesta a sus intenciones. Pero cuando los sujetos se esfuerzan en separar los sonidos la segregación aparece como un proceso de selección provocado por la atención. La existencia de estos dos mecanismos explica los diferentes resultados que se obtienen en las pruebas experimentales cuando se varía alguna característica acústica en la señal de entrada. Las diferencias tímbricas entre sonidos musicales, por ejemplo, facilitan mucho la segregación. Pero si la secuencia se basa en otro factor de agrupamiento diferente, como puede ser la variación de la altura tonal, las diferencias tímbricas no provocan necesariamente la segregación.

El sistema auditivo emplea diferentes datos para separar los objetos auditivos. En lo que sigue describiremos los más destacados. Debe tenerse en cuenta que no siempre se utilizan todos los datos en todos los casos: el agrupamiento perceptual no es un proceso a “todo o nada”. En una situación compleja se observa con frecuencia que el dato saliente enmascara y oscurece a los menos acentuados, mecanismo que nos permite mantener la eficiencia perceptual aún cuando no esté presente toda la información esperada. En el caso del timbre, que depende de una gran cantidad de entradas auditivas concurrentes, el contexto juega un papel determinante a la hora de decidir cuál dato es significativo, y cuál no lo es.

El “efecto de cafetería” (o “cocktail party”) pone de manifiesto la eficacia de nuestra capacidad para separar los diferentes objetos auditivos. Podemos mantener una conversación sin inconvenientes en un ambiente con gran contaminación por ruido. Sin embargo, si registramos con un micrófono dicha conversación y la reproducimos más tarde habremos perdido algunos datos esenciales para separar la figura acústica del fondo de ruido y el discurso se vuelve ininteligible.

A continuación describiremos los rasgos de la información acústica que permiten agrupar o separar los diferentes objetos acústicos entre sí.





**Figura 1.** Cuatro señales separadas reunidas en el último gráfico

## Armonicidad y frecuencia fundamental

Un conjunto de parciales perfectamente armónicos tienden a fusionarse en un solo artefacto perceptual, pues comparten la misma altura tonal, definida por la componente fundamental de la serie. De acuerdo con este criterio, si dos estímulos simultáneos de espectros perfectamente armónicos tienen diferentes frecuencias fundamentales, se oyen separados.<sup>[2]</sup> Una desafinación del 3% es suficiente para que un armónico se perciba segregado del complejo (siempre que su duración exceda los 400 ms). Para el sonido complejo restante, el efecto de la desafinación del armónico es similar

al que ocurre cuando se reduce el nivel del armónico en cuestión hasta volverlo inaudible. La fusión por armonicidad fue utilizada desde el medioevo para combinar los registros de órgano y constituye un recurso clásico en orquestación.

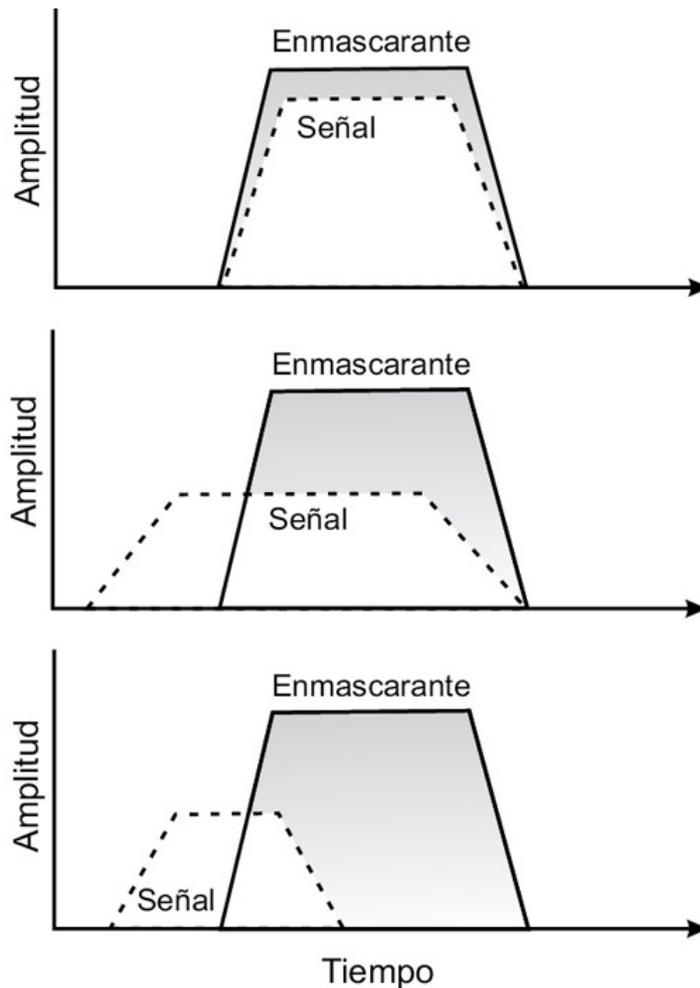
## **Disparidades en el ataque**

Rasch investigó la habilidad para oír un sonido complejo en presencia de otro semejante (Rasch, R. A., 1978). Cuando la única diferencia entre ambos es la frecuencia fundamental, la señal debe estar por encima de los -20 dB respecto del sonido enmascarante. Pero si la señal comienza 30 ms antes del enmascarante, se la percibe aun cuando su nivel llega a apenas -60 dB (figura 2). La señal puede seguir percibiéndose aunque cese de inmediato después del ataque del sonido enmascarante (este efecto se relaciona con el fenómeno de continuidad que veremos más adelante).[\[3\]](#) El valor de 30 ms corresponde aproximadamente a las imprecisiones de ataque que se encuentran con frecuencia en la música instrumental. Esta capacidad de agrupamiento por disparidad en el ataque refuerza la idea tradicional de que las asincronías son un factor importante en la percepción separada de las voces en la música polifónica.

En un experimento relacionado con las disparidades temporales en el ataque, Roberts y Moore demostraron que sinusoides extrañas agregadas a una vocal pueden influir en su calidad tímbrica, pero que esta influencia se reduce notablemente si el elemento extraño aparece antes o después del inicio de la vocal –se segrega de ella (Moore, B., 1997).

## **Contraste con sonidos previos**

La percepción de un estímulo es afectada por los estímulos inmediatamente precedentes. En general, al nuevo estímulo se le resta algo así como el “negativo espectral” de los anteriores.



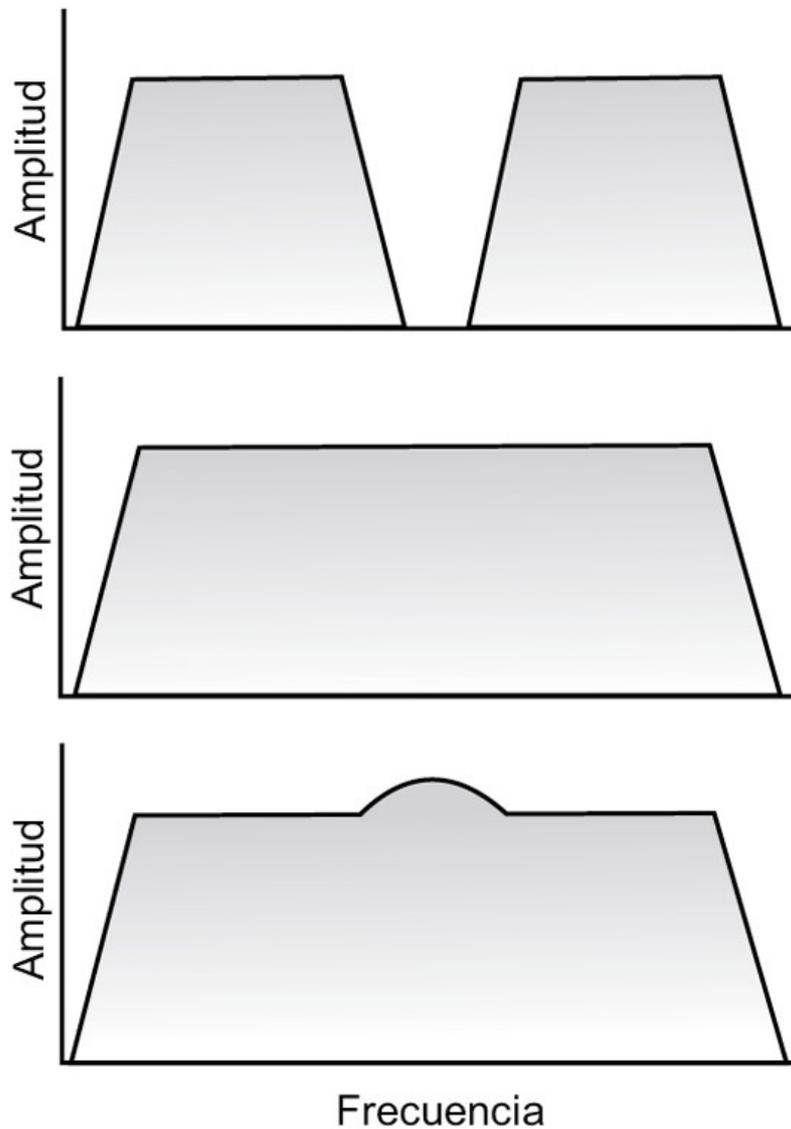
**Figura 2.** Experimento de Rasch

También se verifica un efecto similar cuando ocurre un cambio en el estímulo, que puede ser interpretado como que el estímulo de partida se sustrae del estímulo final. El rasgo modificado se separa perceptualmente del resto y la parte del estímulo que permanece estable se reduce hasta su nivel de adaptación.

Un ejemplo notable de lo anterior se da al oír un estímulo con una estructura espectral particular, seguido inmediatamente por ruido blanco (espectro plano). El ruido blanco aislado puede ser descrito como “no coloreado”. Sin embargo, en la secuencia completa se oye una estructura espectral “coloreada”, que corresponde al espectro de la primera señal invertido (figura 3).

## **Correlación de los cambios en amplitud y frecuencia**

Rasch también notó que, aun cuando dos sinusoides comiencen en sincronía, es posible acrecentar la separación perceptual modulando en frecuencia la más aguda (Rasch, R. A., 1978). Este procedimiento permite reducir el umbral de percepción de la señal de mayor frecuencia en 17 dB. En realidad, el experimento de Rasch no es más que un caso particular de una característica general de la percepción auditiva: los sonidos que presentan cambios “coherentes” tienden a fusionarse perceptualmente, mientras los que presentan variaciones incoherentes entre sí tienden a segregarse.

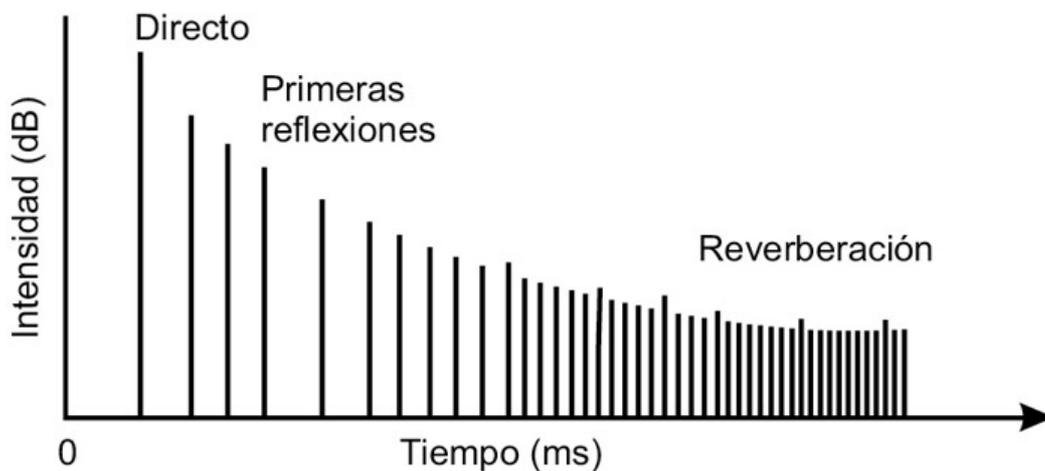
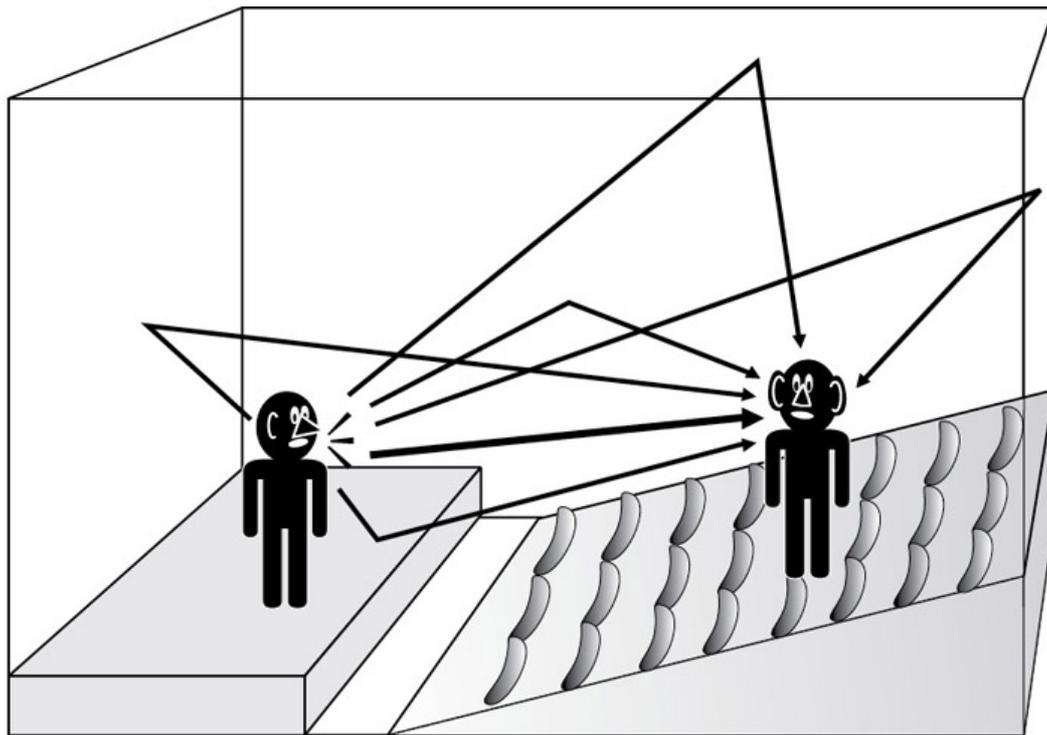


**Figura 3.** Espectro de potencia de dos señales sucesivas. El tercer gráfico representa el espectro percibido por el oyente

### Localización de la fuente

Muchas veces las diferencias espaciales juegan un papel importante en el análisis auditivo del entorno al facilitar o dificultar la segregación basada en otros factores.[\[4\]](#) Sabemos, por ejemplo, que una señal se detecta con mayor facilidad cuando su ubicación espacial difiere de la del ruido enmascarante.

Shackleton y sus colegas investigaron la habilidad para identificar cada vocal en un conjunto de dos vocales concurrentes. Encontraron que una diferencia de frecuencia fundamental entre las dos mejora la identificación en un 22%, mientras que un retardo de 400  $\mu$ s –correspondientes a una diferencia en azimut de 450– la mejora sólo 7% (Shackleton, T. M. *et al.*, 1992). En resumen, el mecanismo de segregación espectral parece ser más poderoso que el espacial. La pregunta es, entonces, ¿por qué los humanos no le dan prioridad absoluta al índice espacial? Y la respuesta se encuentra en la base de nuestra estrategia auditiva, centrada en la percepción de fuentes y no de reflexiones acústicas. Las señales provenientes de la fuente pueden haber sido reflejados muchas veces en el ambiente, o pueden haber sido atenuados por la presencia de un objeto cercano, hechos que deforman la información relativa a su posición. En la figura 4 se ilustra esta aparente confusión de fuentes reales y virtuales.



**Figura 4.** Patrón de reflexiones en una sala mediana y gráfico temporal de la energía que llega a un sector de la platea

El *efecto de precedencia* (o efecto Hass) nace de la necesidad de localizar unívocamente cada fuente acústica aún en presencia de un gran número de reflexiones tempranas.<sup>[5]</sup> En el experimento de Shackleton y Meddis las reflexiones no modifican la frecuencia fundamental ni las relaciones armónicas de cada vocal. De hecho, en un contexto normal las señales no armónicas y de banda ancha son

las que aportan la mayor parte de la información espacial. Aunque la necesidad de atenuar el efecto de las reflexiones explique la menor importancia relativa de los datos espaciales frente a otros más estables, es necesario recordar la capacidad que posee el sistema auditivo para atenuar o enfatizar parte de la información cuando el contexto –o la tarea impuesta– se lo demanda. En algunas ocasiones la localización (o la lateralización) del sonido ocupa un lugar fundamental en la segregación de los objetos auditivos. Un buen ejemplo lo brinda nuestra capacidad para mantener una conversación en ambientes muy ruidosos.[\[6\]](#)

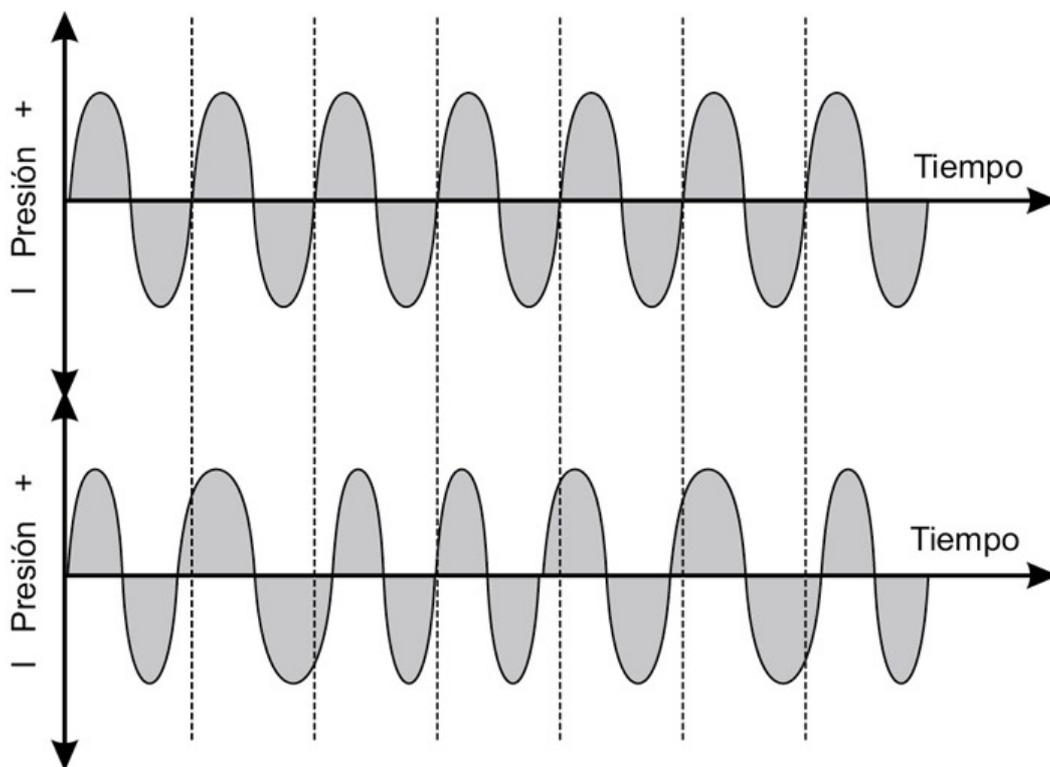
## Periodicidad y fusión perceptual

Un patrón temporal perfectamente regular, que repite exactamente igual en tiempos iguales, es *periódico*, mientras que uno que presenta pequeñas variaciones en su período o amplitud se define como *cuasi periódico*. En la figura 5 se aprecian las diferencias en sus correspondientes gráficos temporales.

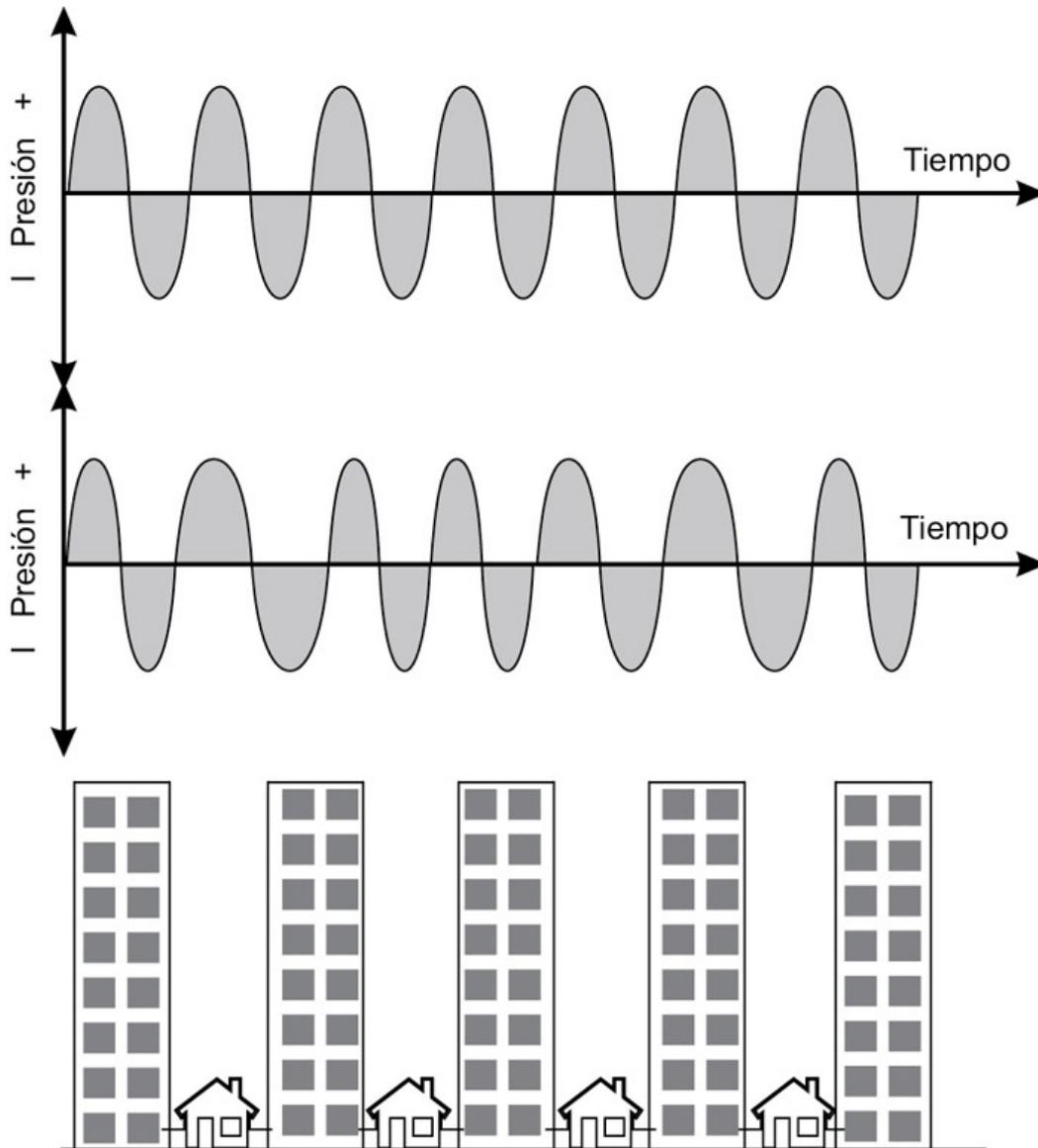
Las señales que se emplean en los experimentos psicoacústicos tradicionales son, a los efectos prácticos, razonablemente periódicas. Están generadas por medios electrónicos y es prácticamente imposible hallarlas en la naturaleza o en las ondas de los instrumentos de música no electrónicos. Las señales cuasi periódicas, por otra parte, son típicas en estos últimos casos.[\[7\]](#) El sistema auditivo es muy sensible a las señales cuasi periódicas al ser capaz de detectar pequeñas variaciones del orden de una fracción de período. Estas variaciones son impuestas por las alinealidades de las fuentes acústicas. En el caso de la música, aparecen tanto en el instrumento como en su ejecución y a veces se las agrega intencionalmente en la forma de vibrato o trémolo. El sistema auditivo emplea de manera provechosa las pequeñas aperiodicidades mencionadas para fundir perceptualmente los estímulos que le llegan. Cuando recibe pequeñas variaciones sincronizadas en paralelo tien-

de a oír las como si estuvieran originadas en la misma fuente, fenómeno relacionado con la *ley de destino común* de la teoría de la Gestalt.

Tomemos por ejemplo la emisión vocal en un cantante. Siempre existen pequeñas irregularidades en el flujo de aire que llega desde los pulmones; hay vibraciones no controladas en los músculos de las cuerdas vocales; aparecen turbulencias no lineales en el tracto vocal y concurren desajustes en la adaptación de impedancias entre el complejo buconasal y el exterior. Como consecuencia de estos factores el conjunto de armónicos es modulado irregularmente en frecuencia (“jitter”) y en amplitud (“shimmer”). Al aplicarse ambas modulaciones al total de la señal, el efecto final es la percepción de un solo objeto auditivo, la voz de tal o cual cantante en particular.



**Figura 5.** Comparación entre una señal periódica y una señal cuasi periódica



**Figura 6.** Percepción visual de la periodicidad y de la simetría. Las ondas son las mismas que en la figura 5, pero sin líneas de referencia

Es muy sencillo realizar el siguiente experimento: se generan digitalmente tres series armónicas con la misma fundamental, que se funden perceptualmente en un solo objeto acústico –se oyen como si estuvieran originadas en la misma fuente. Si a cada serie armónica se le agregan pequeñas fluctuaciones no correlacionadas (que pueden ser tres vibratos diferentes), cada serie se separa de las res-

tantes y se las oye como si provinieran de tres fuentes distintas emitiendo en unísono.

Los sistemas auditivo y visual parecen tratar la periodicidad y la simetría de una manera similar, pero con diferente grado de resolución. Por ejemplo, la vista no detecta con facilidad la aperiodicidad de la segunda señal de la figura 6, al no contar con la ayuda de las líneas que indican los períodos que aparecían en la figura 5. Tampoco nos percatamos inmediatamente de que tres de los edificios son menos simétricos respecto del eje central que los otros dos. El sistema auditivo, en contrapartida, detecta inmediatamente un grado similar de disparidad.

Tanto el sistema visual como el auditivo pierden parte de su capacidad de atención –se podría decir que se “desconectan”– cuando aparecen simetrías o periodicidades perfectas, aún durante un tiempo breve, y fallan al intentar extraerles información relevante.

## **Percepción de secuencias rítmicas**

Numerosas experiencias comprobaron que la fusión y fisión perceptual de eventos repetitivos dependen del intervalo de tiempo que media entre cada ocurrencia. En una escala temporal media, del orden de las décimas de segundo, se habla de *secuencias rítmicas*. En un contexto acústico cotidiano muchas secuencias de sonidos son rítmicas, y es de esperar que existan mecanismos perceptuales específicos que las contemplen. Por ejemplo, si un patrón auditivo es rítmico, entonces ciertos aspectos del mismo resultan temporalmente redundantes. Una vez que los elementos tempranos que definen el patrón son advertidos, los subsiguientes pueden ser anticipados. Handel demostró, en secuencias repetitivas segmentadas por pausas, que el principio de agrupamiento perceptual predominante se basa en la estructura temporal; otros parámetros, como la altura tonal, pasan a segundo plano (Handel, S., 1989). En otras palabras, la estructura temporal –Handel la llamó *timing*– de los

elementos es más importante para determinar el ritmo subjetivo que la estructura interna de cada grupo.

Por el contrario, cuando oímos secuencias rápidas (más de 10 eventos por segundo) los sonidos no se agrupan de acuerdo con su orden temporal físico, sino de acuerdo con ciertas características comunes como la altura tonal o el timbre. Si la secuencia se percibe como un todo se produce *fusión rítmica*, y cuando aparece dividida en varios patrones diferentes se produce *fisión rítmica*. Bregman sugirió que cada agrupamiento producido por la fisión rítmica corresponde a un objeto auditivo diferente. Dowling demostró que la segregación ocurre también cuando las señales sucesivas difieren en intensidad o en ubicación espacial.[\[8\]](#)

Algunos compositores –notablemente Bach– han aprovechado la segregación producida en sonidos muy separados en frecuencia –por ejemplo en las partitas para violín solo–, o la fusión de los sonidos de dos instrumentos en una sola melodía (para que el efecto tenga lugar los instrumentos deben tener timbres y locaciones similares).

## Percepción del orden temporal

Puede ser difícil establecer el orden temporal de sonidos que parecen provenir de diferentes fuentes acústicas. Las personas no entrenadas necesitan duraciones de 200 ms para establecer el orden de cuatro eventos no relacionados y muy diferentes. Este resultado quizá provenga de la segregación en diferentes corrientes auditivas, donde los sonidos parecen flotar unos con respecto a otros en el tiempo subjetivo.

Giovanni Vicario demostró que no es posible ordenar temporalmente dos corrientes de sonidos de alta y baja frecuencia cuando la separación excede las tres octavas.[\[9\]](#) En su experimento, cuando los sonidos sucesivos se unieron con glisados en frecuencia la tendencia a separarse en diferentes corrientes se redujo, aun cuan-

do los glisados no fueran completos. La continuidad perfecta entre tonos no es necesaria para reducir la fisión entre las dos secuencias: si un cambio en frecuencia “apunta” hacia el próximo sonido, el oyente puede seguir el patrón de continuidad más fácilmente. Las transiciones bruscas también ayudan a evitar la segregación en diferentes corrientes auditivas e intervienen considerablemente en la percepción de la palabra.

## Regularidad en los datos físicos

Los teóricos que proponen el *modelo físico* para describir la representación mental de un objeto auditivo, como Claude Cadoz, sostienen que aprovechamos ciertas características de las fuentes físicas para componer el análisis auditivo de escenas (Cadoz, C. *et al.*, 1993). Este mecanismo sería luego internalizado y operaría aun si el auditor no está familiarizado con la señal física en un caso particular. Que el sistema auditivo interpreta una transformación súbita de las propiedades acústicas como el inicio de una nueva señal, o que haber oído previamente un sonido potencia la percepción de su ocurrencia futura, son hechos que nadie cuestiona. Aunque nosotros nos inclinamos por el *modelo acústico*, la propuesta de Cadoz no deja de tener cierto atractivo. Las regularidades físicas han sido corroboradas experimentalmente y son una buena guía práctica, aunque quizás no deriven de una comprensión perceptual profunda de la naturaleza de los procesos físicos involucrados.

De acuerdo con el modelo físico, las regularidades que pueden emplearse para agrupar elementos en un único objeto auditivo son:

- Regularidad 1: es extremadamente raro que dos sonidos, que comiencen y terminen exactamente al mismo tiempo, no posean alguna relación entre sí.
- Regularidad 2: progresión de una transformación: a) las propiedades de un sonido aislado tienden a modificarse de manera continua y lenta; b) las propiedades de una secuencia

de sonidos tomados de la misma fuente tienden a modificarse lentamente.

- Regularidad 3: cuando un cuerpo vibra periódicamente, sus vibraciones dan nacimiento a un patrón acústico donde las frecuencias de las componentes son múltiplos de una misma frecuencia fundamental.
- Regularidad 4: la mayoría de las modificaciones que afectan a una fuente acústica afecta a todos los componentes del sonido resultante, de manera idéntica y simultánea.

El fenómeno de reducción del enmascaramiento por comodulación constituye un buen ejemplo de regularidad del tipo 4.[\[10\]](#) En el experimento clásico de Donald Hall se emite una señal sinusoidal superpuesta a un ruido de bandas múltiples. Si el ruido se mantiene estático el sonido de prueba es enmascarado, pero si el ruido enmascarante fluctúa sincrónicamente en amplitud, la percepción de la señal se restablece. El oyente agrupa a las diferentes bandas de ruido en un solo objeto auditivo a partir de su movimiento común. El sonido de prueba, estático, se segrega del resto.

## **Redundancia de la información**

Como hemos visto, existen muchas estrategias de agrupamiento e interpretación de los datos sensoriales. Esta multiplicidad es necesaria porque cada una de ellas, por separado, está sujeta al error. Por ejemplo, la estrategia que agrupa los sonidos por su origen espacial puede ser ineficaz en ambientes muy reverberantes. La que se basa en relaciones armónicas puede fallar si se encuentra con sonidos inarmónicos o ruidos de banda. Las que buscan similitudes secuenciales pueden ser burladas ante sonidos discontinuos o cambiantes. La utilización en paralelo de varias relaciones de agrupamiento constituye habitualmente una protección contra el fallo de una de ellas. Cuando la información que recibe es contradictoria, el auditor simplemente se puede declarar incapaz de penetrar los

complejos sonoros. La extrema rareza de tales casos testimonia la potencia de análisis del sistema auditivo humano.

### **3. Principios generales de organización perceptual y teoría de la Gestalt**

Junto a sus descubrimientos fundacionales en física y fisiología de la percepción, Helmholtz es conocido también como un precursor de la psicología cognitiva. Formuló el principio de la *inferencia inconsciente*, que establece que nuestro sistema visual “construye” objetos tridimensionales a partir de los datos presentes en el espacio bidimensional de la retina. A partir del principio de Helmholtz se desarrolló la *teoría de la Gestalt*. Ésta se inicia en 1912 con un artículo de Max Wertheimer sobre el movimiento aparente, que explica la ilusión que posibilita la existencia del cine. El autor dedujo que la percepción de un todo (el movimiento) era radicalmente diferente de la percepción de sus componentes (las imágenes estáticas). La teoría tuvo un gran desarrollo durante los siguientes 50 años. Sin embargo, aunque la descripción de los fenómenos perceptuales postulada por los gestaltistas no ofrecía reparos importantes, la argumentación teórica se mostró insuficiente para fundamentar sus leyes y la escuela cayó en cierto descrédito en la década de 1970. En la actualidad, se han retomado sus principios generales de organización perceptual, que describen a grandes rasgos las características de la percepción, pero se han dejado de lado los modelos explicativos originales. Casi todos los psicólogos especializados en la percepción aceptan hoy la operatividad de las leyes de la Gestalt, aunque las justifican a partir de procesos y mecanismos diferentes.

Los gestaltistas, a diferencia de los estructuralistas, entendían que la percepción no podía reducirse a una mera suma o serie de sensaciones, sino que respondía a la configuración total de la que aquellas forman parte. “El todo es distinto que la suma de sus par-

tes” sintetiza uno de sus axiomas básicos, en el que las propiedades emergentes ilustran el concepto gestáltico de organización necesario para explicar por qué los seres humanos vemos el mundo compuesto de objetos distintos. La capacidad de percibir objetos –piedras, árboles, ladridos o viento– debía estar originada en operaciones efectuadas por el sistema nervioso central. Advertir que la percepción de objetos separados no es sólo producto de las sensaciones periféricas fue una de las más importantes contribuciones de la escuela de la Gestalt.

Según la teoría de la Gestalt los hechos psíquicos son formas, unidades orgánicas que se individualizan en el campo espacial y temporal de percepción o de representación. Las formas dependen, en el caso de la percepción, de un conjunto de factores sensoriales objetivos. La percepción de las diferentes clases de elementos y de las diferentes clases de relaciones corresponde a determinados modos de organización de un todo, que depende a la vez de condiciones objetivas y subjetivas. La correspondencia que se puede establecer entre las partes de un todo articulado y ciertos elementos objetivos no se mantiene, en general, cuando esos mismos elementos pertenecen a otro conjunto objetivo. En otras palabras, una parte en un todo es algo distinto a esa parte aislada o en otro todo, a causa de las propiedades que debe a su lugar y función en cada uno de ellos. El cambio de una condición objetiva puede producir un cambio local en la forma percibida, o traducirse en un cambio radical de la forma emergente.

El objetivo inicial de la teoría de la Gestalt consiste en establecer, por medio de la experiencia, un conjunto de condiciones de existencia de las formas perceptuales y de leyes de transformación de dichas formas. Para que éstas se muestren eficaces no es necesario que exista una correspondencia rigurosa entre los elementos del estímulo objetivo y los elementos de la forma –de hecho, la correspondencia directa generalmente no existe y, en todos los casos, no se mantiene constante.

Los teóricos de la escuela afirmaron que el proceso fisiológico que resulta de un conjunto de excitaciones tiende a organizarse espontáneamente siguiendo ciertas leyes de estructura, independientes en principio de las significaciones agregadas por la educación. En la actualidad existen teorías sobre la percepción que contradicen tal afirmación. Serán comentadas más adelante.

Los trabajos de los psicólogos gestaltistas llevaron a enunciar cinco leyes y dos principios generales que permiten agrupar los estímulos que provienen de una misma fuente, y segregar los que no lo hacen. En general, una regla aislada no siempre permite la fusión o fisión perceptual y varias de ellas deben actuar juntas para que emerja la forma final (Köhler, Koffka y Sander, 1962).

En lo que sigue describiremos brevemente las leyes y los principios generales de organización que propone la teoría de la Gestalt.

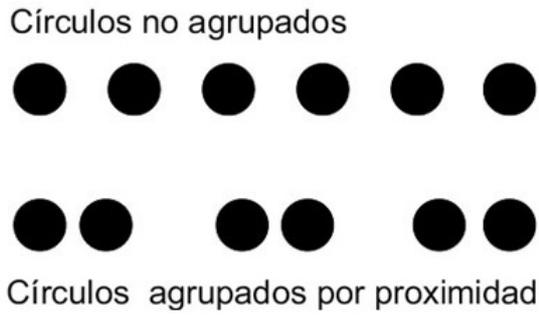
## **Leyes de la Gestalt**

Los psicólogos de la Gestalt centraron sus estudios en el sentido de la vista. Con algunas adaptaciones, sus cinco leyes y dos principios de organización se aplican también a la audición.[\[11\]](#)

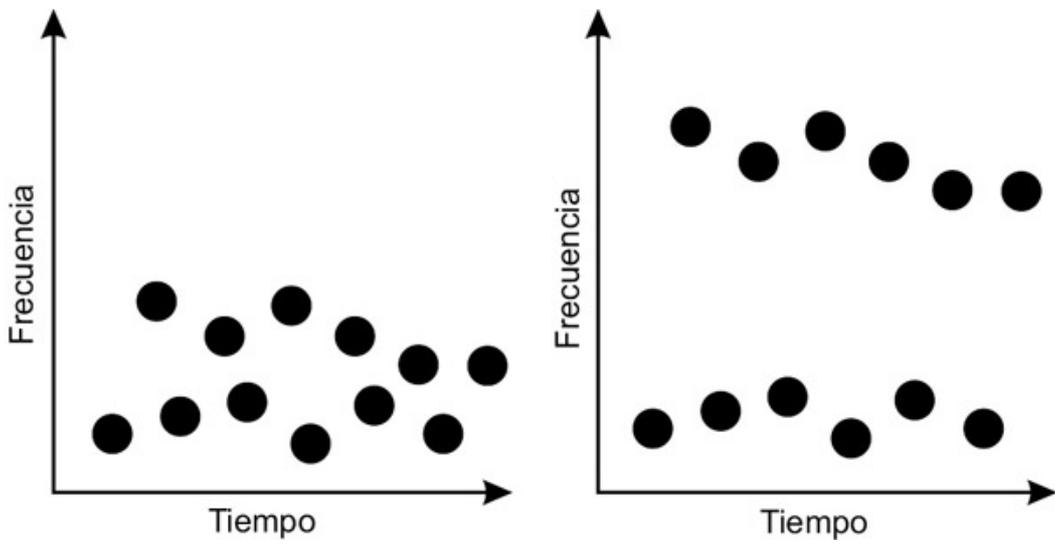
### ***Ley de proximidad***

La ley de proximidad expresa que los elementos cercanos tienden a percibirse como una unidad. En la figura 7 se aprecian pares de círculos agrupados por proximidad espacial.

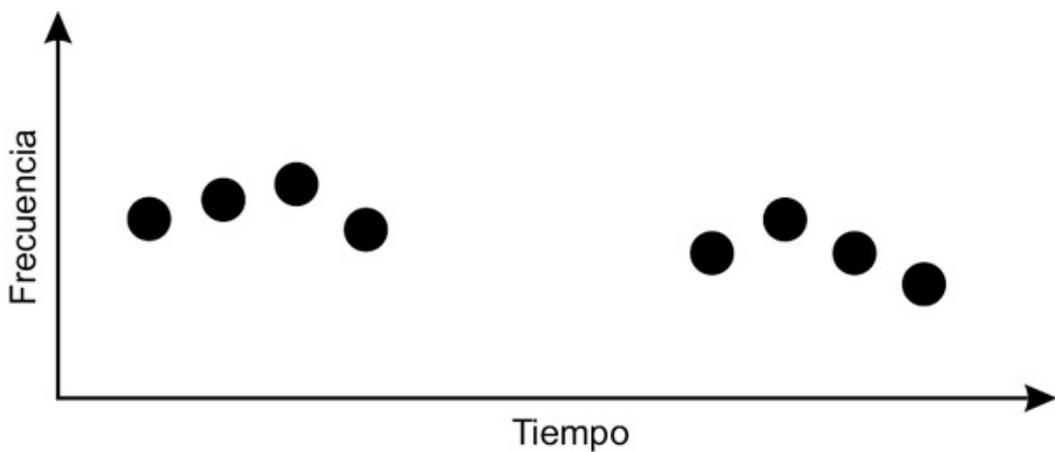
En audición la proximidad puede darse en diferentes dimensiones. Los sonidos se agrupan por proximidad de altura, de tiempo, de timbre o de espacio.



**Figura 7.** Ley de proximidad



**Figura 8.** Agrupamiento de sonidos por proximidad de altura



**Figura 9.** Agrupamiento de sonidos por proximidad temporal

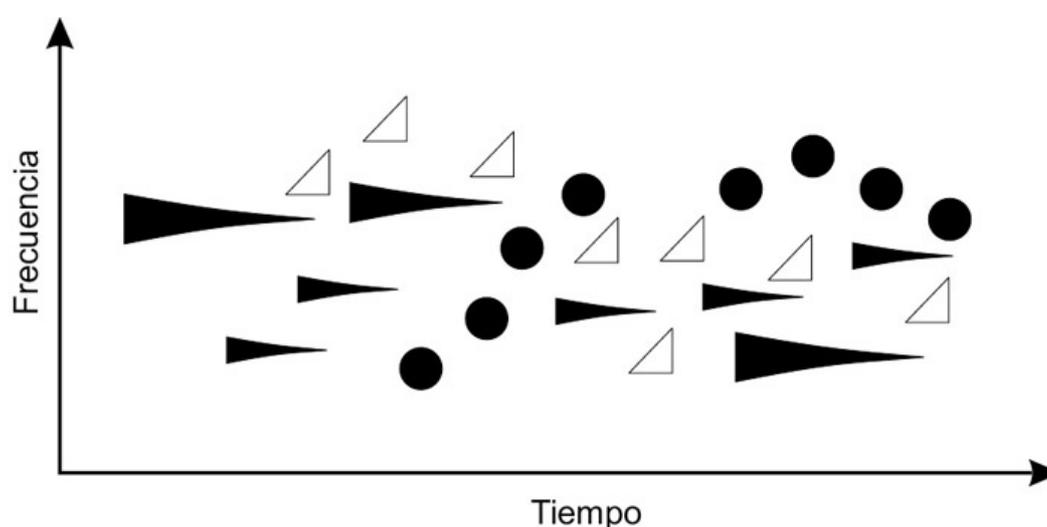
## *Ley de semejanza*

La ley de semejanza establece que elementos idénticos o semejantes (en color, tamaño o forma para el sentido de la vista) tienden a percibirse agrupados. En audición, la semejanza implica similitud de timbre, altura, sonoridad o localización espacial. El rasgo que determina la agrupación perceptual por semejanza depende del contexto o de la tarea impuesta al oyente. Van Noorden demostró que la similitud de timbre puede imponerse a la similitud de altura tonal, al menos bajo ciertas circunstancias (Van Noorden, L., 1975).

En la figura 10 se observa la ley de semejanza aplicada a la vista y en la figura 11 se puede ver el espectrograma de dos corrientes perceptuales agrupadas por semejanza tímbrica.



**Figura 10.** Ley de semejanza



**Figura 11.** Agrupamiento de sonidos por semejanza de timbre

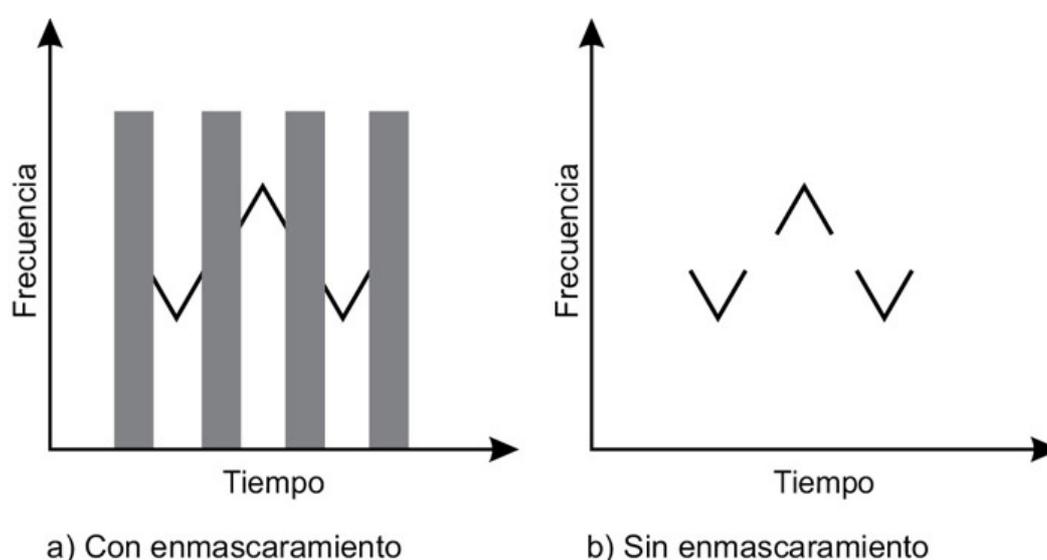
## *Ley de cierre*

La ley de cierre establece que las figuras incompletas tienden a completarse y a ser percibidas como una unidad cerrada. En la figura 12 se ve una serie de rectángulos.

Bajo ciertas condiciones un sonido enmascarado puede oírse como si fuera continuo aún sin serlo (en todo caso, no existe evidencia sensorial de su continuidad). Se lo sigue “oyendo” aún cuando ha cesado. En la figura 13a percibimos que la señal de banda angosta no presenta ninguna solución de continuidad. Puede ocurrir, sin embargo, que resulte discontinua en los tramos enmascarados (figura 13b).



**Figura 12.** Ley de cierre



**Figura 13.** Ley de cierre para una secuencia de sonidos

## *Ley de continuidad o de buena forma*

La ley de continuidad establece que los elementos que forman parte de un patrón continuo tienden a percibirse como una unidad. En la figura 14 vemos dos líneas que se cruzan, aunque son posibles otras configuraciones.

Esta ley, aplicada a la audición, hace uso de nuestro conocimiento de las propiedades físicas de las fuentes acústicas que modifican sus parámetros de manera suave y continua. Los cambios bruscos indican que una nueva fuente se ha activado. Por ejemplo, si la alternancia entre sonidos agudos y graves se intercala con los glisados en la dirección adecuada, se reduce la segregación en dos corrientes perceptuales. En la figura 15 se puede apreciar el correlato auditivo de esta ley.

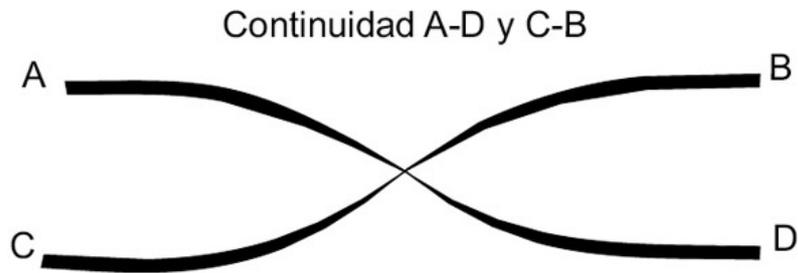
En la voz hablada, grandes cambios inesperados en la frecuencia fundamental dan la impresión de que un nuevo hablante se ha incorporado. Este efecto requiere un “conocimiento” de las reglas normales de entonación. La segregación auditiva produce silencios ilusorios en cada corriente durante la parte de la señal atribuida a la otra corriente.

## *Ley de destino común*

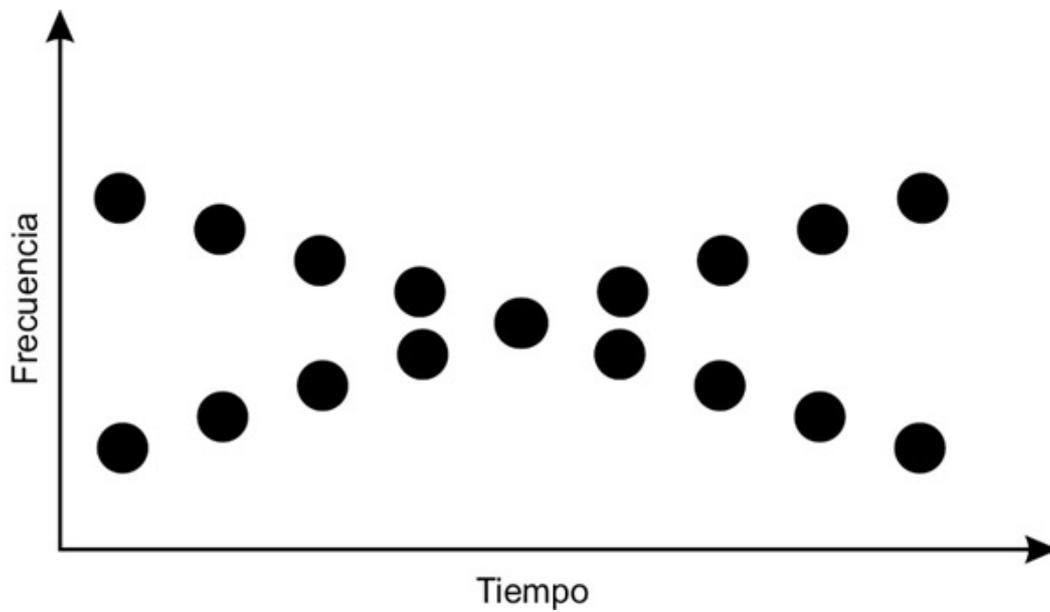
Por la ley de destino común los objetos que presentan un patrón de movimiento común tienden a percibirse como una unidad. Ejemplos clásicos son las bandadas de pájaros o los grupos de maratonistas.

Esta ley es de enorme importancia en audición. Los diferentes componentes en frecuencia provenientes de la misma fuente acústica varían usualmente de una manera coherente. Tienden a comenzar y terminar juntos, cambian de intensidad a la par y alteran sus frecuencias de manera asociada. Por eso si dos o más componentes de un sonido complejo experimentan el mismo tipo de cam-

bios al mismo tiempo, son agrupados y percibidos como provenientes de la misma fuente. En la figura 17 las dos líneas melódicas que se mueven en paralelo tienden a agruparse entre sí.



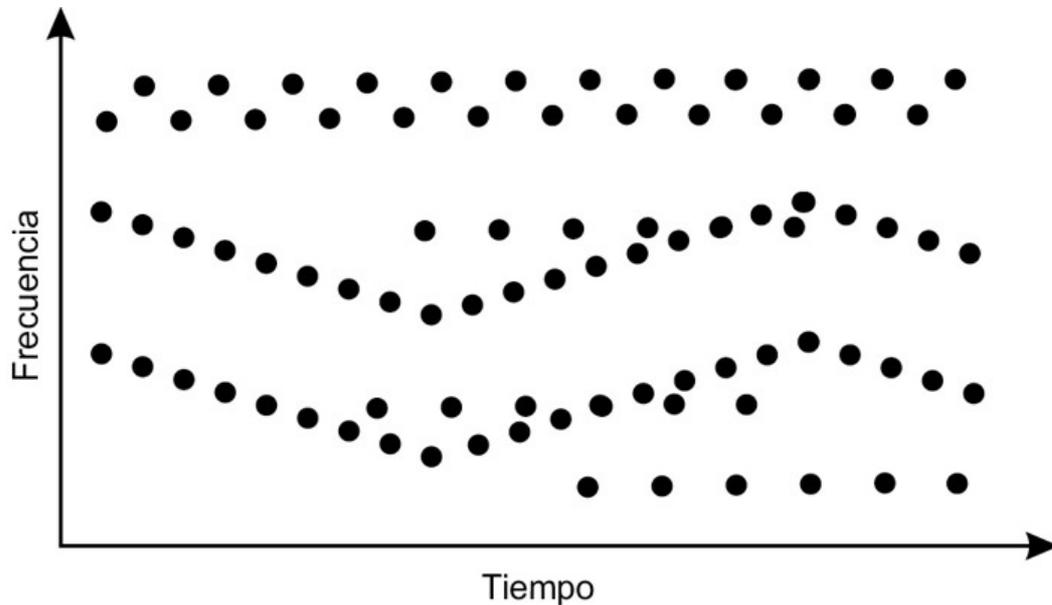
**Figura 14.** Ley de continuidad



**Figura 15.** Ley de continuidad en dos secuencias de sonidos



**Figura 15.** Ley de destino común (las ondas dobles representan movimientos en paralelo)



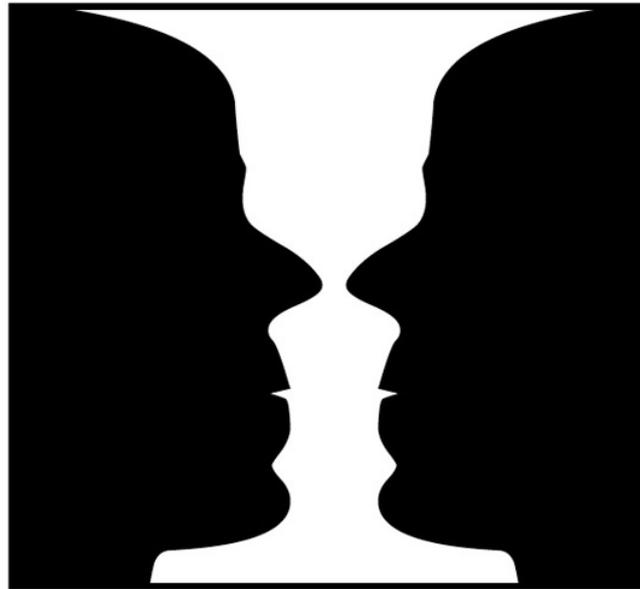
**Figura 17.** Agrupamiento de sonidos por movimiento común

Con relación a la ley de destino común, hay dos factores que se destacan a la hora de sintetizar sonidos para crear nuevos objetos auditivos. El primero tiene que ver con la forma y sincronismo del ataque, y el segundo con la sincronía de las modulaciones (de amplitud o frecuencia) en el desarrollo temporal de los componentes que se pretenden agrupar.

### *Principio de figura y fondo*

Tenemos la capacidad visual de distinguir una figura del fondo contra el cual aparece. También somos capaces de distinguir la melodía que ejecuta un solista destacada por sobre la masa orquestal y de comprender el discurso de locutor de radio a pesar del ruido de fondo. Sin embargo, todos estos estímulos nos llegan desde el campo perceptivo periférico sin jerarquización alguna. En la co-

nocida imagen de la figura 18 se pueden ver alternativamente dos rostros o una copa, según cuál de los campos visuales consideremos como figura y cuál como fondo.



**Figura 18.** Figura y fondo

Gracias al principio de figura y fondo se establece una jerarquía entre los objetos de interés y el medio neutro que, aunque pueda estar constituido por excitaciones complejas y heterogéneas, es desplazado a un grado inferior de diferenciación. La segregación perceptual entre figura y fondo no ocurre de manera mecánica. Al oír una conversación o una melodía en un ambiente con un gran ruido de fondo, seguimos al objeto auditivo que contenga mayor sentido en función del contexto o de nuestro interés particular. Cuando los índices de diferenciación no son evidentes se producen figuras ambiguas o ilusiones figura-fondo, situaciones que se aprovechan en el mimetismo y en el camuflaje.

El *principio de pertenencia* (también conocido como principio de correspondencia) expresa que si un componente simple es usado para la formación de una corriente perceptual, no puede emplearse en la formación de una segunda corriente. En la figura 18 no pue-

den verse a la vez la copa y los rostros. Parece que atendemos a una corriente perceptual a un tiempo, y el resto forma parte del fondo. Algunos autores han aceptado, no obstante, que el principio de pertenencia auditivo puede debilitarse en ciertas ocasiones: un elemento sonoro puede oírse como parte de más de una corriente perceptiva. En el caso de las texturas polifónicas se ha propuesto la posibilidad de percibir más de una figura a la vez.

El papel de la atención es determinante en la selección de los elementos que van a formar parte del objeto acústico que ocupará la función de figura. La atención actúa sobre la entrada auditiva para seleccionar ciertos rasgos del sonido para su análisis consciente. Empleamos nuestra sensibilidad a los cambios en la señal acústica para dirigir la atención hacia nuevos y potencialmente importantes eventos del entorno. Cuando un aspecto del sonido cambia, mientras el resto se mantiene constante, llama la atención del oyente: se convierte en figura. El efecto “cocktail party” o la audición de una voz que se destaca del resto en una pieza polifónica son buenos ejemplos de elecciones atencionales. Neisser sugirió que la atención aparece una vez realizado un análisis preatentivo de las diferentes corrientes perceptivas (Neisser, U., 1967). Sin embargo, este proceso no es de un solo sentido: se sabe que la atención puede influir en la formación de corrientes perceptivas.

### *Principio de pregnancia*

El principio de pregnancia fue formulado por Koffka del siguiente modo: “la organización psicológica será siempre tan excelente como las condiciones dominantes lo permitan” (Guillaume, P., 1975). Se entiende aquí por “excelencia” a la intervención de propiedades como regularidad, simetría, armonía de conjunto, homogeneidad, equilibrio, máxima sencillez, etc. En otras palabras, para agrupar el campo perceptual recurrimos a la menor cantidad posible de in-

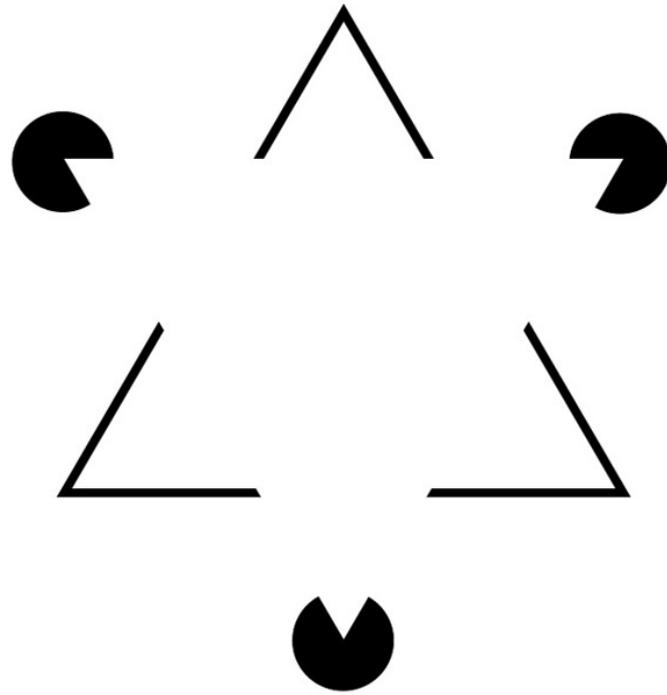
formación. Este principio opera empleando simultáneamente varias leyes de la Gestalt.

En la figura 19 se ha dibujado el triángulo de Kanizsa. La interpretación más económica en términos de información es suponer que existe un triángulo blanco que se superpone a tres círculos y al triángulo delineado, y no la mucho más compleja e improbable alineación de tres ángulos y tres pacman.

En el enmascaramiento por ruido ocurre algo similar. Es más económico en términos de información suponer que el sonido enmascarado continúa en presencia del ruido, que interpretar una compleja operación sincrónica de activación-desactivación.

Se puede apreciar un ejemplo notable de agrupamiento, en el que intervienen varios principios gestálticos a la vez, al comienzo del movimiento final de la sexta sinfonía de Tchaikovsky. En el sistema superior de la figura 20 se ven las partes de los violines I y II tal como figuran en la partitura, mientras que en el segundo sistema está escrito lo que casi todos los oyentes perciben. En este caso se oponen entre sí algunas de las leyes de agrupamiento examinadas. Por semejanza tímbrica y de alturas se fusionan los sonidos provenientes de ambos grupos instrumentales y luego se reagrupan de acuerdo con el principio de pregnancia. En esta segunda fase intervienen las leyes de continuidad y cierre junto a un principio de cohesión estructural llamado *ilusión de escala* por Diana Deutsch (1975).

Es posible acentuar la ambigüedad perceptual del pasaje agregando un componente espacial si se dispone la orquesta en el escenario de tal modo que los primeros y los segundos violines queden alejados entre sí, a ambos lados del director, como ocurre en la formación orquestal “a la alemana”.



**Figura 19.** Triângulo de Kanizsa

Como se escribió:

Violín I

Violín II

This musical score shows two staves, Violín I and Violín II, in 3/4 time with a key signature of two sharps (F# and C#). The first measure contains a sequence of notes: G4, A4, B4, C#5, B4, A4, G4. The second measure contains a whole note G4. The third measure contains a whole note G4 with a fermata above it.

Como se percibe:

Violín I

Violín II

This musical score shows two staves, Violín I and Violín II, in 3/4 time with a key signature of two sharps (F# and C#). The first measure contains a sequence of notes: G4, A4, B4, C#5, B4, A4, G4. The second measure contains a whole note G4. The third measure contains a whole note G4 with a fermata above it.

**Figura 20.** Reagrupamiento melódico en el cuarto movimiento de la sexta sinfonía (“Patética”) de Tchaikovsky

## Isomorfismo

Los creadores de la escuela de la Gestalt trataron de fundamentar su teoría en términos fisiológicos. Postularon una conexión muy directa entre la experiencia perceptiva, la fisiología y el mundo físico a través del *principio de isomorfismo*, que afirma que una experiencia subjetiva y el dispositivo neural asociado tienen estructuras formales similares.

Se trata claramente de una postura monista en la que los procesos físicos, fisiológicos, psicológicos y culturales tienen la misma base real: la forma física. La física se convierte aquí en la ciencia

fundamental a tomar como referencia explicativa última. Esta concepción se manifiesta, por ejemplo, en las numerosas comparaciones entre el mundo físico y el psicológico a las que recurre Köhler: “[...] del mismo modo que en el campo físico (por ejemplo el campo eléctrico), en el campo psicológico hay también una estructuración espontánea que obedece a principios dinámicos” (Köhler, W., 1972).[\[12\]](#)

Para estos investigadores la teoría de las formas físicas y el isomorfismo psicofísico constituyeron el asiento teórico de la doctrina. Al afirmar que en la naturaleza inorgánica se encuentran modos de ser del tipo “formal”, liberaron la barrera que separa los procesos no orgánicos de los procesos fisiológicos en el sistema nervioso central. El marco teórico fue tomado de la teoría del campo electromagnético vigente a principios del siglo XX. Citemos nuevamente a Köhler: “[...] los procesos fisiológicos neurales con que se asocian los hechos perceptuales están en cada caso localizados en un medio continuo, y los hechos que ocurren en una parte de ese medio influyen sobre los que tienen lugar en otras regiones, en la medida directa en que las propiedades de ambos se hallan relacionadas mutuamente [...] a un percepto aislado en medio de un ambiente homogéneo, le atribuimos su campo”. Introduce entonces el principio de isomorfismo: “las propiedades estructurales de las experiencias son al mismo tiempo las propiedades estructurales de sus correlatos biológicos y, en última instancia, del mundo físico” (Köhler, W., 1972).[\[13\]](#)

En la actualidad ya no se cree en el principio de isomorfismo, y hace tiempo que han sido dejados de lado los fundamentos teóricos originales de la psicología de la Gestalt. Sin embargo, las leyes de agrupamiento todavía son empleadas para describir el modo en que percibimos el mundo exterior. De todos los intentos realizados para integrar estas leyes a un modelo teórico general, quizá el más interesante sea el que proponen los creadores de la teoría de la percepción caótica.

# Notas

- [1](#) Es una traducción (quizá no muy feliz) de *auditory stream* (“corriente”, “flujo”). Creemos, sin embargo, que el término “agregado” describe aceptablemente la idea de Bregman.
- [2](#) Las relaciones entre espectro armónico, periodicidad y tonicidad se exponen en el Apéndice I.
- [3](#) Algunos cantantes lo utilizan para conseguir *pianissimos* sorprendentes (y, estrictamente hablando, inaudibles a no ser por el mecanismo citado).
- [4](#) Recordemos que la localización de la fuente se refiere al mundo físico mientras que la lateralización del sonido pertenece al universo de las representaciones mentales (véase el capítulo VII).
- [5](#) Tema desarrollado en el capítulo VII.
- [6](#) Tal es el caso del “efecto de cafetería” (“cocktail party”).
- [7](#) En la actualidad se sabe que poseen un comportamiento caótico y no estocástico, como se creía hace un tiempo.
- [8](#) Dowling (1968), en Moore, B. (1997).
- [9](#) Véase este tema en el punto 4 del capítulo IV.
- [10](#) Véase este tema en el punto 5 del capítulo III.
- [11](#) El orden y aun el número de leyes varía según el autor que se consulte. El que exponemos aquí es el más frecuente en la literatura sobre el tema.
- [12](#) Trad. de Julio Guillén.
- [13](#) Trad. de Julio Guillén.

# Capítulo IX

## Percepción del habla

Resulta evidente que la percepción del habla no depende simplemente de elementos invariantes accesibles directamente en la onda acústica. Por ejemplo, los patrones de la señal acústica que corresponden a una palabra se modifican de manera compleja en función de los sonidos que la preceden y que la suceden, y un fonema que resulta altamente probable en determinado contexto lingüístico va a ser “oído”, aun en ausencia de los datos acústicos correspondientes. El reconocimiento de la palabra depende también de indicadores semánticos, sintácticos y circunstanciales. La importancia de estos indicadores se potencia a medida que aumenta el ruido presente en el canal de comunicación.

Aunque este capítulo trata sobre la percepción del habla, conviene repasar primero brevemente el mecanismo de fonación humano.

### 1. El aparato vocal y la fonación

Las menores unidades del habla son los *fonemas*. El idioma español posee alrededor de 24 fonemas diferentes que se definen a partir de los sonidos que es capaz de producir el aparato vocal.[\[1\]](#) En esencia, producimos tres clases básicas de sonidos diferentes: las *consonantes plosivas*, que son señales transitorias que se producen al bloquear y luego abrir repentinamente el tracto vocal; las *consonantes fricativas*, que dependen de la turbulencia en el aire cuando pasa a través de una pequeña abertura, y las *vocales*, cuyas ondas son periódicas y que poseen altura tonal definida.

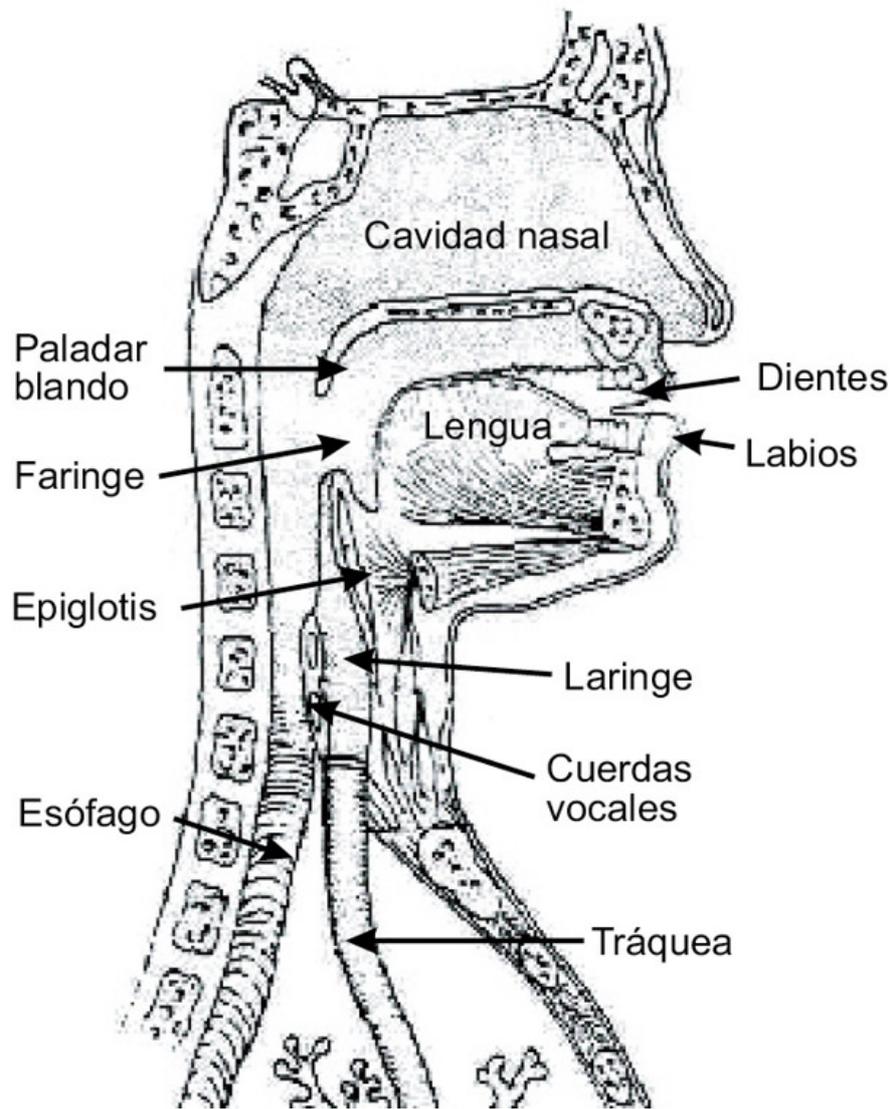
Veamos primero, de manera muy sintética, cómo y con qué producimos esta gran cantidad de sonidos diferentes.

#### El aparato vocal

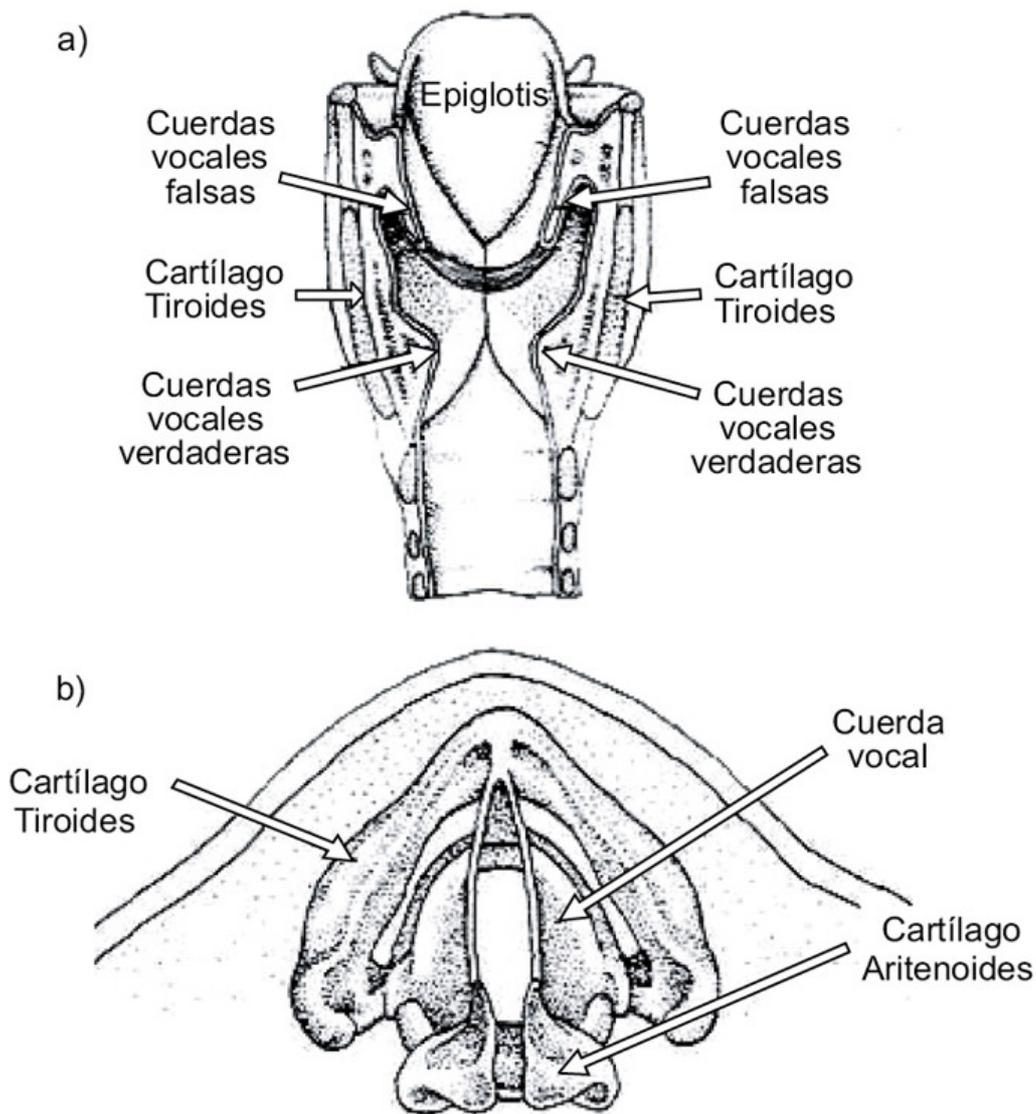
El aparato vocal humano se puede dividir en tres partes principales que son: una reserva de aire a mayor presión que la atmosférica, un canal de salida donde el aire puede ser interrumpido o modulado, y una cavidad resonante que filtra las señales que se emiten al exterior.

La reserva de aire a presión, llamada presión subglótica, se localiza en los pulmones. Su capacidad promedio es de tres a cuatro litros y puede llegar hasta seis durante una inspiración profunda. Movemos alrededor de medio litro durante la respiración pausada, pero un cantante entrenado es capaz de utilizar hasta cinco litros en una sola exhalación.

En el borde superior de la tráquea se ubica una especie de cavidad cartilaginosa llamada laringe, que la comunica con el tracto vocal, término que incluye la garganta y la cavidad naso bucal. Dentro de la laringe se hallan las cuerdas (o membranas) vocales, encargadas de estrechar u ocluir el paso de aire a partir de la acción de varios pequeños músculos. La apertura con forma de V entre las cuerdas vocales se denomina glotis. La glotis mide aproximadamente 2 cm de largo y 1 cm de ancho cuando está totalmente abierta. En la figura 1 se puede ver el corte lateral de la parte superior del aparato vocal, y en la figura 2 se ven dos cortes esquemáticos de la laringe.



**Figura 1.** Corte lateral de la parte superior del aparato vocal humano



**Figura 2.** Vista frontal y superior de una laringe humana

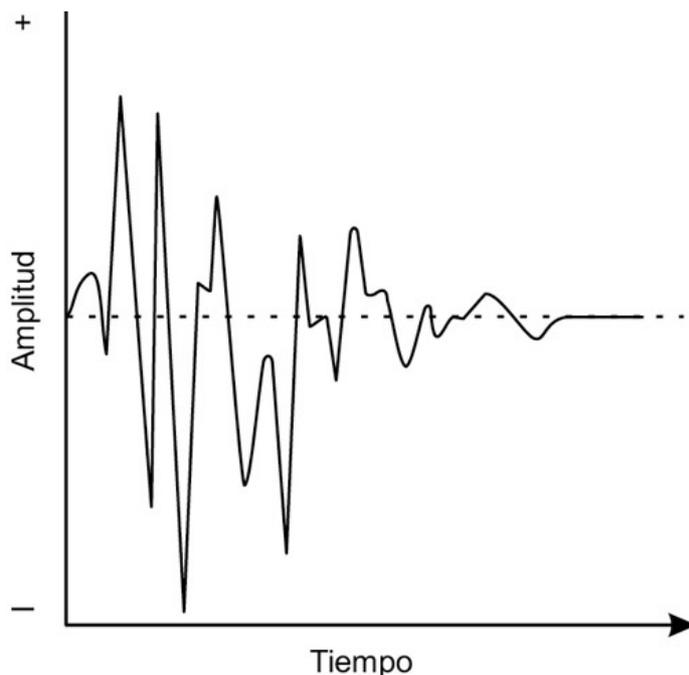
En fonación, la caja torácica aplica una presión subglótica un poco mayor que la presión atmosférica –entre 1,005 y 1,04 atmósferas– que fuerza al aire a pasar por la glotis y que hace vibrar las cuerdas vocales, cuya apertura estática es regulada por el cartílago aritenoides. Las cuerdas vocales, a su vez, vibran a una frecuencia controlada por la presión subglótica y por la tensión aplicada a sus pequeños músculos. Para el habla normal, las frecuencias se extienden en un rango que va de 140 a 400 Hz en las mujeres y de 70 a 200 Hz en

los hombres. Estos rangos pueden extenderse una octava o más en el canto.

## Producción del sonido

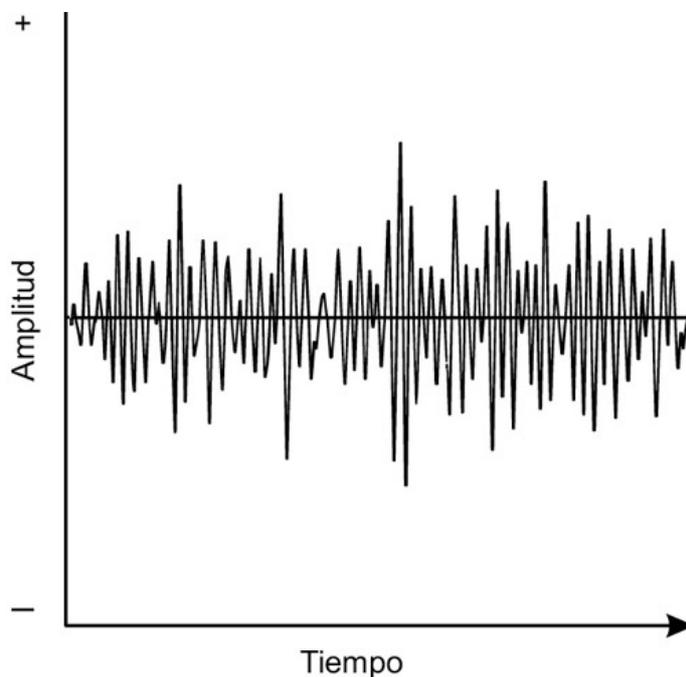
En la fonación participan varios mecanismos diferentes durante la producción de ondas. Las *consonantes plosivas*, que son señales transitorias que se producen al bloquear y luego abrir repentinamente el tracto vocal, tienen un desarrollo temporal en el que primero aparece un ataque similar a un impulso de gran presión, seguido por un breve intervalo en el que la vibración se atenúa rápidamente. En la figura 3 se ve el gráfico temporal de la consonante plosiva /k/.

Esta señal, transitoria y no periódica, posee un espectro de banda ancha y, por supuesto, carece de altura tonal. Las plosivas se dividen en sordas (/p/, /t/ y /k/) y sonoras (/b/, /d/ y /g/). En estas últimas, las cuerdas vocales se ponen en vibración y una vocal suena inmediatamente después –unos 30 ms– de la consonante. Como el carácter transitorio de las plosivas hace que contengan poca energía, los cantantes generalmente las exageran para equilibrarlas en sonoridad con respecto a las vocales.



**Figura 3.** Gráfico temporal de la consonante /k/

Las *consonantes fricativas* (/s/, /z/, /sh/, /j/, /f/, /v/, etc.) producen una turbulencia en el aire que fluye a través de una pequeña abertura a mayor velocidad que la velocidad crítica. Aunque se puede sostener esta clase de señales en el tiempo, son estacionarias solamente en el sentido que lo es el ruido blanco. Sus ondas son no periódicas y el espectro es de banda ancha. En la figura 4 se aprecia el gráfico temporal aislado de la consonante fricativa /sh/.



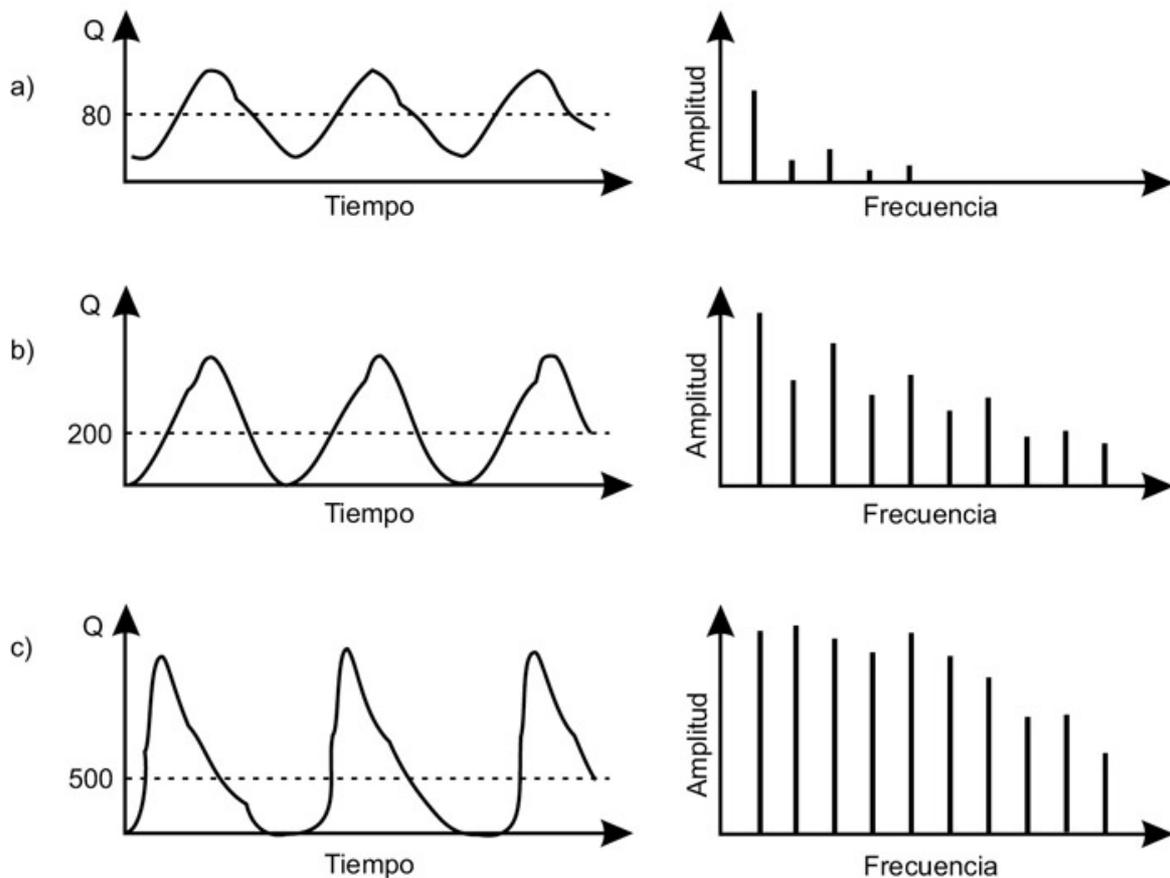
**Figura 4.** Gráfico temporal de la consonante /sh/

Aunque el tracto vocal pueda enfatizar algunas frecuencias de la banda total generada por una consonante fricativa (por ejemplo, 6 kHz para la /s/ o 3 kHz para la /sh/) no se destaca ninguna frecuencia en particular y el sonido resultante es no tónico. Existen otros sonidos como las *consonantes nasales* (/m/, /n/, /ñ/), las *vibrantes* (/r/, /rr/), etc., pero los modos de producción no difieren en esencia de los ya vistos.[\[2\]](#)

A diferencia de las consonantes, las *vocales* se mantienen estables en el tiempo, sus formas de onda son prácticamente periódicas y sus espectros son aproximadamente armónicos. En otras palabras, las vocales sí poseen altura tonal definida. Se producen por la modulación periódica de la apertura glótica. Tanto la frecuencia de vibración de las cuerdas vocales como la forma de onda resultante están determinadas casi completamente por la laringe. La frecuencia fundamental, que define la periodicidad y la altura tonal, depende de la presión subglótica y de la tensión y apertura de las cuerdas vocales. El tracto vocal, tapizado de tejido blando, casi no devuelve energía hacia la laringe.[\[3\]](#) Las diferentes posiciones de la lengua, por ejemplo, modifican la acción de filtrado y determinan cuál vocal se va a emitir, pero no alteran la periodicidad base de la señal que emiten las cuerdas vocales.

En la figura 5 se pueden observar las formas de onda que parten de la laringe para tres niveles de intensidad diferentes. A mayor intensidad aumenta significativamente la cantidad de armónicos en el espectro.

De la laringe surgen ondas aproximadamente periódicas que se parecen, según el caso, a una onda triangular o una onda diente de sierra con gran cantidad de armónicos. El sonido que se registra en la laringe se siente como un zumbido con altura tonal definida, no como los sonidos mucho más agradables que oímos cuando el locutor emite por la boca. Esta gran diferencia acústica se debe a la acción de los filtros de resonancia del tracto vocal.



**Figura 5.** Formas de onda y espectros tomados en la laringe: a) señal débil, b) señal de mediana intensidad y c) señal muy intensa

## Formantes

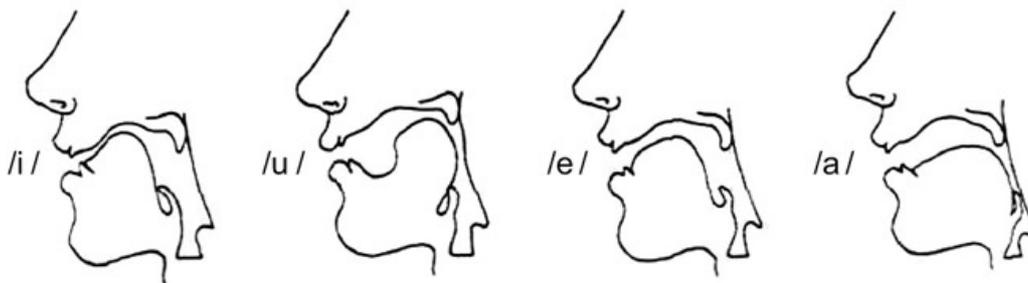
El tracto vocal se puede aproximar a un tubo de 17 cm de largo, cerrado en un extremo por las cuerdas vocales y abierto en el otro a través de la cavidad naso bucal.

Cada vocal define una geometría particular del tracto vocal, como se puede ver en la figura 6.

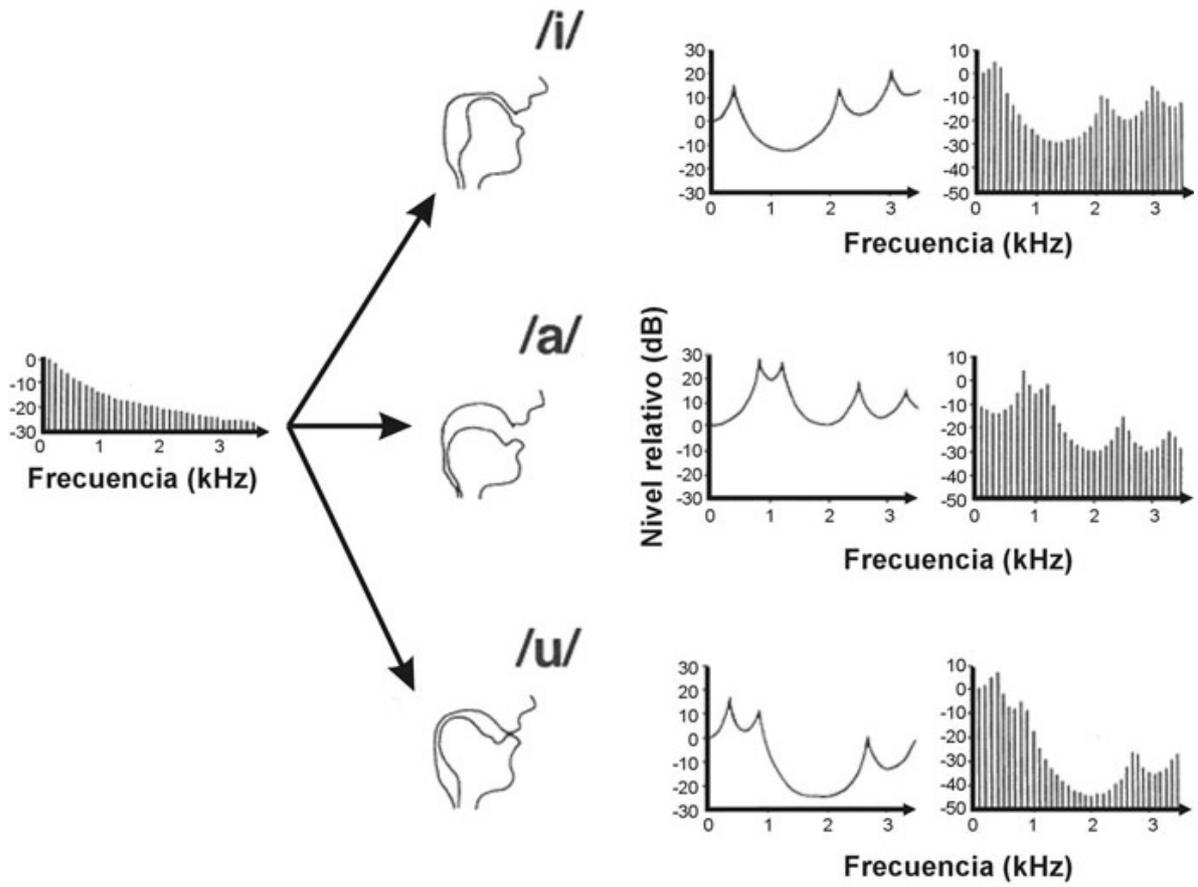
Cada configuración particular del tracto vocal, a su vez, modifica las características espectrales del filtro acústico correspondiente. Lo que oímos es, en definitiva, una señal periódica generada en la laringe y filtrada por el tracto vocal. En la figura 7 se puede ver cómo la misma onda, emitida por la faringe, adopta envolventes espectrales diferentes para cada vocal. En este ejemplo se oye siempre la misma

altura tonal, pero con tres alturas espectrales diferentes (las vocales /i/, /a/ y /u/).

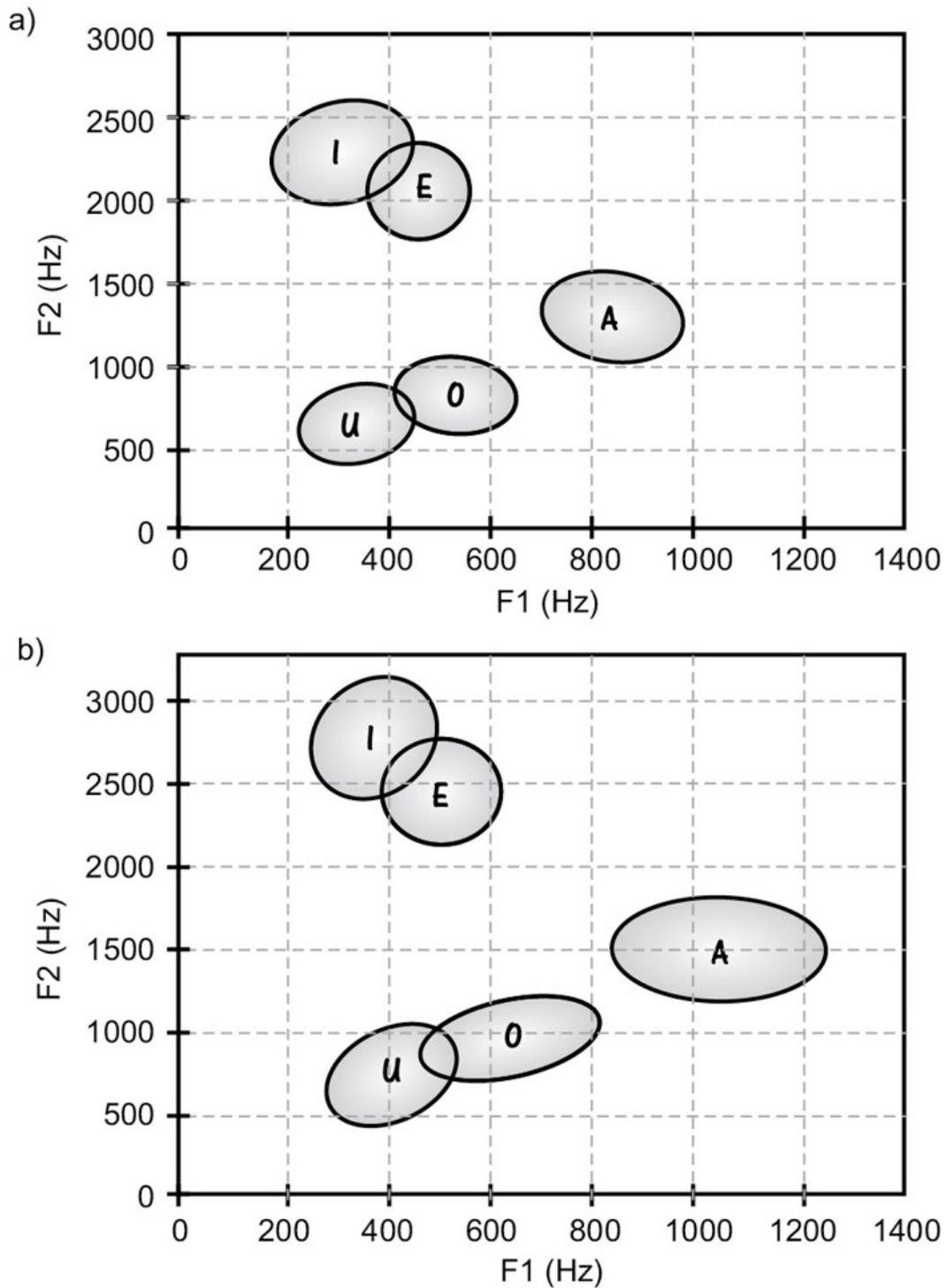
La cavidad naso bucal posee resonancia de banda media porque el tejido blando que la tapiza absorbe gran cantidad de energía acústica. Cada una de estas zonas de resonancia, en las que se refuerza la amplitud de los componentes espectrales, se denomina *formante*. Cada vocal corresponde a una combinación particular de formantes, tres o cuatro según el caso. Sin embargo, es posible distinguirlas con claridad a partir únicamente de las dos primeras. Si se construye un gráfico cartesiano que posea la frecuencia de la primera formante en abscisas y la de la segunda formante en ordenadas, se obtiene un espacio bidimensional en el que cada punto corresponde –idealmente– a una vocal. En la vida real, por supuesto, existe gran variabilidad no sólo entre un hablante y otro, sino hasta en la misma persona en diferentes momentos del día. Por lo tanto hay un rango de valores para cada par de formantes (una región en el gráfico de la figura 8) que contiene información sobre la misma vocal (Aronson, L. *et al.*, 2005).



**Figura 6.** Configuraciones del tracto vocal para las vocales /i/, /u/, /e/ y /a/



**Figura 7.** Efecto de filtro del tracto vocal sobre una señal periódica proveniente de las cuerdas vocales



**Figura 8.** Regiones de reconocimiento vocal para las dos primeras formantes del español rioplatense; a) voces masculinas, b) voces femeninas

Los rangos se superponen en algunos casos y los mismos sonidos pueden percibirse en dos formas diferentes. Por lo general, sólo una

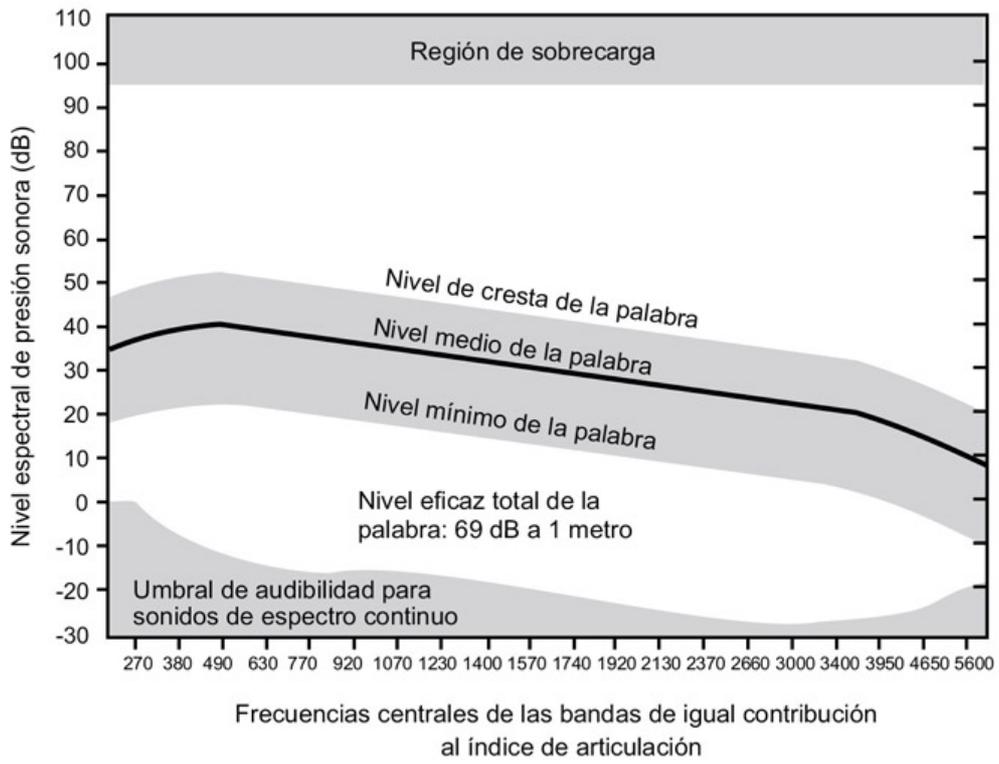
de las interpretaciones tiene sentido en su contexto y la adoptamos automáticamente de manera preconsciente. Una de las razones de la gran extensión de las áreas para cada vocal es que las mujeres, los hombres y los niños no poseen el mismo tamaño de tracto vocal y no pueden producir exactamente las mismas formantes. Los oyentes evalúan las características acústicas generales del locutor antes de decidir las formantes que le van a corresponder a cada vocal. Como ejemplo, en la figura 8 se comparan las regiones de reconocimiento vocal para voces masculinas y femeninas.

### **Características generales de la emisión vocal**

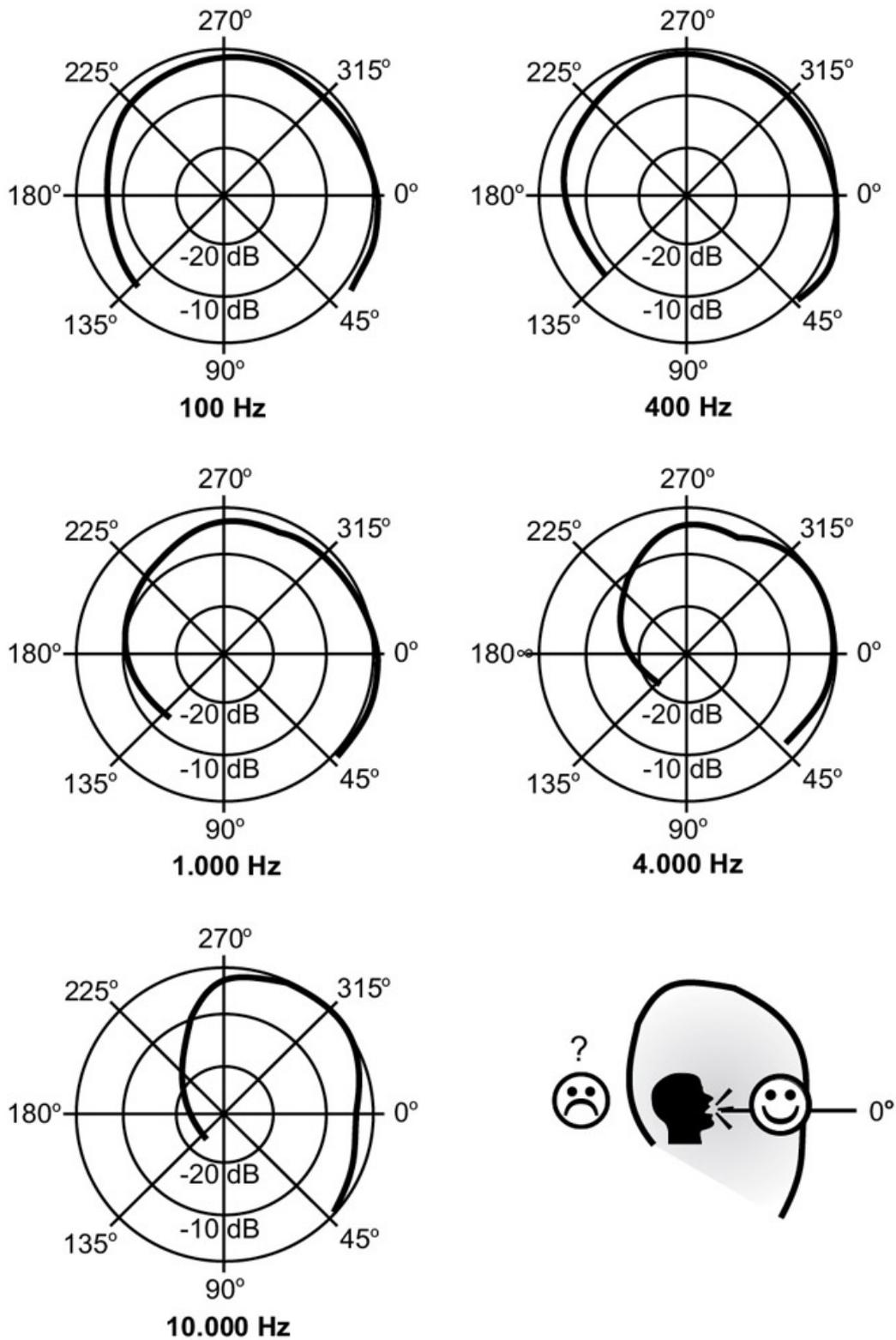
La voz humana genera un espectro de potencia que alcanza un máximo de energía alrededor de los 500 Hz en los hombres y de los 1.000 Hz en las mujeres, para luego decaer a razón de -8 dB por octava hasta una frecuencia comprendida entre 4.000 y 5.000 Hz. En la figura 9 se puede ver el nivel medio de emisión de una voz masculina tomado a 1 m del locutor. El rango dinámico útil para la palabra es del orden de los 30 dB. Las crestas más potentes se encuentran a alrededor de +12 dB por encima y las más débiles a -18 dB por debajo del nivel medio.

En la región de sobrecarga el oído no puede discriminar entre diferentes fonemas y la inteligibilidad cae bruscamente.

Por supuesto, la voz humana no emite de manera omnidireccional a todas las frecuencias. Es marcadamente direccional por encima de los 1.000 Hz. En la figura 10 se muestran los diagramas de directividad en el plano vertical correspondientes a cinco frecuencias diferentes.



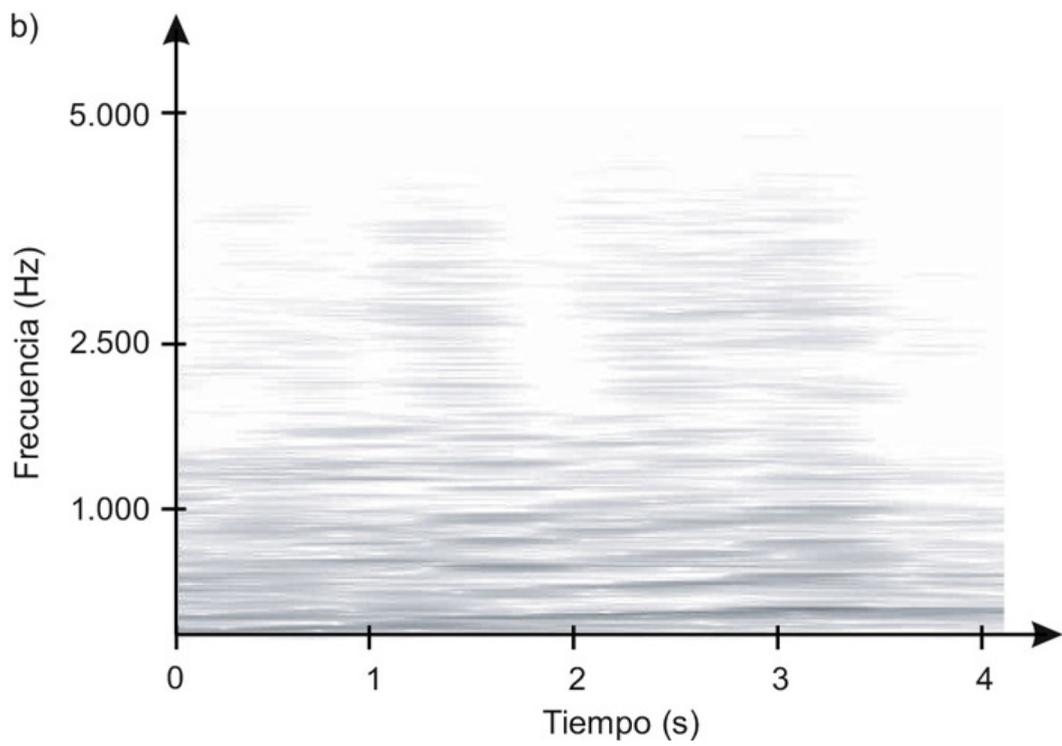
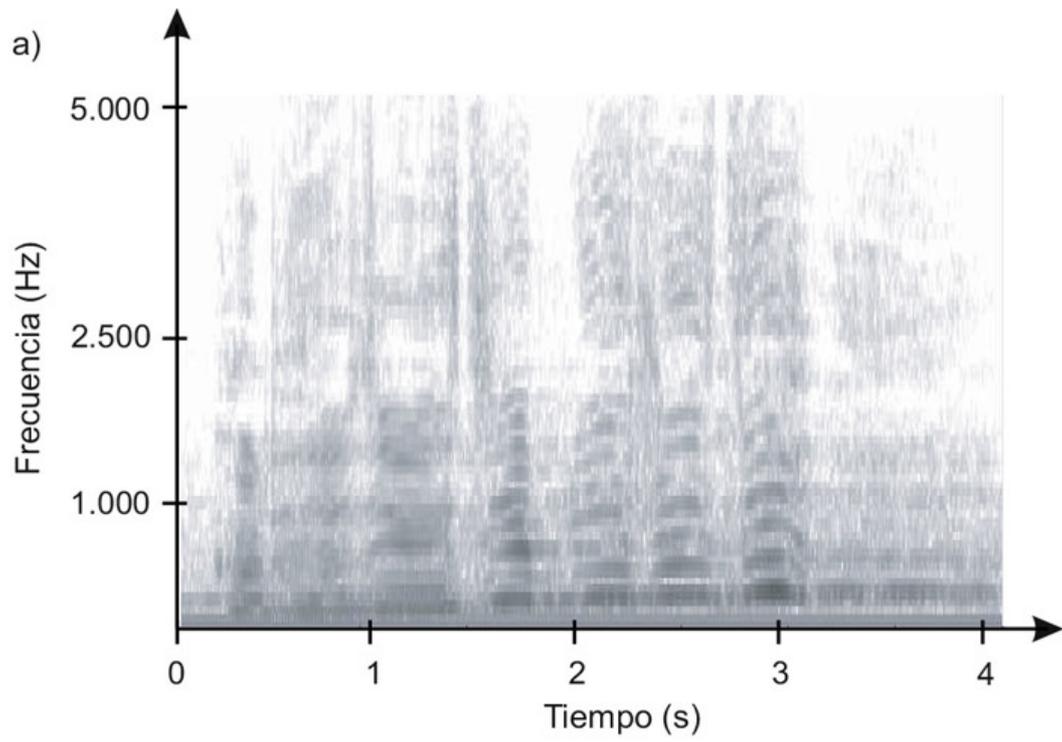
**Figura 9.** Niveles espectrales de emisión para una voz masculina



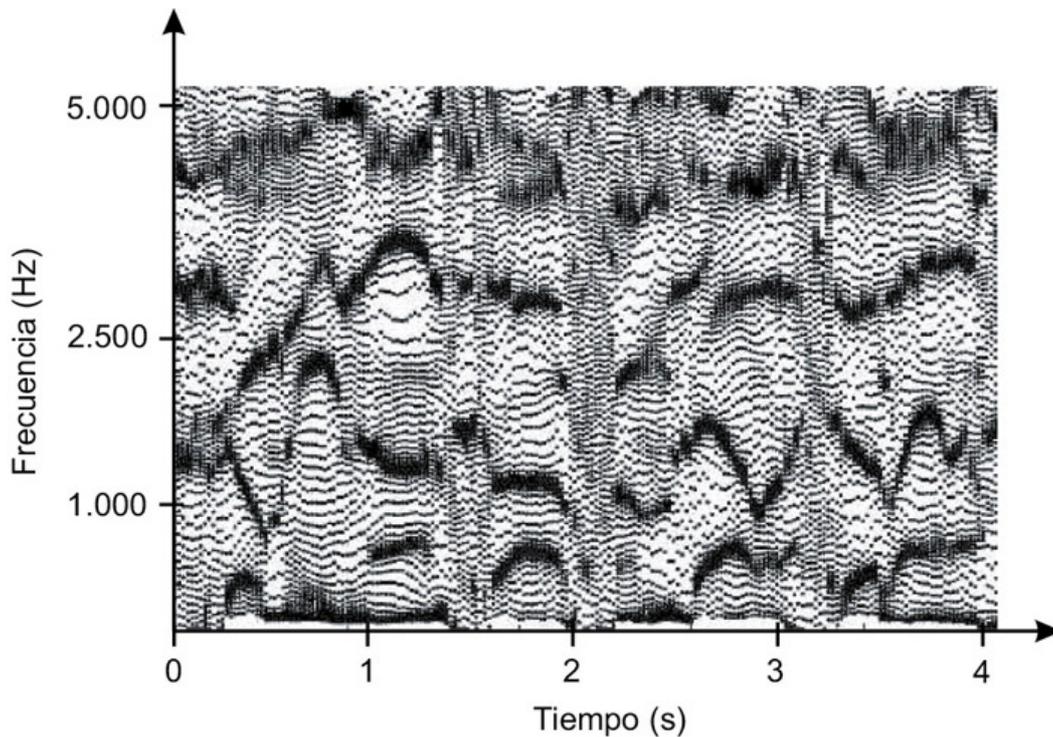
**Figura 10.** Directividad de voz humana en el plano vertical

## 2. Representación gráfica de los sonidos del habla

Como los espectros de los sonidos del habla varían en el tiempo –no son estáticos–, para estudiarlos se usan tradicionalmente gráficos cartesianos tiempo/frecuencia que se denominan *sonogramas*. El problema con estas formas de representación es que, como todo sistema pasivo de análisis de ondas, dependen del límite impuesto por el principio acústico de indeterminación. Resulta imposible obtener simultáneamente gran resolución en el tiempo y la frecuencia. Una solución de compromiso consiste en alternar sonogramas de banda ancha y de banda angosta. En los primeros la banda típica abarca 300 Hz, valor que proporciona una buena resolución temporal ( $1/300$  s), pero que no permite observar los armónicos individuales de una señal periódica. Los sonogramas de banda angosta poseen normalmente bandas de 45 Hz, suficiente para distinguir los armónicos, pero no para diferenciar individualmente los rápidos impulsos glóticos. En la figura 11 se pueden ver dos sonogramas de diferente ancho de banda.



**Figura 11..** Sonogramas de banda ancha y de banda angosta



**Figura 12.** Espectrograma auditivo

Nuestro oído puede resolver con gran detalle y (casi) simultáneamente los aspectos temporales y espectrales de una señal acústica, como ocurre cuando oímos una sucesión rápida de acordes. Sin embargo, no viola el principio acústico de indeterminación porque posee la capacidad de activar y desactivar rápidamente los sensores de la membrana basilar, corrigiendo así el ancho de banda efectivo de acuerdo a la tarea impuesta.[\[4\]](#)

Como ninguno de los sonogramas tradicionales representa lo que el oído detecta realmente, algunos autores crearon los *espectrogramas auditivos*, en los que el ancho de banda de cada frecuencia de análisis iguala a la banda crítica que corresponde a su frecuencia central (Carlson, Granström y Hunnicutt, 1982). En la figura 12 se muestra el espectrograma auditivo del registro de voz hablada de la figura 11. En este caso se pueden observar a la vez los armónicos en bajas frecuencias y las transiciones rápidas a altas frecuencias.

Cuando se compara un registro acústico de voz hablada con su representación gráfica lo primero que llama la atención es que los si-

lencios generalmente no corresponden a los espacios entre palabras. Una de las mayores dificultades a la hora de analizar gráficamente esta clase de señales es determinar el modo en que la secuencia de sonidos se segmenta en sílabas y palabras.

### **3. Características particulares de la percepción del habla**

Aunque el tema sigue siendo motivo de controversia entre neurólogos y lingüistas, es probable que en el ser humano hayan evolucionado mecanismos especiales dedicados a la percepción de los sonidos propios del habla. Existe evidencia de especialización cerebral relacionada con la percepción del habla. Por ejemplo, se ha probado que intervienen diferentes regiones del cerebro en la percepción de los sonidos del habla y en la de los sonidos ambientales en general. Si se aplican simultáneamente a los dos oídos estímulos diferentes –por ejemplo dos frases distintas– el estímulo hablado presente en el oído derecho –que se procesa en el hemisferio cerebral izquierdo– es comprendido con mayor facilidad. Lo contrario ocurre en el caso de la música.

Experiencias con tomografía computada y resonancia magnética demostraron que, mientras que los sonidos ambientales y musicales son procesados en ambos hemisferios cerebrales, los sonidos del habla estimulan casi exclusivamente el hemisferio izquierdo. Esta característica estaría basada en la necesidad de mantener cercanos todos los centros involucrados en el tratamiento de las señales del habla, a fin de permitir la detección de sus pequeñas transiciones –del orden del milisegundo. De otro modo, si la información tuviera que viajar entre hemisferios, estas transiciones se perderían debido al excesivo tiempo involucrado en la comunicación.

A continuación nos referimos a algunas de las características particulares de la percepción del habla, que nos permiten decodificar con eficacia los fonemas en una amplia variedad de situaciones y contextos.

## ***Tasa de ocurrencia de los sonidos del habla***

En el hablar rápido pueden ocurrir 30 fonemas por segundo. Liberman especuló que esta tasa supera la resolución temporal del sistema auditivo (Liberman, A. *et al.*, 1967). Sin embargo, la evidencia reciente indica que es posible identificar secuencias donde cada fonema dura apenas 10 ms, a una tasa de ocurrencia de 100 fonemas por segundo. A esa velocidad no se percibe cada evento por separado, pero se puede seguir auditivamente el patrón sonoro global.

## ***Variación de los datos acústicos***

La percepción del mismo fonema puede estar determinada por diferentes datos acústicos. Por ejemplo, una consonante fricativa “acomoda” sus bandas de ruido de acuerdo con la vocal que la sucede: vocales diferentes dan lugar, inevitablemente, a diferentes transiciones entre formantes. Una sola transición de la señal acústica puede contener información sobre varios fonemas vecinos.

## ***Modo lingüístico***

No hay un *continuum* en la percepción de los sonidos del habla, la percepción es dicotómica. En general, los sonidos son percibidos como lingüísticos o como no lingüísticos. Se puede activar el modo lingüístico dándole ciertas instrucciones al oyente, pero una vez activado es muy difícil revertir el proceso.

Una de las consecuencias de esta característica de la percepción del habla es nuestra incapacidad para oír los sonidos propios del lenguaje en términos de sus parámetros acústicos. Los percibimos como una corriente única de palabras y las descripciones “acústicas” son, casi siempre, confusas. Para lograr una audición limpia de los datos acústicos es necesario aislar cada fonema y sacarlo de contexto. Sólo a partir de ese momento se desactiva el modo lingüístico.

Resulta casi innecesario mencionar la importancia que tiene esta característica en la percepción de canciones y piezas líricas, en las

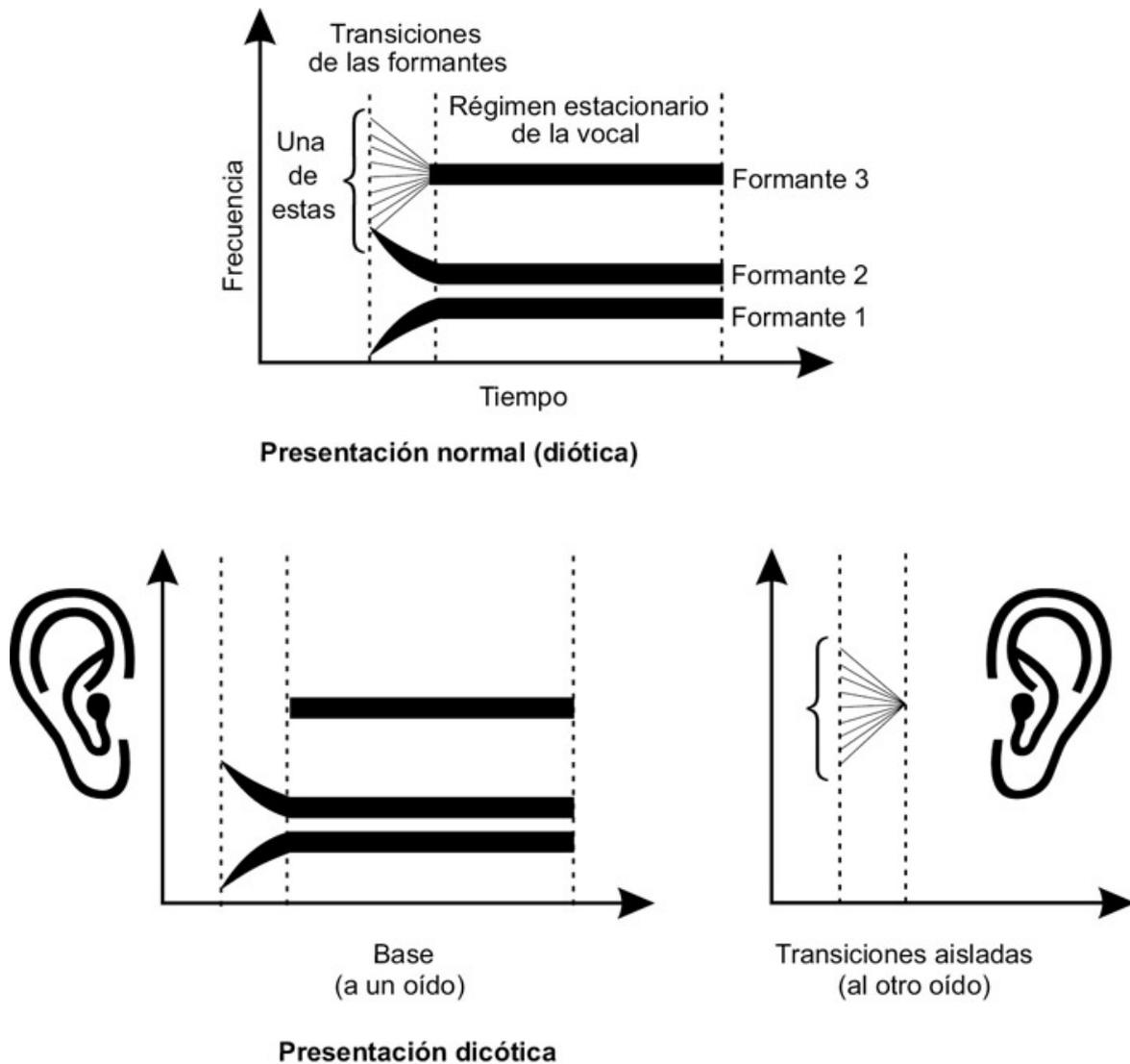
que coexisten elementos propios del modo lingüístico (el texto) y no lingüístico (la música).

### *Percepción dual*

En la década de 1970 Rand realizó por primera vez un experimento que demostró la existencia de lo que se conoce como *percepción dual* (Rand, T. C., 1974). Más adelante, Liberman perfeccionó la experiencia de Rand de la siguiente manera: se presenta a un oído la parte de la señal propia de una sílaba a la cual se le ha quitado la transición del tercer formante, y al otro la parte estática del tercer formante más los dos primeros formantes completos (Liberman, A., 1995). Los oyentes dicen percibir la sílaba completa en el primer oído y sólo un corto zumbido en el otro. En la figura 13 se aprecia la descripción esquemática de dicho experimento.

Según Liberman, los resultados obtenidos indican que existen módulos separados que procesan, por un lado, los sonidos del habla y, por otro, los demás sonidos. Esta conclusión concuerda con los datos obtenidos a partir de la observación directa de la actividad cortical durante la audición.

Una de las consecuencias interesantes asociadas a este fenómeno es que se viola el principio de *separación espacial* que sostiene que un elemento acústico dado no puede ser asignado a más de una fuente a la vez. Parece que el mecanismo de percepción lingüístico es tan poderoso que, bajo ciertas circunstancias, permite agrupar elementos acústicos aún cuando éstos provengan, o aparenten provenir, de fuentes diferentes.



**Figura 13.** Esquema del experimento de Liberman sobre la percepción dual

### *Integración audiovisual*

Al decodificar un mensaje hablado le prestamos atención a una gran cantidad de datos no acústicos. Los movimientos de la cabeza, de las manos y, en particular, de los labios del hablante pueden influir significativamente en la percepción del habla.

Es interesante mencionar aquí un experimento en el que se presenta un video con el audio modificado a un grupo de espectadores (McGurk y MacDonald, 1976). En la imagen aparece un locutor di-

ciendo “tata” junto al audio del mismo locutor pronunciando la palabra “mama”. Los oyentes dicen oír “nana”, sin expresar ninguna duda –oyen “nana” con claridad–, y se sorprenden al oír el audio aislado, libre de toda información visual.

### *Percepción categórica*

Una consecuencia de la percepción en modo lingüístico es nuestra capacidad para ubicar cada fonema en una determinada categoría individual. Cuando varía de manera continua algún parámetro acústico de la señal, por ejemplo la frecuencia de una formante, no oímos una serie de pequeños cambios graduales, sino que percibimos saltos “cuánticos” entre un fonema y otro. Este fenómeno se denomina *percepción categórica* y no ocurre usualmente con los sonidos ambientales. No se distinguen los cambios acústicos de la señal cuando éstos caen en la misma categoría, pero se perciben como muy grandes si provocan un salto de categoría. La percepción categórica se nos presenta como evidente cuando tratamos de comprender un discurso en una lengua que no dominamos completamente: las confusiones frecuentes se deben a una insuficiente competencia respecto de las categorías fonémicas de esa lengua.

La percepción de vocales estacionarias es muy diferente pues los pequeños cambios físicos son percibidos con facilidad. Según Liberman, esto ocurre porque las vocales aisladas están mucho menos codificadas y se las oye como sonidos corrientes, no categorizados lingüísticamente. No obstante, las vocales dentro de una secuencia del habla se comportan de manera similar a las consonantes, es decir, de manera altamente codificada.

Un ejemplo de percepción codificada y categorizada de sonidos no pertenecientes al habla es la percepción de las alturas musicales en un sistema escalístico determinado. En música tonal, las funciones armónicas también están altamente categorizadas.

Existen tres hipótesis principales elaboradas para explicar la percepción categórica. En la primera, Pisoni considera que la memoria

auditiva para los patrones acústicos de las consonantes decae rápidamente y que cuando deberían ser procesados en la identificación acústica del fonema ya se han perdido: la discriminación fina del estímulo dentro de esta categoría de fonemas es imposible (Pisoni, D. B., 1973). Para fonemas más largos e intensos, como los de las vocales prolongadas, los patrones acústicos pueden retenerse en la memoria auditiva por períodos mayores, pudiendo realizarse sobre ellos operaciones de discriminación similares a las operadas en los sonidos en general.

En la segunda hipótesis K. Stevens sugiere que las categorías y los límites en el habla han evolucionado con el fin de explotar la sensibilidad natural del sistema auditivo (Stevens y Blumstein, 1981). Esta hipótesis no da cuenta de muchas de las operaciones características necesarias para la aparición de la percepción categórica.

Una tercera hipótesis, formulada por Kuhl, considera que la percepción categórica surge de la experiencia con el propio lenguaje (Kuhl, P. K., 1993). Cuando aprendemos a entender las palabras de un lenguaje particular, prestamos atención a las diferencias acústicas que afectan su significado e ignoramos las que no lo afectan. Una vez que aprendemos a distinguir entre ambos grupos, nos cuesta oír las diferencias acústicas que no atañen al significado. Esta explicación da cuenta de la dificultad para oír diferencias entre fonemas en un lenguaje no familiar. Kuhl encontró que los niños nacen con habilidades relevantes respecto del lenguaje que parecen depender de mecanismos generales de la percepción auditiva: no responden de manera específica a una lengua en particular. Pero a los seis meses la percepción de los bebés ya está polarizada por la exposición a una lengua específica. Kuhl propuso la teoría del *magnetismo del lenguaje nativo*, en la que la exposición a una lengua en particular resulta en la formación de representaciones almacenadas de categorías fonéticas, que generan “prototipos” de cada categoría. Los prototipos actúan como imanes perceptuales: la percepción categori-

zada es una consecuencia natural. La teoría de Kuhl es compatible con la hipótesis de que el cerebro funciona como un sistema que opera en la frontera del caos.

## **4. Modelos de percepción del habla**

### **Teoría motora**

Esta teoría establece que los objetos de la percepción de la palabra son tratados como gestos fonéticos del hablante, representados en el cerebro como comandos de movimiento invariantes relacionados con ciertas configuraciones lingüísticas significativas (Lieberman y Mattingly, 1985). Estos gestos no se manifiestan directamente en la señal acústica.

El modelo es incompleto pues no especifica cómo se logra la transformación entre la señal acústica y los gestos percibidos. Según Klatt, “es más una filosofía que una teoría de la percepción del habla” (Klatt, D. H., 1989).

### **Teoría de los rasgos invariantes**

A su vez, Stevens y Klatt defienden una teoría en la que la percepción se da en una serie de etapas que incluyen un conjunto de detectores fonéticos (Klatt, D. H., 1989). La primera etapa ocurre en el sistema auditivo periférico (filtrado, supresión lateral, adaptación y sensibilidad a la fase). La segunda consta de una batería de detectores acústicos (detectores de ataque, de cambio espectral, de formantes y de periodicidad). La tercera etapa la forman una serie de detectores de rasgos fonéticos y sus decisiones son específicas de una lengua determinada. Finalmente, aparecen etapas de análisis de segmentos y búsqueda de léxico.

Este modelo está basado en la suposición de que es posible encontrar un mapeo relativamente invariante entre los patrones acústicos y los sonidos del habla. Sus detractores señalan el carácter inevitablemente secuencial de las operaciones involucradas.

## **Modelo de red neuronal**

El modelo de McClelland y Elman es conexionista y está basado en las teorías de las redes neuronales (McClelland y Elman, 1986). Asume que existen tres niveles de representación, cada uno de los cuales contiene unidades de procesamiento altamente interconectadas llamadas nodos. En el nivel más bajo, los nodos representan rasgos fonéticos, en el siguiente representan segmentos fonéticos, y en el último nivel representan palabras. Por ejemplo, el nodo correspondiente a un fonema específico puede disparar hacia dos o tres nodos correspondientes al nivel de los rasgos asociados normalmente con dichos fonemas. La activación excitatoria o inhibitoria puede fluir en ambas direcciones, desde niveles más bajos hacia otros más elevados, o viceversa.

Este modelo permite explicar varios aspectos de la percepción del habla, tales como la restauración perceptiva de fonemas “perdidos”, la percepción categorizada y la armonización de datos. Es uno de los modelos compatibles con la teoría de la percepción caótica.

## **5. La búsqueda de indicadores acústicos invariantes**

Cole y Scott notaron que la percepción del habla involucra la identificación simultánea de, al menos, tres tipos de datos cualitativamente diferentes: indicadores acústicos invariantes, indicadores dependientes del contexto e indicadores asentados en la envolvente dinámica de la onda (Cole y Scott, 1974). En otras palabras, cada sílaba contiene datos invariantes y datos dependientes del contexto. Existe evidencia de que la envolvente dinámica de la onda provee información sobre la composición fonémica de la señal y de que, en cierto grado, el sistema auditivo está capacitado para decodificar los estímulos del habla basándose únicamente en las variaciones de la amplitud en el tiempo. La naturaleza multidimensional de la información acústica permite un alto nivel de redundancia en la señal del

habla. Coexisten diferentes datos que permiten la identificación de un fonema determinado, aunque con uno solo de ellos se lo pueda reconocer. La redundancia sirve para superar las ambigüedades inherentes al discurso hablado, para minimizar la influencia del ruido y de las interferencias, para compensar las distorsiones de la señal (por ejemplo en una línea telefónica) y para superar la mala articulación de un locutor.

## **Resistencia del habla frente al deterioro de la señal**

Es remarcable la resistencia del habla frente a varias formas de distorsión severa. Los métodos de testeo de articulación las han catalogado en: cantidad de ruido de fondo, cambios en el espectro de frecuencia y recorte de picos.

*Cantidad de ruido de fondo.* La inteligibilidad de sílabas aisladas resulta satisfactoria con una relación señal/ruido (S/R) de +6 dB. Sin embargo, las palabras resultan inteligibles aún para relaciones S/R negativas si integran sentencias bien formadas, en particular si el oyente está familiarizado con el tema, o si el habla y el ruido provienen de lugares diferentes.

*Cambios en el espectro de frecuencia.* Se ha demostrado que, contra la afirmación tradicional, no hay componentes en frecuencia esenciales para la comunicación. Se puede cortar una señal de habla con un filtro de banda pasante de 0 a 1.800 Hz, de 1.800 a 20.000 Hz o de 1.000 a 2.000 Hz, y no perder mucho en inteligibilidad.

*Recorte de picos.* Si se coloca un circuito que recorte la señal de entrada al 2% de su amplitud promedio queda una onda casi rectangular de período variable y se pierde la información aportada por la forma de onda original. Sin embargo, se obtienen tasas de articulación entre el 80 y el 90%: son necesarios sólo dos bits para codificar la amplitud sin perder casi nada de inteligibilidad. Por supuesto que el sonido no resulta “natural” y es imposible reconocer al locutor.

En resumen, el sistema de comunicación humano ha evolucionado de tal forma que puede operar bajo una gran variedad de condicio-

nes adversas. Sin estas capacidades el teléfono no existiría, o sería extraordinariamente complejo.

## 6. Inteligibilidad de la palabra

La inteligibilidad del discurso se refiere a la exactitud con que un oyente promedio puede entender una palabra o una frase hablada. En una situación normal parte de la información se extrae del contexto y de las señales visuales que acompañan a la fonación y, por lo tanto, es posible entender el significado incluso si solamente una fracción de las unidades acústicas discretas se oye correctamente. En otros casos el oyente queda limitado a valerse únicamente de la información acústica, como ocurre en los sistemas de comunicación no presenciales, por ejemplo el teléfono o las grandes salas en las que se emplea refuerzo electroacústico.

Un fonema es inteligible cuando se lo oye de forma clara y puede ser diferenciado del resto de los fonemas de la misma lengua. La misma definición puede aplicarse a un discurso articulado, aunque aquí actúan factores contextuales de importancia.

Veamos primero qué ocurre con los fonemas aislados. En este caso el punto crítico en cuanto a la inteligibilidad son las consonantes, pues al ser más débiles que las vocales se pueden enmascarar con mayor facilidad.

### Ensayos de articulación

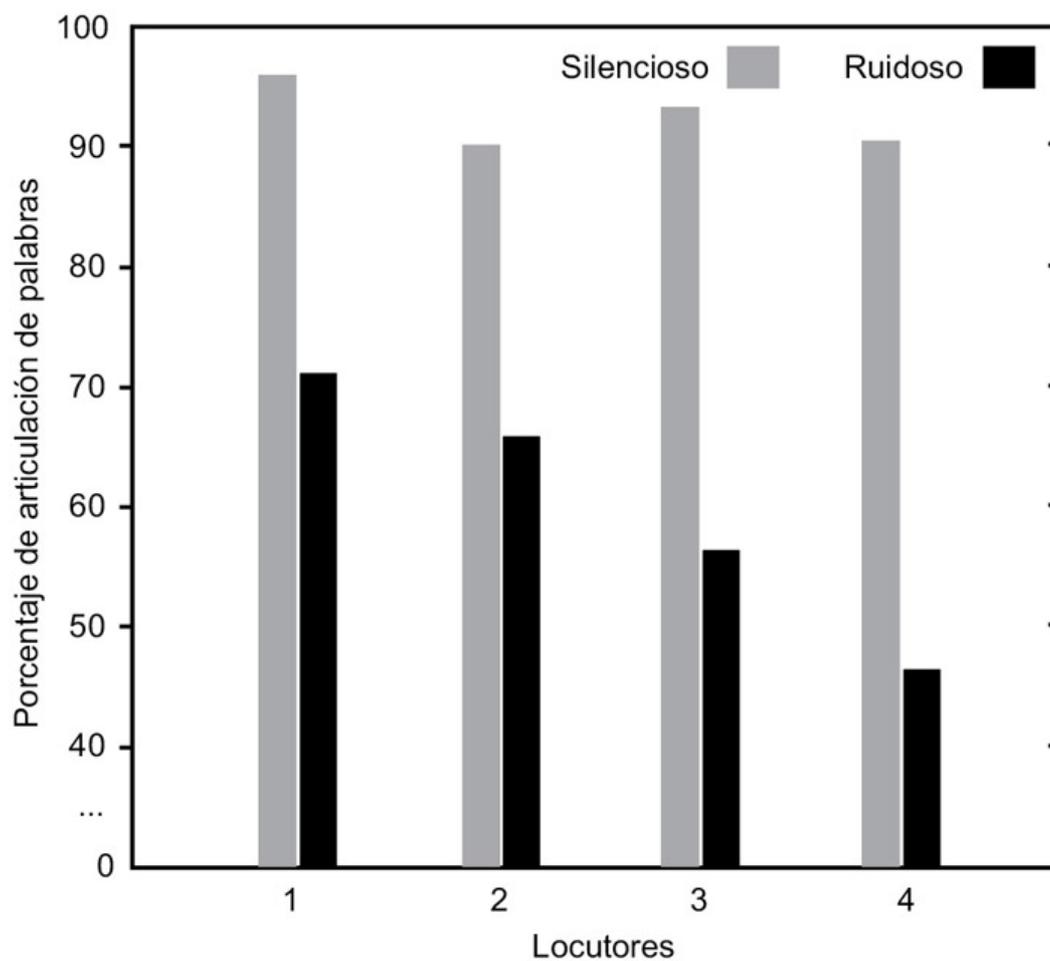
La investigación en esta área comenzó con el desarrollo de los sistemas telefónicos a fines del siglo XIX y de la radiofonía a principios del XX. Los primeros intentos para definir numéricamente la inteligibilidad fueron los *ensayos de articulación*, que consisten en pruebas de campo en las que un locutor lee un conjunto de fonemas o palabras de una lista confeccionada al azar. Para evitar la influencia de elementos no fonéticos, las secuencias carecen de sentido lingüístico y las sílabas están ordenadas de manera estocástica. Es habitual el uso de listas estandarizadas para permitir la comparación entre en-

sayos.<sup>[5]</sup> Los oyentes, distribuidos en la sala de una manera prede-terminada, anotan los fonemas o palabras que creen oír. El resulta-  
do del test es el porcentaje de palabras correctas –que coinciden con  
la lista original– para cada oyente. Además de los factores acústicos  
específicos que analizaremos más adelante, los ensayos de articula-  
ción son muy sensibles a las características particulares de cada lo-  
cutor y de cada oyente. En las figuras 14 y 15 se pueden ver los re-  
sultados obtenidos a partir de dos grupos de locutores y oyentes  
(Beranek, L., 1954).

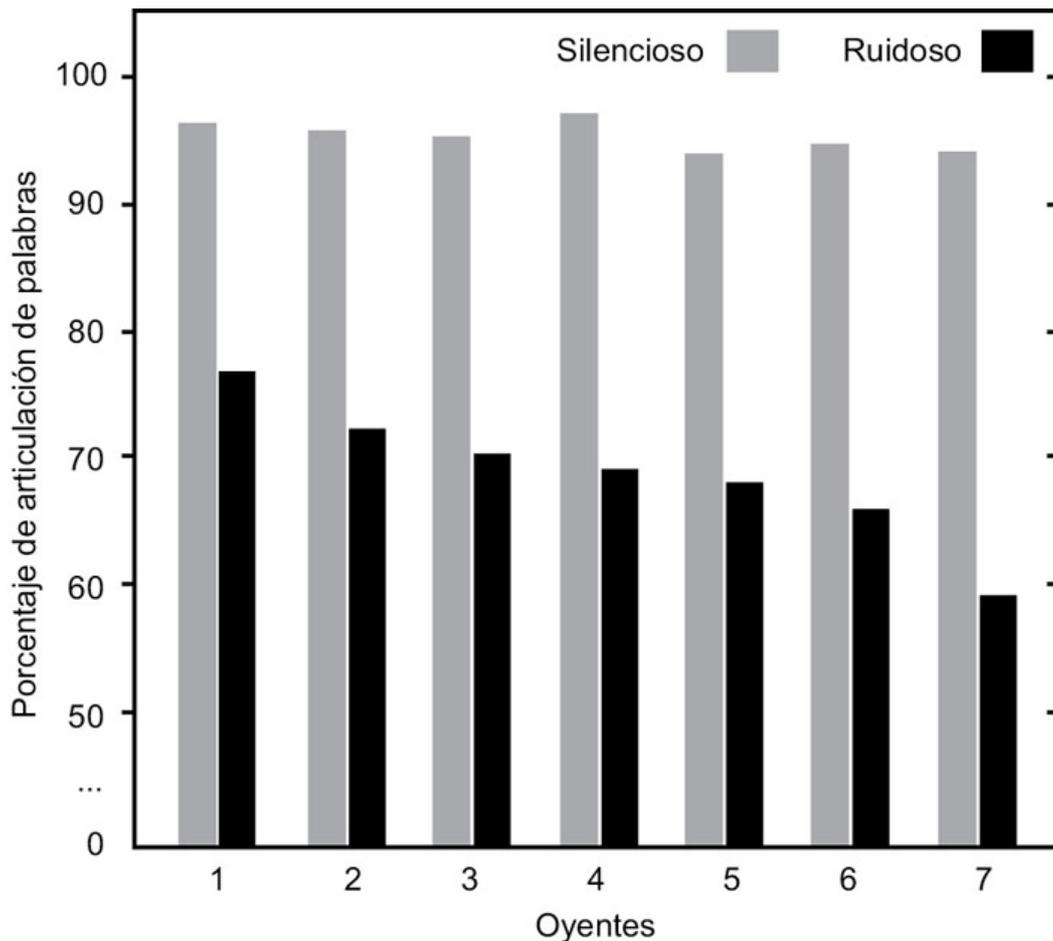
Una característica a destacar de los ejemplos anteriores es la no-  
table disparidad en la resistencia al ruido que presentan las voces de  
los diferentes locutores. En las mismas condiciones de ruido intenso,  
el primer locutor de la figura 14 se comprende correctamente, mien-  
tras que el último de la lista resulta absolutamente ininteligible.

La tasa de articulación se puede considerar una medida válida de  
la inteligibilidad, que es aplicable únicamente a cada caso particu-  
lar. El valor es representativo de la sala en la que se realizó el ensa-  
yo y de las características del locutor y de los oyentes que participa-  
ron en él.

En condiciones acústicas óptimas la articulación casi nunca supe-  
ra el 95% de respuestas correctas. Afortunadamente, un discurso ar-  
ticulado puede comprenderse aunque algunas sílabas resulten ininte-  
ligibles porque el oyente extrae el sentido del contexto. Por esta ra-  
zón, una tasa de articulación del 80% permite comprender el discurs-  
o sin esfuerzo. Cuando baja al 70% los oyentes deben concentrarse  
con esmero para entender las sentencias y, por debajo del 60%, en la  
mayoría de los casos la inteligibilidad resulta insuficiente.



**Figura 14.** Resultado de un ensayo de articulación de palabra a partir de cuatro locutores diferentes



**Figura 15.** Resultado de un ensayo de articulación de palabra obtenido con siete oyentes ubicados en la misma posición. El locutor fue siempre el mismo

Los ensayos de articulación, aunque brindan valiosa información sobre la inteligibilidad, son muy costosos en tiempo y recursos, y en algunos casos pueden resultar imposibles de realizar. En la actualidad se usan principalmente para verificar los modelos durante el desarrollo de métodos objetivos de medición y predicción de la inteligibilidad.

### **Medición indirecta y predicción de la inteligibilidad**

Existen varios métodos disponibles para predecir la inteligibilidad de habla en un sistema de comunicación o en el interior de una sala. Cada uno de estos métodos se basa en el mismo principio funda-

mental, que consiste en hallar alguna relación entre la señal del habla y el nivel del ruido que la interfiere en el canal de comunicación. En esta relación intervienen los siguientes factores mensurables: el nivel y otras características acústicas de la señal emitida, el nivel y composición espectral del ruido de fondo, y la naturaleza del canal de comunicación o de la sala.

En un entorno tranquilo, una conversación produce un nivel de presión sonora entre 55 y 65 dBA medidos a 1 m del locutor, aunque en situaciones especiales el nivel puede alcanzar 96 dBA. Como los niveles varían significativamente entre individuos, para el diseño de sistemas de comunicación se utilizan los casos más desfavorables en lugar de los niveles promedio. La tabla 1 presenta los valores estandarizados para el diseño de sistemas electroacústicos en grandes salas para prosa. Cada valor se obtiene restando un desvío estándar al promedio.

**Tabla 1**

<i>Esfuerzo vocal</i>	<i>dBA</i>
Susurro	32
Suave	37
Relajado	42
Normal (privado)	47
Normal (público)	52
Voz levantada	57
Intenso	62
Muy intenso	67
Grito	72
Máximo	77

Se encontró que un discurso gritado es más difícil de entender que uno a niveles medios, al margen del nivel acústico final en el oído de los oyentes. Este efecto de sobrecarga, que ya analizamos en este capítulo, se debe a los cambios en los transitorios del ataque y en la entonación de las formantes de la señal. Para simplificar los cálculos la sobrecarga se suele tratar de la siguiente manera: en primer lugar, para altos niveles de emisión se reduce la razón señal-ruido permitida en el canal un valor de -4dB cada vez que la señal supera en 10 dB el valor base de 75 dBA (medidos a 1 m del emisor). En segundo lugar, se asume que los niveles de emisión comprendidos entre 45 y 75 dBA que no llegan a los oyentes con valores superiores a 80 dBA no requieren corrección por sobrecarga.

En todo canal acústico hay siempre cierto nivel de ruido de fondo. El efecto más obvio del ruido de fondo es enmascarar la señal del habla. Cuantificar este efecto es difícil dadas las diferentes situaciones posibles, pero como estándar de diseño se considera que cada decibel de ruido que se agrega por encima de un nivel base de 45 dBA debe compensarse con un aumento de 0,6 dB en la señal hablada.

La naturaleza del canal de comunicación también puede afectar decisivamente la inteligibilidad del discurso hablado. Su influencia, beneficiosa o perjudicial, se puede evaluar a partir de la respuesta al impulso del canal. Si se trata de una sala existen dos factores que se analizan por separado: el patrón de reflexiones y la reverberación. En general, el recinto realzará la inteligibilidad cuando proyecte cierta cantidad de energía dentro del período de integración del discurso (de 35 a 50 ms). Dado que el ruido de fondo se distribuye de manera uniforme en el tiempo, toda reflexión temprana mejorará la relación señal-ruido del sistema. Sin embargo, las reflexiones tardías, los ecos y una reverberación excesiva interfieren con la señal y tienden a reducir la relación señal-ruido del canal, deteriorando la inteligibilidad del discurso.

### ***Relación señal-ruido***

Probablemente sea la más simple y sencilla de las relaciones que se usan para medir la inteligibilidad. Se la define como la resta entre el promedio de largo tiempo, o nivel sonoro continuo equivalente, de la señal del habla en el punto de recepción (LAeq. señal) y el nivel promedio de largo tiempo del ruido de fondo (LAeq. ruido). Ambos niveles se miden en dBA:

$$\text{Relación señal-ruido (SR)} = \text{LAeq. señal} - \text{LAeq. ruido (dBA)}$$

Las mediciones de inteligibilidad del discurso conforme a la relación señal-ruido realizadas por Bradley sugieren una meseta de nivelación por encima del valor  $sr = +15$  dBA. En otras palabras, mejorar el sistema más allá de este valor no aumenta la inteligibilidad del sistema. El defecto de este índice es que no contempla la naturaleza acústica del canal de comunicación: relaciones de  $sr$  óptimas pueden corresponder a una inteligibilidad deficiente si, por ejemplo, la sala presenta un eco destacado.

### ***Índice de articulación (AI)***

El primer estudio sistemático y profundo sobre la inteligibilidad del habla fue realizado por Harvey Fletcher durante la década de 1940 en los Laboratorios Bell. Fletcher y su equipo, a partir del estudio de la relación señal/ruido de cada banda de un tercio de octava, crearon el *índice de articulación* (AI), un excelente método para evaluar los efectos del ruido en la inteligibilidad.

El AI es básicamente una medida de la inteligibilidad basada en el cociente señal/ruido en cinco bandas de octava. El cálculo del AI consiste en tres pasos básicos: medición del cociente señal/ruido eficaz para cada banda de octava, aplicación de un factor de compensación tabulado para cada banda y cálculo del valor medio. La siguiente ecuación sintetiza el procedimiento:

$$AI = \frac{G_{[i]}}{30dB} \sum_{i=1}^5 (LA_{eq. señal} - LA_{eq. ruido} + 12) dB$$

en la que el factor de compensación  $G[i]$  vale:

Frecuencia [Hz]	$G[i]$
250	0.072
500	0.144
1000	0.222
2000	0.327
4000	0.234

El índice de articulación no es sensible a los cambios en la reverberación y a la presencia de distorsión temporal en el campo acústico, factores que sí contemplan los ensayos de articulación. Por ello se limita su aplicación a campos acústicos sin defectos ostensibles y con un tiempo de reverberación máximo de 0,5 s.

Durante varias décadas, el AI y los ensayos de articulación fueron los únicos métodos capaces de evaluar la inteligibilidad del habla. Se sabía que el tipo y la complejidad de la señal de prueba (palabras articuladas sin sentido contra palabras conectadas en una sentencia significativa) tienen un impacto importante en los resultados. Sin embargo, la acústica de esos tiempos no podía diferenciar cuantitativamente entre las dos secuencias. Tampoco incluía elementos propios del canal de comunicación, como el patrón de reflexiones o la reverberación, ni contemplaba algunos efectos psicoacústicos relevantes como el enmascaramiento. Era necesario hallar otro método de evaluación que superase estas limitaciones.

### ***Índice de transmisión de la palabra***

Introducido por Steeneken y Houtgast, el *índice de transmisión de la palabra* (STI) [6] es básicamente una versión mejorada del índice de

la articulación que incluye las distorsiones temporales en el canal de comunicación (Steeneken y Houtgast, 1980). El STI se basa en las experiencias de detección de modulación descritas en el punto 3 del capítulo II. Las señales del habla se consideran como un flujo de energía con variaciones espectro-temporales y el grado de preservación de estas variaciones a través del canal de comunicación se toma como una medida de su fidelidad. Los autores argumentaron que la conservación de la envolvente dinámica implica la preservación de sus componentes de Fourier en el punto de recepción. El nivel continuo equivalente de la señal del habla es sustituido por una señal teórica de prueba, cuya intensidad es modulada por una función sinusoidal con un índice de la modulación  $m = 1$ . Cualquier degradación de la señal en el canal de comunicación aparecerá como una reducción en el índice de la modulación en el punto de recepción. Steeneken y Houtgast aconsejan el empleo de señales moduladoras de frecuencias comprendidas entre de 0,4 y 20 Hz para representar el grado de inteligibilidad a todas las tasas de ocurrencia del habla.

Si  $f$  representa la frecuencia de la señal moduladora y  $t$  es el tiempo en segundos, entonces la intensidad de la señal de prueba es modulada por la función  $1 + \cos(2\pi ft)$ . El índice de modulación en el punto de recepción es función de la frecuencia y se calcula con la siguiente expresión, que se deduce de la respuesta discreta al impulso entre los extremos del canal:

$$m(f) = \frac{\sum_n \frac{a_n e^{-j\pi f \frac{r_n}{c}}}{r_n^2}}{\sum_n \frac{a_n}{r_n^2}}$$

en la que  $c$  es la velocidad del sonido y  $r_n$  la trayectoria entre el emisor y el receptor. Si se desea incluir el efecto del ruido de fondo se la debe modificar de la siguiente manera:

$$m'(f) = m(f) \frac{\sum_n \frac{I_S a_n}{r_n^2}}{I_S \sum_n \frac{a_n}{r_n^2} + I_R}$$

aquí  $I_S$  es la intensidad promedio de la señal del habla en el punto de recepción e  $I_R$  la intensidad del ruido de fondo.

En esta función se manifiestan los efectos del patrón de reflexiones y de la reverberación. También es posible deducir de manera sencilla un cociente señal/ruido aparente:

$$(S/R)_{ap} = 10 \log \left( \frac{m(f)}{1 - m(f)} \right)$$

Si se recorta cada cociente individual a un valor limitado entre +15 y -15 dB, se calculan las relaciones S/R aparentes por bandas de octava y se promedian los valores de las 14 bandas involucradas, queda:

$$\overline{(S/R)_{ap}} = \frac{\sum_{oct.1}^{14} (S/R)_{ap}}{14}$$

El índice de transmisión de la palabra (sti) representa una razón linealizada de la forma:

$$STI = \frac{\overline{(S/R)_{ap}} + 15}{30}$$

Usando mediciones empíricas de las señales reales de prueba, es posible medir el STI de cualquier sala o sistema de comunicación. Si se reduce el número de bandas de octava se puede construir un sistema mucho más sencillo y rápido denominado RASTI,<sup>[7]</sup> que permite determinar la inteligibilidad del habla en tiempo real.

### ***Pérdida de articulación en consonantes***

Como la inteligibilidad de la palabra depende en gran medida de la audición correcta de las consonantes, a veces se emplea un índice denominado *pérdida de articulación en consonantes* (% Alcons) que se calcula únicamente en la banda de 2.000 Hz. Es un índice apropiado para predecir la inteligibilidad por software durante la etapa de diseño de una sala. Por ejemplo, a distancias mayores que el radio de reverberación, la pérdida de articulación en consonantes se calcula de la siguiente expresión:

$$\%Alcons = 9 TR \frac{\overline{S\alpha}}{1-\alpha}$$

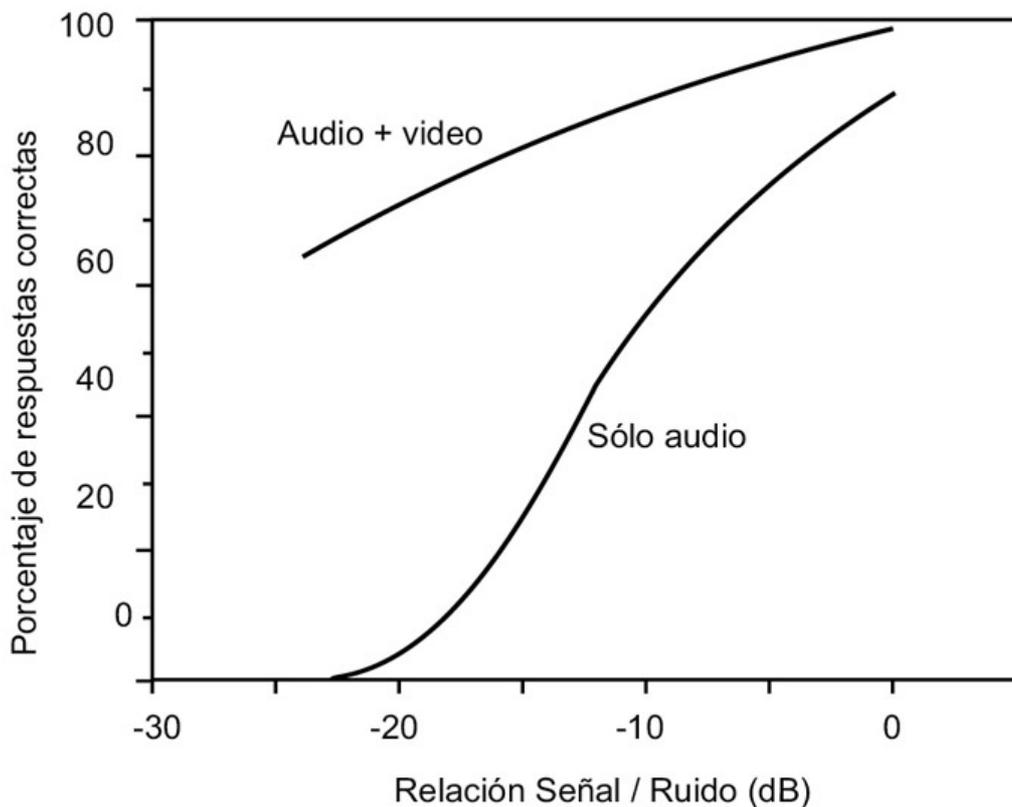
en la que TR es el tiempo de reverberación global en segundos, S es la superficie interior total de la sala y  $\alpha$  es el coeficiente de absorción medio de la sala.

### **Ayuda visual**

Se sabe que la lectura de los labios permite que las personas hipoacúsicas participen en una conversación. Las personas con oído normal también se benefician con la información visual y, si bien no les permite la comprensión completa del discurso hablado, la reconstrucción fonética en presencia de ruido se facilita al observar el rostro del locutor.

Se ha comprobado que la información lingüística se comprende mejor cuando se la transmite por televisión que por radio (Reisberg y McLean, 1987). MacLeod y Summerfield encontraron que el “be-

neficio de la lectura de labios” corresponde a un aumento promedio de 11 dB en la relación señal-ruido del sistema y Ostberg probó que el tamaño del monitor de video también contribuye en el mismo sentido (McLeod y Summerfield, 1987). Estas observaciones confirman la idea intuitiva de que la información visual contribuye a mejorar la inteligibilidad, sobre todo en condiciones adversas provocadas por la existencia de ruido en el canal de comunicación.



**Figura 16.** Aumento de la inteligibilidad en condiciones de ruido al agregar información visual

En la figura 16 se observa el resultado del experimento de Benoît, en el que se midió la inteligibilidad en función de la razón señal/ruido en dos situaciones diferentes: sólo audición y audición más visión (Benoît, C. et al., 1994).

La visión es innecesaria en condiciones acústicas óptimas ( $S/R > 0$  dB), pero resulta decisiva en condiciones acústicas altamente de-

gradadas ( $S/R = -24$  dB) en las que la información auditiva no alcanza para entender el discurso lingüístico.

## Notas

- [1](#) La cantidad de fonemas del español depende de cada región lingüística y varía levemente, por ejemplo, entre Madrid, Buenos Aires y La Habana. Como referencia comparativa, el inglés posee alrededor de 40 fonemas.
- [2](#) No pretendemos cubrir aquí el panorama completo de los fonemas del español. Tampoco empleamos los símbolos recomendados por la International Phonetic Association (IPA), que exigirían una descripción detallada. Quien desee ampliar en el tema deberá recurrir a textos específicos de fonética y fonología.
- [3](#) Aquí falla la difundida analogía entre el aparato vocal humano y los instrumentos de viento del grupo de los bronce: la campana de estos últimos devuelve hacia la boquilla un gran porcentaje de energía en cada ciclo. Así se establecen las ondas estacionarias necesarias para su funcionamiento.
- [4](#) El control activo de la membrana basilar se vio en el capítulo I, punto 4.
- [5](#) En idioma inglés se utiliza la lista definida en el “Harvard P.B.50 word score”.
- [6](#) Del inglés *Speech Transmission Index*.
- [7](#) Del inglés *Rapid Speech Transmission Index*.

# Apéndice I

## Teorema de Fourier

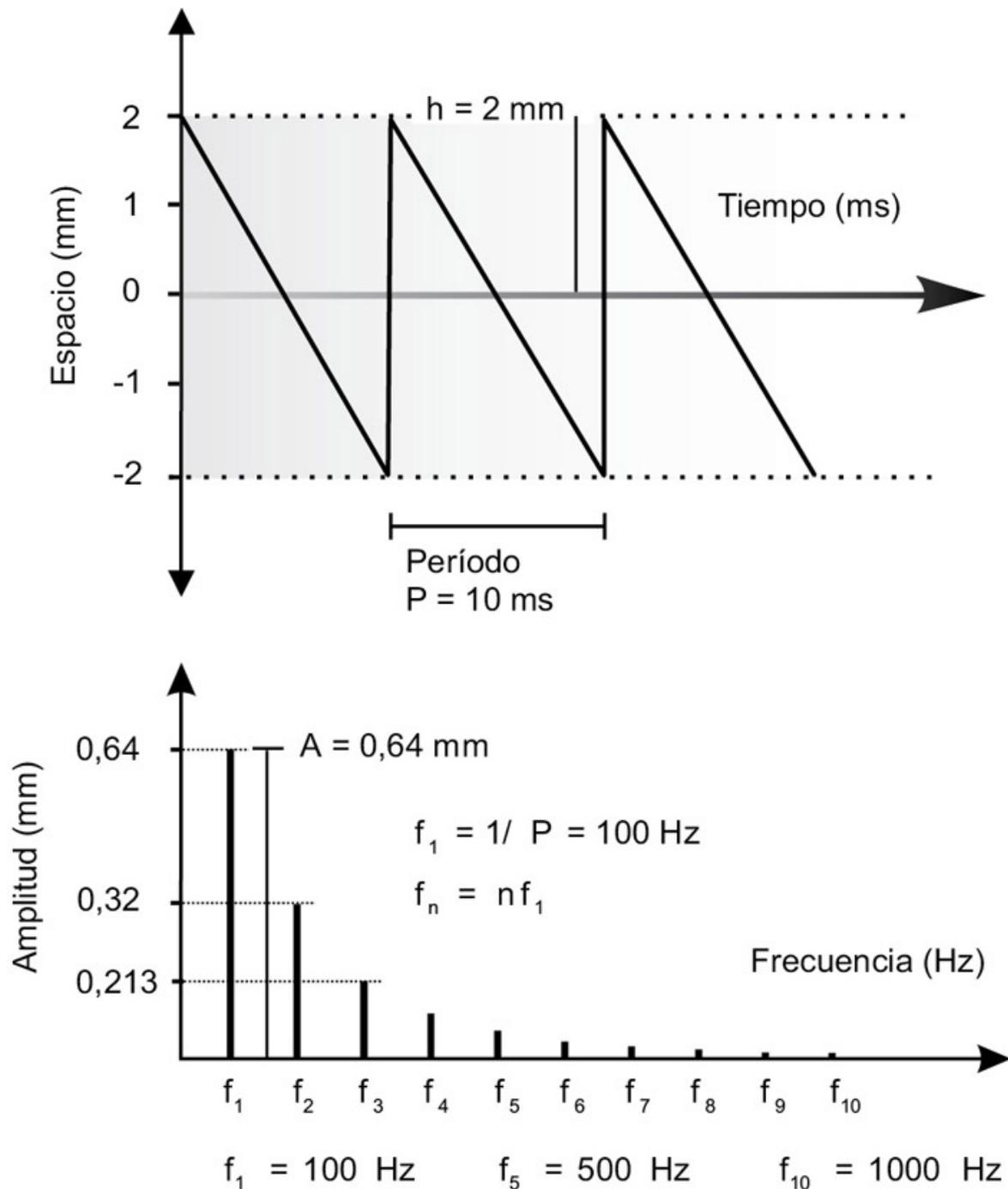
El teorema de Fourier se puede enunciar de la siguiente manera: “Toda función periódica de período  $P$  puede descomponerse en una suma de sinusoides armónicas, de amplitudes y fases adecuadas, cuyo primer armónico o fundamental posea período  $P$ ”.[\[1\]](#) En otros términos, una suma de sinusoides armónicas genera una onda periódica cuyo período coincide con el período de la senoide de menor frecuencia, llamado *primer armónico* o *fundamental* de la serie. El teorema permite descomponer y analizar cualquier función periódica y habilita la posibilidad de construir señales periódicas complejas a partir de una suma de sinusoides puras.

Recordemos que una sucesión es armónica cuando presenta una base y todos sus múltiplos. Es un caso particular de sucesión aritmética en la que la base  $a_0$  coincide con la *razón*  $k$  ( $a_n = a_0 + n k$  con  $k = a_0$ ). Por ejemplo, una sucesión armónica de base 5 es: 5, 10, 15, 20, 25, ... ,  $n \times 5$ ; y una de base 300: 300, 600, 900, 1200, ... ,  $n \times 300$ . La diferencia entre dos valores sucesivos, o razón de la sucesión armónica, es igual al valor de la base. Así en el último de los ejemplos  $1.200 - 900 = 300$ ;  $900 - 600 = 300$ , etcétera.

La disposición de las amplitudes y frecuencias de las sinusoides involucradas en la suma se denomina *espectro de Fourier*, y cada una de ellas toma el nombre de *componente de Fourier*.

Aunque la complejidad de las operaciones matemáticas involucradas en el cálculo de los armónicos es grande, en la actualidad existen numerosos programas de computación que los realizan con gran eficiencia. Aun sin realizar grandes esfuerzos matemáticos podemos extraer numerosas conclusiones cualitativas que son de utilidad. Por ejemplo, el teorema dice que una función periódica posee un espectro armónico. Si tenemos una señal periódica y cono-

ceamos su período, sabemos inmediatamente: 1) que el espectro correspondiente será armónico, y 2) cuáles serán las frecuencias de cada uno de los armónicos (la correspondiente al período de la señal original y sus múltiplos). Desconoceremos, hasta que se aplique completo el teorema de Fourier, las amplitudes de cada uno de esos armónicos.



**Figura 1.** Gráfico temporal y espectral de una onda diente de sierra

Como ejemplo aplicaremos el análisis de Fourier a la función *diente de sierra* dibujada en la figura 1. El espectro resultante nos muestra que posee todos los armónicos posibles desde la fundamental hasta el infinito. La amplitud de cada uno de ellos va decreciendo a medida que aumenta la frecuencia según la ley  $A_n = A_1 / n$ , siendo  $n$  el número del armónico,  $A_1$  la amplitud de la fundamental y  $A_n$  la amplitud del armónico número  $n$ . Por ejemplo, el quinto armónico tendrá una amplitud  $A_5 = A_1 / 5$ , cinco veces menor que la fundamental.

En teoría, se deben sumar todos los infinitos armónicos para sintetizar exactamente una onda diente de sierra. Es esta gran riqueza armónica uno de los motivos de su empleo en gran cantidad de aplicaciones prácticas. La voz humana y algunos instrumentos, como el violín o el oboe, generan señales que se le aproximan.

## Reconstrucción de parciales por batido

Dos sinusoides de distinta frecuencia baten a una tasa igual a la diferencia de frecuencias entre ambas. Como las componentes de Fourier son sinusoides, es de esperar que este fenómeno tenga lugar para todo par de componentes armónicos que se consideren. Si  $f_n$  y  $f_m$  son las frecuencias de dos armónicos cualquiera, la frecuencia de batido será  $f_b = f_m - f_n$ . La frecuencia de batido entre armónicos siempre va a coincidir con la frecuencia de algún armónico de la serie. En particular, dos componentes consecutivas baten a la frecuencia de la fundamental. Por ejemplo  $f_2 - f_1 = f_1$ ;  $f_3 - f_2 = f_1$ ; ... ,  $f_n - f_{n-1} = f_1$ .

La armonicidad de un espectro refuerza la periodicidad dada por su fundamental. Incluso, aunque falte la fundamental, la frecuencia que le corresponde se reconstruye a partir del batido de

sus armónicos. Este fenómeno, de gran importancia musical, se conoce como *fundamental perdida* o *fantasma* de la onda.

Un ejemplo de reconstrucción de la fundamental se da cuando se reproduce música a través de un altavoz pequeño, incapaz de radiar ondas de baja frecuencia. Como la mayoría de los instrumentos graves –contrabajos, fagotes o bajos eléctricos– producen señales espectralmente ricas en armónicos, se pueden oír los bajos de la música porque los primeros armónicos se reconstruyen por batido a pesar de no estar presentes a la salida del altavoz. El fenómeno de fundamental reconstruida también puede originarse durante el proceso de percepción. Ocurre, por ejemplo, cuando dos señales acceden, a través de auriculares, de manera independiente a cada uno de los oídos: en ese caso no existe batido físico, de manera que la fundamental se reconstruye en algún lugar de la cadena neural de procesamiento superior.

## Notas

- <sup>1</sup> En realidad existen ciertas restricciones matemáticas: las funciones deben ser seccionalmente continuas además de periódicas. Afortunadamente, las que intervienen en acústica musical cumplen con estos requisitos.

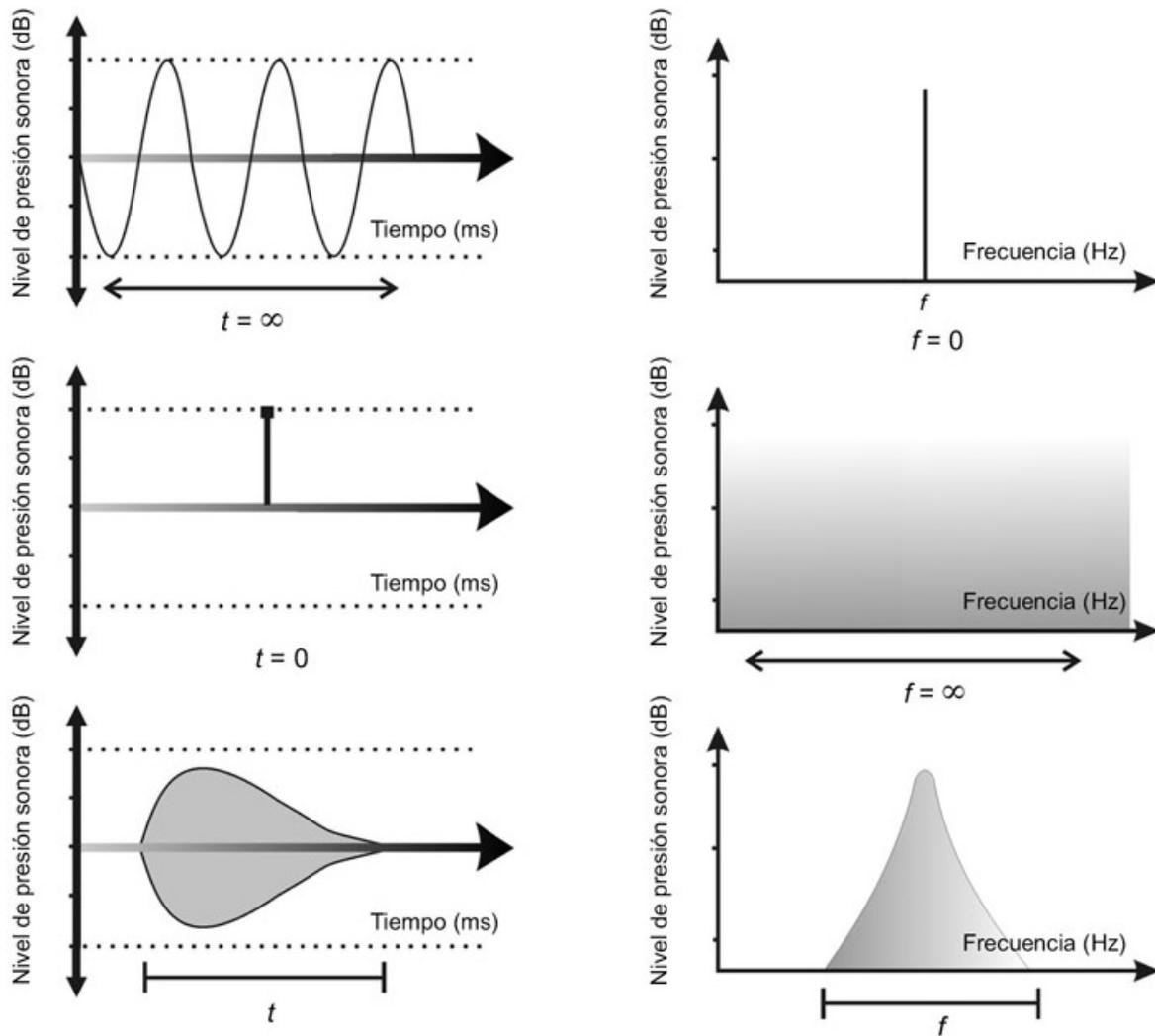
## Apéndice II

### Principio acústico de indeterminación

La serie de Fourier se puede aplicar únicamente a señales permanentes periódicas cuyos espectros son siempre armónicos y de líneas. No es posible analizar con ella el *estado transitorio* de una señal, que se puede definir como el intervalo temporal en el que una onda evoluciona de un estado permanente, o estacionario, a otro. Así como hay señales totalmente transitorias –la generada por el golpe sobre una mesa es un buen ejemplo–, existen otras que presentan características transitorias en el comienzo y en la extinción, permaneciendo relativamente estables dentro de esos límites.

Por oposición al estado permanente, en el que toda oscilación es una combinación de sinusoides de duración infinita, el estado transitorio posee una duración finita y necesariamente ocurre por lo menos al comenzar y finalizar una señal. Cualquier modificación de las condiciones de oscilación en medio del desarrollo de una onda, como el cambio en la amplitud o la frecuencia de alguna componente, forma parte del estado transitorio. El efecto de vibrato, que consiste en una variación periódica de la frecuencia y/o de la amplitud de aproximadamente 7 Hz, origina en muchos casos la supresión del estado permanente en la porción de señal a la que se aplica. En señales sin estado permanente el estado transitorio coincide con su duración.

J. B. Fourier desarrolló la *transformada o integral de Fourier* para extender el alcance de sus series a todo tipo de señales, sean permanentes o transitorias. Una de las derivaciones más importantes de la transformada de Fourier establece que las señales limitadas en el tiempo poseen espectros de banda. Se puede concluir, por simetría, que los espectros de líneas, armónicos o no, corresponden a señales no limitadas en el tiempo.



**Figura 1.** Aplicación del principio de indeterminación a diferentes señales

Todas las señales reales, obviamente limitadas en el tiempo, poseen anchos de banda no nulos. La onda que genera un diapasón posee una gran duración y un pequeño ancho de banda, mientras que la que produce un tambor se extingue rápidamente y ocupa un ancho de banda considerable. Podemos preguntar por la existencia de alguna relación general entre duración y ancho de banda. De acuerdo con el desarrollo de Fourier las representaciones temporales y espectrales son dos formas de interpretar el mismo fenómeno. El vínculo que las relaciona se deduce sin dificultad a partir de la transformada de Fourier y se denomina *principio acústico de*

*indeterminación*. Se lo puede enunciar de la siguiente manera: “el producto entre la duración del estado transitorio de una señal y el ancho de banda que le corresponde no puede ser inferior a un cierto valor mínimo”. Si  $\Delta f$  es el ancho de banda y  $\Delta t$  la duración del estado transitorio, el principio acústico de indeterminación establece, para el caso general:

$$\Delta f \times \Delta t \geq 1$$

Si se aplica la relación anterior a una senoide, como en ella la duración  $\Delta t$  tiende a infinito el ancho de banda correspondiente ( $\Delta f \geq 1/\Delta t$ ) será nulo y el espectro contendrá una sola línea (como era de esperar). Si se analiza un impulso de duración  $\Delta t \rightarrow 0$ , el ancho de banda  $\Delta f$  será 1 dividido (casi) cero que equivale a infinito. Este resultado confirma que el espectro de un impulso es un ruido blanco con un ancho de banda infinito. En el caso intermedio de una señal limitada en el tiempo –de duración finita– el principio se aplica sin inconvenientes. En la figura 1 se ilustra lo anterior con los diagramas temporales y espectrales de una señal sinusoidal, de un impulso y de una señal de duración intermedia.

El principio acústico de indeterminación impone un límite inferior, pero no uno superior: el producto de la duración por el ancho de banda puede alcanzar valores considerables. Por ejemplo, el ruido en una avenida muy transitada posee a la vez una duración ( $\Delta t$ ) y un ancho de banda ( $\Delta f$ ) muy grandes y no viola por ello la relación  $\Delta f \times \Delta t \geq 1$ .

Es importante destacar que el principio de indeterminación no se refiere a una limitación en la tecnología de generación de ondas o en el mecanismo de percepción humano, sino que describe el comportamiento físico propio de los fenómenos ondulatorios. No existe la posibilidad de construir una señal que a la vez esté bien determinada en la frecuencia (pequeño  $\Delta f$ ) y en el tiempo (pequeño  $\Delta t$ ), cualquiera sea la tecnología utilizada. En otras palabras,

no se puede construir un sonido tan corto como el de un tambor y a la vez tan tónico como el de una flauta.

## Referencias bibliográficas

- Ando, Yoichi (1998), *Architectural Acoustics. Blending Sound Sources, Sound Fields, and Listeners*, Nueva York, Springer-Verlag.
- *et al.* (1997a), “Acoustic Design of a Concert Hall Applying the Theory of Subjective Preference, and the Acoustic Measurement after Construction”, *Acustica & Acta Acustica*, 83.
- y Dennis Noson (1997b), *Music and Concert Hall Acoustics*, Londres, Academic Press Limited.
- (1985), *Concert Hall Acoustics*, Berlín, Springer-Verlag.
- Aronson, L. *et al.* (2005), “Características acústicas de la vocales del español rioplatense”, publicado en internet por la Fundación Arauz.
- Attneave, F. y R. Olson (1971), “Pitch as a Medium: A New Approach to Psychophysical Scaling”, *American Journal of Psychology*, vol. 84.
- Backus, J. (1969), *The Acoustical Foundations of Music*, Nueva York, W.W. Norton & Company Inc.
- Barton S. (1994), “Chaos, Self-Organization and Psychology”, *American Psychologist*, vol. 49.
- Basso, G. (2005), “Orden generativo, fractales y música”, en *Revista del Instituto Superior de Música UNL*, N<sup>o</sup> 10, Santa Fe.
- (2004), “Variables de diseño en campos acústicos no exponenciales”, en *Actas de las Terceras Jornadas de Acústica Electroacústica y Áreas Vinculadas (CADA E)*, Buenos Aires.
- (2002), “Fractales y estructuras musicales autosemejantes”, en *Actas de XV Conferencia Anual de la Asociación Argentina de Musicología*, Buenos Aires.

- (2001a), *Análisis espectral. La transformada de Fourier en la música*, La Plata, Ed. Al Margen - UNLP.
  - (2001b), “Fractales, ruido  $1/f$  y música”, en *Actas del Encuentro Nacional 2001 de Investigación en Artes y Diseño (ENIAD 2001)*, La Plata.
  - (2000), “Percepción auditiva. Nuevos enfoques”, Bernal, UNQ, mimeo.
  - (1999), “El principio de indeterminación en acústica musical”, en *Actas de la XIII Conferencia Anual de la Asociación Argentina de Musicología*, Buenos Aires.
  - (1998a), “Modelos estáticos y dinámicos en acústica musical”, en *Actas de la 2<sup>a</sup> Conferencia Iberoamericana de Investigación Musical*, Lanús.
  - , J. Guillén y M. Liut (2000), “Condiciones necesarias para el uso musical eficaz del desplazamiento acústico temporal”, *Actas del II Congreso Iberoamericano de Acústica*, Madrid.
  - *et al.* (1998a), “Límites de aplicación del desplazamiento acústico temporal”, en *Actas de la XII Jornadas Argentinas de Musicología*, Buenos Aires.
  - *et al.* (1998b), “El desplazamiento acústico temporal. Ajuste de la experiencia y uso musical”, en *Actas de las III Jornadas del Instituto Julio E. Payró*, Buenos Aires.
- Békesy, G. Von (1960), *Experiments in Hearing*, Nueva York, Wiley.
- Benade, A. H. (1976), *Fundamentals of Musical Acoustics*, Nueva York, Oxford University Press
- (1960), *Horns, Strings and Harmony*, Nueva York, Anchor Books.
- Benoît, C., T. Mohamadi y S. Kandel (1994), “Effects of phonetic context on audio-visual intelligibility of French speech in noise”,

*Journal of Speech & Hearing Research.*

- Beranek, L. (1996), *Concert and Opera Halls: How they Sound*, Nueva York, Acoustical Society of America.
- (1988), *Acoustical Measurements*, Woodbury, Acoustical Society of America.
- (1962), *Music, Acoustics, and Architecture*, Nueva York, Wiley.
- (1954), *Acoustics*, McGraw Hill [trad., (1961), *Acústica*, Buenos Aires, Editorial Hispano Americana].
- Bismark, G. von (1974), “Sharpness as an attribute of the timbre of steady sounds”, *Acustica*, vol. 30.
- Blauert, J. (1997), *Spatial Hearing. The psychophysics of human sound localization*, Cambridge, MIT Press.
- Bregman, A. (1994), *Auditory Scene Analysis*, Cambridge, MIT Press.
- Brownell, W. E. *et al.* (1985), “Evoked Mechanical Response of Isolated Cochlear Outer Hair Cells”, *Science*, vol. 227.
- Butler, R. A. (1969), “Monaural and binaural localization of noise bursts vertically in the median saggital plane”, *Journal of Auditory Research*, vol. 3.
- Butler, D. (1992), *The Musician’s Guide to Perception and Cognition*, Nueva York, Schirmer Books.
- Cadoz, C. *et al.* (1993), “CORDIS-ANIMA: a Modeling and Simulation System for Sound and Image Synthesis-The General Formalism”, *Computer Music Journal*, vol. 17.
- Carlson, R., B. Granström y S. Hunnicutt (1982), “A multi-language text-to-speech module”, *Proceedings of ICASSP*, vol. 82.
- Carlyon, R. y B. Moore (1984), “Intensity discrimination: a severe departure from Weber’s Law”, *Journal of the Acoustical Society of America*, vol. 76.

- Chalmers, D. (1996), “El problema de la conciencia”, *Investigación y Ciencia*, N<sup>o</sup> 233.
- Chion, M. (1998), “Du son à la chose : hypothèses sur l’objet sonore”, *L’analyse musicale*, vol. 11.
- Colburn, H. S. (1995), “Computational models of binaural processing”, *Auditory Computation*, Nueva York, Springer-Verlag.
- Cook, P. (1999), *Music, Cognition, and Computerized Sound*, Cambridge, MIT Press.
- Chowning, J. (1999), “Perceptual Fusion and Auditory Perspective”, *Music, Cognition, and Computerized Sound*, Cook (ed.), Cambridge, MIT Press.
- Cole, R. y B. Scott (1974), “Toward a theory of speech perception”, *Psychological Review*, vol. 81.
- Crowder, R. G. (1989), “Imagery for musical timbre”, *Journal of Experimental Psychology: Human Perception and Performance*, vol. 15.
- Culver, C. A. (1947), *Musical Acoustics*, Filadelfia, Blackiston.
- Delgute, B. (1987), “Peripheral auditory processing of speech information”, en *The Psychophysics of Speech Perception*, Países Bajos, Nijhoff.
- Dennett, D. C. (1993), *Coscienza*, Milán, Rizzoli.
- Deutsch, D. (1999), *The Psychology of Music*, Nueva York, Academic Press.
- (1995), *Musical Illusions and Paradoxes*, La Jolla, Philomel R. Inc.
- (1992), “Paradojas de la tonalidad musical”, *Investigación y Ciencia*, N<sup>o</sup> 193.
- (1975), “Two-channel listening to musical scales”, *Journal of the Acoustical Society of America*, vol. 57.

- Di Liscia, O. P. (2000), "Spatial listening and its computer simulation in electronic music", en <http://depts.washington.edu/smccweb/courses/401-3/402examps/spatial.html>.
- Divenyi, P. (1992), "Binaural suppression of nonechoes", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 91.
- y Zakarauskas, P. (1992), "The effect of bandwidth on auditory localization: An estimation theory model", *Proceedings of the 12th International Symposium on Hearing*, Londres, Pergamon Press.
- y Danner, W. (1977), "Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra", *Percept. Psychophys.*, vol. 21.
- Durlach, N. I. y L. D. Braida (1969), "Intensity perception. I. Preliminary theory of intensity resolution", *J. Acoust. Soc. Amer.*, vol. 46, pp. 372-383.
- Ehresman, D. (1978), "Perception of timbral analogies", *Rapports IRCAM N° 13/78*.
- Erickson, R. (1975), *Sound Structure in Music*, Berkeley, University of California Press.
- Fechner, G. (1860), "Elementos de psicofísica", publicado en internet.
- Feyman, R. (1971), *Física*, Panamá, Fondo Educativo Latinoamericano.
- Fine, P. A. y B. Moore (1993), "Frequency analysis and musical ability", *Music Perception*, vol. 11.
- Fletcher, H., W. A. Munson (1933), "Loudness, its definition, measurement and calculation", *Journal of the Acoustical Society of America*, vol. 5.

- Fletcher, H. (1995), "Speech and Hearing in Communication - ASA Edition", reimpresso, Woodbury, American Institute of Physics.
- Fraisse P. (1967), *Psychologie du temps*, París, PUF.
- Fraser, J. T. (1992), *Il Tempo, una Presenza Sconosciuta*, Milán, Feltrinelli.
- Freed, D. J. (1990), "Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events", *Journal of the Acoustical Society of America*, vol. 87.
- Freeman, W. (1991), "Fisiología de la percepción", *Investigación y Ciencia*, N<sup>o</sup> 175.
- Gardner, M. (1968), "Historical Background of the Hass and/or Precedence Effect", *Journal of the Acoustical Society of America*, vol. 43.
- Garner, W. R. y G. A. Miller (1947), "The masking threshold of pure tones as a function of duration", *J. Exp. Psychol.*, vol. 37.
- Glasberg, B. y B. Moore (1990), "Derivation of auditory filter shapes from notched noise data", *Hearing Research*, vol. 47.
- Goertzel, B. (1993), *The evolving mind*, Nueva York, Gordon and Breach Sci. Pub.
- Goldstein, J. L. (1973), "An optimum processor theory for the central formation of the pitch of complex tones", *Journal of the Acoustical Society of America*, vol. 54.
- Goode, R. L. *et al.* (1994), "New knowledge about the function of the human ear: development of an improved model", *Am. J. Otol.*, vol. 15.
- Grantham, D. (1995), "Spatial hearing and related phenomena", en *Hearing*, ed. B. Moore, Nueva York, Academic.
- Green, D. M. (1973), "Temporal acuity as a function of frequency", *Journal of the Acoustical Society of America*, vol. 54.

- Grey, J. (1975), *An Exploration of Musical Timbre*, CCRMA, Stanford University.
- Guillaume, P. (1975), *Psicología de la forma*, Buenos Aires, Psique.
- Haas, H. (1951), “Über den einfluss eines einfaches auf die harsamkeit von sprache”, *Acustica*, vol. 1 [trad. al inglés, (1972), “The influence of a single echo on the audibility of speech”, *Journal of the Audio Engineering Society*, vol. 20].
- Haftner, E. y T. Buell (1990), “Restarting the adapted binaural system”, *Journal of the Acoustical Society of America*, vol. 88.
- Hall, J. W. y R. W. Peters (1981), “Pitch from nonsimultaneous successive harmonics in quiet and noise”, *Journal of the Acoustical Society of America*, vol. 69.
- Hall, J. W. *et al.* (1984), “Detection in noise by spectro-temporal pattern analysis”, *Journal of the Acoustical Society of America*, vol. 76.
- y M. Fernández (1984), “The role of monaural frequency selectivity in binaural analysis”, *Journal of the Acoustical Society of America*, vol. 76.
- Hall, D. (1991), *Musical Acoustics*, California, Brooks/Cole Publishing Company.
- Handel, S. (1995), “Timbre perception and auditory object identification”, *Hearing*, ed. B. Moore, Nueva York, Academic Press.
- (1989), *Listening*, Cambridge, MIT Press.
- Helmholtz, H. von (1885), *Die Lehre von den Tonempfindungen* [trad. en inglés, (1954), *On the Sensation of Tone*, trad. A. J. Ellis, Nueva York, Dover Publications].
- Henning, G. B. (1970), “Comparison of the effects of signal duration on frequency and amplitude discrimination”, en

- Frequency analysis and periodicity detection in hearing*, ed. Plomp, Leiden, Sijthoff.
- Hirsch, I. J. (1959), “Auditory perception of temporal order”, *Journal of the Acoustical Society of America*, vol. 31.
- Houtgast, T. (1976), “Subharmonic pitches of a pure tone at a low S/N ratio”, *Journal of the Acoustical Society of America*, vol. 60.
- (1972), “Psychophysical evidence for lateral inhibition in hearing”, *Journal of the Acoustical Society of America*, vol. 51.
- y Steeneken, H.. (1984), “A multi-language evaluation of the RASTI-Method for estimating speech intelligibility in auditoria”, *Acustica*, vol. 54.
- Houtsma, A. y J. Goldstein (1972), “The central origin of the pitch of pure tones: evidence from musical interval recognition”, *Journal of the Acoustical Society of America*, vol. 51.
- , T. Rossing y W. Wagenaars (1987), *Auditory Demonstrations*, ASA (cd y libro).
- Iverson, P. y C. L. Krumhansl (1989), “Pitch and timbre interaction in isolated tones and in sequences”, *Journal of the Acoustical Society of America*, vol. 86.
- Iverson, P. y C. L. Krumhansl (1993), “Isolating the dynamic attributes of musical timbre”, *Journal of the Acoustical Society of America*, vol. 94.
- Javel, E. (1980), “Coding of AM tones in the chinchilla auditory nerve: implications for the pitch of complex tones”, *Journal of the Acoustical Society of America*, vol. 68.
- Johnson, D. L. y H. Gierke (1974), “Audibility of infrasound”, *Journal of the Acoustical Society of America*, vol. 56.
- Katz, D. (1967), *Psicología de la forma*, Madrid, Espasa-Calpe.
- Kemp, D. T. (1978), “Stimulated Acoustic Emissions from within the Auditory System”, *Journal of the Acoustical Society of*

- America*, vol. 64.
- Kendall R. A. y E. C. Carterette (1993), “Verbal attributes of simultaneous wind instrument timbres: II. Adjectives induced from Piston’s Orchestration”, *Music Perception*, vol. 10.
- Kendall, R. A., E. C. Carterette y J. M. Hajda (1999), “Natural vs. synthetic instrument tones”, *Music Perception*, vol. 16.
- Klatt, D. H. (1989), “Review of selected models of speech perception”, en *Lexical Representation and Process*, Cambridge, MIT Press.
- Klumpp, R. y H. Eady (1956), “Some measurements of interaural time difference thresholds”, *Journal of the Acoustical Society of America*, vol. 28.
- Köhler, W. (1972), *Psicología de la forma. Su tarea y últimas experiencias*, Madrid, Biblioteca Nueva.
- , K. Koffka y F. Sander (1962), *Psicología de la forma*, Buenos Aires, Paidós.
- Konishi, Masakazu (1993), “Audición binaural”, *Investigación y Ciencia*, N<sup>o</sup> 201.
- Krumhansl, C. (1990), *Cognitive Foundations of Musical Pitch*, Nueva York, Oxford University Press.
- y R. Shepard (1979), “Quantification of the Hierarchy of Tonal Functions Within a Diatonic Context”, *Journal of Experimental Psychology: Human Perception and Performance*, vol. 5.
- Kubovy, M. (1981), “Integral and separable dimensions and the theory of indispensable attributes”, en *Perceptual Organization*, Erlbaum, Hillsdale.
- Kuhl, P. K. (1993), “Innate predispositions and the effects of experience in speech perception: The native language magnet theory”, en *Developmental neurocognition: Speech and face*

- processing in the first year of life*, Países Bajos, Kluwer Academic Publishers.
- Lieberman, A. *et al.* (1967), “Perception of the speech code”, *Psychological Review*, vol. 74.
- Lieberman, A. y Mattingly, I. (1985), “The motor theory of speech perception revisited”, *Cognition*, vol. 21.
- Lieberman, A. M. (1995), *Speech perception*, Cambridge, MIT Press.
- Lo, Y. O. (1987), “Toward a Theory of Timbre”, Stanford, CCRMA, Stanford University.
- McAdams, S. y E. Bigand (1994), *Penser les sons. Psychologie cognitive de l’audition*, París, PUF.
- McAdams, Stephen (1995), *Localisation Auditive*, París, IRCAM.
- McClelland, J. L. y J. L. Elman (1986), “The TRACE Model of Speech Perception”, *Cognitive Psychology*, vol. 18.
- McFadden, D. (1989), “Spectral differences in the ability of temporal gaps to reset the mechanisms underlying overshoot”, *Journal of the Acoustical Society of America*, vol. 85.
- McGill, W. J. y J. P. Goldberg (1968), “A Study of the near-miss involving Weber’s Law and pure-tone intensity discrimination”, *Perception and Psychophysics*, vol. 4.
- McGurk, H. y J. MacDonald (1976), “Hearing lips and seeing voices: A new illusion”, *Nature*, vol. 264.
- McLeod A. y Q. Summerfield (1987), “Quantifying the contribution of vision to speech perception in noise”, *British Journal of Audiology*, vol. 21.
- Majerník, V. y J. Kaluzn (1979), “On the auditory uncertainly relations”, *Acustica*, vol. 43.
- Malham D. G. (1998), “Spatial hearing mechanisms and sound reproduction”, trad. al español O. P. Di Liscia, autorizada por el autor, en <<http://www.unq.edu.ar/cme/personales/odiliscia>>.

- Matthew, Robert (1975), "Acoustics in perspective", en  
*Auditorium Acoustics*, Londres, Applied Science Publishers.
- Mathews, M. y J. Pierce (1980), "Harmony and nonharmonic  
partials", *Journal of the Acoustical Society of America*, vol. 68.
- Miller, G. y W. Taylor (1948), "The perception of repeated burst  
of noise", *Journal of the Acoustical Society of America*, vol. 20,  
pp. 171-182.
- Miyara, F. (s/f), "Introducción a la psicoacústica", mimeo.
- Moore, B. (1997), *An introduction to the Psychology of Hearing*,  
Londres, Academic Press.
- (ed.) (1995), *Handbook of Perception and Cognition: Hearing*,  
Londres, Academic Press.
- *et al.* (1993), "Comodulation masking release in subjects with  
unilateral and bilateral cochlear hearing impairment", *Journal of  
the Acoustical Society of America*, vol. 93.
- y B. Glasberg (1983), "Suggested formulae for calculating  
auditory-filter bandwidths and excitation patterns", *Journal of  
the Acoustical Society of America*, vol. 74.
- Neisser, U. (1967), *Cognitive psychology*, Nueva York, Appleton-  
Century-Crofts.
- Olson, H. (1952), *Music, Physics and Engineering*, Nueva York,  
Dover Publications.
- Ostberg O., B. Lindstrom y P. O. Renhall (1988), "Contribution to  
speech intelligibility by different sizes of videophone displays",  
*Proc. of the Workshop on Videophone Terminal Design*, Turín.
- Patterson, J. y D. Green (1970), "Discrimination of Transient  
Signals Having Identical Energy Spectra", *Journal of the  
Acoustical Society of America*, vol. 48.
- Patterson, R. D. (1976), "Auditory filter shapes derived with noise  
stimuli", *Journal of the Acoustical Society of America*, vol. 59.

- Pickles, J. (1988), *An Introduction to the Physiology of Hearing*, San Diego, Academic Press.
- Pierce, J. R. (1991), "Periodicity and Pitch Perception", *Journal of the Acoustical Society of America*, vol. 90.
- (1985), *Los sonidos de la música*, Barcelona, Ed. Labor.
- Pisoni, D. B. (1973), "Auditory and phonetic memory codes in the discrimination of consonants and vowels", *Percept. Psychophys.*, vol. 13.
- Plenge, G. (1974), "On the difference between localization and lateralization", *The Journal of the Acoustical Society of America*, vol. 56.
- Plomp, R. (1976), *Aspects of Tone Sensation*, Londres, Academic Press.
- (1967), "Pitch of complex tones", *Journal of the Acoustical Society of America*, vol. 41.
- y Levelt, J. M. (1965), "Tonal Consonance and Critical Bandwidth", *Journal of the Acoustical Society of America*, 38.
- Pollack, I. (1968), "Auditory pulsed Doppler discrimination", *Journal of the Acoustical Society of America*, 44, pp. 592-598
- (1953), "The Information of Elementary Auditory Displays", *Journal of the Acoustical Society of America*, 25, pp. 765-769.
- Rand, T. C. (1974), "Dichotic release from masking for speech", *Journal of the Acoustical Society of America*, vol. 55.
- Randell, M. (1993), "Mental models and complex systems: the adaptive dynamics of cognition", en internet.
- Rasch, R. A. (1978), "The perception of simultaneous notes such as in polyphonic music", *Acustica*, vol. 40.
- Rasch, R. y R. Plomp, (1999), "The perception of musical tones", en *The psychology of music*, ed. D. Deutsch, San Diego, Academic Press.

- Rayleigh, L. (1945), *The Theory of Sound*, Londres, Dover.
- Repp, B. H. (1987), “The sound of two hands clapping”, *Journal of the Acoustical Society of America*, vol. 81.
- Reisberg D. y J. McLean (1987), “Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli”, en *Hearing by eye: The psychology of lip-reading*, New Jersey, Lawrence Erlbaum Associates.
- Risset, J. C. (1986), “Pitch and Rhythm Paradoxes: Comment on Auditory Paradoxes Based on Fractal Waveform”, *Journal of the Acoustical Society of America*, vol. 80.
- (1978a), “Musical Acoustics”, París, Rapports IRCAM N° 8.
- (1978b), “Paradoxes de Hauteur”, París, Rapports IRCAM N° 10.
- (1978c), “Hauteur et Timbre des Sons”, París, Rapports IRCAM N° 11.
- y M. V. Mathews (1969), “Analysis of Musical Instrument Tones”, *Physics Today*, vol. 22.
- Risset, J.C., y D. L. Wessel (1999), “Exploration of timbre by analysis and synthesis”, en *The Psychology of Music*, ed. D. Deutsch, San Diego, Academic Press.
- Ritsma, R. (1967), “Frequencies dominant in the perception of the pitch of complex sound”, *Journal of the Acoustical Society of America*, vol. 42.
- (1962), “Existence region of the tonal residue”, *Journal of the Acoustical Society of America*, vol. 34.
- Rock, I. y S. Palmer (1991), “El legado de la psicología de la forma”, *Investigación y Ciencia*, N° 173.
- Roederer, Juan (1997), *Física y psicofísica de la música*, Buenos Aires, Ricordi.
- Ruiz, Ricardo (1995), *Ideas para el estudio de la organización psíquica*, La Plata, UNLP.

- Schaeffer, P. (1966), *Traite des Objets Musicaux*, París, Edition du Seuil.
- (1952), *A la Recherche d'une Musique Concrete*, París, Edition du Seuil [trad. en castellano, (1959), *¿Qué es la música concreta?*, Nueva Visión].
- Schouten, J. F. *et al.* (1962), “Pitch of the residue”, *Journal of the Acoustical Society of America*, vol. 34.
- Sek, A. (1994), “Modulation detection thresholds and critical modulation frequency based in random amplitude and frequency changes”, *Journal of the Acoustical Society of Japan*, vol. 15.
- Sethares, W. A. (1993), “Local Consonance and the Relationship between Timbre and Scale”, *Journal of the Acoustical Society of America*, vol. 94.
- Sethares, W. A. (1998), *Tuning, Timbre, Spectrum, Scale*, Londres, Springer-Verlag.
- Shackleton, T. M., *et al.* (1992), “Across frequency integration in a model of lateralization”, *Journal of the Acoustical Society of America*, vol. 91.
- Shailer, M. y B. Moore (1987), “Gap detection and the auditory filter: phase effects using sinusoidal stimuli”, *Journal of the Acoustical Society of America*, vol. 81.
- Sheppard, R. N. (1964), “Circularity in Judgments of Relative Pitch”, *Journal of the Acoustical Society of America*, vol. 35.
- Slawson, W. (1968), “Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency”, *Journal of Acoustical Society of America*, vol. 43.
- Soderquist, D. R. (1970), “Frequency analysis and the critical band”, *Psychonomic Science*, vol. 21.
- Spiegel, M. F. (1981), “Thresholds for tones in maskers of various bandwidths and for signals of various bandwidths as a function

- of signal frequency”, *Journal of the Acoustical Society of America*, vol. 69.
- Steeneken, H. y T. Houtgast (1980), “A physical method for measuring speech-transmission quality”, *Journal of the Acoustical Society of America*, vol. 67.
- Stern, R. y C. Trahiotis (1995), “Models of binaural interaction”, en *Hearing*, ed. B. Moore, Nueva York, Academic Press.
- Stevens, K. y Blumstein, S. (1981), “The search for invariant acoustic correlates of phonetic features”, *Perspectives on the study of speech*, eds. P. Eimas y J. Miller, Hillsdale, Lawrence Erlbaum Associates.
- Stevens, S. S. (1975), *Psychophysics: Introduction to its Perceptual, Neural and Social Prospects*, Nueva York, Wiley & Sons.
- (1959), “On the validity of loudness scale”, *Journal of the Acoustical Society of America*, vol. 31.
- (1935), “The relation of pitch to intensity,” *Journal of the Acoustical Society of America*, vol. 6.
- *et al.* (1937), “A scale for the measurement of the psychological magnitude of pitch”, *Journal of the Acoustical Society of America*, vol. 8.
- Sundberg, J. (1991), *The Science of Musical Sound*, San Diego, Academic Press.
- (1987), *The Science of the Singing Voice*, DeKalb, Northern Illinois University Press.
- y Lindqvist, J. (1973), “Musical octaves and pitch,” *Journal of the Acoustical Society of America*, vol. 54.
- y Verrillo, V. (1980), “On the anatomy of the retard: A study of timing in music,” *Journal of the Acoustical Society of America*, vol. 68.

- Terhardt, E. (1974), "Pitch, consonance and harmony," *Journal of the Acoustical Society of America*, vol. 55.
- *et al.* (1982a), "Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions", *J. Acoust. Soc. Am.*, 71.
- y W. Dixon Ward (1982b), "Recognition of musical key: Exploratory study", *J. Acoust. Soc. Am.*, 72.
- Unemoto, T. (1989), "The psychological structure of music", *Proceedings of the First International Conference on Music Perception and Cognition*, Kyoto.
- Vanderveer, N. J. (1979), "Ecological Acoustics. Human perception of environmental sounds", Doctoral Thesis, University Microfilms No. 8004002.
- Van Dommelen W. (1990), "Acoustic parameters in human speaker recognition", *Lang Speech*, vol. 33.
- Van Noorden, L. (1975), "Temporal Coherence in the Perception of Tone Sequences", *Instituut voor Perceptie Onderzoek*, Eindhoven.
- Vicario, G. B. (1997), "Time in Psychology", *Sci. Am.*
- (1967), *Tempo Psicologico ed Eventi*, Florencia, Giunti-Barbèra.
- Viemeister, N. F. (1988), "Psychophysical aspects of auditory intensity coding", en *Auditory Functions*, ed. Edelman, Nueva York, Wiley.
- Wallach, H. *et al.* (1949), "The Precedence Effect in Sound Localization", *American Journal of Psychology*, vol. 57.
- Warren, R. M. (1970), "Elimination of biases in loudness judgements for tones", *Journal of the Acoustical Society of America*, vol. 48.
- Warren, W. H. y R. R. Verbrugge (1984), "Auditory Perception of Breaking and Bouncing Events: A Case Study in Ecological

- Acoustics”, *Journal of Experimental Psychology: Human Perception and Performance*, vol. 10.
- Wegel, R. y C. Lane (1924), “The auditory masking of one sound by another sound and its probable relation to the dynamics of the inner ear”, *Physical Review*, vol. 23.
- Wier, G. C., W. Jesteadt y D. M. Green (1977), “Intensity discrimination as a function of frequency and sensation level”, *Journal of the Acoustical Society of America*, 61, pp. 169-177.
- Wessel, D. (1978), “Low Dimensional Control of Musical Timbre”, París, Rapports IRCAM N° 12.
- Whittle, L. S. *et al.* (1972), “The audibility of low frequency sounds”, *J. Sound Vib.*, vol. 21.
- Zwicker, E. (1965), “Temporal effects in simultaneous masking by white-noise burst”, *Journal of the Acoustical Society of America*, vol. 37.
- , G. Flottorp y S. S. Stevens (1957), “Critical Bandwidth in loudness summation”, *Journal of the Acoustical Society of America*, vol. 29.
- y H. Fastl (1999), *Psychoacoustics. Facts and Models*, Berlín, Springer-Verlag.
- y H. Fastl (1990), *Psychoacoustics*, Berlín, Springer-Verlag.
- Zwislocki, J. J. y H. N. Jordan (1986), “On the relation of intensity jnds to loudness and neural noise”, *Journal of the Acoustical Society of America*, vol. 79.