



FACULTAD DE INGENIERÍA Y CIENCIAS AGROPECUARIAS

DESARROLLO DE UN SOFTWARE DE PROCESAMIENTO PARA OBTENER
SEÑALES DE AUDIO BINAURALES A PARTIR DE SEÑALES DE AUDIO
MONOFÓNICAS

Trabajo de Titulación presentado en conformidad con los requisitos
establecidos para optar por el título de
Ingeniero en Sonido y Acústica

Profesor guía

Ing. Marcelo Darío Lazzati Corellano

Autor

Juan Antonio León Buitrón

Año

2013

DECLARACIÓN DEL PROFESOR GUÍA

“Declaro haber dirigido este trabajo a través de reuniones periódicas con el estudiante, orientando sus conocimientos y competencias para un eficiente desarrollo del tema escogido y dando cumplimiento a todas las disposiciones vigentes que regulan los Trabajos de Titulación.”

Marcelo Darío Lazzati Corellano

Ingeniero en Ejecución de Sonido

CI: 171163573-8

DECLARACIÓN DE AUTORÍA DEL ESTUDIANTE

“Declaro que este trabajo es original, de mi autoría, que se han citado las fuentes correspondientes y que en su ejecución se respetaron las disposiciones legales que protegen los derechos de autor vigentes.”

Juan Antonio León Buitrón

CI: 010360527-5

AGRADECIMIENTOS

A mi director, el Ing. Marcelo Lazzati, así como a Hugo Jácome, Juan Pazmiño, Christoph Scheuermann y Rui Penha. A todos ellos por su tiempo y colaboración

DEDICATORIA

A mi vieja y mi viejo.

RESUMEN

Este trabajo se enfoca en la investigación de las teorías de la psicoacústica orientadas a la localización de fuentes sonoras en la percepción binaural del sonido y las técnicas de procesamiento digital de señales para su implementación en forma de un software que se encarga de transformar señales monofónicas de entrada en señales binaurales de salida.

El software es diseñado y desarrollado en lenguaje de programación MAX/MSP, de una manera intuitiva que permite al usuario el control de la posición de la fuente monofónica de entrada y su convolución con Funciones de Transferencia Relativas a la Cabeza (HRTF) de varios individuos, considerando los fenómenos psicoacústicos presentes en la audición humana en la etapa de procesamiento. Finalmente se llevan a cabo pruebas con el objetivo de evaluar el procesamiento que efectúa el software para una misma señal de entrada por medio de una comparación análisis espectral de sus salidas binaurales resultante para distintas posiciones.

ABSTRACT

This work focuses on researching psychoacoustics theories oriented for sound source localization in binaural perception of sound and the digital signal processing techniques required for the implementation in the form of a software that is responsible for transforming monophonic input signals into binaural output signals.

The software is designed and developed in MAX/MSP's programming language, in an intuitive way that allows the user to control the position of the monophonic input source and its convolution with the Head Related Transfer Functions (HRTF) of several individuals, considering the psychoacoustic phenomena present in human audition in the processing stage. Finally, tests are conducted with the objective of evaluating the software that performs processing for a single input signal by the processing being done by the software for a same input signal by the comparison of their spectral analysis binaural outputs for different positions

ÍNDICE

Introducción.....	1
Antecedentes	1
Alcance	2
Justificación.....	3
Objetivo General	3
Objetivos Específicos	3
Hipótesis	4
1. Marco Teórico	5
1.1. Conceptos Básicos.....	5
1.1.1. Producción y Propagación del Sonido	5
1.1.2. Características de la Onda Sonora	6
1.1.2.1. Amplitud.....	6
1.1.2.2. Envolvente Dinámica	6
1.1.2.3. Forma de Onda.....	7
1.1.2.4. Longitud de Onda (λ).....	8
1.1.2.5. Período (T).....	8
1.1.2.6. Frecuencia (f).....	9
1.1.2.7. Fase.....	9
1.1.3. Características del Medio de Propagación	11
1.1.3.1. Impedancia Acústica (Z)	11

1.1.3.2. Reflexión y Difracción	11
1.1.4. Percepción Sonora	14
1.1.4.1. Percepción de Amplitud.....	14
a) Nivel de Presión Sonora (NPS).....	14
b) Sonoridad	15
1.1.4.2. Percepción de la Frecuencia	17
a) Espectro Sonoro	17
b) Timbre.....	19
1.1.5. Análisis Espectral mediante Transformada de Fourier	19
1.2. Sistemas de Sonido.....	21
1.2.1. Sonido Binaural.....	21
1.2.2. Sonido Estéreo y <i>Surround</i>	24
1.3. Sistema Auditivo	25
1.3.1. Oído Externo.....	26
1.3.2. Oído Medio	27
1.3.3. Oído Interno.....	28
1.3.4. Deficiencias de la Audición	29
1.3.4.1. No Linealidad.....	29
1.3.4.2. Efecto Haas	30
1.4. Psicoacústica y Escucha Binaural	33
1.4.1. Localización en el Plano Horizontal (Acimut).....	36
1.4.1.1. Diferencias Interaurales.....	37

a) Diferencias de Tiempo Interaural (ITD).....	37
b) Diferencias de Nivel Interaural (ILD).....	38
1.4.1.2. Sombra Acústica de la Cabeza.....	38
1.4.1.3. Sombra Acústica del Pabellón Auditivo	39
1.4.2. Localización en el Plano Vertical (Elevación)	39
1.4.2.1. Pabellón Auditivo o Auricular	39
1.4.3. Percepción de Rango o Distancia.....	42
1.4.4. Otras Contribuciones	44
1.4.5. Fuentes Móviles.....	45
1.5. Función de Transferencia	46
1.5.1. Funciones de Transferencia Relativas a la Cabeza (HRTF).....	47
1.5.2. HRTF No Individualizadas	50
1.5.3. Medición, Grabación y Reproducción de HRTF.....	51
1.5.4. Densidad Espacial e Interpolación de HRTF	53
1.5.5. Procesamiento de HRTF	54
1.5.5.1. En el Dominio del Tiempo.....	54
1.5.5.2. En el Dominio de la Frecuencia	54
1.6. Síntesis Binaural o Sonido 3D	56
1.6.1. Procesamiento Digital	57
1.6.1.1. Señal Impulso	58
1.6.1.2. Filtrado de Señales.....	59
1.6.1.3. Transformada Discreta de Fourier (DFT).....	59

1.6.1.4. Convolución.....	60
1.6.1.5. Función Ventana.....	62
a) Ventana de <i>Hanning</i>	64
1.6.2. Simulación de Distancia y Externalización.....	66
1.6.3. Auralización	67
1.6.4. Moldeamiento Estructural de HRTF	67
1.6.4.1. Modelado de ITD	68
1.6.4.2. Modelado de ILD	68
1.6.4.3. Teoría Dúplex y Modelo Estructural de Cabeza Esférica	69
1.6.4.4. Cono de Confusión.....	70
1.6.4.5. Modelo de Cabeza Elipsoidal	71
1.6.4.6. Modelo Estructural Monoaural del Pabellón Auditivo.....	72
1.6.4.7. Modelo de Entorno Acústico (Room Model)	72
1.6.5. Movimientos de Cabeza (<i>Head-Tracking</i>).....	73
1.7. Entorno de Programación MAX/MSP.....	74
1.7.1. Configuración de Entradas y Salidas de Audio en MAX/MSP	76
1.7.2. Manejo de Audio En MAX/MSP	77
1.7.3. Manejo de Archivos en MAX/MSP.....	78
2. Desarrollo del Software	80
2.1. Prototipo del Software	80
2.1.1. Base de Datos de HRIR.....	81

2.2. Configuración de Objetos y Funcionamiento del Software.....	84
2.2.1. Sección de Selección de Entrada	84
2.2.2. Sección de Posición de la Fuente.....	86
2.2.3. Sección de Selección de Individuo	88
2.2.3.1. Subpatch Hrir.....	92
2.2.3.2. Denominación de Archivos HRIR en el Software.....	95
2.2.4. Sección de Convolución	96
2.2.5. Sección de Salida de Audio	98
2.2.6. Configuraciones Adicionales: Presentación del Software	100
2.2.7. Configuraciones y Requerimientos del Sistema	101
2.3. Análisis del Software	102
2.3.1. Diseño de Pruebas de Análisis del Software	102
2.3.2. Presentación y Análisis de Resultados.....	104
2.3.2.1. Señal de Entrada: Ruido Blanco Monoaural.....	106
2.3.2.2. Pruebas de Elevación.....	106
a) Elevación -30°, Acimut 0°.....	107
b) Elevación 0°, Acimut 0°.....	108
c) Elevación 30°, Acimut 0°	109
d) Elevación 60°, Acimut 0°.....	111
e) Elevación 90°, Acimut 0°.....	112
2.3.2.3. Resumen y Comparación de Posiciones de Elevación.....	113
2.3.2.4. Pruebas de Acimut	118

a) Acimut 0°, Elevación 0°	119
b) Acimut 45°, Elevación 0°	120
c) Acimut 90°, Elevación 0°	121
d) Acimut 135°, Elevación 0°	123
e) Acimut 180°, Elevación 0°	125
f) Acimut 225°, Elevación 0°	126
g) Acimut 270°, Elevación 0°	128
h) Acimut 315°, Elevación 0°	129
2.3.2.5. Resumen y Comparación de Posiciones de Acimut	130
2.3.3. Pruebas Adicionales	139
2.3.3.1. Resumen y Comparación de Posiciones de Elevación.....	139
2.3.3.2. Resumen y Comparación de Posiciones de Acimut	143
3. Análisis Costo-Beneficio	151
3.1. Análisis Costo-Beneficio	151
4. Proyecciones.....	153
4.1. Capacidad de Expansión y Potencial del Proyecto.....	153
5. Conclusiones y Recomendaciones	156
5.1. Conclusiones	156
5.2. Recomendaciones	162
Referencias	164

Anexos	168
--------------	-----

Introducción

Antecedentes

La historia del procesamiento de señales comienza con el desarrollo y los avances en las tecnologías de grabación, post-producción y transmisión de sonidos que alcanzan un gran éxito a mediados del siglo XX como consecuencia de la investigación altamente promovida y financiada por las guerras. De éste gran avance tecnológico, gran parte sería luego aprovechado para el creciente campo del audio aplicado a la producción musical.

De ahí que se ve la necesidad de explorar con el procesamiento de señales con el objetivo de brindar herramientas útiles enfocadas a la manipulación de las señales de audio en una creciente industria musical que fomenta la producción discográfica de calidad.

Los primeros procesadores y efectos de señal para aplicaciones musicales fueron realizados con tecnologías analógicas. Con el avance de la tecnología digital en el campo del Procesamiento Digital de Señales (DSP), los procesadores son llevados a éste dominio, permitiendo no solo el procesamiento y la creación efectos únicos, si no la emulación de fenómenos acústicos (reverberación por ejemplo) mediante el empleo de algoritmos que modelan los fenómenos físicos que dan lugar a dicho fenómeno.

La psicoacústica es el campo de la acústica que estudia la percepción del sonido, el cómo afecta psicológicamente a los individuos, y el impacto de la música y el sonido en los humanos. La audición humana presenta características únicas de percepción que ha sido parte de estudios científicos por décadas. Con éste campo en constante desarrollo, compañías y fabricantes han logrado el desarrollo de unidades de procesamiento y efectos basados en éstas características de la audición con resultados que les ha permitido convertirse en estándares en la industria de producción musical, cuyo origen se establece a mediados de los años 70.

Éstas herramientas si bien no han llegado a establecerse en el mercado nacional por facilidad de acceso (de distribución y por razones económicas), y

por lo que no han sido comercializadas ampliamente en nuestro país, razón por la cual la tecnología digital es la más establecida. Por otro lado, los algoritmos únicos que representan herramientas tanto novedosas como de gran utilidad en producción audiovisual, no llegan a ser difundidas por su potencial empleo en software de fabricantes que explica la limitación de bibliografía sobre técnicas binaurales.

Una solución viable para ésta realidad en el campo de producción audiovisual del Ecuador representa el desarrollo de herramientas a modo de software capaces de llevar a cabo un procesamiento de señales en base a parámetros psicoacústicos de la audición, como en éste caso, de las técnicas para obtener sonidos de forma binaural.

El avance de software y lenguajes de programación enfocados al audio y su facilidad de acceso permiten la implementación de fundamentos técnicos establecidos y aplicados al procesamiento digital de señales y promueve el desarrollo de éstas aplicaciones enfocadas al audio, que representa una necesidad y carencia en una industria casi inexistente en el Ecuador.

Alcance

La finalidad del proyecto es el desarrollo de un software que simule los fenómenos acústicos asociados a la propagación del sonido y la influencia de cabeza humana en la percepción de los mismos, capaz de la manipulación de señales de audio que han sido capturadas con técnicas de microfónica monofónica con el fin de obtener señales de audio digitales para su reproducción de forma binaural. Dicho software establece la relación entre parámetros de la psicoacústica asociados al timbre para el procesamiento de las señales de audio, como una herramienta de producción audiovisual.

Se presenta en forma de un prototipo del procesador, desarrollado en un lenguaje de programación orientado a objetos que resulte adecuado para el tratamiento de señales de audio. El prototipo cuenta con una interfaz gráfica para crear un software a nivel de usuario.

Justificación

El proyecto fomenta el desarrollo de nuevas herramientas de producción audiovisual carentes en el país, aprovechando la ventaja de las tecnologías de programación y DSP.

El procesamiento que establece las relaciones entre los fenómenos asociados a la audición humana y las señales de audio se basa en fundamentos teóricos de la psicoacústica y el audio.

Como resultado, se obtendrá una herramienta de producción de audio útil en varias de sus etapas, con la posibilidad de emplearla como herramienta educativa para estudiantes de ingeniería acústica y de sonido, producción musical, publicidad, psicología y músicos con resultados auditivos.

Objetivo General

Desarrollar un prototipo de software de procesamiento de señales de audio que simule una cabeza humana para obtener señales binaurales a partir de señales monofónicas de entrada, basado en los fenómenos que afectan al timbre de las propiedades psicoacústicas de la audición.

Objetivos Específicos

- Determinar las claves psicoacústicas relevantes en la recepción de sonidos de forma binaural.
- Determinar el tipo de análisis de los parámetros de una señal de audio monofónica con el fin de establecer las propiedades y de qué manera estos serán manipulados.
- Establecer el algoritmo de procesamiento en base a fundamentos del sonido para el procesamiento de señales de audio digital almacenadas en memoria con el fin de obtener señales de audio binaurales.
- Desarrollar un prototipo de software de procesamiento de señales que considere fenómenos psicoacústicos propios de la audición que influyen en el timbre del sonido percibido.
- Realizar una comparación de espectro de señales procesadas con el software prototipo para la evaluación del procesamiento realizado.

Hipótesis

En la actualidad existen técnicas para obtener señales estéreo binaurales empleando un arreglo de micrófonos especial que simule la influencia de la cabeza en la recepción sonora.

Este tipo de grabación requiere un montaje y ciertas condiciones de grabación aplicables únicamente a estudios de grabación además de una limitación en la adquisición de dichas herramientas por costos elevados y de empleo en ciertas condiciones de grabación de sonido por su delicadeza de sus componentes y tamaño de los modelos.

Gracias a las posibilidades que brindan las herramientas de procesamiento digital de señales y la disponibilidad de estudios científicos que definen la influencia de la cabeza humana en la percepción sonora, se busca obtener señales binaurales a partir de señales monofónicas mediante procesamiento de señales una vez grabadas.

La hipótesis plantea que el prototipo de software desarrollado será capaz de realizar procesamiento de señales monofónicas, estableciendo la relación de los parámetros que la definen y los de los fenómenos psicoacústicos propios de la audición humana, para obtener señales estéreo binaurales y será desarrollado en el lenguaje de programación MAX/MSP.

Al obtener una herramienta cuya validez queda demostrada mediante las pruebas de comparación de timbre por comparación de espectro obtenido por análisis FFT, se obtiene una solución al problema de limitación del empleo de técnicas de microfonía binaural tradicionales.

De ésta manera, se contribuye con una herramienta de ingeniería en sonido para producción de audio en grabaciones como conciertos o sonido directo, donde exista la limitación de grabación binaural y son ampliamente utilizadas técnicas de grabación monofónicas.

1. Marco Teórico

1.1. Conceptos Básicos

1.1.1. Producción y Propagación del Sonido

El sonido se produce por una vibración mecánica de partículas que se propagan por el medio. Dicho movimiento es periódico, lo que quiere decir que la oscilación se completa un número de veces por segundo.

Para que las partículas entren en movimiento, una fuente (a la que se denomina fuente sonora) deberá impartir energía a éstas, lo que sucede al entrar en movimiento la fuente, generando una zona de presión donde las partículas oscilan alrededor de su posición de equilibrio que es mayor a la presión de equilibrio (cuando el medio es el aire).

El movimiento oscilatorio produce compresión de las partículas (zonas de mayor presión) del medio en los sectores próximos a donde fueron generados (cerca de la fuente) y posteriormente, zonas de descompresión (zonas de menor presión). Esta perturbación es la que se propaga por el medio en forma de una onda sonora, (mientras tenga la suficiente energía) como contracción y dilatación del medio gracias a la masa y la elasticidad del mismo.

Los parámetros fundamentales que definen y permiten diferenciar entre distintas ondas sonoras son su amplitud, frecuencia y fase. Su propagación ocurre a una velocidad conocida como velocidad del sonido, que depende del medio de propagación, su temperatura y presión. Para ser clasificada como una onda sonora, esta deberá ser capaz de producir una sensación auditiva, por lo que su presión y frecuencia deberán estar en un rango audible.

La propagación eventualmente llega a su fin debido a las pérdidas energéticas que sufre la onda por las condiciones del medio (humedad y temperatura) y las características de la onda (frecuencia). Se produce mayor amortiguación (mayor pérdida energética por roce de partículas del medio) en ambientes más secos, de baja temperatura y para sonidos de alta frecuencia.

1.1.2. Características de la Onda Sonora

1.1.2.1. Amplitud

Al tratarse de un movimiento oscilatorio, se produce alrededor de su posición de equilibrio en el medio en el cual se desplaza. La distancia entre dicha posición y la que recorre sobre o debajo de la misma, describe la amplitud de la onda sonora.

La amplitud indica las variaciones de nivel que presenta la onda sonora. Sin embargo, existen varias formas de indicar dichas variaciones en distintos dominios (acústico, eléctrico), que dependen de varios factores, por lo que no existe una relación lineal entre estos indicadores y la amplitud.

La amplitud se puede medir ya sea por amplitud pico o por amplitud eficaz (*Root Mean Square* o RMS). El valor pico, indica el valor máximo que alcanza la amplitud, sea positivo o negativo (sobre o bajo el punto de equilibrio respectivamente).

El valor eficaz, entrega un promedio de la amplitud en el tiempo, que “aproxima de forma más cercana el nivel que es percibido por nuestros oídos y entrega una mejor medida del mundo real de la amplitud global de la señal” (Miles y Runstein, 2005, p.36). Este se encuentra calculando la raíz cuadrada de la amplitud de la onda y obteniendo un promedio matemático de la misma.

1.1.2.2. Envolvente Dinámica

La variación de la amplitud en el tiempo permite obtener una envolvente dinámica, que puede ser visualizada como una línea que une los picos de amplitud. Dicha envolvente es generalmente dividida en cuatro partes, y hace referencia a cuatro momentos, especificados como tiempos (en milisegundos) que abarcan la duración total del sonido.

- Ataque: tiempo entre el inicio (amplitud 0) y amplitud máxima.
- Decaimiento: tiempo entre la amplitud máxima hasta que decae a un valor que se mantiene constante.

- Sostenimiento: tiempo en que la amplitud se mantiene considerablemente constante.
- Relevo: tiempo en el que el sonido presenta una caída hasta una amplitud 0.

1.1.2.3. Forma de Onda

La variación de amplitud en el tiempo para un ciclo de la onda considera todos los puntos en su desplazamiento y presentan la forma de onda. Las formas más simples, a pesar de que no son producidas por fuentes sonoras en la naturaleza, son las ondas sinusoidales, cuadradas, triangulares, y pulsos, son implementados regularmente en pruebas de audio y acústica, así como varios tipos de señales no periódicas (ruido).

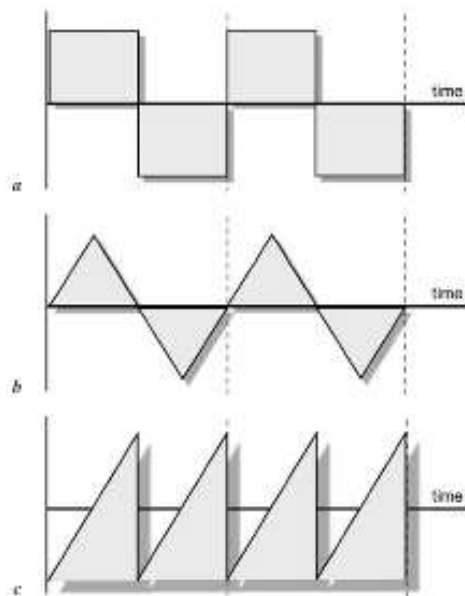


Figura 1. Ejemplos de formas de onda.

Tomado de Miles y Runstein, 2005, p. 47.

a. Ejemplo a es una onda cuadrada, ejemplo b una onda triangular y ejemplo c, una onda diente de sierra.

1.1.2.4. Longitud de Onda (λ)

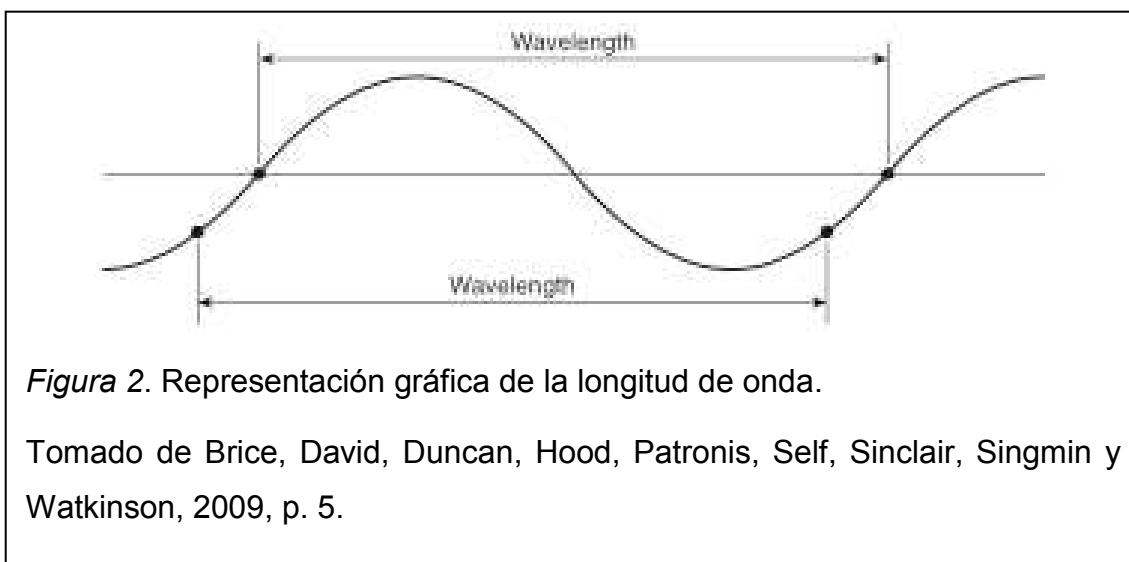
En el tiempo de un periodo, representa la distancia entre el mismo punto para dos ciclos consecutivos de onda, es decir que recorrería la onda en el espacio, por lo que depende del periodo y la velocidad del sonido en el medio (c):

$$\lambda = cT \text{ [m]} \quad \text{(Ecuación 1)}$$

Donde:

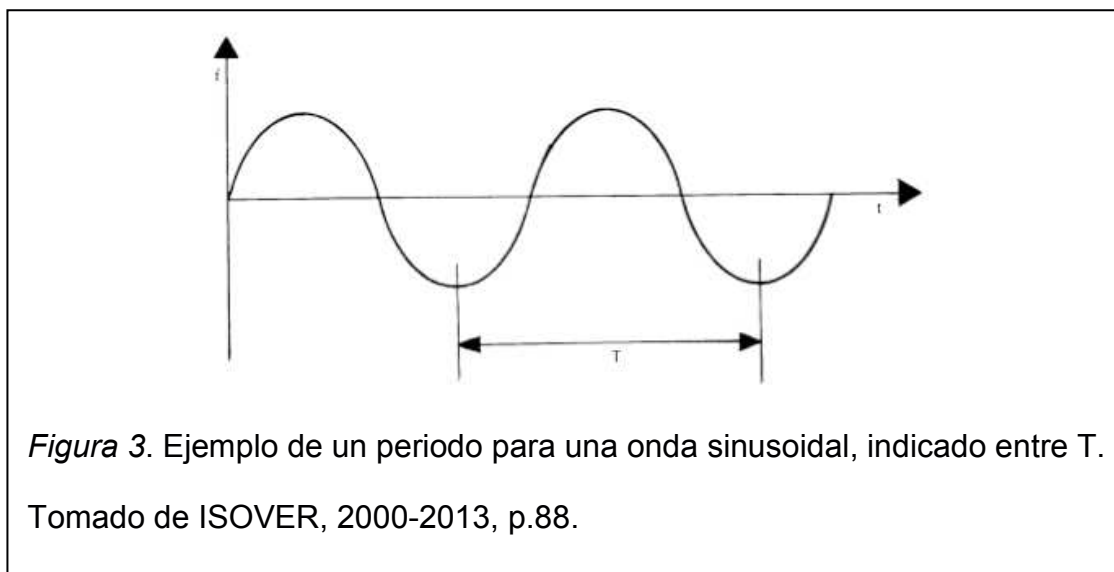
c: es la velocidad del sonido en el medio en el cual se propaga [m/s].

T: es el período de la onda sonora [s].



1.1.2.5. Período (T)

Es indicado en segundos y representa el tiempo que demora en completar una oscilación o ciclo.



1.1.2.6. Frecuencia (f)

Su unidad son los *Hertz* (Hz). Especifica la cantidad de oscilaciones que se completan en una unidad de tiempo (un segundo [s]), definida como:

$$f = \frac{1}{T} = \frac{1}{s} \text{ [Hz]} \quad (\text{Ecuación 2})$$

Donde:

T: es el período de la onda sonora [s].

La frecuencia de la onda está asociada a la percepción de cuan alto o bajo es un sonido con respecto a otro o de manera absoluta, lo que se denomina como percepción de “altura” o *pitch*. Sin embargo, el concepto de altura es subjetivo y el concepto de frecuencia es estrictamente físico.

1.1.2.7. Fase

Un ciclo u oscilación representa un movimiento de 360°. La fase hace referencia a un momento en el tiempo (o un punto cualquiera en la forma de la onda) en el ciclo de la misma.

Para un sonido complejo, formado de varias componentes de frecuencia, cuando dicho punto es el mismo para todas, se consideran que se encuentran “en fase” (0°) entre sí. Si estas componentes comienzan en distintos momentos

(distintos a 0°), es decir, si se presenta un retardo temporal entre componentes, se produce un “desfase”.

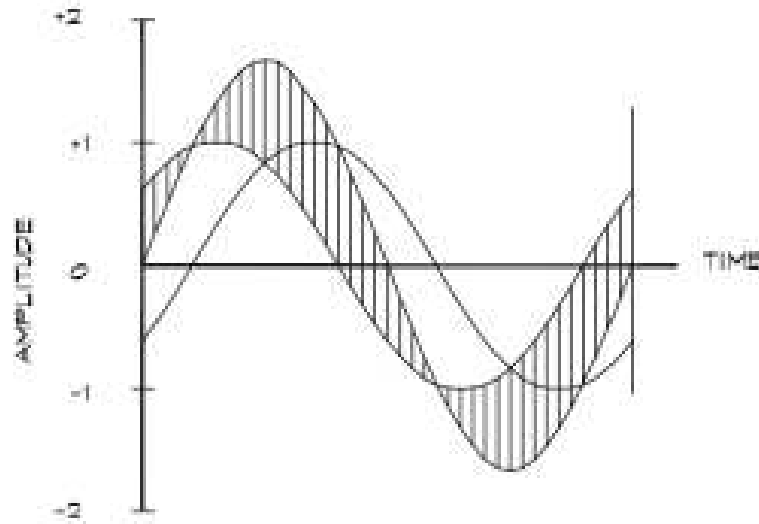


Figura 4. Suma de amplitudes de dos ondas sinusoidales con misma frecuencia y distinta fase.

Tomado de Long, 2006, p. 48.

a. La onda resultante de la suma es la de mayor amplitud.

Para ondas en fase, al poseer la misma amplitud, forma y frecuencia, las amplitudes se combinarán y sumarán en el resultado final. Pero en caso de ondas desfasadas o fuera de fase, la suma de las ondas en el resultado producirá picos y cancelaciones en el espectro que definen el timbre del sonido percibido. Las cancelaciones serán de amplitud cero cuando la diferencia de fase sea de 90° entre dos ondas idénticas.

1.1.3. Características del Medio de Propagación

1.1.3.1. Impedancia Acústica (Z)

Impedancia posee varias definiciones, como la “resistencia al movimiento en cierto punto” (Long, 2006, p.55) o la “facilidad para más o menos grande para la propagación del sonido” (ISOVER, 2000-2013, p.89). La impedancia acústica, se define matemáticamente como relación entre presión sonora (P) y velocidad de la partícula o movimiento vibratorio (u) en un punto.

$$Z = \frac{P}{u} \left[\frac{\text{N}\cdot\text{s}}{\text{m}^3} \right] \quad (\text{Ecuación 3})$$

Donde:

P: es la presión sonora [Pa].

u: es la velocidad de las partículas $\left[\frac{\text{m}}{\text{s}}\right]$.

El concepto está relacionado con el material o el medio y dicta una propiedad del mismo, como Long (2006, p.57) afirma que “El valor de la impedancia es frecuentemente utilizado para caracterizar el medio conductor” por lo que se puede decir que, por ejemplo el aire, posee una impedancia acústica baja.

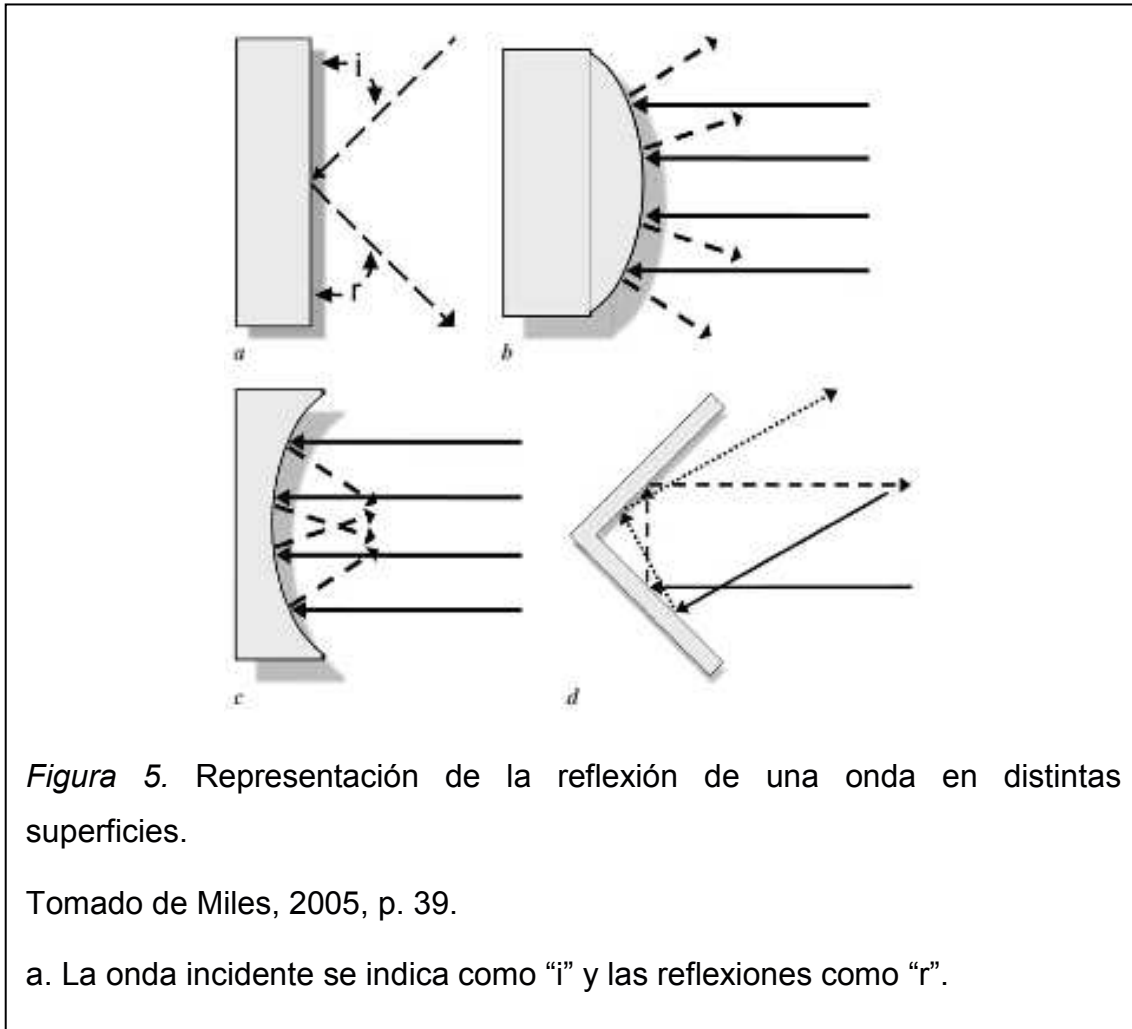
1.1.3.2. Reflexión y Difracción

En campo libre, el sonido presenta un decaimiento de intensidad (potencia sonora por unidad de área) al aumentar la distancia a la fuente, lo que se describe en la Ley del Inverso Cuadrado como “la intensidad del sonido en campo libre es inversamente proporcional al cuadrado de la distancia a la fuente” (Everest, 2001, p.10).

En un ambiente cerrado una onda sonora se propaga hasta alcanzar el límite físico del espacio, y dependiendo de las características de la superficie con la que se encuentra, parte de la misma es reflejada, absorbida, y transmitida.

Parte de la energía de la onda incidente es transformada en energía calórica (lo que se entiende como absorción), parte es transmitida a través de la superficie

o por la estructura de la misma, y parte es reflejada de vuelta, tantas veces mientras la onda todavía posea la energía para generar una nueva onda reflejada.

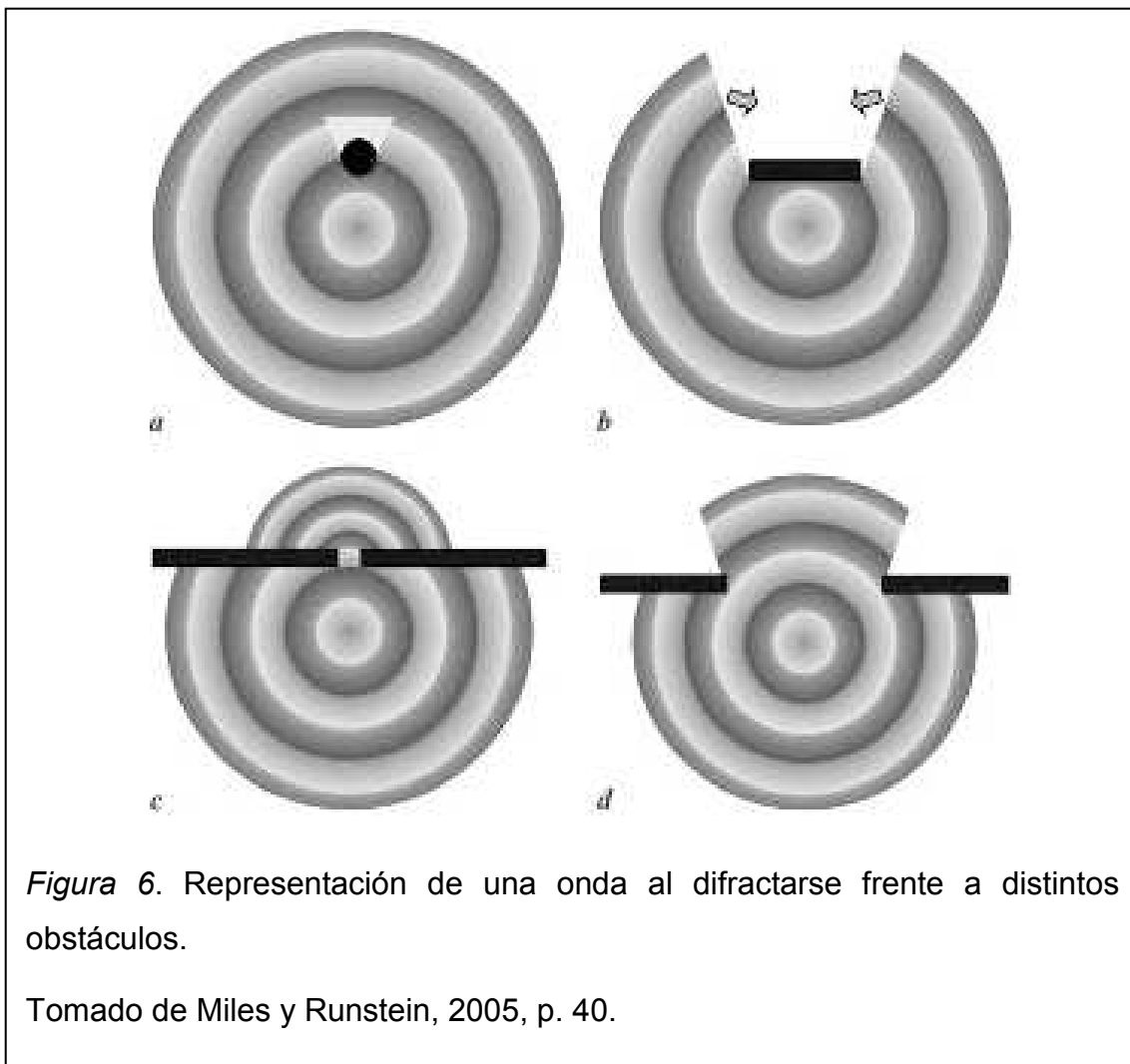


De esta manera, en un campo cerrado, se obtiene un campo directo representado por el sonido que llega a un punto de forma directa desde la fuente, y un campo reverberante formado por todas las ondas reflejadas en las superficies que limitan el recinto y en las que incide la onda original.

Debido a este fenómeno, se define al tiempo de reverberación (T_{60}), que indica el tiempo en segundos necesario para que la energía sonora generada por las reflexiones en el recinto decaiga a una millonésima (60 [dB]) de la energía producida por la fuente, es decir, el tiempo en que el sonido persiste una vez

que ésta ha dejado de sonar, debido a las reflexiones que refuerzan el sonido que producido.

Otra característica de las ondas sonoras es su capacidad de difractarse cuando se encuentran con un objeto cuya dimensión es menor a su longitud de onda, por lo que no representan un obstáculo en la propagación y por lo tanto no se reflejan en el mismo y son capaces de rodearlo y continuar su recorrido.



En el caso de que el objeto posea una dimensión comparable o mayor a la longitud de onda de la onda, se producirá una sombra acústica en el lado opuesto al lugar en el cual la onda “choca” con el objeto.

1.1.4. Percepción Sonora

1.1.4.1. Percepción de Amplitud

a) Nivel de Presión Sonora (NPS)

Las variaciones de presión ambiental son percibidas como sonido dentro de un rango audible. Esta presión se denomina Presión Sonora o Acústica. Su límite inferior se establece por la Presión de Referencia (20 [μ Pa]), y su límite superior, generalmente aceptado alrededor de los 20 [Pa].

Debido a que la presión varía en el tiempo, se emplea el valor eficaz para obtener un valor único. Ésta presión generalmente se expresa de forma cuantitativa como Nivel de Presión Sonora, expresada matemáticamente como:

$$\text{NPS} = 20 \text{ Log} \left(\frac{P}{P_0} \right) [\text{dB}] \quad (\text{Ecuación 4})$$

Donde:

P: es la presión ejercida por la onda sonora.

P₀: es la presión de referencia.

Debido a que la respuesta del oído no es lineal frente a los estímulos sonoros (percepción de sonoridad), y se aproxima más a una función logarítmica, se emplea dicha relación. El empleo del logaritmo también permite obtener la relación entre dos valores (presión eficaz y presión de referencia), además de reducir la escala de valores del rango auditivo.

La unidad del NPS son los decibeles [dB]. De ésta manera, el rango audible, que presenta una relación demasiado extensa para ser cuantificada empleando unidades de presión, se comprime y limita únicamente para valores positivos, entre los 0 [dB] (ausencia de sonido) y 120 [dB] (o hasta 140 [dB] dependiendo de la referencias bibliográficas consultada), al que se denomina “Umbral de Dolor”, donde la sensación auditiva se vuelve dolorosa y se presentan daños al sistema auditivo, probablemente de forma permanente.

El factor 20 que multiplica a la relación logarítmica hace referencia al hecho de que los “Niveles de Presión Sonora cambian con el cuadrado de la distancia” (Miles y Runstein, 2005, p.54). Cada vez que la distancia a la fuente es duplicada, el NPS disminuye 6 [dB].

b) Sonoridad

La magnitud de un campo sonoro puede ser cuantificada para su estudio y análisis, como en el caso de Nivel de Presión Sonora que representa un valor en base a una medida física. Pero la percepción de dicha magnitud, conocida como Sonoridad o *Loudness* depende tanto de la amplitud como de la frecuencia del sonido, y se trata de una magnitud subjetiva.

El trabajo de Fletcher y Munson publicado en 1933, considerado seminal en el campo, demuestra dicha dependencia en la percepción. Fue realizado para varios individuos a quienes se pidió comparar la sonoridad o que tan alto se percibe un tono puro de determinada frecuencia, con respecto a un tono de 1000 [Hz], en un rango entre 0 y 120 [dB] en pasos de 10 [dB].

De ésta manera, se obtuvieron las Curvas Isofónicas o Contornos de Igual Sonoridad, que indican que en cualquier punto en una curva, dos tonos se perciben con la misma sonoridad, y se establece que poseen un valor en la unidad de sonoridad Fones (*Phones*) que equivale a la sonoridad en los 1000 [Hz] para dicha Curva Isofónica.

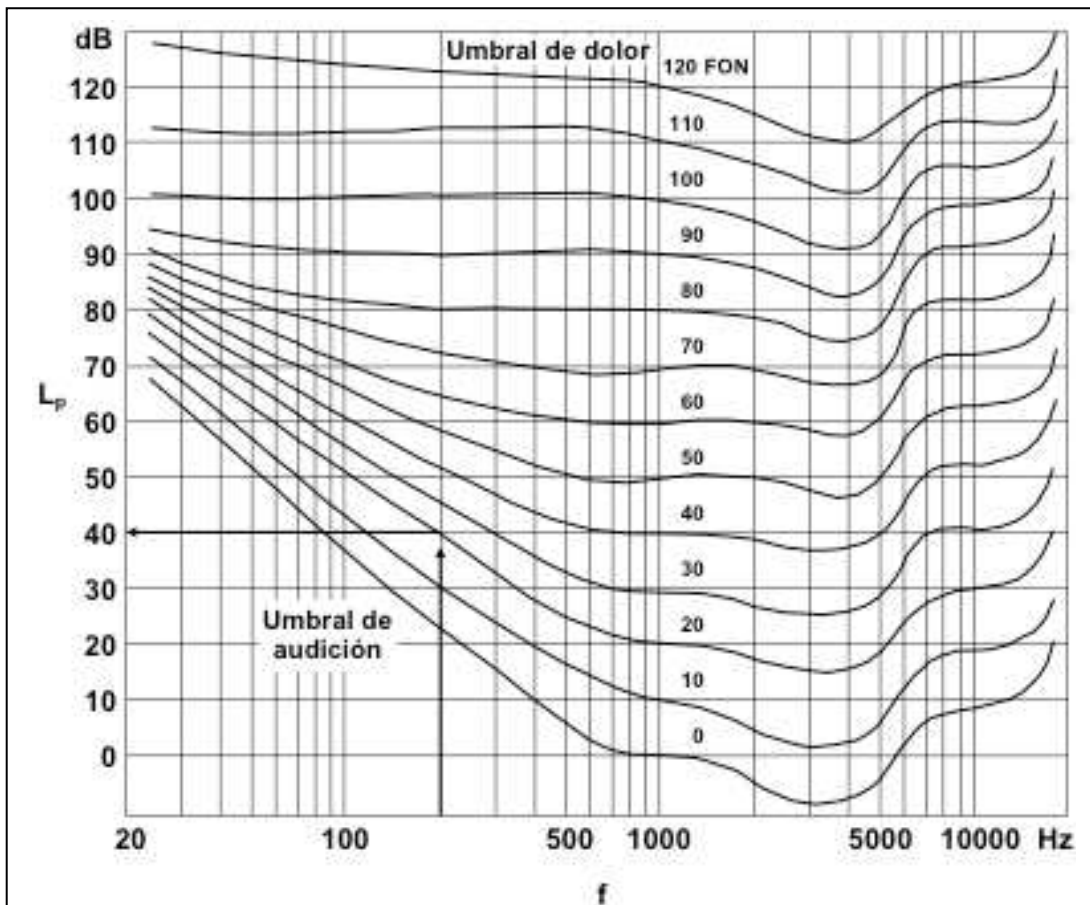


Figura 7. Curvas Isofónicas obtenidas por Fletcher y Munson.

Tomado de Miyara, 2000, p. 1-30.

a. La escala de fones se indica alrededor de los 1000 [Hz] como "Fon".

Estas curvas también presentan de forma clara ciertas características asociadas a la audición humana, demostrando que la percepción no es lineal para el rango de frecuencias audibles. Para niveles de presión bajos, existe una pronunciada deficiencia en percepción de frecuencias bajas, así como una capacidad pronunciada en la zona de los 4000 [Hz], donde el oído presenta una mayor sensibilidad y otra deficiencia en frecuencias altas.

Al aumentar el nivel y acercarse al umbral del dolor, la sonoridad a lo largo del espectro se vuelve más plana debido a que las deficiencias disminuyen hasta formar curvas más homogéneas.

El trabajo de Fletcher y Munson ha sido repetido por varios autores, y en 1956, Robinson y Dadson, empleando altavoces en una cámara anecoica en lugar de audífonos, fue adoptado por la Organización Internacional de Estandarización (ISO) como ISO:226.

Otra unidad conocida como Son (*Sone*) ha sido adoptada para evaluar y comparar la sonoridad entre distintos sonidos, ya que se puede considerar que la unidad de Fon es limitada en la información que entrega.

El son se define como la sonoridad experimentada por un individuo para un sonido percibido a 40 [fonos]. Por lo que 2 [sones], significarían que el doble de sonoridad es percibida. 0,5 [sones], equivaldría a la mitad de sonoridad percibida.

Este aumento y decremento de nivel, que es obtenido de forma experimental, en el caso de duplicarse la sonoridad, en el ejemplo de 2 [sones], es de 10 [dB], a pesar de que algunos teóricos afirman que es debería ser de 6 [dB].

Sin embargo, los sones (por las mismas razones que los fonos) pueden seguir siendo insuficientes en la información que entregan, además de que depende de éstos, por lo que existen varios métodos que pretenden cuantificar el fenómeno subjetivo de la sonoridad y han sido adoptados para diferentes propósitos de evaluación.

1.1.4.2. Percepción de la Frecuencia

a) Espectro Sonoro

El sonido más simple se compone de una sola frecuencia y se conoce como tono puro (forma de onda sinusoidal). Sin embargo, la mayoría de sonidos se consideran complejos, y están constituidos por varias frecuencias, como es el caso de la música, o el habla.

En los sonidos complejos, las frecuencias que los componen se denominan armónicos, que se encuentran en el rango audible generalmente aceptado entre 20 [Hz] y 20000 [Hz]. El primero de estos, se denomina frecuencia

fundamental o primer armónico, seguido por una cantidad de armónicos que pueden ser considerados consonantes cuando son múltiplos pares de la fundamental, o disonantes, al ser múltiplos impares de ésta. Otra denominación para las frecuencias que componen los sonidos complejos es “parciales”, que no indican la relación armónica con la fundamental.

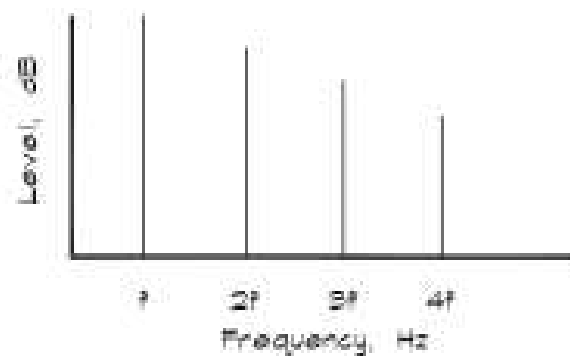


Figura 8. Ejemplo de Espectro de una nota generada por un piano.

Tomado de Long, 2006, p. 42.

a. La frecuencia fundamental se denota como f , y sus armónicos acompañados por un número que indica el múltiplo de la misma.

Dependiendo de la fuente y la forma en que ésta genera sonido, los armónicos, con sus distintas amplitudes y relaciones de fase entre sí, define el espectro sonoro del sonido dentro del rango audible y es único para cada sonido.

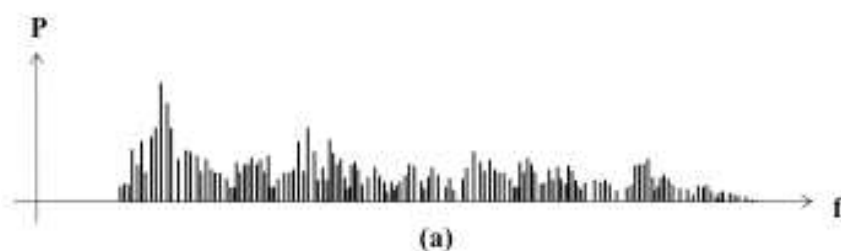


Figura 9. Ejemplo del espectro sonoro para un sonido complejo.

Tomado de Miyara, 2000, p. 1-12.

El espectro se representa la distribución energética en el rango de frecuencias audible, presentado como una gráfica de NPS vs. frecuencia, y es el resultado de llevar una forma de onda que se encuentra en el dominio del tiempo al dominio de la frecuencia, mediante un análisis de sus componentes de frecuencia conocido como Transformada de Fourier.

Esto gracias a que los sonidos periódicos complejos pueden ser reducidos a sus componentes como una suma de tonos puros que representen sus armónicos, sus amplitudes relativas y relaciones de fase.

Por lo general, para sonidos complejos, su espectro presenta un mayor NPS para su frecuencia fundamental, con un decaimiento de nivel para el resto de armónicos que lo componen.

b) Timbre

Sin embargo, la percepción de un sonido complejo difiere del concepto de espectro, siendo éste una medida o visualización estrictamente física del sonido. El oído percibe un sonido de forma subjetiva, que si bien depende de las componentes armónicas que componen el espectro, no tiene una relación directa con el mismo.

La percepción del espectro sonoro se denomina timbre y también resulta es único para cada sonido, pero bien puede diferir para un mismo sonido entre dos individuos, por lo que su una evaluación objetiva de un fenómeno perceptivo no posible de ser cuantificada.

1.1.5. Análisis Espectral mediante Transformada de Fourier

Una señal u onda periódica puede ser representada como una suma de sus componentes armónicos, con sus amplitudes y fases relativas, es decir, es posible llevar una señal en el dominio del tiempo al dominio de la frecuencia para el análisis de su contenido armónico para un momento determinado en el tiempo.

Demostrado por Joseph Fourier, por lo que lleva el nombre de Transformada de Fourier, el análisis se realiza para un pequeño fragmento de la señal especificado mediante un número de muestras (conocido como *frame*) que deberá ser exactamente igual a las muestras en un ciclo de la onda para el análisis correcto sin la aparición de componentes falsos debido a discontinuidades.

Para señales de audio digital, el análisis se realiza para un número finito de muestras, por lo que el proceso se denomina Transformada Discreta de Fourier (DFT) y es posible emplear el algoritmo más eficiente conocido como Transformada Rápida de Fourier (FFT) donde el tamaño del segmento o *frames* (también conocido como *FFT size*) deben ser igual a una potencia de 2.

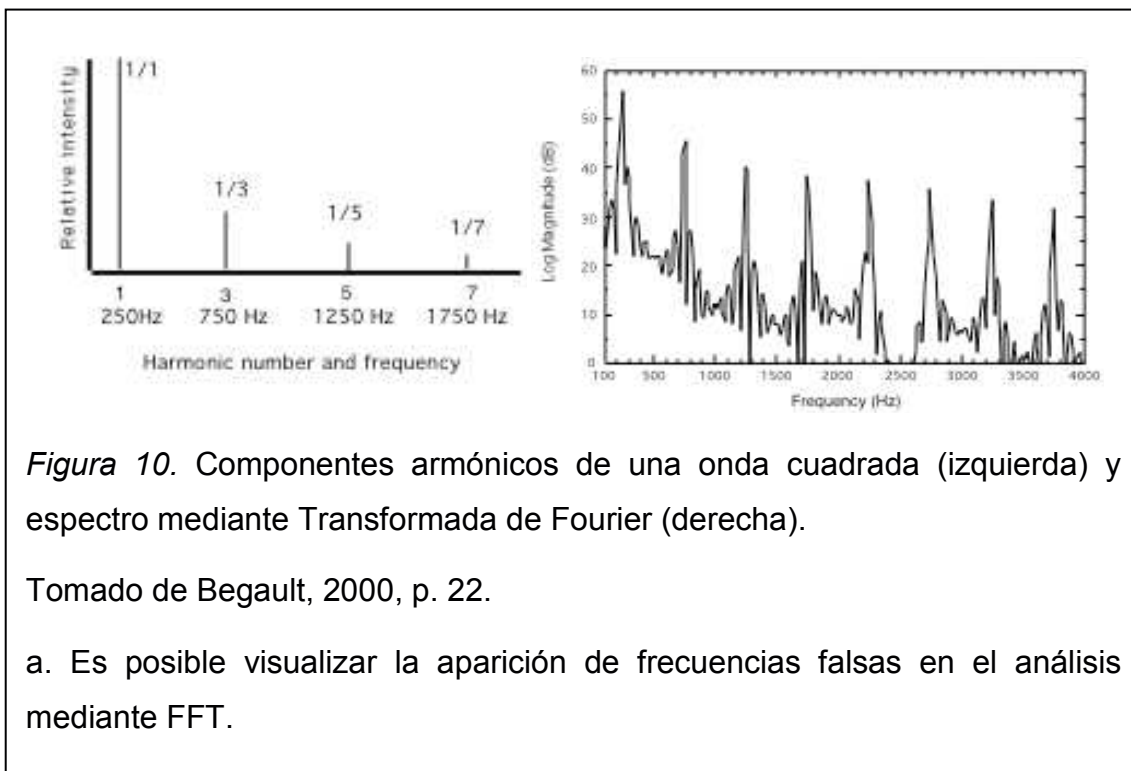


Figura 10. Componentes armónicos de una onda cuadrada (izquierda) y espectro mediante Transformada de Fourier (derecha).

Tomado de Begault, 2000, p. 22.

a. Es posible visualizar la aparición de frecuencias falsas en el análisis mediante FFT.

Los componentes falsos o inexistentes pueden evitarse en un proceso denominado *windowing* y *overlapping*. Se emplea una envolvente de amplitud que “suaviza” el final de cada *frame*, que a su vez, deberán sobreponerse sucesivamente y con un retardo en el tiempo igual a la mitad de muestras que poseen los *frames* analizados.

Para sonidos no periódicos como sonidos complejos, que no poseen una frecuencia conocida, se puede aplicar el análisis a *frames* consecutivos de la señal para obtener un análisis más representativo del espectro y su variación con respecto al tiempo.

Mediante FFT se obtienen las amplitudes y fases relativas (como un par de números reales e imaginarios) para cada banda de frecuencia en las que se divide al espectro sonoro, que son iguales al número de muestras por *frame* analizado.

En el dominio de la frecuencia, la señal puede ser empleada para procesamiento, manipulación o visualización del espectro en forma de un gráfico de frecuencia vs. magnitud, pero deberá ser transformado una vez más al dominio del tiempo para poder ser escuchada.

1.2. Sistemas de Sonido

1.2.1. Sonido Binaural

El primer experimento de sonido binaural se dio, de acuerdo a Thebinauralsource (s.f.), con la transmisión vía telefónica de un concierto de la Opera de Paris en 1881 por una idea de Clement Ader, quien utilizó dos transmisores de carbón que emitían a dos líneas telefónicas a los hogares de los subscriptores para poder disfrutarlas.

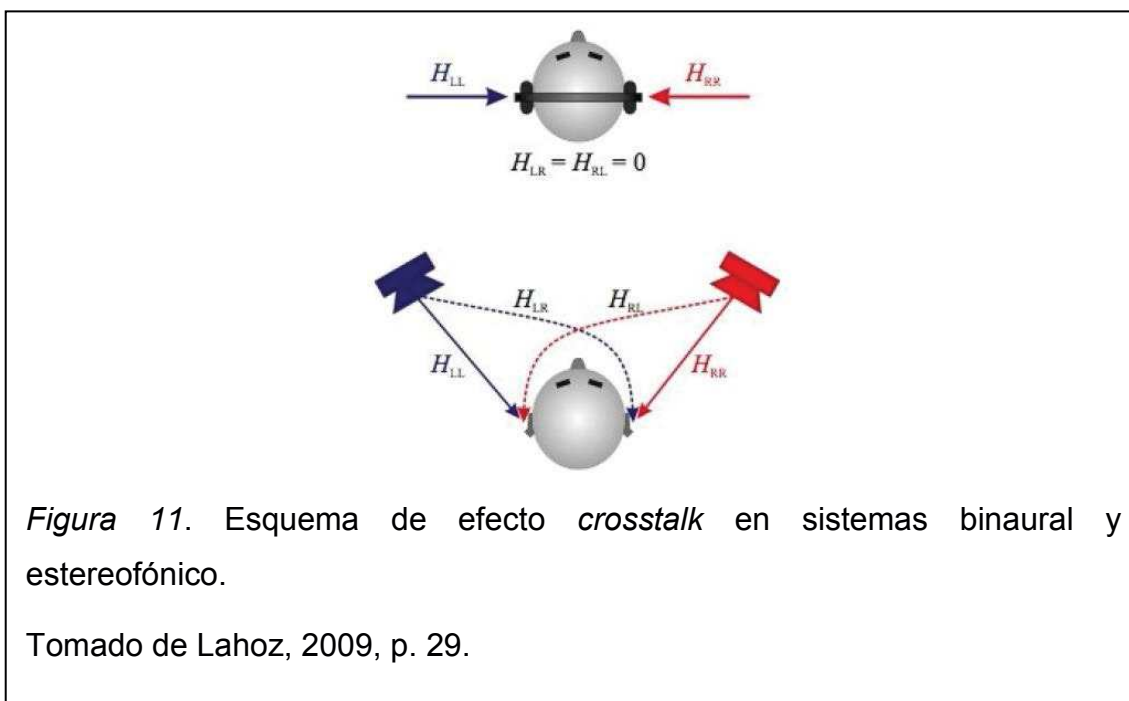
Sonido binaural hace referencia a una técnica de grabación y reproducción sonora. El objetivo de ésta técnica es involucrar al oyente exactamente donde se realizó la grabación, equivalente a la técnica visual POV (*point of view*), donde los eventos sonoros que suceden se localizan de forma tridimensional (360°) alrededor del oyente (es decir, más allá de una simple configuración estéreo).

En grabación se utiliza dos micrófonos omnidireccionales principalmente sensibilidad y respuesta en frecuencia (aunque puede llegar a necesitar de curva de ecualización compensación tanto en grabación como en *playback*) ubicados en la entrada del canal auditivo o en una cabeza artificial o *Dummy*

Head, donde no solo se consideran dimensiones promedio y materiales que simulen las características físicas de una cabeza humana (o inclusive estándares definidos) sino la contextura similar a la del cuerpo humano, incluyendo por lo general cuello, parte del torso (busto), e inclusive cabello en ciertos modelos, pues cada parte influye en la recepción de sonido dependiendo de la dirección de procedencia del sonido emitido de la fuente.

En la reproducción se utiliza audífonos estereofónicos, donde la señal captada por el canal izquierdo llega directamente al oído izquierdo del oyente, y lo mismo en el lado derecho sin ser mezclado (como el material estereofónico). Esto se debe a que ambos canales se consideran como discretos y deben mantenerse independientes entre sí antes de llegar a los oídos del receptor, para evitar perder el efecto pretendido al emplear una técnica de sonido binaural.

Sin embargo, al ser una señal que contiene información para dos canales, izquierdo y derecho, también es posible su reproducción en un sistema de altavoces estereofónicos, con la limitación de que, una vez que se encuentre en el dominio acústico, se producirá una mezcla de las señales de ambos canales, lo que se conoce como diafonía o *crosstalk*.



En 1953, Harvey Fletcher describe la diferencia entre sistema estereofónico y binaural:

“(…) éste último necesita únicamente de dos canales para una reproducción perfecta pero involucra el uso de un par de receptores en la cabeza sostenidos estrechamente a los oídos de cada oyente. Todos los oyentes con dicho sistema pueden tener la ilusión de estar en el mejor asiento de la sala de conciertos” (Thebinauralsource, s.f.).

Cabe recalcar que distintas fuentes bibliográficas concuerdan que el hecho de reproducir señales binaurales a través de altavoces son “casi idénticas a las obtenidas con técnicas de microfónica estéreo convencionales, con incrementada profundidad de campo” (Everest, 2007, p.183).

El uso de audífonos simplifica el sistema pero también introduce limitaciones. Al tener los transductores tan cercanos al tímpano, la fuentes suelen parecer más cercanas (por la proximidad física) o inclusive dentro de la cabeza, a lo que también se lo conoce como *In Head Location*, además de en varios modelos, suelen atenuar y aislar al oyente del entorno. La respuesta en frecuencia propia de los audífonos al no ser plana, modifican el espectro de la señal y los resultados de la señal binaural pueden comprometer los resultados.

Sin embargo, existen varios sistemas de procesamiento digital de señales disponibles comercialmente para “transformar” la información binaural para ser reproducidas en sistemas estereofónicos (como *Crosstalk Cancelled Stereo*) utilizando una matriz de decodificación de forma sencilla y sin comprometer la imagen (se preserva de mejor forma la información de fase entre L y R al ser canales discretos).

Existen también técnicas que permiten preservar el “efecto binaural” en su reproducción mediante altavoces estéreo como *Ambiophonics* (colocando un baffle entre los altavoces) o utilizando técnicas de *Crosstalk Cancelled* (matriz de codificación para evitar mezcla acústica de canales discretos), además de varios otros métodos incluidos en procesadores de señales digitales.

La importancia de la técnica radica en que la tendencia a consumir música en dispositivos móviles y portátiles, donde los transductores tradicionales son reemplazados por audífonos estereofónicos en gran parte. Considerando que señales binaurales son señales que poseen dos canales, su reproducción en sistemas estereofónicos es fácilmente aplicable.

1.2.2. Sonido Estéreo y Surround

El sonido estéreo ideado y patentado por primera vez por Alan Blumlein y patentado en 1931, consiste en un sistema de dos canales para su reproducción en dos altavoces. El sistema ganó una gran aceptación a nivel comercial y se impuso desde la década de 1950 en la industria musical. Sin embargo, las limitaciones del sistema se deben a la simplicidad del mismo.

Si bien es posible la ubicación de fuentes en un plano horizontal entre ambos altavoces y la generación de fuentes fantasma en cualquier punto intermedio entre estos, también resulta posible, debido a fenómenos de percepción psicoacústica de la audición, ubicar los elementos de una mezcla estéreo en el plano vertical según los componentes espectrales del sonido.

Sin embargo, la manipulación del espacio que ocupan las fuentes con respecto al oyente se ven limitadas por el campo que ocupan los altavoces y el tamaño de éstos, por lo que resulta casi imposible lograr una sensación de que una fuente se encuentra fuera del rango físico delimitado por el sistema.

Con la llegada de sistemas más complejos denominados “envolventes” o *surround*, es posible generar un sistema convincente pero limitado, pues se ve necesario incrementar el número de altavoces y diseñar una mejor distribución alrededor del oyente para mejorar la percepción tridimensional de material.

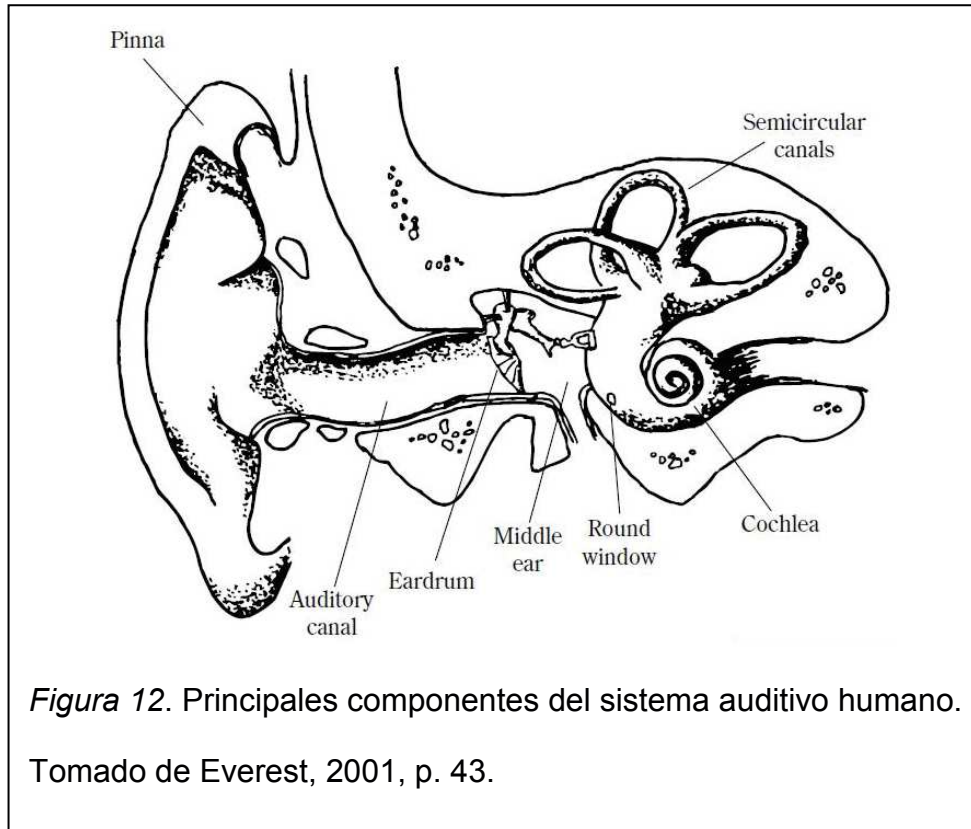
Éstos sistemas actualmente, y en cierta medida estándares en salas de cine e incluso disponibles para el mercado no profesional (como sistemas de “teatro en casa”), son cada vez más asequibles, pero siguen resultando inconvenientes y caros mientras mayor sea su complejidad y por lo tanto, la capacidad de “tridimensionalización” que sean capaces de alcanzar.

Por ejemplo, existen opciones de sistemas de sonido envolvente que varían desde 5.1 o 7.1 hasta 10.1 o 22.2, donde el primer número especifica la cantidad de altavoces de rango amplio, y el número después del punto, la cantidad de *subwoofers* necesarios en el sistema.

1.3. Sistema Auditivo

El sistema auditivo humano está dividido en tres partes y es el responsable del sentido de la audición. El oído se encarga de captar, amplificarlo o adaptar, y transformar la información sonora en impulsos que puedan ser interpretados por el cerebro, por lo que es considerado como un potente analizador de señales.

Es un complicado mecanismo que se que tiene la tarea de convertir energía acústica del medio en energía mecánica y finalmente en energía eléctrica, a modo de un transductor y que "a pesar de que el oído no es eficaz como instrumento de medida entregando valores absolutos, es muy perspicaz en la comparación de las frecuencias, niveles, o calidad sonora" (Everest, 2001, p.80).



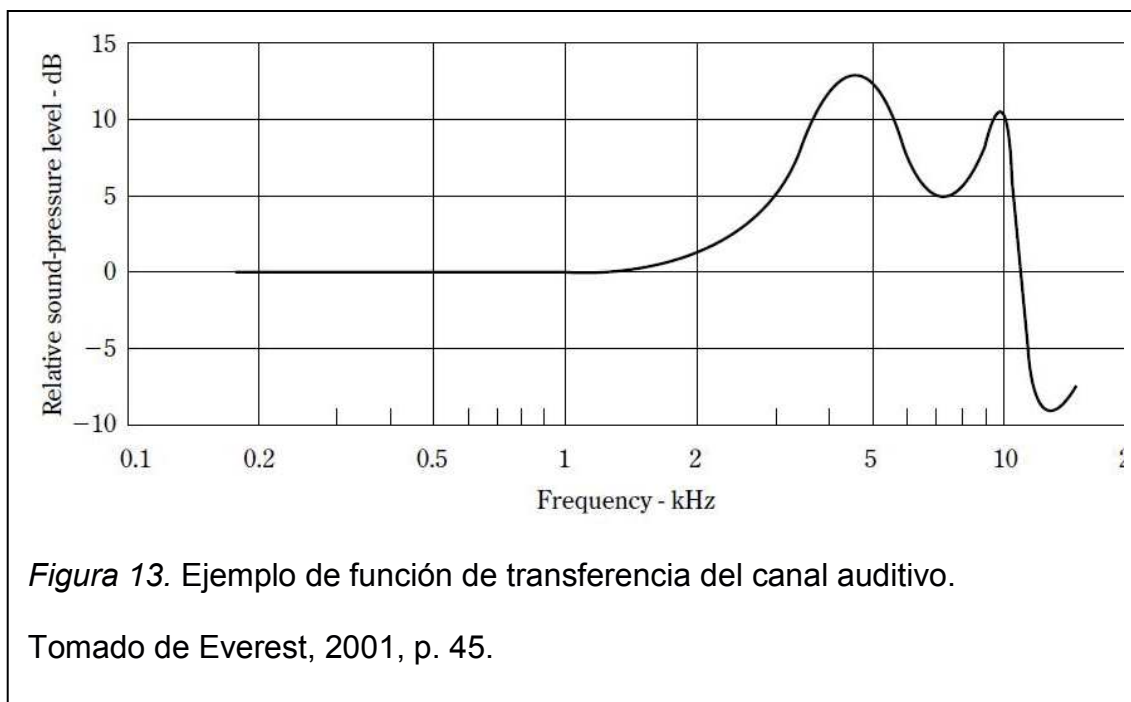
1.3.1. Oído Externo

Compuesto por el pabellón auditivo, el canal auditivo y el tímpano generando una cavidad acústica que modifica al sonido que llega al tímpano.

El pabellón, “recoge” el sonido del ambiente y lo dirige hacia el canal auditivo. Al incrementar su superficie efectiva, la respuesta de frecuencia varía significativamente.

El canal actúa como un tubo resonante cerrado en un lado, que produce una resonancia (de hasta +20 [dB] alrededor de 3000 [Hz] aproximadamente) propia para sonidos cuyo cuarto de longitud de onda (cuando ésta alcanza mayor velocidad) es igual al de la longitud del canal (aproximadamente 3 [cm] largo y 0,7 [cm] de diámetro). Un poco de ésta amplificación también se debe a la difracción por la cabeza. Otras resonancias se producen alrededor de los 4000 [Hz] (+12 [dB] aproximadamente) y una segunda resonancia de menor amplitud alrededor de los 9000 [Hz]. Esta es la primera fuente de no linealidad del sistema auditivo.

La función de transferencia propia del canal auditivo (Figura 16) es fija y se combina con la función de transferencia del sonido en la entrada del canal, que será único para cada dirección de procedencia, dando como resultado, una función compleja en el tímpano. Esto debido a que “El cerebro pasa por alto la función fija para interpretar las funciones de transferencia como percepción de dirección” (Everest, 2001, p.67).



El tímpano es una membrana que representa el fin del oído externo y su conexión con el interno. Al existir una masa de aire en la cavidad del canal auditivo, el tímpano representa una suspensión acústica que actúa contra su compliancia.

1.3.2. Oído Medio

Compuesto por tres huesecillos: el martillo, el yunque y el estribo. El martillo está apoyado en el tímpano, así como el estribo en la ventana oval, y todos los huesecillos entre sí, generando un vínculo mecánico para la transmisión de las ondas que llegan al tímpano como variaciones de presión hacia el oído interno, a modo de palanca.

Estos huesecillos cuentan con músculos que protegen de altos niveles de presión, como mecanismo de defensa, que actúa a su vez como limitador, lo que representa un elemento no lineal en el sistema auditivo.

La función del oído medio es la de transformar energía entre dos medios (de acústica en el oído externo a mecánica en el oído interno), por lo que la bibliografía sugiere que su efecto, más allá de ser un amplificador, es el de un adaptador de impedancia necesario para su conversión eficiente entre medios donde su densidad incrementa (siendo mayor en un medio líquido, como en el oído interno).

Según Everest (2001, p.46), la acción de tipo palanca del oído medio produce una relación de impedancias entre 1.3:1 a 3.1:1 (que representa la reducción del movimiento del tímpano en la ventana oval) a lo que se suma la relación de áreas entre el tímpano (aprox. 80 [mm²]) y ventana oval (3 [mm²]), reduciendo la fuerza producida en el tímpano en un radio de 80/3, formando un mecanismo de adaptación de impedancia complejo y completo.

Como parte del oído medio, sin poseer una función en la audición, se encuentra el Tubo de Eustaquio, que regula la presión de aire en el oído medio (al estar atrapado entre dos membranas. Al tragar, el tubo se abre y permite el paso del aire y es un mecanismo de protección y drenaje de infecciones del oído medio.

1.3.3. Oído Interno

El estribo se apoya sobre la ventana oval (se considera parte de ésta), una membrana que determina la entrada al oído medio y a la cóclea. La cóclea es un órgano en forma de caracol, llena de fluido y rodeada de hueso en su exterior, que se encarga del análisis del sonido y la generación de impulsos nerviosos para el cerebro, además de encargarse del balance del cuerpo (mediante los canales semicirculares). Su operación no ha sido totalmente comprendida.

El estímulo que recibe esta la ventana oval pone a vibrar el fluido que se encuentra dentro de la cóclea, produciendo ondas estacionarias en la membrana basilar, que presentan picos de amplitud en distintas partes de ésta, dependiendo de la frecuencia del sonido que recibe.

Para bajas frecuencias, el pico se producirá al final de la membrana basilar, y para altas frecuencias, se presentará al inicio, cerca de la ventana oval. Como la mayoría de sonidos que se reciben poseen varios componentes espectrales, se producirá un cambio constante de la posición en la membrana basilar donde se produzca el pico de amplitud.

Dentro de la cóclea, de acuerdo a Everest (2001, p.49) se encuentran distribuidas alrededor de 15000 células ciliadas externas juntas a 140 estereocilios y 3500 células ciliadas internas ligadas a 40 estereocilios más. Los picos de amplitud de las ondas en el fluido “doblan” los cilios que son terminales nerviosos, lo que causa que los estereocilios encargados de la transducción eléctrica, produzcan una descarga eléctrica (neural) a la corteza auditiva. Estas descargas son llevadas por el nervio auditivo hacia el cerebro.

Es en la ventana basilar, al ser éste el segundo transductor en el sistema, una segunda fuente de “no linealidad compresiva” de la audición. No actúa como un compresor o limitador como en los huesecillos pero en comprimiendo todo el rango audible, debido a su acción como transductor de energía entre medios.

1.3.4. Deficiencias de la Audición

1.3.4.1. No Linealidad

Como todo sistema de sonido existente, la audición no está libre de distorsiones, es decir, no siempre se escucha o percibe exactamente lo que se recibe debido a la naturaleza y composición del sistema, considerado como no lineal, principalmente en el oído medio e interno. La señal de entrada (onda acústica) se ve alterada hasta su salida (impulso nervioso e interpretación del mismo como sonido).

Ejemplos de no linealidad son los batimentos, como una variación de amplitud entre señales cuyas frecuencias son cercanas. Otro es la suma y diferencia de componentes espectrales o tonos puros, generando componentes resultado de la combinación de estas interacciones (o intermodulación), también conocidos como armónicos aurales.

Estas distorsiones, imposibles de medir al ser fenómenos psicoacústicos, comprueban la capacidad de análisis de fase del sistema auditivo en forma de impulsos nerviosos.

Estas distorsiones pueden crear todavía más alteraciones al interactuar entre sí a pesar de que ocurren a niveles bajos, generalmente perceptibles a niveles altos, donde la distorsión se vuelve más evidente.

1.3.4.2. Efecto Haas

La “Ley del Primer Frente de Onda” describe el fenómeno de percepción subjetiva de la ubicación de la fuente, dictado por el primer sonido en llegar al oído. Si el mismo sonido llega a ambos oídos con un ligero retardo de tiempo en uno de ellos, causara la impresión de que la fuente se localiza más cercana al oído donde el sonido llegó primero.

En condiciones naturales, siempre existirá un sonido que llega primero y de forma directa, y una o más reflexiones producidas en el medio, que llegan ligeramente después y de distintas direcciones. El sonido directo entrega la información de la dirección de la fuente.

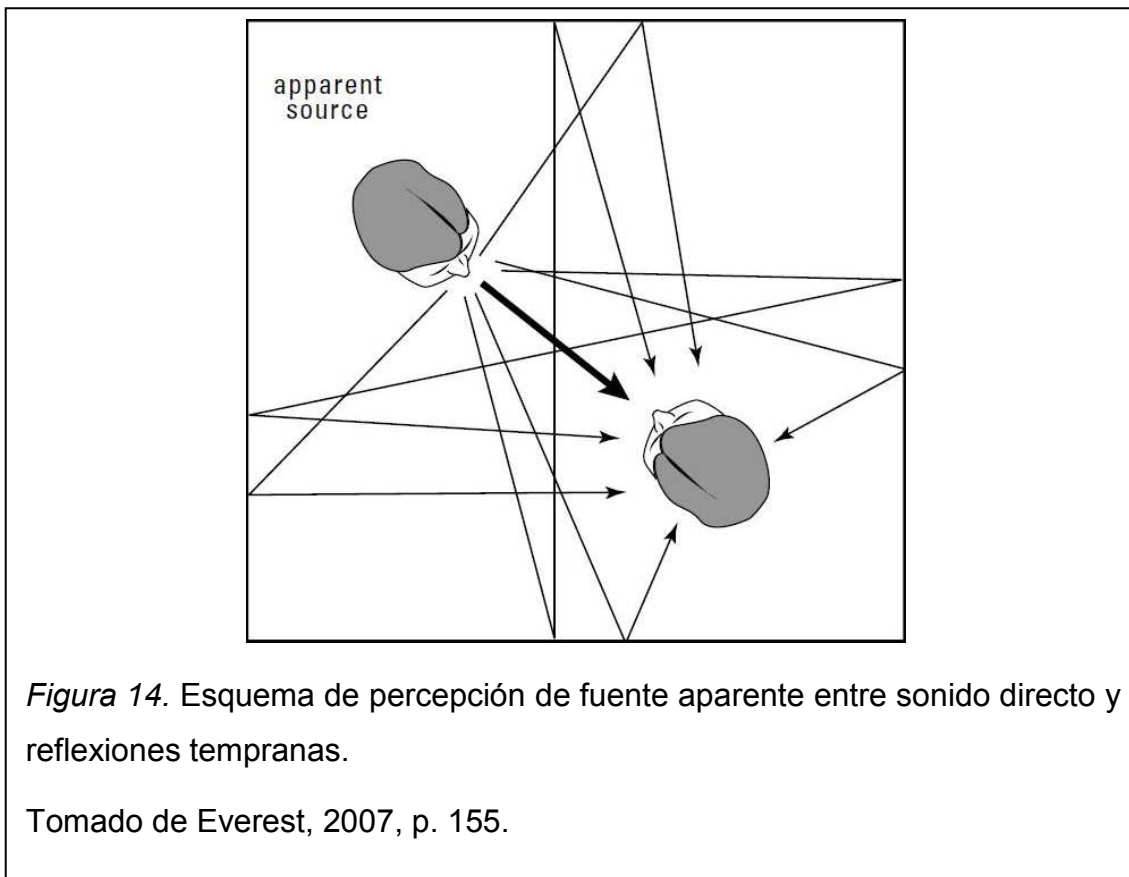


Figura 14. Esquema de percepción de fuente aparente entre sonido directo y reflexiones tempranas.

Tomado de Everest, 2007, p. 155.

Esto debido al efecto de integración de la audición conjunto con el cerebro, que integra sonidos en lugar de interpretarlos como independientes (ecos). Esto dentro de un tiempo de integración generalmente aceptado alrededor de los 30 y 35 [ms].

Al integrarse las energías de todas las ondas que son recibidas, se produce una sensación de mayor sonoridad y mayor espacialidad. Helmut Haas, quien estudio el efecto y cuyo apellido se ha vuelto sinónimo de dicho fenómeno (Efecto Haas) lo describe como "...una modificación agradable de la impresión sonora en el sentido de ampliación del ancho de la fuente de sonido primaria, mientras que la fuente de eco no se percibe acústicamente" (Everest, 2001, p.74).

El tiempo de integración dependerá de la envolvente del sonido. El límite de 35 [ms] es aceptado generalmente para señales de habla o musicales, al ser similares en cierta medida, pero no es válido para señales impulsivas (ataques cortos y decaimientos rápidos), donde el tiempo necesario para la fusión de

sonidos se limita a aproximadamente a 4 [ms]. Sonidos con sostenimiento prolongado podrán contar con tiempos de integración de hasta 80 [ms].

Dentro del rango de tiempo de integración, una señal retardada puede ser percibida independientemente si se presenta un incremento de nivel con respecto al primer fuente de onda (sonido directo). Este incremento necesario para que se produzca un eco deberá sobrepasar el “Umbral de Detección de Reflexión” (*Threshold of reflection detection*) que depende del tiempo de retardo.

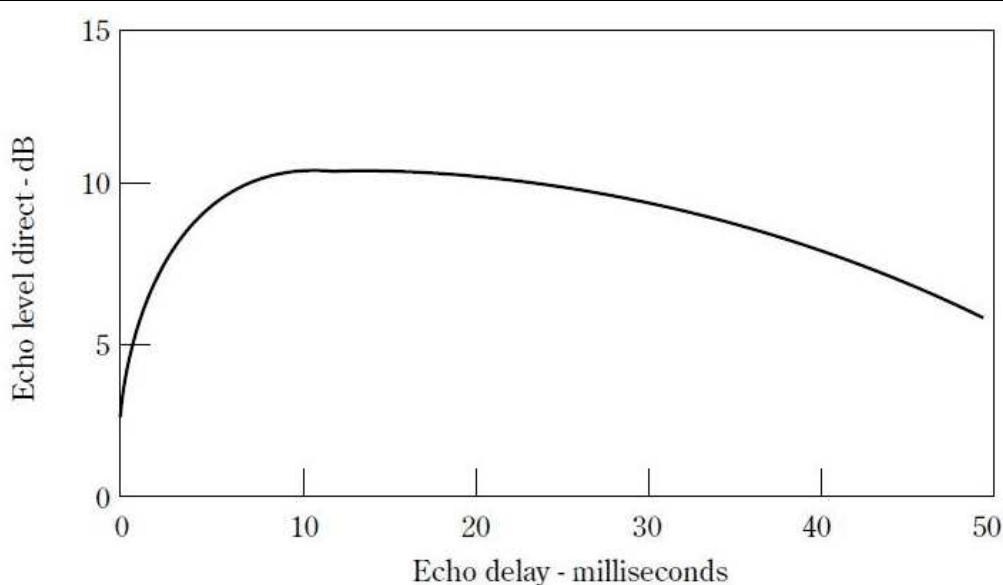
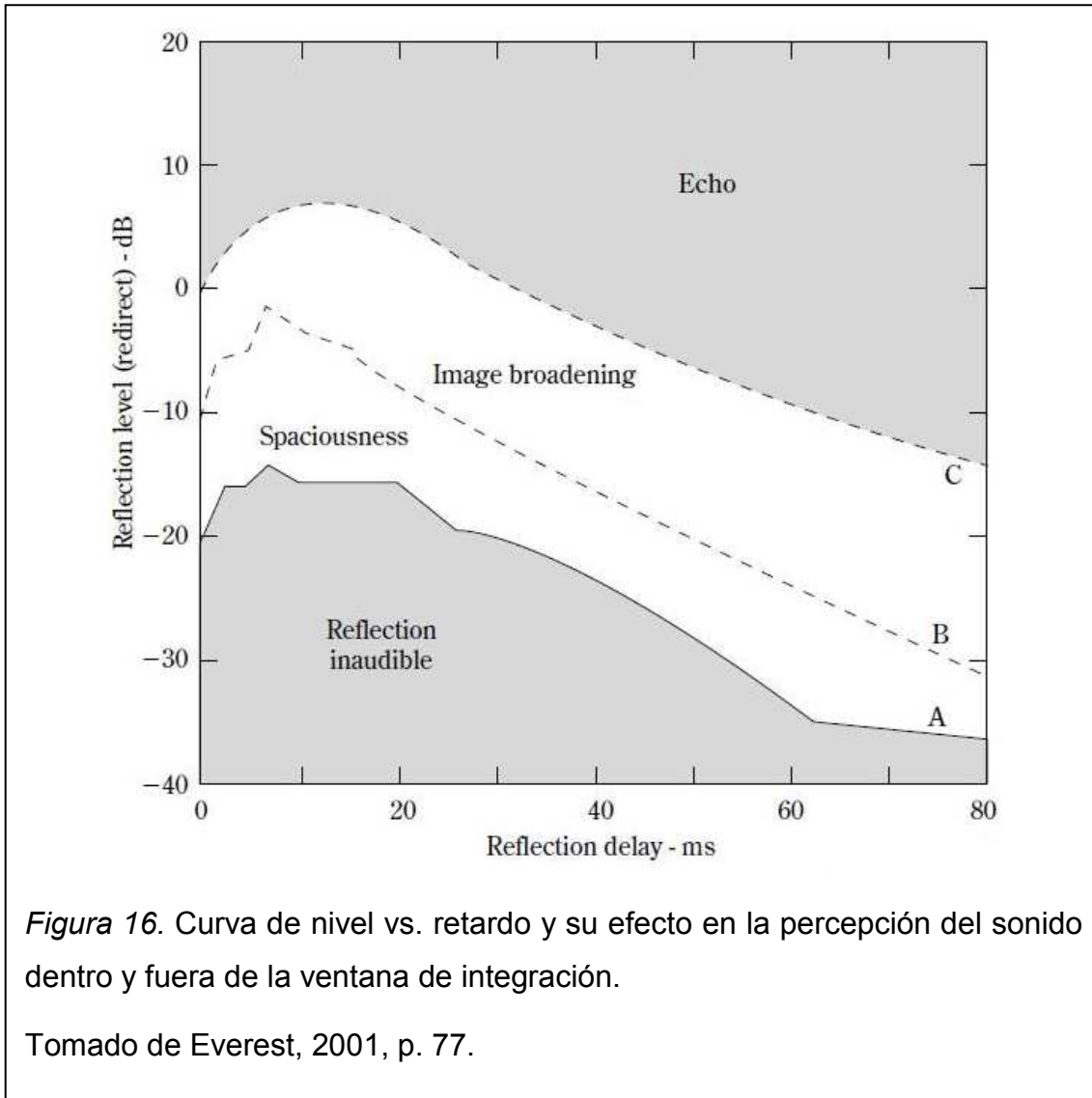


Figura 15. Curva nivel vs. Retardo necesario para la identificación de ecos según Haas.

Tomado de Everest, 2001, p. 75.

Sobre el umbral, las reflexiones se vuelven audibles en cierta medida, y la percepción de la fuente es afectada. Al aumentar gradualmente el nivel, incrementa la “espacialidad” del sonido” recibido. Al sobrepasar el umbral por 10 [dB] aproximadamente, se presenta un “ensanchamiento de la imagen sonora y un posible desplazamiento de la imagen” (Everest, 2001, p.76) hacia el oído que recibe el sonido directo, y a 20 [dB] sobre el umbral, el sonido es escuchado como ecos.



1.4. Psicoacústica y Escucha Binaural

La rama de la acústica que se encarga del estudio de anatomía del sistema auditivo oído y la percepción del sonido, tanto en su estructura como función, se denomina Psicoacústica.

Resulta muy complicado tratar de comprender en su totalidad a la audición ya que la percepción ocurre a niveles de procesamiento elevados en el cerebro, e involucra procesos cognitivos donde interviene el conocimiento adquirido y cierto grado de subjetividad.

En la parte sensorial, tampoco está totalmente comprendido el funcionamiento del mecanismo de la audición, así como del aporte de los órganos de del cuerpo y cabeza en la escucha tridimensional y ha sido sujeto de estudio por décadas, dejando algunos vacíos todavía por descubrir. Información de la naturaleza de la fuente, del espacio donde se produce y recibe el estímulo, la experiencia del oyente en localización de sonidos, además de las contribuciones del cuerpo son claves en la experiencia auditiva, en lo que respecta a la sensación del oyente.

Uno de las grandes capacidades es la de discernir información que llega de forma simultánea, siendo capaz de distinguir y concentrarse en estímulos concretos para su análisis según sea conveniente.

El hecho de poseer dos oídos y el estar distribuidos de una manera específica, es lo que permite contar con un sistema de “escucha binaural”, es decir, el ser humano es capaz de percibir sonidos provenientes de cualquier dirección y ser capaz de localizar la precedencia del mismo gracias a las claves o indicadores entregadas por el cuerpo e interpretadas por el cerebro.

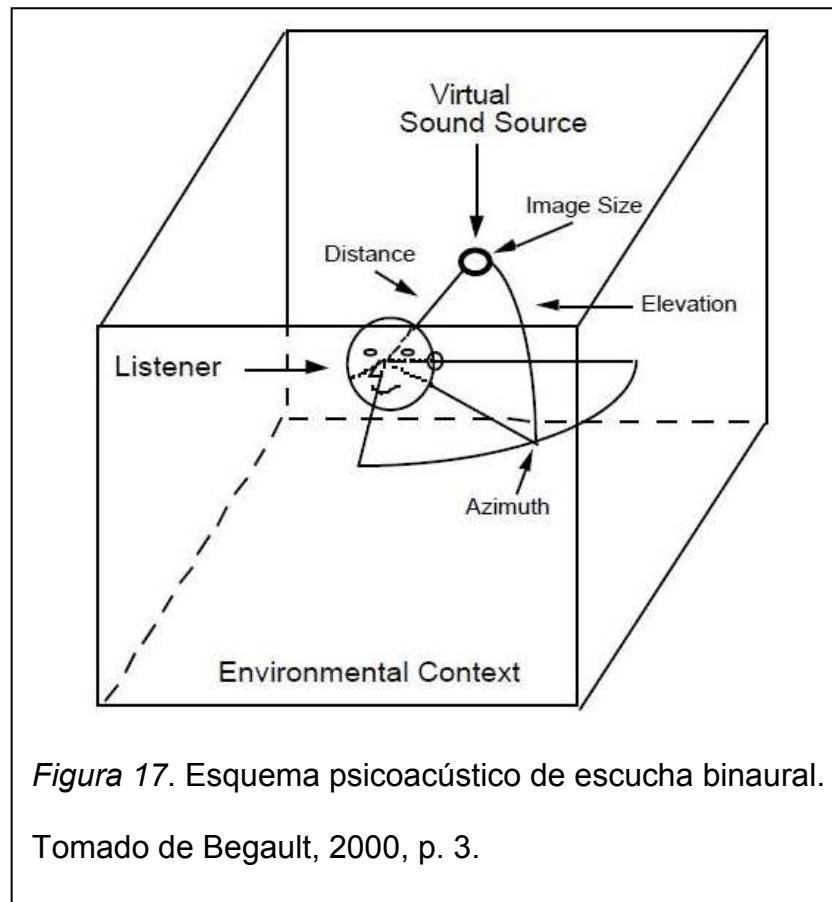
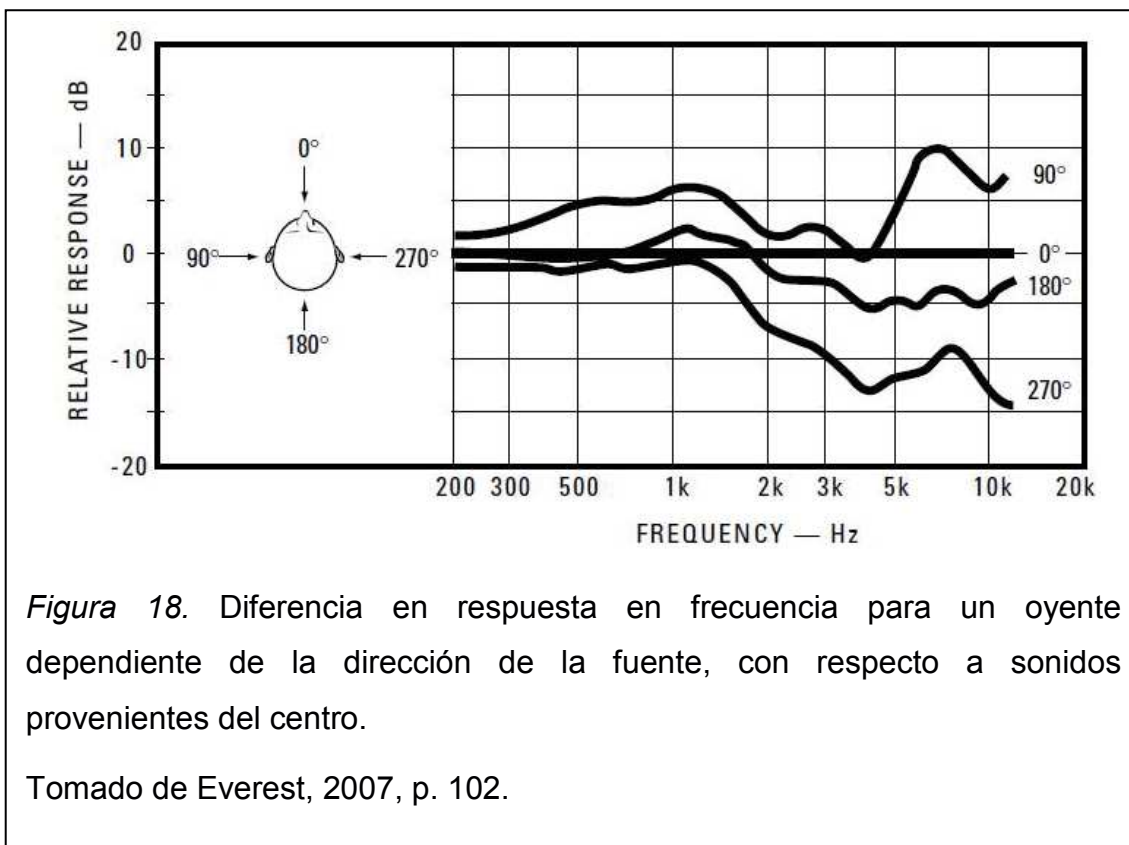


Figura 17. Esquema psicoacústico de escucha binaural.

Tomado de Begault, 2000, p. 3.

Las principales contribuciones son superpuestas en el contenido del sonido mismo en el oído externo para la localización de sonidos utilizadas por el cerebro para discernir la procedencia de los mismos, han sido estudiadas numerosas veces y comprobadas de forma experimental con test psicoacústicos y son introducidas por el del cuerpo (torso), la cabeza y el oído externo.

Debido a las diferencias interpersonales con respecto a tamaño y geometría de los órganos involucrados “la experiencia auditiva es altamente personal” (Lahoz, 2009, p.8).



La presión sonora que llega al tímpano, resultado de sonido directo y reflejado y modificado por la interacción del cuerpo, lleva información sobre la naturaleza de la fuente y su localización, ya que para cada ángulo de precedencia resulta diferente con respecto al sonido que llega de manera directa.

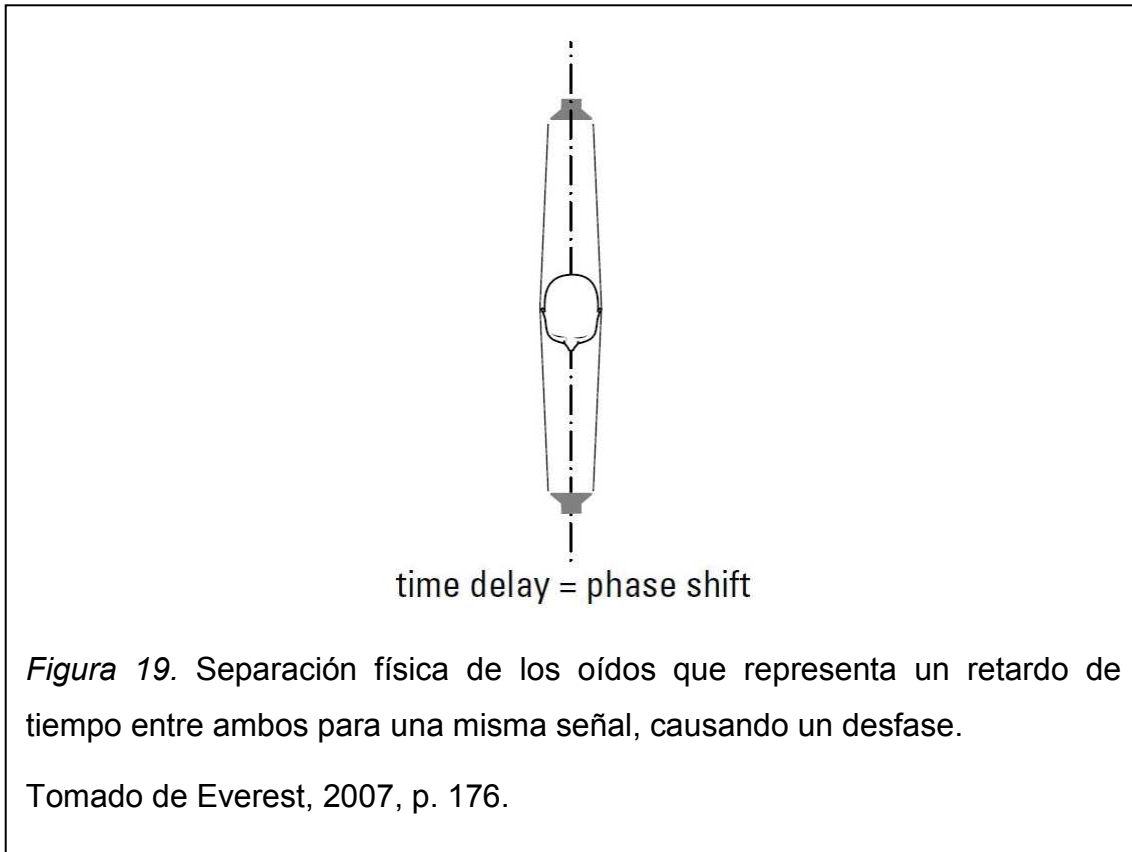
1.4.1. Localización en el Plano Horizontal (Acimut)

Cues o indicadores para el plano vertical o acimut, se deben al hecho de poseer dos oídos en la cabeza y al estar éstos separados por una cierta distancia y ubicados en el eje interaural. Gracias a esta configuración, se introducen diferencias interaurales, siendo éstos indicadores de “primer orden” para ubicación de la fuente en éste plano. La percepción para éste plano es la más eficiente en la audición.

Debido a ésta separación (aproximadamente 20 [cm]), la onda sonora incidente emitido por la fuente, toma mayor tiempo para llegar al oído más lejano a la fuente (oído contralateral) que para el oído más cercano (oído ipsilateral)

presentando diferencias de fase entre ambas señales que son analizadas en el cerebro.

Las ondas llegan al oído contralateral por dos caminos, por la frente y alrededor, por la parte posterior de la cabeza.



1.4.1.1. Diferencias Interaurales

a) Diferencias de Tiempo Interaural (ITD)

Es la diferencia de tiempo de llegada entre ambos oídos para un único frente de onda. Al recorrer distancias distintas para llegar al oído ipsilateral y contralateral y ser dividida por la velocidad del sonido en el medio de propagación, se puede encontrar el retardo de tiempo generado, que depende de la localización de la fuente.

b) Diferencias de Nivel Interaural (ILD)

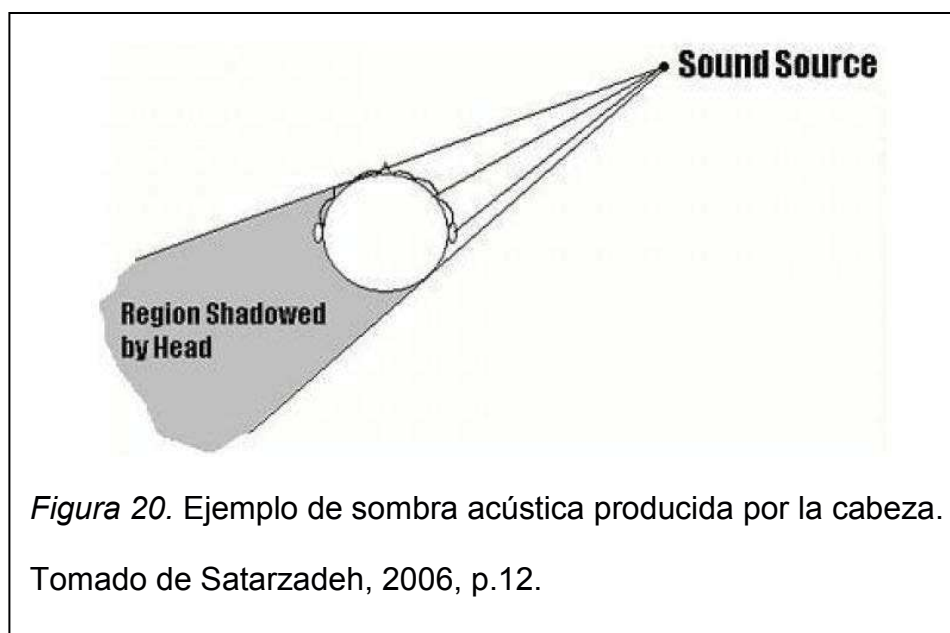
Es la diferencia de nivel de presión sonora entre ambos oídos, siendo mayor para el oído ipsilateral y menor para el contralateral, debido al efecto de atenuación y filtrado producido por la sombra acústica de la cabeza.

El rango efectivo del ITD está comprendido entre 0,005 y 1 [ms], y entre 1 y 10 [dB] para ILD, aunque los valores dependerán de acuerdo al sonido que produce la fuente.

1.4.1.2. Sombra Acústica de la Cabeza

Otra comparación entre sonidos que llegan a ambos oídos, es la atenuación de nivel que produce la cabeza. Ésta actúa como un obstáculo que se interpone en la propagación de la onda generando una sombra acústica, siempre y cuando la longitud de onda del sonido sea menor con respecto a su circunferencia.

Para longitudes de onda mayores a la circunferencia, la onda se difractará alrededor de la cabeza, por lo que su efecto se volverá más débil a medida que la frecuencia del sonido sea más baja.



1.4.1.3. Sombra Acústica del Pabellón Auditivo

La confusión “frontal/trasera” producida por la inexistencia de diferencias interaurales, muchas veces se resuelven por medio del pabellón auditivo. Para sonidos de frecuencias altas, cuyas longitudes de ondas se vuelven menores al área del pabellón, este puede actuar como una barrera, produciendo una sombra acústica, es decir, un filtrado de frecuencias altas.

Está claro que para poder lograr una diferenciación frontal/trasera en base a la sombra acústica producida por el pabellón, el sonido incidente deberá poseer componentes espectrales de alta frecuencia, y de ésta manera, las diferencias sean apreciables para el oyente.

1.4.2. Localización en el Plano Vertical (Elevación)

En el plano vertical o elevación, existen varias partes que contribuyen para su localización, siendo generalmente aceptada como la principal y más importante la del pabellón auditivo (o *pinna*).

1.4.2.1. Pabellón Auditivo o Auricular

Su forma, tamaño y complejidad (las depresiones en su superficie frontal), introducen cambios espectrales en el sonido incidente. Sus pliegues y crestas, se distribuyen de forma asimétrica, y debido a que se recibe sonido de forma directa además de sonido reflejado (y retardado en el orden de los microsegundos) en la superficie del pabellón y son combinados en la concha (entrada al canal auditivo), se producen interferencias constructivas y destructivas (diferencias de fase).

Esto lo convierte en lo que algunos se refieren como un dispositivo de localización monoaural de sonido, ya que no es necesaria la interacción de ambos oídos para su función como decodificador de elevación.

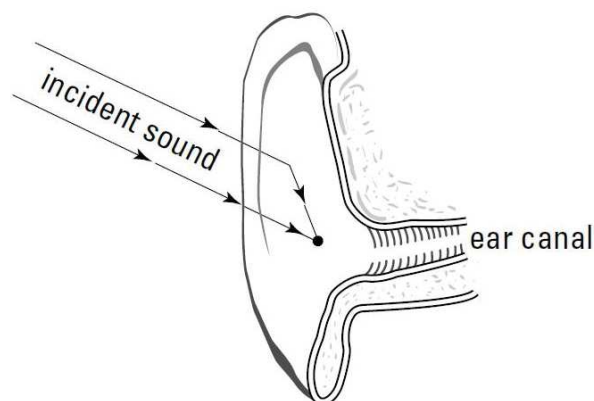


Figura 21. Sonido directo y reflexiones producidas en el pabellón antes de ingresar al canal auditivo.

Tomado de Everest, 2007, p. 180.

Esto produce valles (conocidos como *Pinna Notch*) que varían dependiendo principalmente de la elevación de la fuente a modo de un filtro peine o *comb*, para frecuencias donde el tiempo de retardo de acuerdo con las conclusiones de Batteau “en el rango de los 0 a 300 [μs]” (Begault, 2000, p.41), entre el sonido directo y el reflejado sea la inversa de la distancia entre el sonido directo y la reflexión, para una determinada velocidad del sonido.

Las frecuencias de los *notchs* se producen en el rango entre 6 [kHz] y 16 [kHz] aproximadamente, por lo que se lo considera un equivalente a un ecualizador gráfico. Esto indicaría que el sonido deberá poseer componentes espectrales sobre los 6 [kHz] aproximadamente para poder percibir los cambios en el espectro que son introducidos por el pabellón.

Sin embargo, el tamaño y forma del pabellón varía de persona a persona, por lo que los rangos no se pueden determinar con exactitud, aunque se consideran rangos recurrentes y generalmente aceptados en las referencias consultadas.

Además, no se conoce cómo ni en qué medida cada uno de estos picos o valles representan una clave o indicador de localización psicoacústica, e inclusive, teorías como la de Noble (1987) concluye bajo sus propios estudios,

que “alrededor del 25% de la población exhibe poco o nada de juicio de elevación basado estrictamente en claves espectrales” (Begault, 2000, p.52).

También juega un rol importante en resolver la “Confusión Frontal/Trasera”, debido al filtrado y atenuación que introduce y se muestra en la función de transferencia para fuentes provenientes del frente y detrás.

No todas las claves de elevación se pueden considerar monoaurales o de alta frecuencia. Debido a sus dimensiones, la cabeza y el torso producen difracción y reflexión al interponerse al sonido incidente, producirán cambios espectrales (como filtrado tipo peine) en bajas frecuencias (considerado como menor a 3 [kHz] aproximadamente) donde sus tamaños comienzan a ser comparables con las longitudes de onda.

Pruebas y análisis de HRTF demuestran la existencia de dichas claves por lo que sonidos sin gran cantidad de energía en la parte “alta” del espectro podrán ser localizadas en el plano vertical.

Sin embargo, Algazi, Avendaño y Duda (2001, pp.1119-1120) afirman que los efectos producidos por el torso, no solo son poco conocidos pero débiles y notorios en el oído ipsilateral, al contrario de los efectos producidos por la cabeza, más claros para el oído contralateral (debido a la sombra que produce).

La ITD dependiente de la elevación introduce cambios en baja frecuencia, además de otras partes como piernas o rodillas, pero en un rango muy limitado de localizaciones espaciales de la fuente.

Sumados los efectos tanto del pabellón, como de la cabeza y del torso, se obtiene un sistema de localización completo en el espectro audible, aunque todavía no esté totalmente comprendido.

Su complejidad (pliegues, distribución, tamaño, etc.) varía significativamente entre individuos y “se ha demostrado que estas diferencias son significativas perceptualmente” (Satarzadeh, 2006, p.17), por lo que el definir modelos de

pabellón para síntesis de sonido en el plano vertical es mucho más compleja y difícil de definir y muchos modelos han sido planteados y estudiados.

Combinando los planos horizontal y vertical, es posible ubicar a una fuente en una superficie esférica que rodea a la cabeza del receptor, pero sigue siendo necesaria una tercera dimensión para lograr un efecto convincente de distancia.

1.4.3. Percepción de Rango o Distancia

Para el plano de la distancia o “rango”, es decir, que tan lejos o cerca se encuentra la fuente del oyente, el cerebro se basa en los fenómenos acústicos del entorno donde sucede la producción y recepción del sonido.

La percepción de distancia es la más defectuosa. Begault (2000, p.2) la describe como multidimensional, por lo que puede resultar confusa en la determinación, con claves contradictorias, al poder ser relativa o absoluta, como lo es la percepción del tamaño que ocupa la fuente dentro del espacio donde se encuentra.

El primer indicador se debe a la relación entre sonido directo y sonido reverberante que llega al oído, siendo mayor el sonido directo al estar más cercana la fuente, y mayor el reverberante al estar más alejada ésta.

Las características de la reverberación darán cuenta de las condiciones acústicas (contexto) del lugar para así poder generar un espacio necesario para ubicar a la fuente conocido como Externalización (ya que no existe una condición anecoica de forma natural).

El tiempo de reverberación o T_{60} está relacionado con la percepción del tamaño del recinto (sobre todo gracias a las reflexiones tardías), la difusión (dispersión de las reflexiones) establece la relación entre fuente, oyente y realidad física del entorno y, el *predelay* (tiempo entre señal y primera reflexión), la distancia entre fuente y oyente.

Estos parámetros, dependen en cierta medida también de las propiedades de emisión de la fuente. Las propiedades acústicas de los materiales que recubren los límites del entorno determinarán las características del timbre de la reverberación.

A mayor distancia entre fuente y oyente, se produce un decaimiento de nivel. El fenómeno conocido como “Ley del Inverso Cuadrado”, describe la caída de intensidad con el cuadrado de la distancia. La percepción de parámetros físicos como nivel o intensidad se relacionan con el *Loudness* o Sonoridad, que si bien pueden ser medidos, no se limita a ser un indicador numérico. Esto pues la sonoridad es un fenómeno que depende de la amplitud y la frecuencia del sonido.

La respuesta del oído a niveles bajos presenta deficiencias en ambos extremos del espectro y para sonidos a niveles altos, la respuesta tiende a igualarse. Esto no significa que llega a ser plana ni que existe una relación directa con el nivel o intensidad. El conocer la fuente y sus propiedades de emisión sonora (aunque sea de manera inconsciente) y estar familiarizado con el sonido que produce, ayudan a poder diferenciar entre una fuente que se encuentra a distintas distancias según la respuesta que es emitida y percibida.

Las diferencias interaurales, especialmente la ILD, juegan un papel en la percepción de distancia. Para fuentes muy cercanas, pequeños movimientos de la cabeza producirán notables diferencias entre ambos oídos. Para fuentes lejanas, ligeras diferencias no significan un cambio que aporte información útil en este plano, por lo que se la considera a la ILD como un indicador importante para distancias dentro de un radio de 1 [m] alrededor del oyente

Existen ciertas modificaciones espectrales presentes en el plano de la distancia. Al estar la fuente más cercana, el nivel incrementará y de acuerdo a los contornos de sonoridad, la respuesta tenderá a aplanarse (con una notable compensación en altas y bajas frecuencias). Para fuentes muy lejanas, también existirán deficiencias en altas frecuencias, por su incapacidad física (mayor pérdida de energía) de llegar hasta el oyente.

La relación entre sonido directo y sonido reverberante puede ser considerada como la más importante al considerar que es eficiente en grandes y cortas distancias, exista o no diferencia de nivel o movimientos de cabeza y diferencias interaurales. El problema también radica en que la percepción tiende a ser relativa más que absoluta, es decir, no es tan fácil establecer una distancia exacta (en metros por ejemplo) a la fuente pero sí afirmar si se encuentra lejos o cerca en relación al oyente o al entorno.

1.4.4. Otras Contribuciones

El torso y la cabeza también contribuyen a la localización al producir difracción y actúan como barreos generando sombras y reflexiones a las ondas que se propagan entre la fuente y el tímpano, generando indicadores que se pueden considerar como secundarios, ya que si bien son pocos o más recientes los estudios que demuestran dichas contribuciones y resultan no ser tan relevantes o importantes en la localización, en un modelo para simular dichos aspectos físicos, no son estrictamente necesarios para generar un sistema de síntesis de sonido 3D.

Se ha comprobado tras varios estudios que existen problemas de localización de fuentes y que ubicar sonidos en el plano horizontal, vertical y de distancia es respectivamente más complicado para el oído y sus mecanismos son respectivamente menos comprendidos, además de una falla común de la audición en general al ser un sistema todavía menos preciso en diferenciar entre sonidos ubicados al frente o detrás (conocido como confusión frontal/trasera).

Esta confusión, común en la mayoría de personas, se debe a la similitud en la respuesta para fuentes ubicadas atrás y adelante. Al analizar las HRTF, se observa que la respuesta frontal es ligeramente mayor en nivel a la trasera en el rango cercano a los 4 [kHz]. Claves espectrales introducidas por el pabellón ayudan a solucionar éste problema y movimientos de cabeza.

De acuerdo a Blauert, la diferenciación para señales provenientes desde el frente y desde atrás se debe a la resonancia que se produce en la concha del

pabellón con “una diferencia de alrededor de 5 [dB] [que] ocurre en los 10 [kHz]” (Begault, 2000, p.43).

Al emplear moldes de pabellón auditivo en ciertos experimentos Lahoz (2009, p.9), la percepción de la elevación resultó menos eficiente ya que al alterar su forma, las claves de localización que son aprendidas por el cerebro cambian, lo que demuestra que el cerebro necesita aprender dichas claves para cada individuo y su propio pabellón, además de estar relacionadas con otros estímulos sensoriales. Si bien es posible adaptarse a nuevas claves, esto involucra un proceso de adaptación y aprendizaje.

El sistema auditivo también se ve complementado por otros estímulos que recibe el cuerpo. Para el caso de fuentes ubicadas al frente del oyente, pequeños movimientos del cuerpo son los que introducen cambios interaurales al parecer insignificantes que logran hacer una diferencia entre los sonidos dentro y fuera de la cabeza (error presente en sistemas de sonido 3D).

Para confusión de fuentes frontales/traseras a la cabeza, en el caso de no existir diferencias interaurales, el cerebro espera a un estímulo visual que represente la fuente emisora de sonido, poniendo a trabajar a ambos sentidos en conjunto (y al no encontrarla, ubicarla atrás de la cabeza por ejemplo).

El mover la cabeza e introducir pequeñas diferencias interaurales, puede ayudar a eliminar el problema. Si la fuente esta cerca del oyente, el cambio de acimut será más evidente que para fuentes distantes. “Es probable que el cerebro use éste fenómeno, conocido como paralaje, para precisar mejor la distancia a la que se encuentra la fuente” (Monte, Ramírez y Vallez, 2003, p.41).

1.4.5. Fuentes Móviles

Para fuentes móviles, de acuerdo con Begault (2000, p.40), en condiciones optimas, se afirma que el Mínimo Angulo de Movimiento Audible es de 3°, en comparación con 1°, establecido para fuentes. El principal encargado de entregar información para fuentes en movimiento es el fenómeno *Doppler*,

donde existe un cambio de frecuencia del sonido emitido por la fuente, al estar esta en movimiento con respecto al observador.

1.5. Función de Transferencia

Una onda sonora que llega al tímpano depende de un ángulo de incidencia y la distancia a la fuente. Si dicha distancia está en campo lejano (mayor a 1 [m]), la función deja de depender de ésta al volverse una onda plana. Al recibir la onda, existirá presión sonora tanto en el canal auditivo como en el tímpano.

Al obtener una relación entre la presión producida por dicha onda sonora en un punto P_i , con respecto al punto de origen denominado P_1 , se obtiene una función de transferencia (P_i/P_1).

“Una onda sonora proveniente de una dirección y distancia resulta en dos presiones sonoras, una en cada tímpano. La transmisión se describe en términos de dos funciones de transferencia que incluye cualquier distorsión lineal como coloración y tiempo interaural y diferencias espectrales” (Moller, 1992, p.176).

Estas funciones fueron definidas por primera vez por Blauert como *Free Field Transfer Function* (FFTF). Las funciones que define Blauert son varias y “entregan una descripción completa de la transmisión sonora, incluyendo difracción alrededor de la cabeza, reflexión en los hombros, reflexiones en el canal auditivo, etc.” (Moller, 1992, p.181).

Dentro de este grupo de funciones, se encuentran las Funciones de Transferencia Relativas a la Cabeza (HRTF), válidas para ciertos puntos P_i y donde el punto de origen P_1 , es la presión sonora en el medio de la cabeza cuando el oyente se encuentra ausente.

Las HRTF son únicas y representan una “estampa” espectral de escucha en condiciones normales (naturales, no artificiales) que puede ser superpuesta en un sonido, lo que la hace válida para simulación de sonido 3D.

1.5.1. Funciones de Transferencia Relativas a la Cabeza (HRTF)

El grupo de funciones denominadas como *Head Related Transfer Functions* (HRTF) en el dominio de la frecuencia, o bien HRIR (*Head Related Impulse Responses*) en el dominio del tiempo (donde es posible visualizar las reflexiones), pueden ser para cualquiera de las relaciones entre $P2/P1$, $P3/P1$ o $P4/P1$, donde $P2$ es la presión sonora en el canal auditivo bloqueado, $P3$ la presión a la entrada del canal auditivo abierto y $P4$, la presión en el tímpano.

Para una grabación binaural, se puede emplear cualquiera de estos puntos con la compensación necesaria en la reproducción para asegurar su correcta representación.

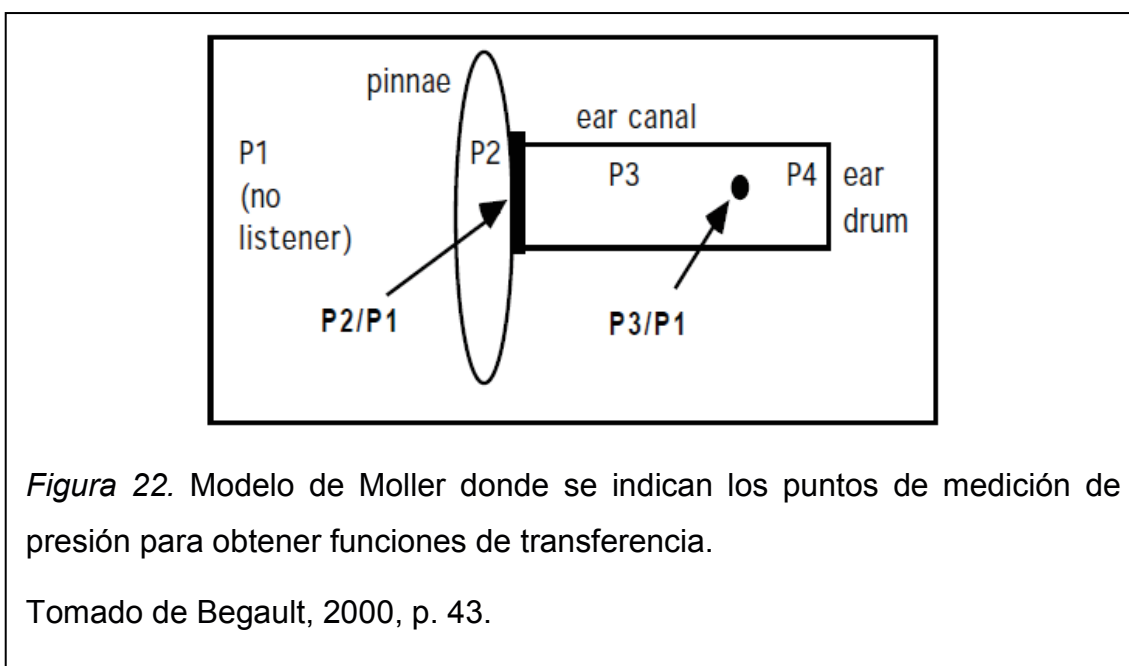


Figura 22. Modelo de Moller donde se indican los puntos de medición de presión para obtener funciones de transferencia.

Tomado de Begault, 2000, p. 43.

En algunas de éstas HRTF descritas, sus variables dependen de la anatomía del oyente, a partir del punto de medición empleado. La entrada del canal auditivo como punto de medición no está claramente definido, únicamente se asume que a partir de dicho punto, la transmisión es unidimensional (existen diferencias entre individuos y no dependientes de la posición de la fuente).

$$[P_2/P_1](\text{subject, side, } r, \phi, \theta)$$

$$[P_3/P_2](\text{subject, side})$$

$$[P_4/P_3](\text{subject, side})$$

Figura 23. Dependencia de las variables para cada Función de Transferencia de la Cabeza HRTF.

Tomado de Moller, 1992, p. 182.

En un intento de determinar a partir de qué punto se puede asumir la transmisión unidimensional, Henrik Moller llega a la conclusión que el mismo “puede ser escogido en cualquier lugar entre el tímpano y la entrada del canal auditivo. Posiblemente, algunos puntos unos pocos milímetros fuera del canal y en línea con el mismo pueden ser usados” (Moller, 1992, p.183).

Los efectos del canal auditivo son “esencialmente no direccionales” (Begault, 2000, pp.42-43) y son demostrados por varios experimentos. En cuanto a la concha (pabellón), la direccionalidad (que depende del rango de frecuencia medido y el lugar donde se realiza la medición) están presentes en las HRTF, por lo que la relación entre presión sonora en P2 y P1, es la única que solo incluye aspectos direccionales y contiene la menor cantidad de información unidimensional posible al estar el canal físicamente bloqueado, y lo que resta de la transmisión, considerada unidimensional, sin influir en la grabación.

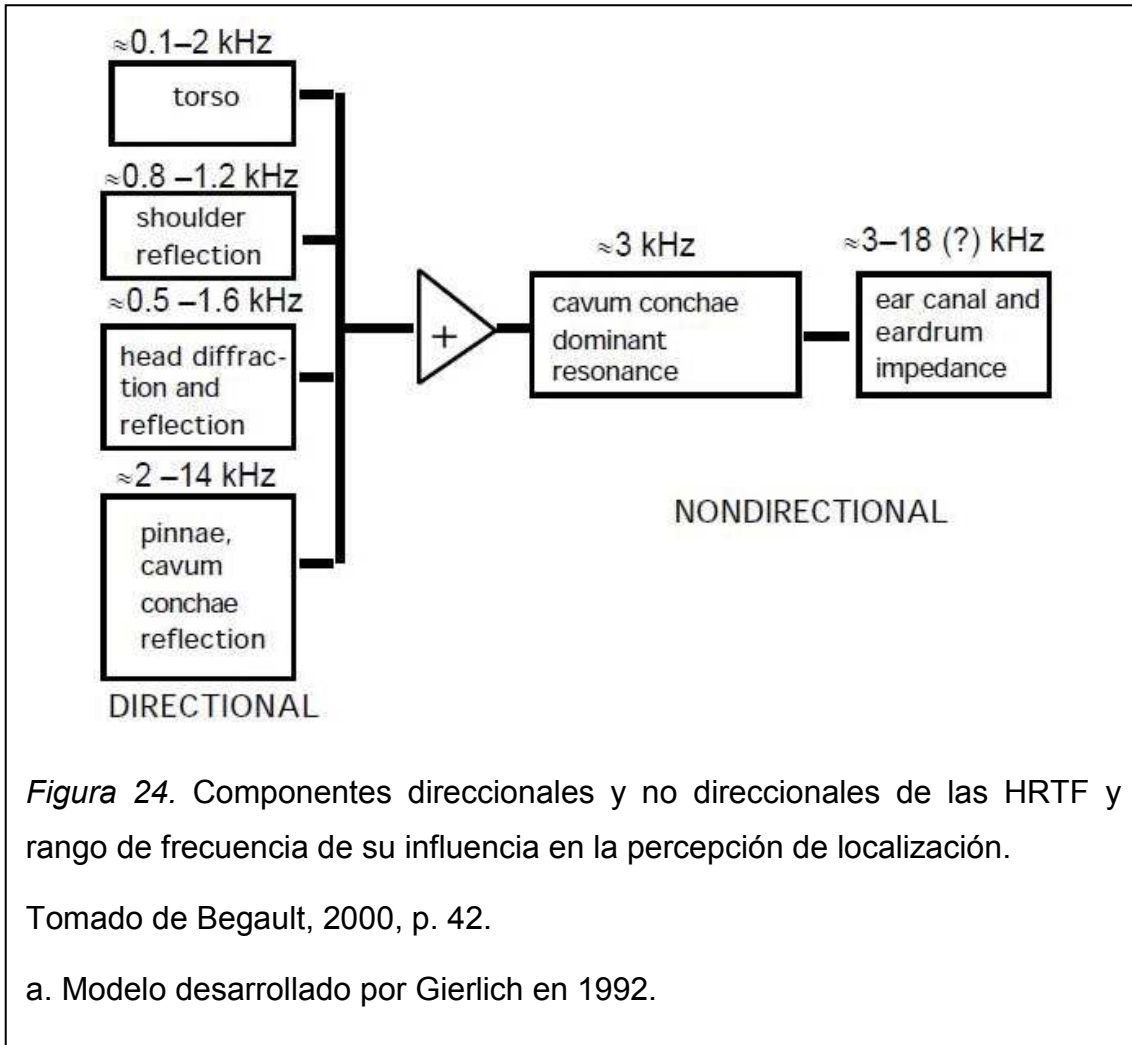


Figura 24. Componentes direccionales y no direccionales de las HRTF y rango de frecuencia de su influencia en la percepción de localización.

Tomado de Begault, 2000, p. 42.

a. Modelo desarrollado por Gierlich en 1992.

Todos los cambios en los parámetros (amplitud, frecuencia y fase) que definen la onda sonora, además de las modificaciones espectrales debido a reflexiones y difracciones (son captados y caracterizados matemáticamente por cualquiera de las HRFT definidas, siendo la diferencia la cantidad de información unidimensional que poseen.

Al realizarlas para cierta posición de la fuente con respecto al punto medido, y obteniendo una función para cada oído, se puede hablar de HRTF binaural, pues se obtienen dos HRTF discretas para cada punto.

Ejemplos de su uso generalmente implican tablas de HRIR binaurales medidas y almacenadas en memoria que son convolucionadas con una señal

monofónica de entrada para cada oído dependiendo de la ubicación de la fuente deseada.

Las HRTF pueden ser empleadas para sistemas de sonido 3D logrando los mejores resultados posibles siempre y cuando, las HRTF utilizadas en el filtrado de señales en la reproducción sean las propias del oyente, debido a que involucran su estructura anatómica propia, por lo que se consideran imprácticas al emplearse para un oyente cualquiera e ineficientes al ser generalizadas.

Por ésta razón, existe un gran interés en sistemas de síntesis de sonido 3D en base a modelos de HRTF, que ofrezca resultados convincentes al ser sintéticos y ser generados a partir de un modelo físico o estructural, y que sea adaptable a distintos individuos, y así evitar la inconveniencia de ser medidos para cada individuo.

El empleo de HRTF o HRIR tiene su ventaja en simulación de elevación (donde los modelos estructurales no han alcanzado mayor éxito debido a la complejidad y lo poco que se conoce del asunto). Además, de acuerdo a Begault (2000, pp.49-50), el rol principal de HRTF es el de diferenciar entre errores comunes, como confusión frontal/ trasero o arriba/abajo (aunque depende del estímulo utilizado para su evaluación), además de ayudar en la Externalización de imágenes sonoras.

Una desventaja en términos prácticos también es el hecho de que sigue sin existir un estándar para la grabación de los impulsos-respuesta, determinación de puntos o espaciado, filtrado o ecualización de compensación (de ser necesario) entre otros que entorpecen procesos de comparación o evaluación de distintos trabajos, o el uso de distintas base de datos entre otros.

1.5.2. HRTF No Individualizadas

El uso de HRTF no individualizadas es bastante común. Muchos de los estudios buscan generalizar los resultados obtenidos e inclusive sintetizar nuevas funciones de transferencia basadas o derivadas de las obtenidas con un segmento de la población, aunque su resultado no garantiza que las nuevas

funciones posean información espectral adecuada. Confusión frontal/trasera, así como *reversals* es más notoria con su empleo.

Además, el hecho de emplear HRTF no individualizadas, degrada la capacidad de escucha del oyente, especialmente en el plano vertical, como es el escuchar con otro set de pabellones, con los cuales no se aprendió a localizar fuentes y es necesario adaptarse.

1.5.3. Medición, Grabación y Reproducción de HRTF

No existe un procedimiento estandarizado para la medición de HRIR. Varias formas para su obtención han sido propuestas en base a las necesidades y condiciones de los autores y sus investigaciones. Sin embargo, es posible establecer ciertas ventajas, como el medir en la entrada del canal auditivo bloqueado. De ésta forma se evita capturar la resonancia del mismo y al momento de ser escuchados mediante audífonos, se evita una resonancia doble (la del oyente) durante la reproducción.

Para la grabación de HRIR, una fuente genera una señal de prueba (como un impulso, a veces llamado “impulso analítico” o un barrido de frecuencia conocido como *sine sweep*) dentro de una cámara anecoica y se mide, ya sea en la entrada del canal, en el canal auditivo o en el tímpano, de forma simultánea se captan las presiones en ambos oídos y se digitalizan para el almacenamiento.

El hecho de ser realizadas bajo condiciones anecoicas, mantiene a los impulsos-respuesta cortos y, en caso de ser implementados para síntesis de señales monofónicas, menor carga en el procesador y no las limita a ser válidas para simular únicamente el entorno donde fueron obtenidas. El problema está en que carecen de Externalización. Es posible la implementación de un modelo de reverberación externo para evitar este problema en un sistema de sonido 3D.

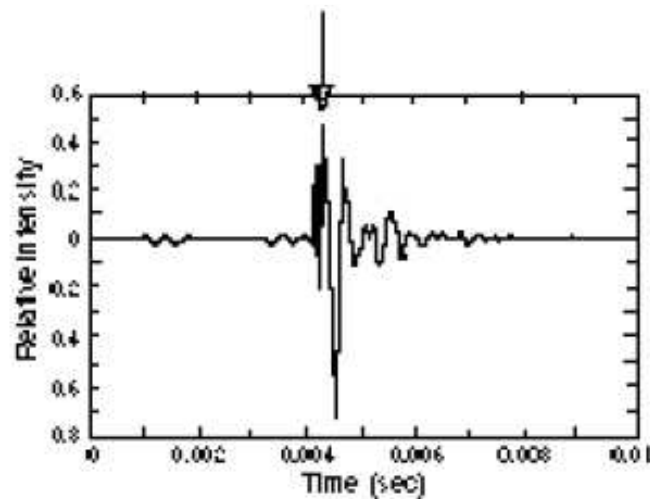


Figura 25. Ejemplo de un impulso-respuesta en el dominio del tiempo medido en cámara anecoica.

Tomado de Begault, 2000, p. 110.

Se emplean micrófonos “de sonda” (*probe microphone*), por su tamaño reducido y facilidad de colocarlos en las cavidades del oído, y que según Moller (1992, p.185) presentan problemas de baja sensibilidad, dificultad de calibración y respuesta poco plana típicamente (por ondas estacionarias en el tubo).

Para la grabación se emplea o bien un individuo o una cabeza artificial, y el micrófono de sonda puede ubicarse en cualquier de los puntos P_i antes descritos para la obtención de impulsos respuesta y posteriormente, HRTF, las cuales deberán ser compensadas con la ecualización correcta previo a la etapa de reproducción de acuerdo al método (posición del micrófono durante la medición) empleado.

La compensación en la reproducción es necesaria para sobrellevar problemas de la sensibilidad de micrófono, la función de transferencia del audífono desde sus terminales hasta la presión sonora P_i (según donde se grabó) y únicamente en el caso de grabar con los micrófonos ubicados en el canal auditivo bloqueado (P_2), se deberá compensar por la diferencia en la transmisión causada por diferencias en la impedancia acústica de las fuentes en las dos

situaciones de escucha, únicamente en caso de no emplearse audífonos abiertos en la reproducción, de acuerdo a Moller (1992, p.191).

Todos los métodos requieren calibración de audífonos (determinación de su función de transferencia en el punto donde se realiza la grabación) además de la colocación de audífonos en dichos puntos en sus oídos, lo que puede llegar a ser inconveniente (incomodo y se interpone en el camino de propagación de la onda, que perjudica la grabación), lo que representa la menor ventaja para método del canal auditivo bloqueado (P2), aparte de poseer la menor cantidad de información individual (transmisión unidimensional) y facilidad de obtener datos para cabezas artificiales con pabellones removibles como forma de estudio de la influencia de las mismas en la localización.

1.5.4. Densidad Espacial e Interpolación de HRTF

Es necesario 1° de separación para la que un adulto mayor pueda distinguir entre dos posiciones de una fuente sonora en el plano frontal, lo que resulta “prácticamente imposible medir las HRIR con esta densidad espacial” (Araneda, Ramos y Tommasini, 2010, p.3).

Debido a que es posible obtener HRIR es realizarlo con individuos que se ubican frente a la fuente mientras ésta se mueve para cubrir las posiciones de las cuales se deseen obtener los impulsos, el individuo naturalmente se moverá durante el proceso, por lo que tratar de asegurar 1° de precisión desde el punto de vista práctico y desde el la medición es muy ambicioso, por lo que muchos investigadores optan por comprometer la resolución espacial de los impulsos a favor de una medición más precisa y segura.

Muchos desarrolladores optan por interpolación de las HRTF para obtener puntos intermedios entre puntos medidos y así mayor número de grados en aplicaciones donde se requiera impulsos respuesta para su convolución por ejemplo. Si bien es una solución al problema de encontrar valores intermedios, la interpolación de las HRTF entrega valores a partir de un cálculo matemático por lo no asegura una correcta representación de las claves psicoacústicas reales de la nueva posición obtenida.

Las posiciones generadas por interpolación tienen valores intermedios de espectro entre sus HRTF más próximas, resultando en un proceso totalmente objetivo. Wenzel y Foster determinaron en 1993 que al usar HRTF no individualizadas, “el error inherente a una HRTF interpolada es sumergido por el error total de localización del sujeto” (Begault, 2000, p.135).

1.5.5. Procesamiento de HRTF

Muchas veces es necesario post procesamiento de los impulsos obtenidos en la medición, generalmente referidos como “crudos” (*RAW*). Los procesos se llevan tanto en el dominio del tiempo y frecuencia, con el fin de eliminar errores o información no deseada. Es importante que los procesos sean aplicados a todas los impulsos obtenidos.

1.5.5.1. En el Dominio del Tiempo

Desechar porciones en blanco, es decir, el tiempo que toma al impulso en llegar al micrófono desde la fuente. El proceso se debe llevar a cabo para pares de HRIR (izquierda y derecha), para evitar eliminar la ITD. Normalización para mejor uso de rango dinámico, donde “la muestra de mayor sonoridad tendrá en valor de cuantización mas alto” (Begault, 2000, p.121).

1.5.5.2. En el Dominio de la Frecuencia

Ecuilización para compensar las deficiencias y no linealidad de los elementos de la cadena electroacústica empleada en la medición de las HRIR. Algunas mediciones requieren ecualización para añadir la resonancia típica del canal auditivo para su estudio, o remover dicha resonancia ya que si los impulsos serán implementados para su escucha mediante audífonos, la resonancia del usuario estará presente en la reproducción, y de ésta manera, evitar una doble resonancia.

Si en la medición los micrófonos fueron ubicados en el canal auditivo, las HRTF resultantes presentarán información unidimensional al incorporar ésta respuesta extra como parte de la función de transferencia. “Una manera de eliminar todos los factores que no varían en función de la dirección es ecualizar

las HRTF en referencia a un campo difuso” (Casey, 1996), mediante filtrado de las mismas con una nueva HRTF que presenta un espectro plano y es generada al obtener un promedio energético de todas las HRTF medidas (y que deberían haber sido obtenidas de forma uniforme alrededor de la cabeza).

Otra compensación se debe al tipo de audífonos a utilizarse en la reproducción. Audífonos comerciales son considerados como de ecualización de campo libre (*free-field equalization*) y se basan en el principio de que el espectro producido en el tímpano es exacto al que se escucha en una cámara anecoica, a 0° de elevación y acimut (para ruido blanco). En este caso, “la reproducción es óptima para una posición de 0° acimut y elevación, pero no óptima para una fuente sonora virtual que incide desde otra dirección” (Begault, 2000, p.125), lo que los vuelve incompatibles para un sistema de sonido 3D, como lo estableció Thiele en 1983 y 1986, “debido a modificación espectral incompatible entre curva de campo libre y HRTF de posiciones distintas a 0°” (Begault, 2000, p.125).

Para ecualización de campo difuso (*diffuse field equalization*), la compensación se basa en el hecho que, si se reciben tanto sonido directo como indirecto, el campo sonoro se vuelve aleatorio o difuso, en su dirección de incidencia con respecto al oyente, al acercarse y sobrepasar la distancia crítica. Así, la reproducción es óptima para varias direcciones (aleatorias) de incidencia de la fuente.

Ya que existen audífonos disponibles con dicha ecualización, aplicarlas a las HRTF dependerá de cuales se utilizarán. En el caso de emplear audífonos con ecualización de campo libre, es posible implementar la ecualización de campo difuso a los impulsos. Para esta técnica, la curva de ecualización se obtiene al generar una HRTF promedio de las mediciones, donde se consideran todos los ángulos de incidencia.

Es posible una tercera etapa de ecualización a elección del diseñador del sistema y se debe principalmente al tipo de material que será empleado en el sistema como es el caso de programas musicales, donde se puede preferir una

respuesta en bajas frecuencias más pronunciada que la que entrega un micrófono típico utilizado en grabación de HRIR. Un ejemplo es el caso del software Binaural Panning desarrollado por Deepak Natarajan para la Universidad de California Davis, que emplea un resonador para añadir bajas frecuencias que no se obtienen en el proceso de convolución empleado en el sistema.

1.6. Síntesis Binaural o Sonido 3D

Debido a que el tamaño y la forma de todas las partes encargadas de entregar información al cerebro sobre la localización de los sonidos varían de persona a persona, se busca encontrar modelos de síntesis de sonido que funcionen de manera general para lograr ubicar a una fuente virtual dentro de un espacio tridimensional convincente, que resulta un reto, al tratar de lograr controlar de la percepción auditiva de otra persona (el usuario del sistema) y lograr de ésta forma, optimizar recursos de procesamiento.

Para un sistema de sonido 3D, el control del sonido para ubicación de las fuentes no solo deberá considerar fenómenos físicos, si no psicoacústicos involucrados en la percepción auditiva, para el procesamiento de señales de audio que pueden dar lugar a la generación desde espacios virtuales o reemplazo de existentes, hasta la creación de realidades acústicas totalmente novedosas.

También existe el problema del tamaño o “Ancho Aparente de la Fuente” (*Apparent Source Width*) que ocupará la fuente dentro del espacio generado por el sistema, conocido como *Auditory Spaciousness*. Si bien no existe una manera de determinarlo o una relación para establecerlos dentro de un sistema virtual, están fuertemente relacionadas con el contexto del entorno y los efectos de la reverberación en la percepción auditiva.

Los problemas con las pruebas psicoacústicas o el conocimiento que se tiene de las mismas radica en que para localización, se presentan claves como picos o valles en el espectro, dependiendo de la dirección de la fuente. Por lo que, para sonidos de anchos de banda angosta, la percepción de la ubicación puede

verse alterada en base a su frecuencia central. Dependiendo del tamaño del pabellón (que varía de persona a persona), la modificación espectral también depende de la persona.

La mayoría de contribuciones establecidas mediante estudios y necesariamente mediante pruebas y test psicoacústicos a individuos (pues se considera fenómenos perceptibles que van más allá de la teoría), también sugiere un prejuicio o tendencia (*bias*) en la prueba.

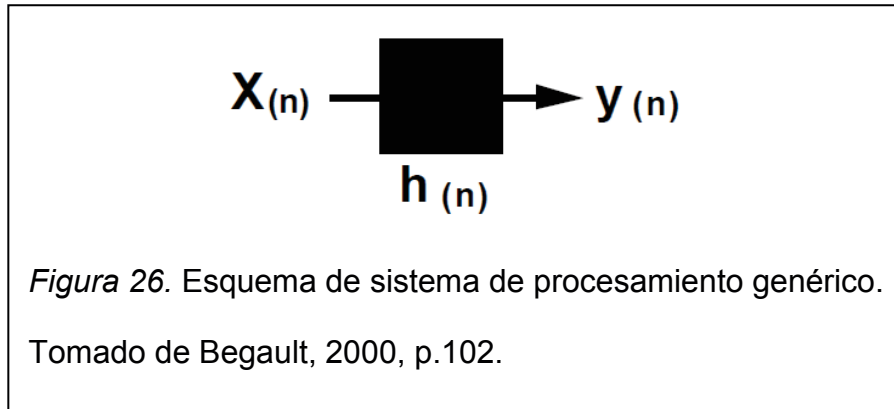
Mientras más específica resulte la prueba, más alejada de la realidad resulta. Esto debido a que al tratar de controlar el espacio, eliminando las variables asociadas a éste de forma natural, las conclusiones son “relevantes en la medida que dilucidan los mecanismos específicos de la localización e indican habilidades de éstos mecanismos en condiciones controladas” (Begault, 2000, p.26).

La opción de realizarlos desde un punto de vista personal del usuario, dando la opción de ubicar a la fuente sonora en un campo a su alrededor. “Según algunos estudios psicoacústicos, el punto de referencia está localizado en un punto de origen entre los oídos, aproximadamente al nivel de los ojos y al centro de la cabeza” (Begault, 2000, p.2).

Además, es común el asumir fuentes omnidireccionales, para la implementación de la ley del inverso cuadrado (encontrada por lo general en sistemas de sonido 3D).

1.6.1. Procesamiento Digital

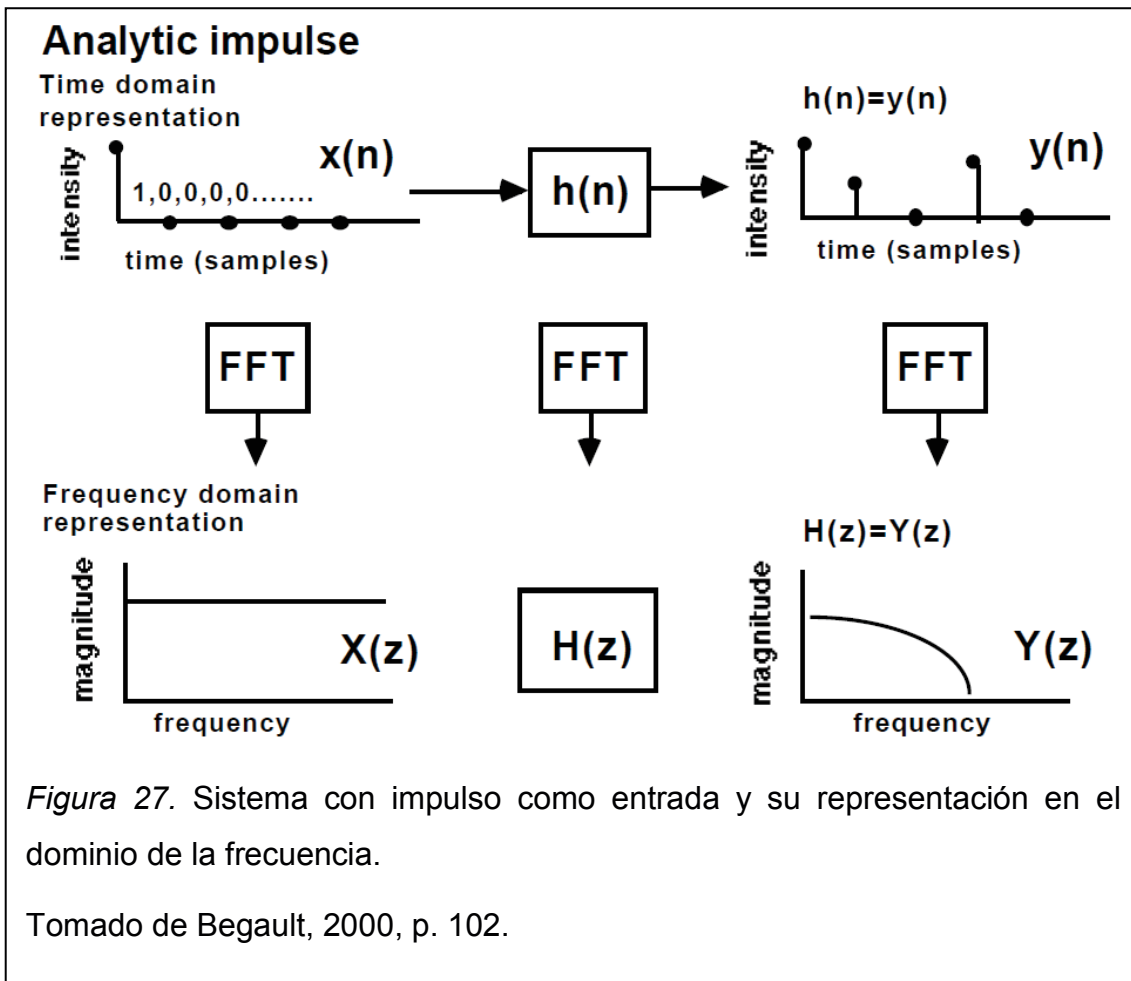
Una señal digital cualquiera se compone de una secuencia de n muestras, que tienen un valor asociado $y(n)$, para cada instante donde exista una muestra (que dependerá de la frecuencia de muestreo).



1.6.1.1. Señal Impulso

En un sistema, la señal $x(n)$ que ingresa (o señal de entrada) es afectada por $h(n)$ en su camino hacia la salida $y(n)$. Si se desea conocer cómo afecta (que representa $h(n)$) a la señal de entrada, es necesario alimentar al sistema con un impulso. Si este es afectado o modificado al llegar a la salida, se puede determinar el efecto o modificación que sufre, gracias a que se conoce la señal de entrada.

La señal impulso es conocida y caracterizada en el dominio del tiempo como una señal cuya primera muestra posee un valor de 1 seguido por muestras de valor 0, y al emplear la Transformada de Fourier, presenta una respuesta en frecuencia plana en todo el espectro.



1.6.1.2. Filtrado de Señales

Una de las herramientas del procesamiento de señales más empleadas son los filtros. Estas se emplean para eliminar o aislar el contenido frecuencial de una señal o bien, para dividir el espectro de la señal en rangos constantes.

Los filtros pueden ser analógicos o digitales y existen varios tipos de éstos, cuyo comportamiento varia y son empleados de acuerdo con la aplicación.

En su diseño, es posible definir a que rango del espectro afectan (las frecuencias que lo determinan) y cómo influyen en el mismo.

1.6.1.3. Transformada Discreta de Fourier (DFT)

La Transformada Discreta de Fourier (DFT) es la función matemática que representa el proceso de transformar una señal del dominio del tiempo al de la

frecuencia mediante la transformada de Fourier de una secuencia finita y periódica de valores de una secuencia total de N números (donde solo se definen valores donde existen muestras).

Para una señal digital discreta secuencia de duración finita $x(n)$, de longitud N (hasta N-1 sus valores son distintos de 0) se puede utilizar secuencia periódica con periodo N y describirla entre $0 \leq k \leq N-1$ como una serie de Fourier:

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j\left(\frac{2\pi kn}{N}\right)} \quad (\text{Ecuación 5})$$

1.6.1.4. Convolución

Señales pueden ser expresadas como funciones que generan una nueva mediante su convolución, por lo que este proceso puede ser descrito matemáticamente como la “integral que representa la cantidad de superposición entre una función g mientras cambia a otra función f” (Weisstein, s.f.).

Matemáticamente se describe en un producto de funciones para un rango finito (0, x):

$$f(x) * g(x) = \sum_0^x f(x)g(u - x)dx \quad (\text{Ecuación 6})$$

Donde:

*: Convolución de funciones.

Un sistema invariante en el tiempo, donde k es un entero positivo o negativo y n representa las muestras en el tiempo de la señal, “está completamente caracterizado por su respuesta al pulso unitario $h(n)$ ” (De La Fraga, 2001). En una secuencia digital, x se reemplaza por n y las funciones ahora son $x[n]$ y $h[n]$ en un rango entre (0, n), expresado como:

$$y[n] = x[n] * h[n] = \sum_0^n x[n]h[n - k] \quad (\text{Ecuación 7})$$

La transformada de Fourier resultante de la suma de las transformadas (suma de convolución) representa la “respuesta en frecuencia del sistema” (De La

Fraga, 2001) y para una secuencia expresada como una exponencial compleja, se define como:

$$Y(e^{j\omega}) = H(e^{j\omega}) X(e^{j\omega}) = H(Z) X(Z) \quad (\text{Ecuación 8})$$

Donde:

$H(e^{j\omega})$: Transformada de Fourier (Z) de la respuesta a la muestra unitaria $h(n)$.

$X(e^{j\omega})$: Transformada de Fourier (Z) de secuencia de entrada compleja $x(n)$.

La escucha binaural en un sistema de síntesis de sonido 3D, busca implementar los cambios espectrales y temporales que son captados en las HRTF mediante filtrados de señales, tal como lo hace el pabellón y el resto del cuerpo en la realidad.

Lo que dicho filtrado logra es “multiplicar los espectros de dos formas de onda, lo que es equivalente a convolución en el dominio del tiempo de las formas de onda” (Begault, 2000, p.106), lo que resulta matemáticamente equivalente.

La convolución para señales binaurales se puede llevar a cabo en el dominio del tiempo, conocida como convolución discreta, que emplea los impulsos respuesta de la cabeza o HRIR, o bien en el dominio de la frecuencia, al pasar dichos impulsos a funciones de transferencia (HRTF) mediante Transformada rápida de Fourier FFT (*Fast Fourier Transform*). El proceso puede o no ser realizado en tiempo real.

“Si se considera a las HRIR como las salidas de un sistema lineal e invariante en el tiempo (...) entonces, en teoría, pueden ser representadas por filtros no recursivos de respuestas finitas (*Finite Impulse Response, FIR*)” (Araneda et al., 2010, p.2). Por lo que es posible esta aproximación, así como mediante el empleo de filtros IIR obteniendo sus coeficientes en base a una respuesta impulsiva.

Para convolucionar una señal infinita de entrada (idealmente obtenida en una cámara anecoica) con una señal impulsiva (HRIR), se emplea el método

solapamiento-suma (*overlap-add*), que divide a la señal infinita para convolucionarse con el impulso.

En la convolución discreta de una señal de entrada $x[n]$ de N muestras y un impulso $h[n]$ se realiza la convolución de ambas señales y se requiere de N multiplicaciones por cada punto de salida. En la convolución por FFT se emplea un algoritmo DFT (Transformada Discreta de Fourier) y se realiza para un segmento (grupo de N muestras) la transformada. Una vez en el dominio de la frecuencia se produce la multiplicación compleja de los espectros de las señales, y se los regresa al dominio del tiempo con la DFT inversa.

“La convolución discreta no es una implementación viable cuando se requieren resultados en tiempo real. La implementación de la convolución por FFT es un 72.5% más rápida, produciendo exactamente los mismo resultados” (Araneda et al., 2010, p.8). Filtros IIR binaurales llegan a ser inclusive más efectivos (más rápidos y menos carga al procesador) que la técnica por FFT en rendimiento, pero carece la precisión de ésta última.

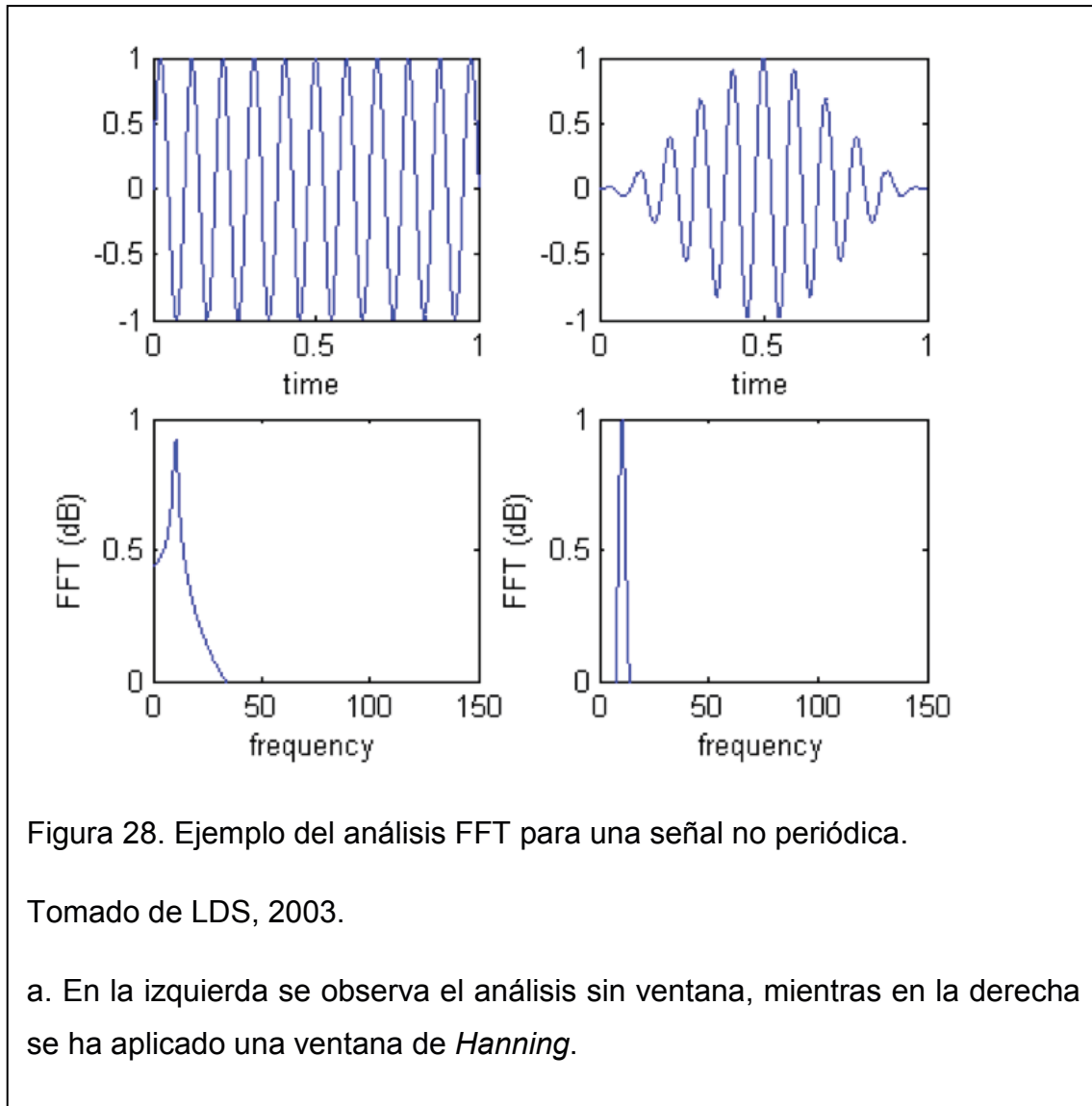
1.6.1.5. Función Ventana

Fuga o filtraje espectral, del término en inglés *leakage*, denota un error típico del análisis FFT de ondas no periódicas, resultando en una lectura incorrecta de la señal analizada y presentando el contenido espectral (amplitud y frecuencia) “expandido” o “disperso” en zonas en las que la señal no posee energía.

Al analizarse la señal de entrada por bloques de datos (*frames*), es posible aplicar una función denominada ventana, que fuerza a que el primer y último valor sea cero en cada *frame* (analizados uno a la vez), asemejándolos a los de una onda periódica que se espera en un análisis mediante DFT.

Así se evita las discontinuidades entre *frames* y se obtiene una representación más fiel de la señal analizada. Esta ventana “se multiplica luego con el bloque de datos (...) un factor de ponderación especial también debe aplicarse de

modo que el nivel de amplitud correcto de la señal FFT se recupera después de aplicar la ventana” (LDS, 2003).



La desventaja de aplicar estas ventanas (proceso denominado *windowing*) es de la atenuación que se produce en la señal. Mediante un proceso denominado *overlapping*, se puede aprovechar el análisis realizado para el *frame* anterior y utilizar una porción del mismo, sobreponiéndolos en la salida con el siguiente, evitando distorsión por dicha atenuación y sin necesidad de calcular el mismo bloque varias veces.

Las funciones controlan pero no eliminan por completo el problema, por lo que existen varios modelos propuestos y cuya parte intermedia (entre los ceros inicial y final del *frame*) difieren entre sí. Distintas ventanas se recomiendan en base a su aplicación: tipo de señal y precisión de que se espera obtener del análisis.

Tabla 1. Comparación de Ventanas en base a distintos criterios de acuerdo a LDS (2003).

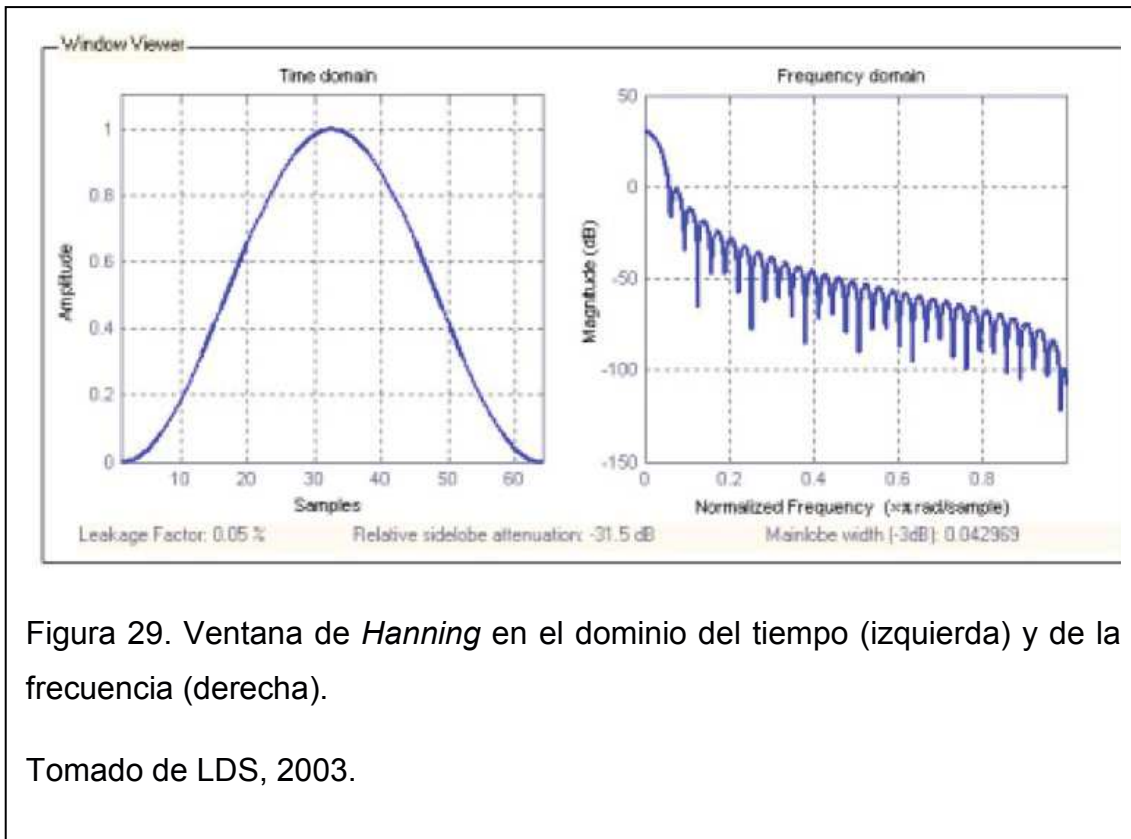
Window	Best for these Signal Types	Frequency Resolution	Spectral Leakage	Amplitude Accuracy
Barlett	Random	Good	Fair	Fair
Blackman	Random or mixed	Poor	Best	Good
Flat top	Sinusoids	Poor	Good	Best
Hanning	Random	Good	Good	Fair
Hamming	Random	Good	Fair	Fair
Kaiser-Bessel	Random	Fair	Good	Good
None (boxcar)	Transient & Synchronous Sampling	Best	Poor	Poor
Tukey	Random	Good	Poor	Poor
Welch	Random	Good	Good	Fair

Tomado de LDS, 2003.

a. Se comparan los tipos de ventana de acuerdo al Tipo de Señal y se evalúa su rendimiento en Resolución en Frecuencia, Filtraje Espectral y Precisión en Amplitud.

a) Ventana de *Hanning*

Entre las funciones comunes se encuentra la Ventana de *Hann* o *Hanning*, recomendada para señales aleatorias y con un rendimiento generalmente bueno de acuerdo a LDS (2003) o para “ondas sinusoidales o combinación de ondas sinusoidales” (National Instruments Corporation, 2013).



En un análisis espectral de una ventana, se observa un pico en la frecuencia aplicada (de mayor nivel) y varios picos de menor amplitud a sus lados, denominados lóbulos laterales (*side lobes*). “Las ventanas más estrechas en el dominio del tiempo tienen lóbulos principales más amplios en el dominio de la frecuencia, y viceversa” (National Instruments Corporation, 2013). La ventana de *Hanning* en el dominio del tiempo “tiene la forma de un ciclo de una forma cosenoidal, a la que se agrega 1 para que así siempre sea positivo” (White, 2010) y su función se expresa matemáticamente como:

$$W(x) = \frac{1}{2} \left(1 + \cos \frac{\pi x}{a} \right) \quad (\text{Ecuación 9})$$

Donde:

a: distancia entre dos puntos de la curva donde la función alcanza la mitad de su valor máximo.

También se recomiendan su uso para casos donde “no se conoce la naturaleza de la señal pero se desea aplicar una ventana suavizante” National Instruments Corporation (2013). “La ventana de *Hanning* es satisfactoria el 95% de los casos. Tiene una buena resolución en frecuencia y filtrado espectral reducido” (National Instruments Corporation, 2013).

1.6.2. Simulación de Distancia y Externalización

La percepción de la distancia es la menos precisa y la de mas difícil implementación. Para un sistema de sonido 3D que se basa en HRTF obtenidos en cámara anecoica donde no existe Externalización, es posible determinar distancia y dirección al contar únicamente con la fuente presente, pero al sumarse a ésta las reflexiones del espacio donde se encuentra “el oyente obtiene la experiencia sónica total” (Moller, 1992, p.176).

Una manera de lograr una simulación simple de distancia es la de escalar la salida binaural mediante un control de ganancia global. El decremento en la señal dependerá de principio se emplea para determinar dicha disminución, sea ley del inverso cuadrado o una escala perceptual y su control inclusive, no deberá indicar la cantidad de unidades de la escala que se disminuyen necesariamente, sería posible implementar una indicación de cómo es percibida la disminución al manipular el control de ganancia.

El incluir un modelo de reverberación además de las HRTF para obtener una sensación de Externalización y contexto ambiental “provee un sentido de contexto ambiental que rodea al oyente y la fuente y puede entregar al oyente una imagen de la extensión de la fuente” (Begault, 2000, p.139). Sin embargo, “la percepción de las fuentes virtuales son afectadas significativamente (...) y pueden causar deterioración de la precisión en la localización de acimut y elevación” (Begault, 2000, p.69).

1.6.3. Auralización

En caso de realizar convolución de HRIR con impulsos respuesta de distintas salas, se obtienen impulsos binaurales de las mismas, en un proceso conocido como Auralización, que es una representación artificial de un campo sonoro para escucha binaural en cualquier punto de dicho campo donde se ubica una fuente sonora. El resultado, denominado a veces como “reverberación espacial” permite obtener una simulación del contexto ambiental.

La medición de los impulsos de salas (a veces mediante un *Dummy Head*) puede ser reemplazada por la síntesis o modelado de los mismos. Ambos procesos, tanto medición como síntesis cuentan con desventajas.

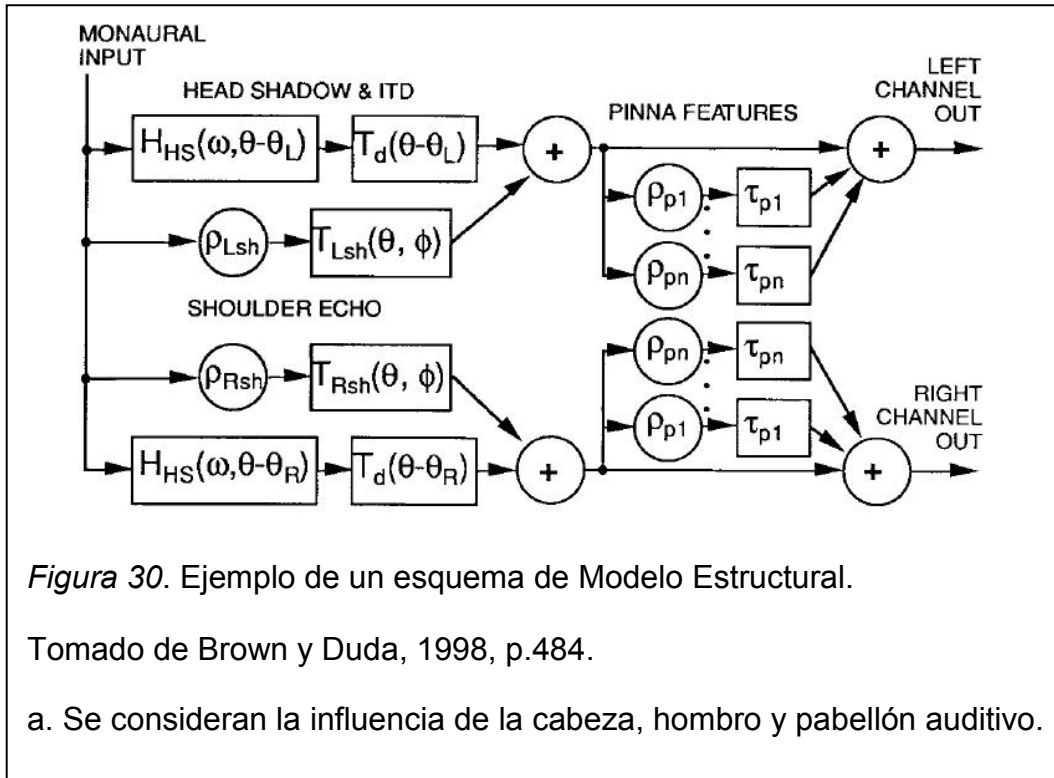
1.6.4. Moldeamiento Estructural de HRTF

La idea de modelar los claves espectrales presentes en las HRTF aparecen en la década de 1960, inicialmente implementada mediante filtrado analógico de señales a fines de la década de 1970 y posteriormente empleando filtros digitales a inicios de la década de 1980 por distintos investigadores.

Posteriormente, se han planteado varias formas de modelar HRTF, como Modelos *Pole/Zero*, Expansión de Series (Análisis de Componentes Principales) y Modelos Estructurales (Duda, 2011).

Éstos últimos, tratan de emplear funciones de transferencia que simulen la contribución de cada mecanismo que influye en la escucha binaural, desde modelos sencillos y limitados hasta modelos más complicados, resultado de la combinación de modelos o de experimentar con otras formas de mayor o menor complejidad como el “cabeza esférica”, “cabeza elipsoidal”, “*snowman*”, etc., e inclusive añadiendo modelos que contribuyan con la función de hombros (*shoulder reflection*) y torso (*torso diffraction*) en la escucha.

De acuerdo a Brown y Duda (1998, p.476) las ventajas del modelo es la de que cada componente representa y es responsable de la simulación del fenómeno físico, su implementación es mas económica y puede realizarse en tiempo real, es posible relacionar parámetros con medidas antropométricas.



Las aproximaciones, desde las más simples, pueden ser muy eficaces debido a su sustento físico y teórico, además de ser ligeros en el procesamiento, aumentando su capacidad (y carga) de simulación de condiciones realidad al aumentar la complejidad del modelo. Se explica brevemente algunos de los modelos estructurales básicos para síntesis de sonido binaural.

1.6.4.1. Modelado de ITD

Se puede implementar con un filtro FIR. Introduce un retardo de tiempo entre los oídos (diagonalmente opuestos) al mover la fuente. Si bien produce movimiento de la fuente en el plano horizontal (entre -90° y $+90^\circ$), carece de Externalización y presenta confusión frontal/trasera. Una sombra producida por la cabeza puede arreglar el efecto de dos fuentes aparentes (la energía sigue siendo la misma en ambos oídos).

1.6.4.2. Modelado de ILD

Se puede determinar con un filtro IIR obtenido partir de la solución analítica de ILD para una esfera rígida (Lord Rayleigh). La función de transferencia (*one*

pole / one zero), cuando la fuente se encuentra justo al frente, produce un énfasis en altas frecuencias que será opacado (filtrado) al ubicarse la fuente justo detrás como el producido por la sombra de la cabeza y produce el movimiento de la fuente al compensar el acimut a las posiciones del oído. Presenta los mismos problemas que el modelo de ITD además de introducir un *group delay* en bajas frecuencias que recae en el problema de *split image*.

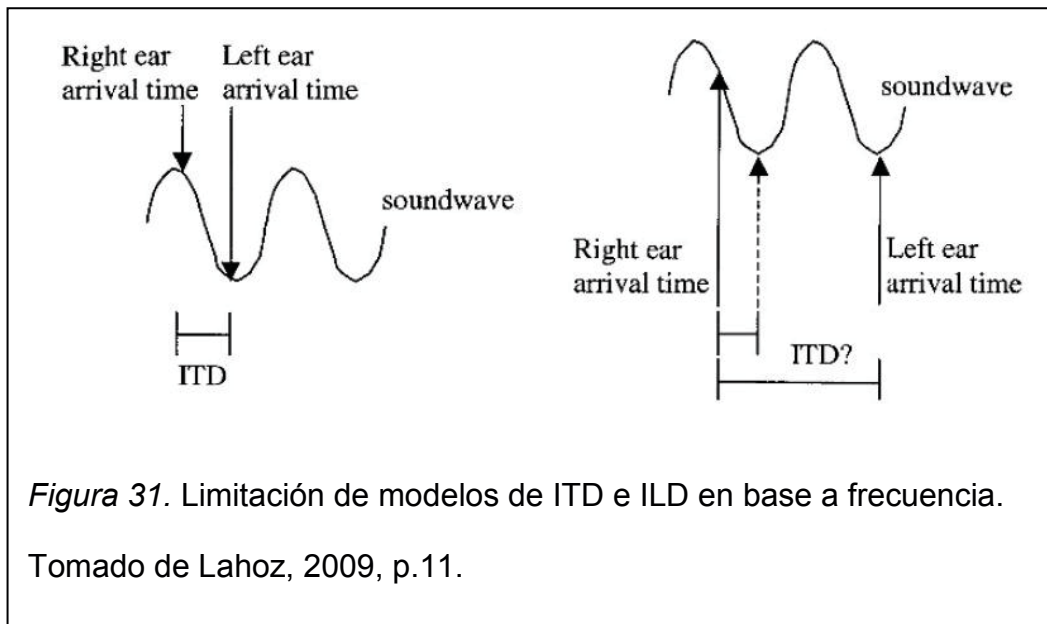
1.6.4.3. Teoría Dúplex y Modelo Estructural de Cabeza Esférica

La mejor forma de disminuir los problemas (*split image*) y generar un modelo más completo y eficiente es combinar el modelo de ITD e ILD (en serie) en un modelo de Cabeza Esférica.

De ésta manera, la fuente se mueve de forma más realista y se mantiene al modelo simple. Además, es posible emplear un modelo sencillo de reverberación para añadir la Externalización necesaria.

Tanto ITD como ITL, asumiendo un modelo de cabeza esférica y perfectamente simétrica, fueron planteados por Lord Rayleigh en su “Teoría Dúplex” en 1907 de forma combinada. Debido a la dependencia de la frecuencia de cada uno, de ésta forma, es posible poder usarlos de efectivamente cubriendo todo el espectro audible.

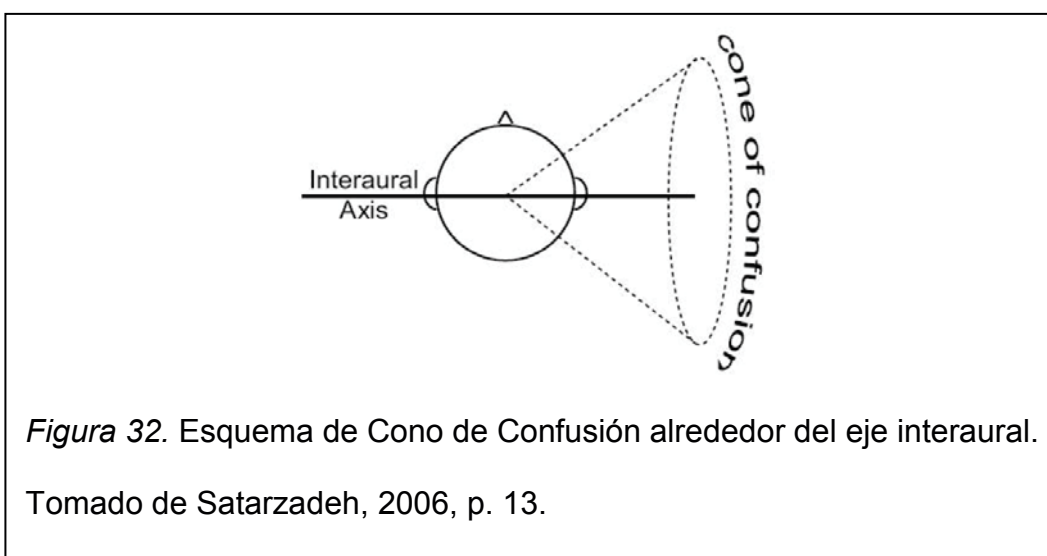
Para altas frecuencias, la longitud de onda del sonido incidente es menor con respecto al de la circunferencia de cabeza, y mayor para bajas frecuencias, donde la sombra producida por la cabeza es mínima y poco influyente. Esto resulta en una ITD poco efectiva para sonidos con componentes espectrales en altas frecuencias y una ILD poco efectiva para sonidos con componentes en bajas frecuencias, siendo aproximadamente 1500 [Hz] como el límite aceptado del rango de validez para ambos modelos.



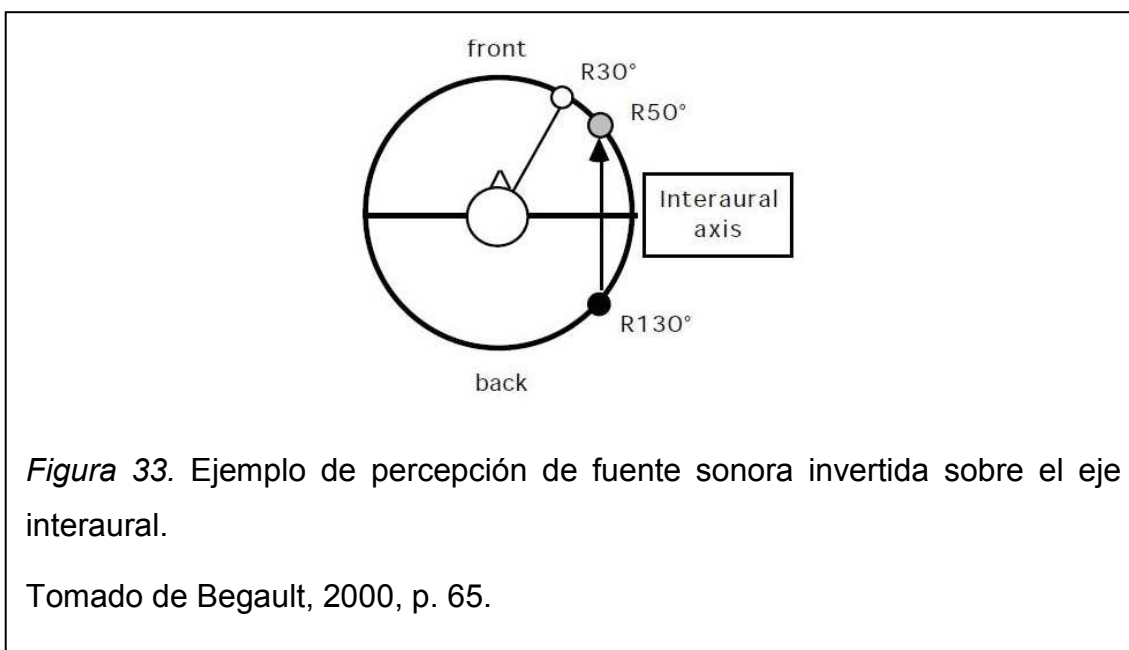
La Teoría Dúplex presenta ciertos problemas debidos a su modelización. Sonidos provenientes directamente desde el frente y desde atrás, no presentan diferencias interaurales, lo que genera un problema conocido como “Confusión Frontal/Trasera”.

1.6.4.4. Cono de Confusión

El cono de confusión se presenta al asumir un modelo de cabeza esférica, por lo que no representa una situación real ya que la cabeza nunca será una esfera. Fue presentado por primera vez por Hombostel y Wertheimer en 1920.



Para todos los puntos del cono de confusión, es decir, donde existan ITD e ILD constantes y sean simétricas alrededor del eje interaural, se produce una confusión de precedencia del sonido. En el caso de un modelo esférico, se presenta para todos los puntos en el cono, generando una confusión al oyente que se puede percibir como si la fuente está atrás en lugar de adelante, arriba en lugar de abajo, etc. A éste problema también se lo conoce como *reversals*, pues crea una imagen invertida de la ubicación real de la fuente.



Sin embargo, ésta teoría es lo suficientemente útil como una aproximación válida de ubicación de fuentes en el plano horizontal (no introducen ningún indicador en las otras dimensiones), al considerar ITD e ILD para distintos rangos de frecuencia donde los obstáculos no representan un problema, evitando el problema del cono de confusión.

1.6.4.5. Modelo de Cabeza Elipsoidal

El modelo de cabeza puede tomar una forma elipsoidal para aproximarse a situaciones más realistas, eliminando los errores producidos por modelos esféricos, al considerar que los oídos no se ubican en el eje horizontal exacto al medio de la cabeza pero ligeramente atrás y abajo del mismo.

De ésta manera, el ITD varia alrededor del cono de confusión, mejorando la percepción en el plano horizontal ya que, "ITD es una función de la elevación como del acimut" (Algazi, Avendaño y Duda, 1999, p.1). Esta dependencia de la elevación se debe que la distancia entre la fuente y el oído ipsilateral permanece constante, la distancia con el oído contralateral varia con la elevación de la fuente.

1.6.4.6. Modelo Estructural Monoaural del Pabellón Auditivo

La función del pabellón como dispositivo que capta las ondas sonoras ha sido descrita de distintas formas y por varios autores, como en los trabajos de Algazi, Avedano y Duda (2001), Monte, Ramirez y Vallez (2003) o Satarzadeh (2006) por nombrar algunos, pero un modelo exacto que permita simular su contribución en la escucha binaural todavía no ha sido definida. Su papel es fundamental en la localización de fuente en el plano vertical, por lo que se han sugerido modelos considerando los ecos que se producen en su superficie.

De acuerdo al modelo propuesto por Monte y otros (2003, p.42) es necesario únicamente cinco ecos, cada uno con un tiempo de retardo propio pero con coeficientes de ganancia que afirman, pueden permanecer constantes en base a tests auditivos realizados.

Tal como sucede en la audición misma, en sistemas digitales para sonido binaural, el control y simulación de acimut es fácilmente alcanzable, lo que no puede afirmarse para el control y simulación de elevación y distancia, que todavía presentan limitaciones.

1.6.4.7. Modelo de Entorno Acústico (*Room Model*)

Es posible generar un modelo simple que simule la acústica del entorno para introducir Externalización adicional y que trabajen en conjunto con modelos que carezcan de ésta cualidad. Lo que se tiene en cuenta al modelar la reverberación, son cinco o seis de las reflexiones producidas en los límites del entorno, su cantidad y el nivel de las mismas, que dependerán de las características acústicas de los acabados o materiales en los que se refleja el

sonido. El uso de filtrado de altas frecuencias para las reflexiones consiguen una aproximación todavía más realista, al conseguir captar la pérdida de energía en dichas frecuencias.

Si bien el método no se aproxima a una condición completamente real pues busca modelar únicamente reflexiones tempranas (siendo estas solo una parte de la reverberación, aproximadamente dentro de los primeros 80 [ms]), el método resulta eficaz y puede ser complementado por algoritmos para modelado de reflexiones tardías más complejos (imposibles de realizarlos mediante este método).

Modelos de reverberación basados en impulsos-respuesta representan alternativas más complejas y realistas, al considerar características espectrales como ángulos de incidencia entre límites y el oyente, orientación y dirección de las reflexiones y temporales: distancia y atenuación de las reflexiones.

No importa el modelo, el objetivo es evitar terminar con un algoritmo capaz de ubicar una fuente en un plano bidimensional pero en condiciones anecoicas. Existe la condición de establecer un límite, si bien no se conoce exactamente cuál es, de la cantidad de reverberación que es necesaria de agregar a un sistema de síntesis 3D.

Exceso de reverberación puede disminuir la precisión del algoritmo de localización horizontal y vertical, al “esparcir” las señales sintetizadas con las reflexiones.

1.6.5. Movimientos de Cabeza (*Head-Tracking*)

Estos movimientos pequeños e insignificantes proveen claves de localización de fuentes. Un reflejo natural e inconsciente es el del oyente que tiende a “apuntar” con la cabeza hacia la fuente al escuchar un sonido, minimizando diferencias interaurales, evitando confusiones y cerciorándose de su ubicación.

El tener en consideración cada movimiento de la cabeza del usuario, forma parte de los sistemas de síntesis binaurales más ambiciosos. No solo es necesario un dispositivo capaz de calcular e informar en tiempo real al software

la posición relativa de la cabeza (*head-tracking*), pero requiere de una gran capacidad de procesamiento para llevar a cabo dicha tarea en tiempo real.

Sin embargo, el hecho de contar con un modelo que tenga en cuenta estos movimientos, se encargara de forma sencilla de eliminar problemas como confusión frontal/trasera, *reversals* e inclusive mejoría en percepción de distancia o problemas de fuente aparentemente dentro de la cabeza, causados por el uso de audífonos.

1.7. Entorno de Programación MAX/MSP

El software MAX/MSP es un lenguaje de programación orientado a objetos ideal para el desarrollo de aplicaciones de audio y video. Es posible la creación de *patches* donde objetos son colocados, sean los incluidos en el paquete como los desarrollados por otros usuarios (*externals*), son conectados mediante cables (*patchcords*) para crear una red o sistema de procesamiento de datos (MAX) o de audio (MSP) o sistemas híbridos complejos.

MAX/MSP funciona como un entorno modular, por lo que dentro del *patch* principal, pueden existir un o varios *subpatch* que realizan procesos de forma simultánea y contienen objetos encapsulados en sí mismos, lo que pueden facilitar el arreglo en la programación. Los *subpatches* deben llevar objetos llamados *inlets* y *outlets* (entradas y salidas) para recibir y enviar mensajes al *patch* u otros *subpatches*.

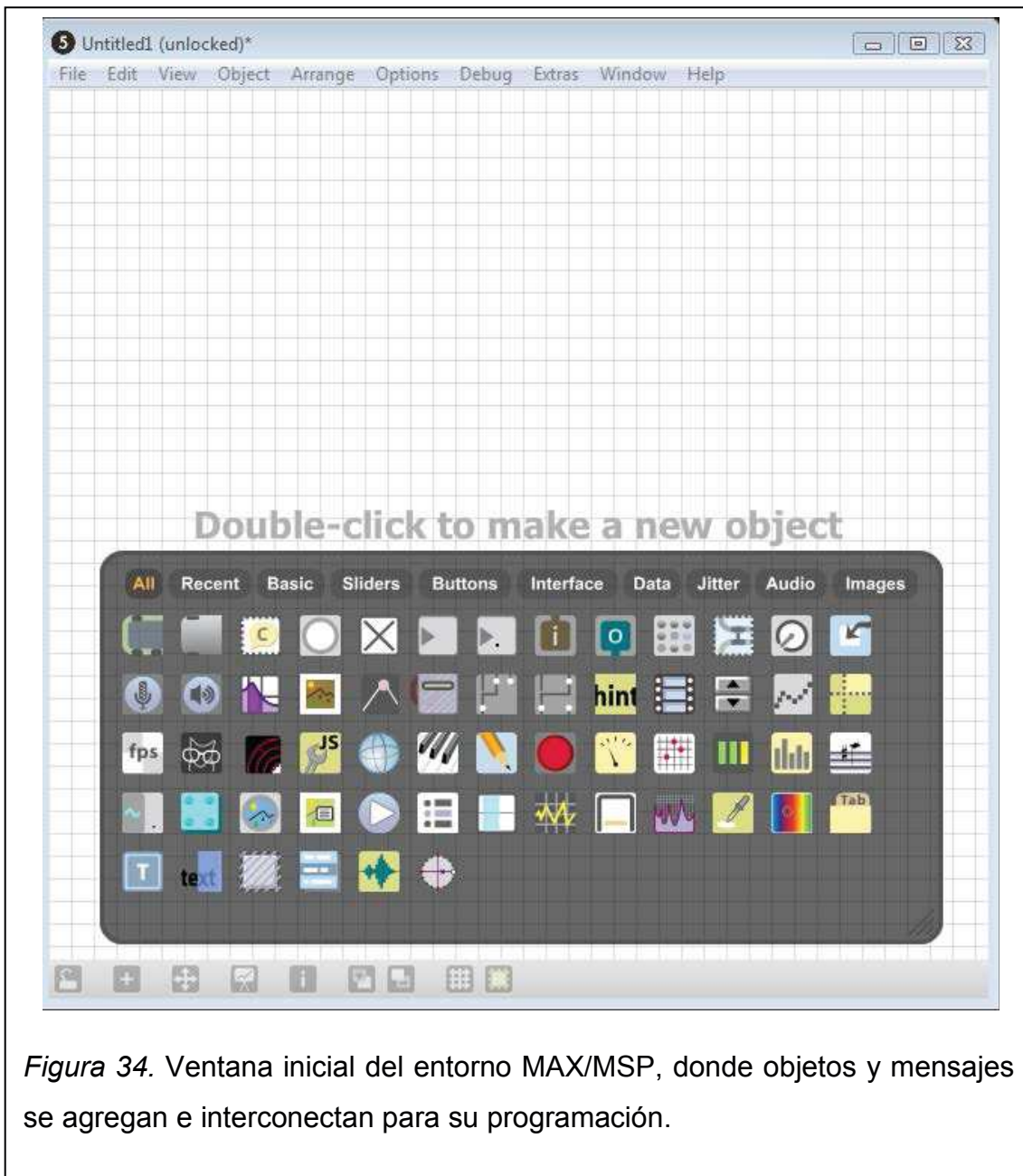


Figura 34. Ventana inicial del entorno MAX/MSP, donde objetos y mensajes se agregan e interconectan para su programación.

Los objetos de MAX y MSP se diferencian en que los últimos llevan el símbolo “~” al final del nombre. Estos pueden incluir argumentos después de su nombre que especifican ciertas propiedades que pueden ser necesarias (dependiendo del comportamiento necesario). El inspector permite modificar ciertos atributos de los objetos.

Los objetos tienen varias entradas, ubicadas en su parte superior y salidas, ubicadas en su parte inferior, que dependen de su función y de sus

argumentos. Para objetos MSP, el *patchcord* en su salida es de color verde, lo que indica que lleva información de audio.

Los mensajes llamados *symbol* (símbolos) también son instrucciones específicas para cada objeto, y pueden ser generados como una casilla *message* o bien, en la salida de otros objetos, como es el caso del mensaje “Bang” funciona como un *trigger* para que el objeto que lo recibe, genere un mensaje en su salida, común para muchos objetos en el entorno y de gran utilidad.

1.7.1. Configuración de Entradas y Salidas de Audio en MAX/MSP

Los objetos *adc~* y *dac~* o sus equivalentes, *ezadc~* y *ezdac~* respectivamente, funcionan como interfaz entre el software y el hardware mediante *drivers*.

Los *drivers* en MAX son objetos especiales que contienen el código necesario para comunicarse con el hardware de sonido. También se encarga de manejar las entradas y salidas lógicas (en *MSP* mediante los de los objetos mencionados) y las entradas y salidas físicas integradas del computador, o del hardware empleado. Cualquier entrada o salida lógica puede ser asignada a una entrada o salida física en un proceso denominado *mapping*, que puede llevarse a cabo en la ventana *I/O Mappings*.

El tipo de *driver* depende de la plataforma (Windows o Macintosh) y se establece en la ventana denominado *DSP Status* en el menú “*Options*”. Una vez seleccionado el *driver*, que quedara configurado por defecto, es posible el procesamiento de señales en redes *MSP*.

Para plataformas Macintosh, como el empleado para el desarrollo del software, los *drivers* disponibles por defecto son denominados *Core Audio*. Al seleccionarlos en MAX/MSP, las entradas y salidas lógicas se comunican con las de la interfaz de audio integrada del computador.

Es posible emplear *drivers ASIO* en sistemas Macintosh cuando ciertas interfaces como hardware externo, proveen el software para su funcionamiento en éstas plataformas debido a su incompatibilidad con *Core Audio*.

Para procesamiento en tiempo real, el software depende de las capacidades del computador y del procesamiento que sea capaz de manejar su *CPU*. Al emplear tarjetas de sonido o interfaces de audio como unidades hardware externas, la capacidad dependerá de sus características.

Sin embargo, existe la posibilidad de realizar procesamiento que no se realiza en tiempo real, realizándolos en el disco duro, liberando de esta manera la carga del procesador si este no es capaz de manejarla. Para esto existen los *driversad_nonreal*.

También es posible implementar los *driversad_rewire*, que permite el envío y recepción de audio y *MIDI* bajo el estándar *Rewire*, desarrollado por Propellerhead Software para su comunicación con otras aplicaciones de audio que soporten dicho estándar.

Dentro de la ventana *DSP Status* es posible la activar y desactivar audio, profundidad de muestras y frecuencia de muestreo, así como monitoreo estimada de procesamiento es utilizado por el *CPU* (en *CPU Utilization*) y la posibilidad de establecer un límite, en porcentaje, del uso de *CPU*, en *CPU Limit*).

1.7.2. Manejo de Audio En MAX/MSP

MSP maneja el audio de entrada, los cálculos necesarios para procesamiento de señales, y las salidas de audio en bloques de muestras denominados vectores. Es necesario asignar un valor a dichos bloque en la ventana *DSP Status* para que el comportamiento del programa sea óptimo dependiendo del resultado deseado y los objetos empleados.

El primer bloque hace referencia al número de muestras que se toman y se envían a la interface de audio, sea ésta la integrada en el computador o una externa, denominado *I/O Vector Size* (Tamaño de Vector de Entrada y Salida). El tamaño de este vector influye en la latencia y el rendimiento del software, pero no influye en la calidad del audio producido.

Valores pequeños (bloques de muestras pequeños) producen menor latencia entre entrada y salida pero con más carga al procesador. Esto debido a que se realizan cálculos para bloques pequeños, pero es el número de vectores que tienen que ser calculados es mayor.

Para valores altos (bloques de muestras mayores), el computador tiene que interrumpir cualquier otro proceso que se encuentre llevando a cabo para realizar el cálculo, y reanudarlo una vez que se ha completado, lo que MAX/MSP, denomina como "*interrupts*". Si el tamaño del vector es muy alto, el computador puede ralentizarse considerablemente hasta finalizar todos los cálculos necesarios, poniendo en espera a todo el resto de procesos fuera de MAX que debe llevar a cabo, forzando sus capacidades.

Por ésta razón, debe encontrarse un compromiso entre el uso del CPU en base al procesamiento necesario y su capacidad (además de las capacidades y necesidades del computador) y el procesamiento necesario para el funcionamiento del *software* o *patch*.

El segundo vector o bloque se denomina *Signal Vector Size* y determina cuantas muestras de la señal se procesan o calculan en los objetos MSP a la vez, por lo que no pueden ser mayores al valor de *I/O Vector Size*. Éste no tiene un efecto en la latencia pero si en el desempeño, dependiendo de los objetos externos utilizados y las limitaciones que presentan éstos (generalmente especificadas por sus autores).

1.7.3. Manejo de Archivos en MAX/MSP.

Para el manejo de archivos incluidos en el paquete del software y cualquier otro archivo externo necesario en el diseño de un *patch*, MAX/MSP dispone de una ruta predeterminada denominada *Search Path*, en la cual se determinan las carpetas que contienen dichos archivos y objetos y así especificar sus rutas cada vez que sea necesario llamarlos en un *patch*.

Las carpetas que sean necesarias se pueden añadir en el *Search Path*, al añadir la ruta de la misma en el disco (dependiendo de la plataforma) mediante

en la ventana *File Preferences* en el menú *Option*, donde los objetos y librerías necesarias para el funcionamiento del software se encuentran incluidos por defecto.

2. Desarrollo del Software

2.1. Prototipo del Software

El software desarrollado es un prototipo de procesador para convertir señales monofónicas en señales binaurales. El lenguaje de programación escogido es MAX/MSP versión 5, desarrollado por Cycling '74.

Las señales de entrada son señales mono o estéreo digitales (de las cuales se utiliza únicamente un solo canal) y se encuentran almacenadas en la memoria del computador.

Dichas señales son convolucionadas con impulsos respuesta (HRIR) de cinco individuos distintos y dependiendo de la posición de acimut y elevación, ambas opciones seleccionables por el usuario.

De ésta forma, se obtiene en la salida una señal estereofónica que deberá ser escuchada mediante audífonos mediante el usuario para obtener una señal binaural.

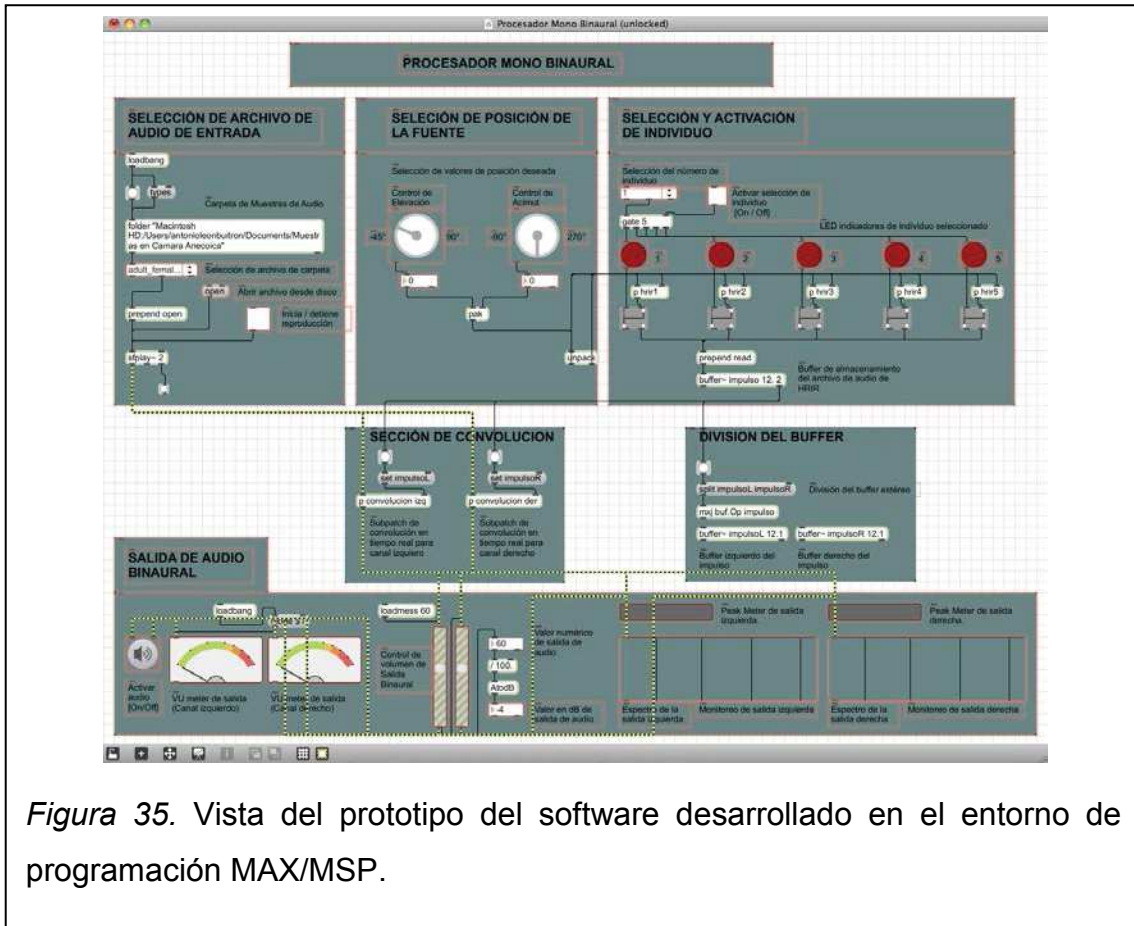


Figura 35. Vista del prototipo del software desarrollado en el entorno de programación MAX/MSP.

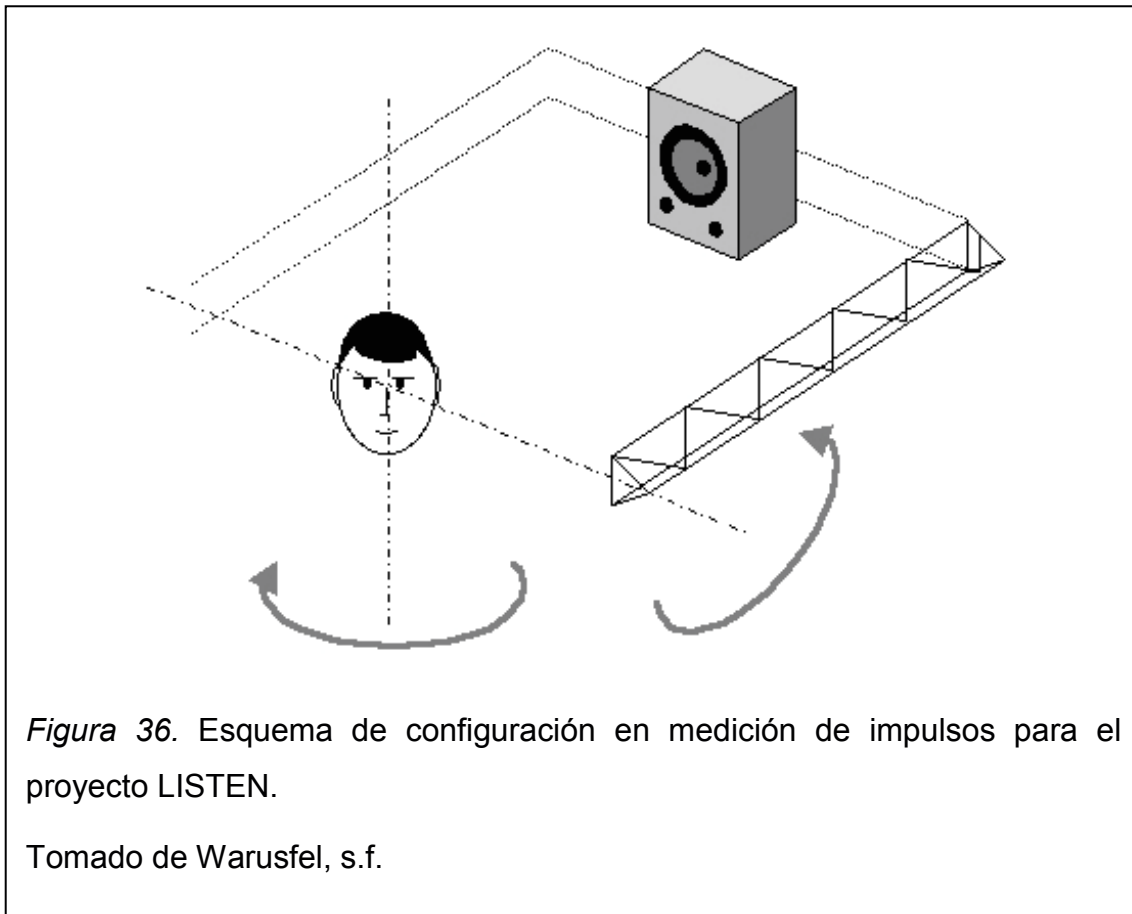
Debido a la decisión de implementar HRTF para el procesamiento de señales en la etapa de diseño del software, se consideran que todas las señales de entrada (que funcionan a modo de fuente) se encuentran en el centro, directamente al frente del oyente.

2.1.1. Base de Datos de HRIR

La base de datos de HRIR es tomada del proyecto LISTEN, donde IRCAM en asociación con AKG llevaron a cabo la medición de los impulsos empleados en el software, de acuerdo a Warusfel (s.f.).

Las mediciones fueron realizadas en una cámara anecoica y empleando sistemas de rotación tanto para el oyente como para el altavoz para cubrir todas las posiciones deseadas, con un sistema de *head-tracking* computarizado para asegurar la correcta posición de los individuos en cada

posición requerida. La información técnica sobre el equipamiento empleado en la medición se adjunta en el Anexo 3.



Las mediciones se realizaron en el punto P2, es decir, en la entrada del canal auditivo con éste bloqueado con el fin de evitar información de transmisión unidimensional. Para cada individuo, se utilizó una masilla de silicón para bloquear el canal y asegurar la posición del micrófono.

Los impulsos fueron registrados en un computador en un software personalizado realizado en MAX/MSP, encargada de medir un barrido logarítmico de frecuencias de 8912 muestras a 44,100 [kHz] (de frecuencia de muestreo) y deconvolucionar la señal del micrófono en tiempo real para ambos oídos.

El uso del barrido largo en lugar de ruido (a veces utilizado para medición) se justifica en la página web del proyecto que se debe a:

“Considerablemente menos vulnerable a distorsión y variación en el tiempo. Usando un barrido algo más largo que el impulso respuesta de la sala a ser medida permite la exclusión de todos los productos por distorsión armónica, prácticamente dejando solo ruido de fondo como la limitación de relación señal ruido alcanzable” (Warusfel, s.f.).

Se midieron 187 puntos en total para cada individuo, en una configuración de semiesfera alrededor de los mismos. Comenzando en -45° de elevación hasta 90° de elevación. Los pasos de elevación son de 10° y los pasos de acimut de 15° . Sin embargo, no existen todos los puntos de elevación para todos los puntos de acimut, como se aprecia en la siguiente figura.

Tabla 2. Tabla de puntos de medición de HRIR para cada individuo.

Elevation (degrees)	Azimuth increment (degrees)	Points per elevation
-45	15	24
-30	15	24
-15	15	24
0	15	24
15	15	24
30	15	24
45	15	24
60	30	12
75	60	6
90	360	1

Nota: Las columnas hacen referencia a elevación, incremento en acimut (en grados) y puntos por elevación respectivamente.

Tomado de Warusfel, s.f.

De todos los individuos para los cuales se realizaron las mediciones, se encuentran disponibles como archivos de audio para descarga las HRIR para 51 de ellos, como archivos de audio WAV a 24 [bits] y 44,100 [kHz].

De los 51 individuos, se han empleado los 5 primeros de ellos para el software. Se ha optado por los archivos denominados como “Compensados” (*compensated*), que poseen ecualización por las deficiencias de la cadena electroacústica empleada además de estar ecualizados para campo difuso. Estos impulsos son reducidos a 512 muestras.

2.2. Configuración de Objetos y Funcionamiento del Software

El prototipo desarrollado se ha dividido en tres etapas o secciones en base a su función para una visualización más clara de los objetos y sus conexiones.

2.2.1. Sección de Selección de Entrada

Esta es la primera sección que se encarga de seleccionar el archivo de audio que será procesado con las HRIR. El archivo deberá ser idealmente una muestra de audio registrada en cámara anecoica para mejores resultados, por lo que se provee 7 muestras que se especifican en el Anexo 4 y han sido obtenidas de la librería OpenAIR, de acuerdo con Audiolab (2009). También se facilita la opción de escoger cualquier otro archivo que se encuentre almacenado disco duro.

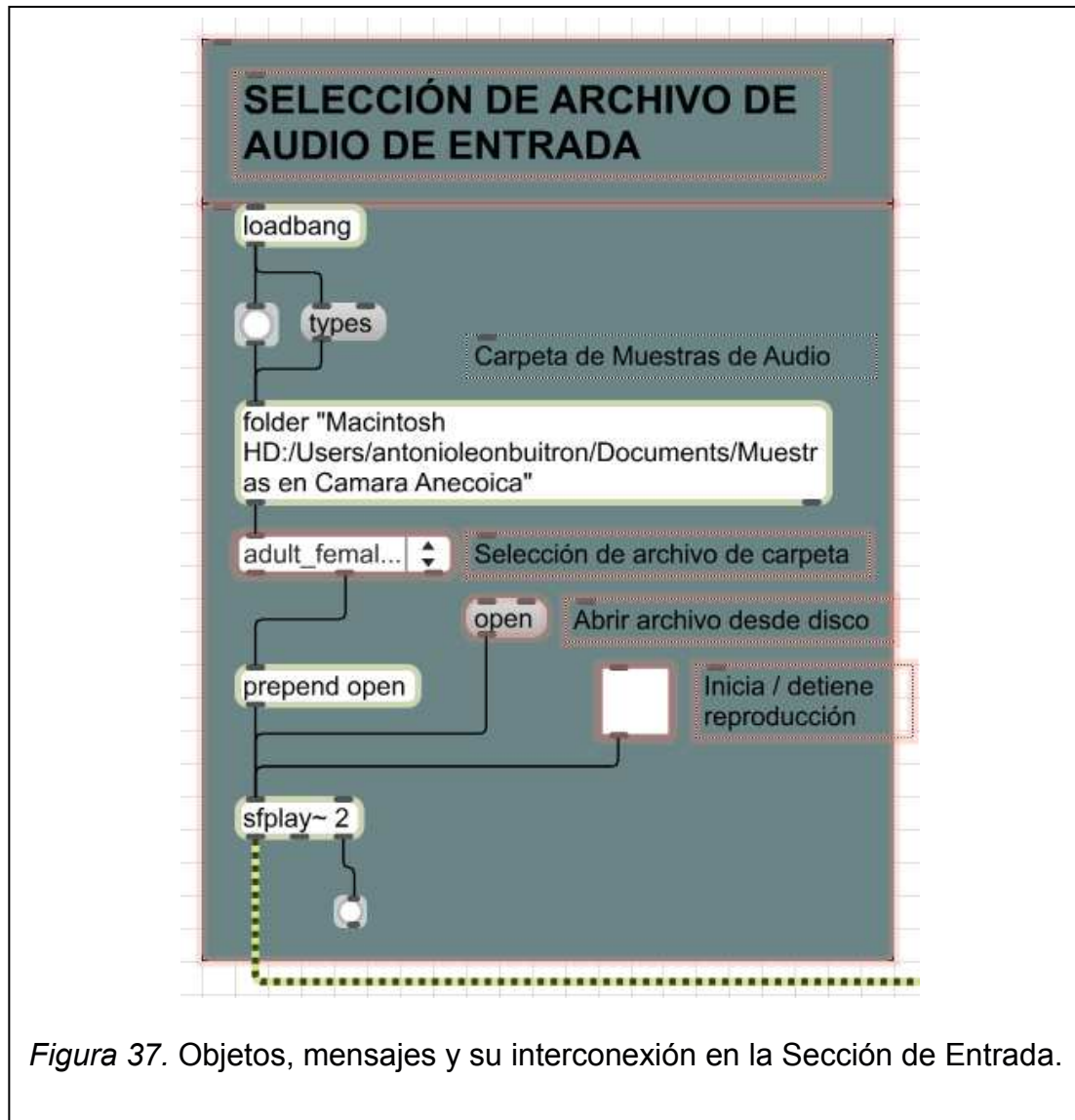


Figura 37. Objetos, mensajes y su interconexión en la Sección de Entrada.

El objeto *loadbang* envía un mensaje tanto al objeto *button* como al mensaje "types". Éste último se emplea para que todos los archivos que se encuentran en la carpeta especificada en el objeto *folder* se visualicen en el objeto *umenu*.

El objeto *folder* especifica la carpeta en el sistema donde se encuentran las muestras anecoicas proporcionadas. El objeto tipo botón *button*, se emplea para refrescar el mensaje de salida de *folder*, en caso de ser necesario cambiar dicha ruta (para una carpeta distinta).

El objeto *umenu* se encarga de presentarlos archivos de audio en cámara anecoica en forma de un menú despegable para su visualización y selección.

Una vez seleccionado el archivo deseado, se envía un mensaje al que se antepone el mensaje “open” mediante el objeto *prepend*, lo que completa el mensaje que necesita el objeto *sfplay~* para reproducir el archivo de audio.

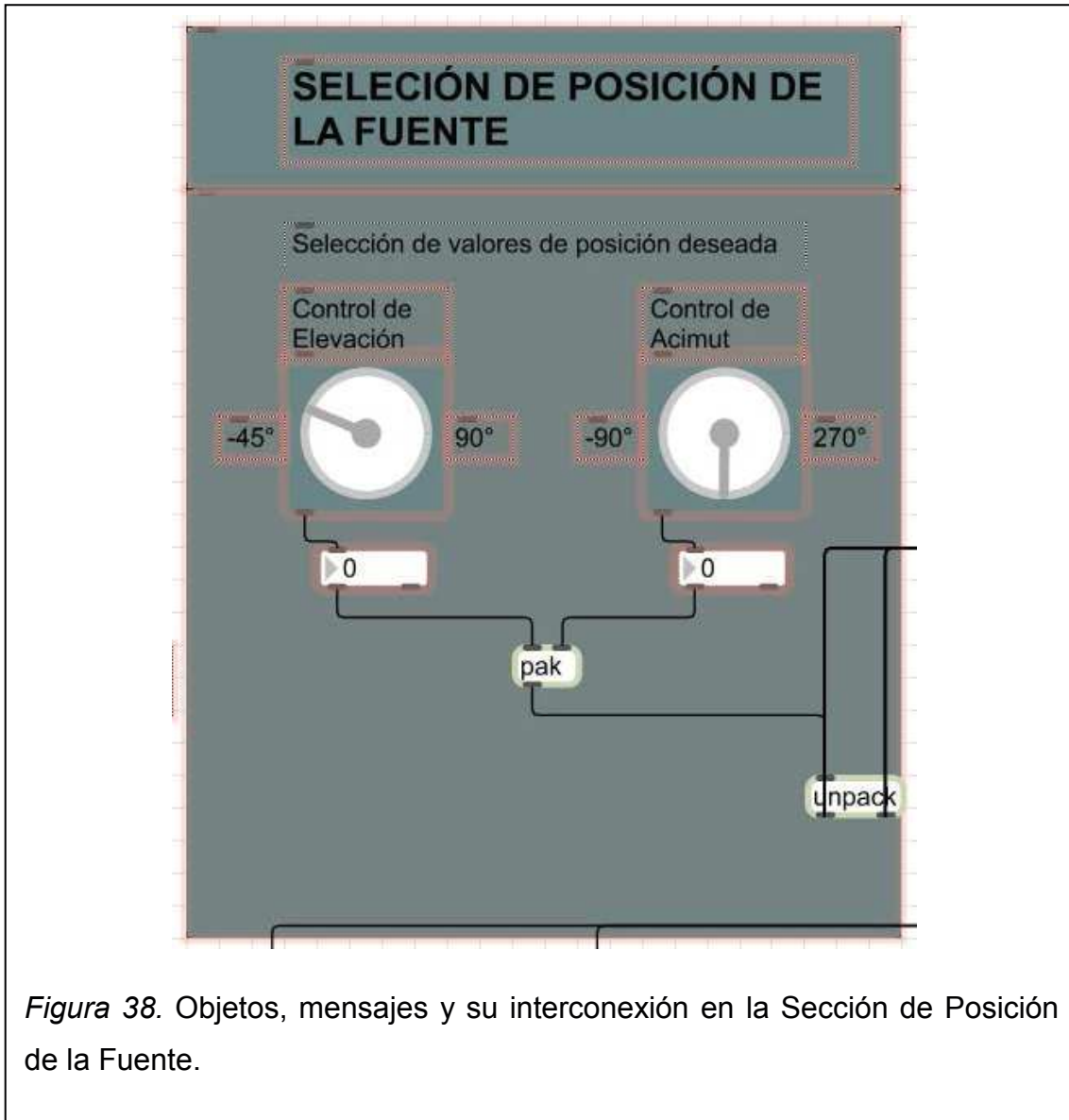
El mensaje “open” que se encuentra por separado, permite seleccionar cualquier otro archivo de audio almacenado en el disco duro del computador.

El objeto *sfplay~* se encarga de reproducir el archivo seleccionado ya sea mediante el menú, o mediante open (y siempre reproducirá, entre cualquiera de éstos, el último en ser escogido). La reproducción se controla mediante un *click* en el objeto *toggle*. El objeto *button* conectado a la salida derecha de *sfplay~*, se iluminara una vez finalizada o detenida la reproducción.

Es posible cargar archivos estéreo a pesar de que el programa solo se encarga del procesamiento de señales monofónicas. Esto se debe a que solo se toma un canal de salida del objeto *sfplay~* que será enviado a la primera entrada (entrada izquierda) de los *subpatch convolucion izq* y *convolucion der* encargados de llevar a cabo el procesamiento.

2.2.2. Sección de Posición de la Fuente

El software presenta la opción al usuario de seleccionar en que punto de la semiesfera quiere ubicar a la señal de entrada (seleccionada en la Sección de Entrada). La selección de la posición representa la selección de los puntos para los que se dispone de una HRIR en el proyecto LISTEN, por lo que no existen todos las posiciones de elevación para las posiciones de acimut.



Se disponen dos controladores tipo potenciómetros para la selección del valor de elevación y acimut (indicados por *comment boxes* sobre las mismas), mediante objetos *dial*. Sus valores son enviados respectivamente a dos objetos *number*, que muestran los valores escogidos.

Los pasos de los diales han sido modificados dentro del *inspector*. Para el dial de elevación, la rotación está configurada para pasos de 10°, y para el dial de acimut, para pasos de 15°, de acuerdo a los pasos de la base de datos de las HRIR.

Los valores de elevación y acimut son enviados a las entradas del objeto *pak*, que genera una lista de los mismos cada vez que recibe un nuevo valor. Dicha lista es recibida por un objeto *unpack*, y la descompone en sus elementos (dos en éste caso) y los envía por separado mediante sus salidas 1 y 2 hacia las entradas 1 y 2 respectivamente de los *subpatch* *hrir1*, *hrir2*, *hrir3*, *hrir4* y *hrir5*.

El punto 0°, tanto en elevación como en acimut, representa la posición al frente, con el eje en el centro de la cabeza y equidistante de ambos oídos. Para valores de elevación negativos, la posición se encuentra bajo dicho eje. Un valor de 90° de elevación representa la posición sobre la cabeza, al centro de la misma. Para valores de acimut, los valores entre 0° y 180° son para posiciones a la izquierda, y entre 180° y 345°, para posiciones a la derecha.

2.2.3. Sección de Selección de Individuo

Debido a que el empleo de HRTF no individualizadas no es el ideal, se han empleado todas las posiciones disponibles de cinco individuos de la base de datos del proyecto LISTEN. De ésta forma, el usuario es capaz de realizar una prueba puramente auditiva (ya que los aspectos anatómicos de dichos individuos son desconocidos) de estos grupos de impulsos y realizar el procesamiento de acuerdo a la que le entregue un mejor resultado en la localización binaural.

Los valores de posición escogidos en la Sección de Posición de la Fuente, son enviados a los *subpatch* de la Sección de Selección de Individuo. Estos *subpatch* se denominan *hrir1*, *hrir2*, *hrir3*, *hrir4*, *hrir5* para los 5 individuos respectivamente, contienen el proceso lógico para enviar un mensaje que almacena el impulso respuesta correspondiente al valor de posición escogido (recibido desde el objeto *unpack*) en objetos *buffer~* para su procesamiento con la señal de entrada.

Todos los *subpatch* reciben los mensajes de valor de posición de la fuente cuando estos son seleccionados, pero solo aquel que corresponda al individuo seleccionado será leído desde el disco y almacenado en el *buffer*, por lo que las salidas de los mismos pasan a través de un objeto tipo puerta denominado

gswitch2, cuyo control depende de los objetos *umenu*, *gate* y *toggle* presentes en esta sección.

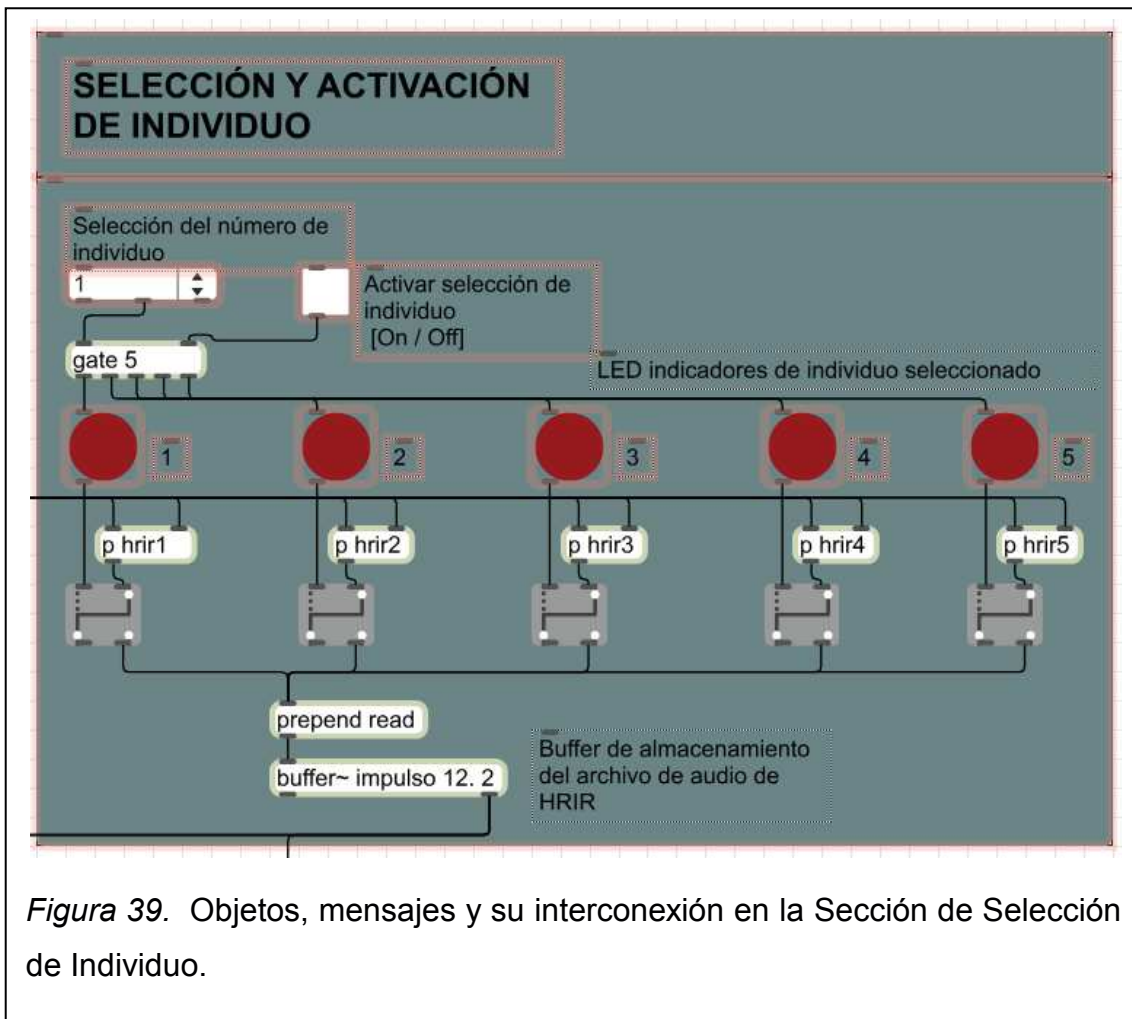


Figura 39. Objetos, mensajes y su interconexión en la Sección de Selección de Individuo.

La selección del individuo se realiza mediante el menú despegable en forma del objeto *umenu*, cuya lista ha sido ingresada manualmente, para que sean únicamente los números del 1 al 5. Para activar la selección, se debe dar un *click* en el objeto *toggle* (indicado como "On/Off") que se encuentra a la derecha del menú.

El objeto *gate* posee cinco salidas y determina cual puerta se tiene que abrir y cerrar. Éste objeto recibe el número de individuo seleccionado desde el menú y envía un mensaje por la salida correspondiente al realizarse un *click* sobre el objeto *toggle*.

El mensaje de salida de *gate* llega hacia la entrada de control del objeto *gswitch2*. Cuando se ha seleccionado y activado un individuo, *gswitch2* permite el paso del mensaje de la salida del *subpatch* correspondiente a dicho individuo por su salida derecha, enviándolo hacia el objeto *prepend*. Por defecto, los objetos *gswitch2* no se conectan a ninguna objeto (mediante sus salidas izquierdas).

Los objetos *led* realizan una función indicativa. Al seleccionar y activar un individuo mediante el objeto *toggle*, un *led* correspondiente al individuo seleccionado (acompañado por un indicador numérico) se enciende y al ser desactivado el *toggle*, se apaga.

El objeto *prepend* con argumento "*read*", antepone dicha palabra al mensaje que obtiene de la salida del *subpatch* activo. Por ejemplo, si el mensaje que se obtiene del *subpatch* es "*11a000e000.wav*" (nombre del archivo para el impulso respuesta del primer individuo, cuya posición y acimut son 0°), el mensaje de salida de *prepend* será "*read 11a000e000.wav*", necesario para ser entendido como mensaje por el objeto *buffer~*.

El *buffer~* de nombre impulso, como indica su primer argumento, también posee dos argumentos adicionales. "12" indica el tiempo en milisegundos en los que almacena información, suficiente para los impulsos de 512 muestras, de 11,60997732 [ms] (debido a su frecuencia de muestreo de 44100 [Hz]). El tercer argumento, indica el número de canales del *buffer*, que deberá ser "2", para almacenar los impulsos binaurales (de dos canales cada uno).

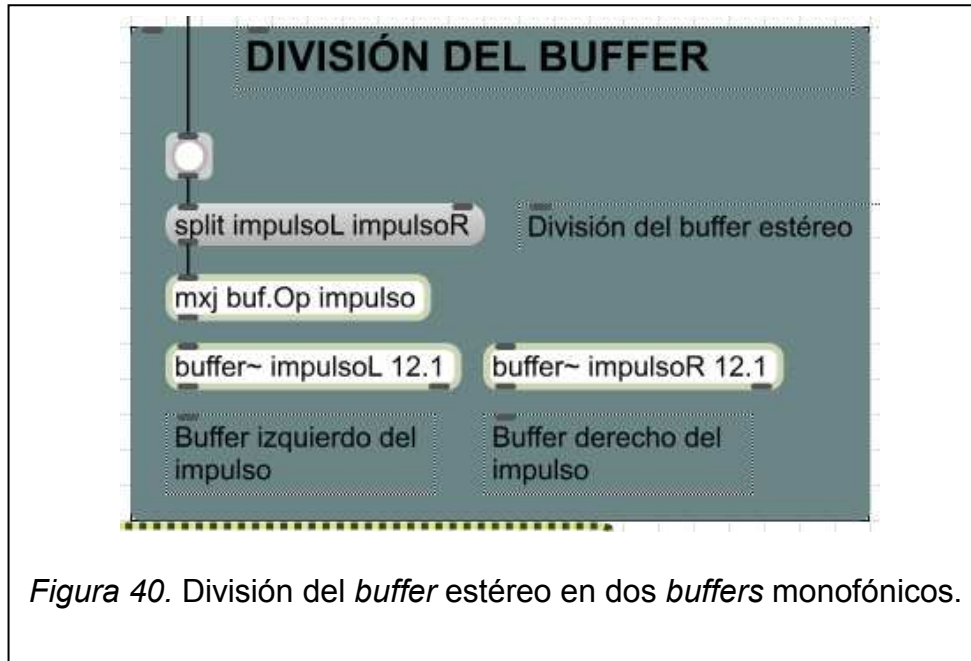


Figura 40. División del *buffer* estéreo en dos *buffers* monofónicos.

Debido a los objetos empleados para la convolución necesitan de *buffers* monofónicos, los impulsos binaurales que se almacenan en el objeto *buffer~impulso*, son divididos en dos *buffers* monofónicos para los canales izquierdo y derecho. Ambos *buffers*, poseen el mismo tiempo de almacenamiento (12 [ms]) pero solo 1 canal, por lo que su tercer argumento es "1".

El objeto *mxj buf.Op* es capaz de realizar varias operaciones adicionales sobre los objetos *buffer*, que almacenan archivos en memoria RAM, por lo que están limitados a la capacidad del propio computador. En el objeto *buffer~impulso* se almacena el archivo de audio original del impulso correspondiente a la posición seleccionada. Al recibir el mensaje "*split impulsoL impulsoR*", el objeto *mxj buf.Op* divide al archivo que se encuentra en dicho *buffer* en dos archivos monofónicos independientes, uno por cada canal. La señal dividida se almacena en dos objetos *buffer~* monofónicos, denominados *buffer~impulsoL* y *buffer~impulsoR* para el canal izquierdo y derecho respectivamente.

Éste proceso de división se realiza cada vez que el objeto *button* recibe un mensaje del objeto *buffer~impulso* al cual está conectado, lo que sucede cada vez que se escoge una nueva posición. Esto se debe a que *button* fuerza la

generación del mensaje “*split impulsoL impulsoR*” cuando *buffer~ impulso* termina de leer y almacenar el archivo.

2.2.3.1. Subpatch Hrir

Los *subpatch hrir* son los encargados de enviar un mensaje al *buffer~ impulso* para cargar el impulso respuesta de la posición seleccionada. Existe uno por cada individuo. Todos estos reciben la misma información, poseen la misma estructura, cantidad y tipo de objetos, así como su conexión, además de producir el mismo tipo de mensaje en su salida, con pequeñas modificaciones, explicadas posteriormente.

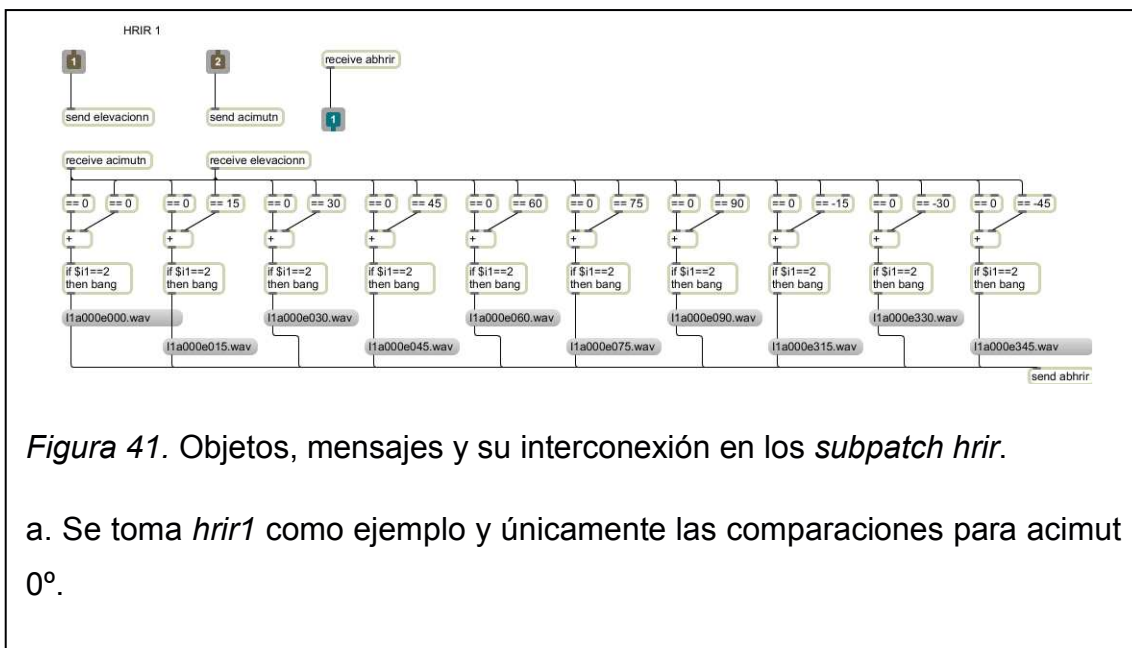
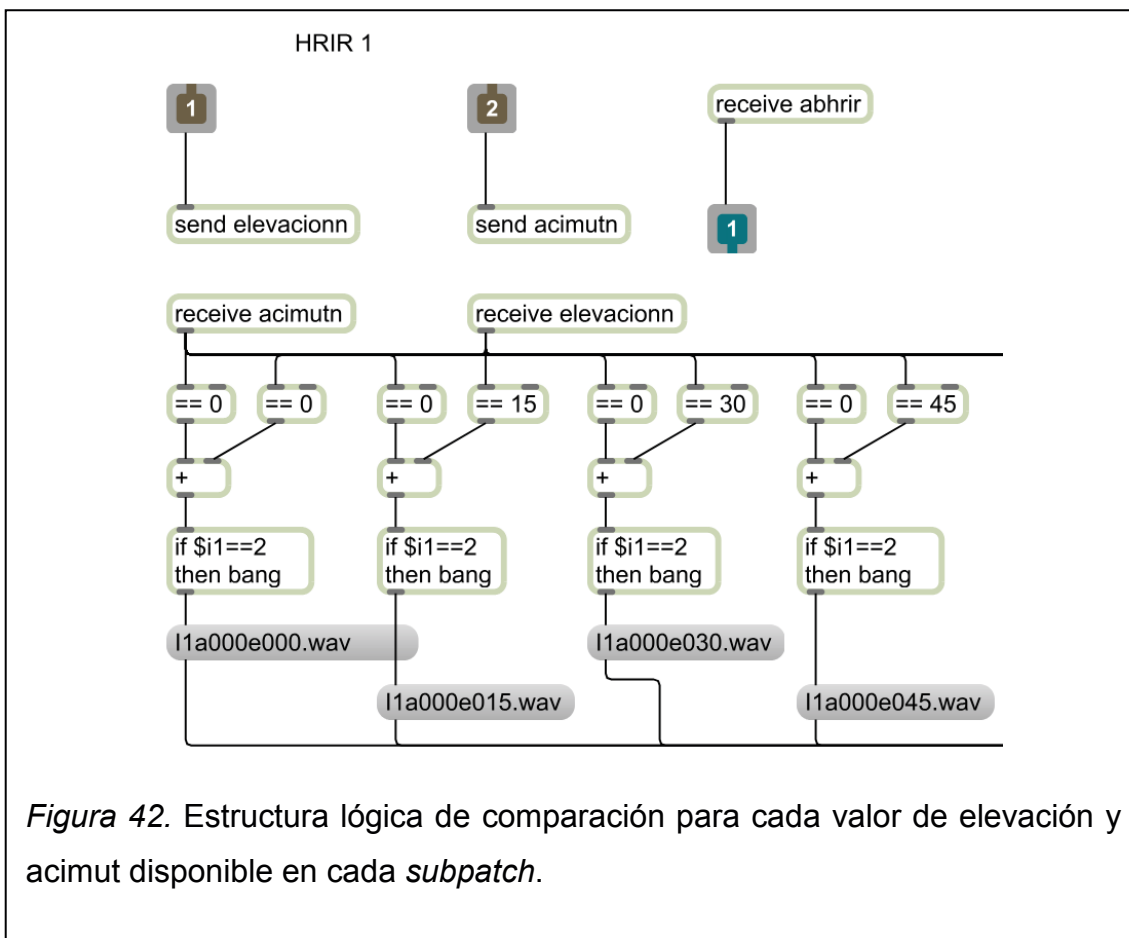


Figura 41. Objetos, mensajes y su interconexión en los *subpatch hrir*.

a. Se toma *hrir1* como ejemplo y únicamente las comparaciones para acimut 0°.

Cada *subpatch* posee dos entradas y una salida. Recibe los valores de posición de elevación y acimut mensajes de las salidas del objeto *unpack* y son enviados a través objetos *send* a los objetos *recieve*, para facilitar su envío a los 184 comparadores que posee cada *subpatch*. Existen un objeto *send elevacionn* y *send acimutn* en cada uno, que envían los valores a los objetos *recieve elevacionn* y *recieve acimutn* respectivamente.

El mensaje que recibe el objeto *recieve* es enviado hacia una serie de objetos que forman una estructura lógica de comparación, que se funciona de la siguiente manera.



Para cada impulso, existen dos objetos ==, uno para elevación y otro para acimut. Si el número seleccionado en los diales (en la Sección de Posición de la Fuente) es una de las posiciones disponibles como impulso, entonces un par de objetos de comparación ==, enviarán cada uno un valor de 1 por sus salidas y serán sumadas en el objeto +. El resultado de la suma, es enviada al objeto *if*.

If realiza una segunda comparación mediante el argumento “*if \$i1==2 then bang*”. Esto quiere decir que si la entrada 1 del objeto *if* (“*\$i1*”) es igual a “2” (únicamente cuando las dos posiciones han sido comparadas y están disponibles en la base de datos de los impulsos), el objeto *if* produce un mensaje “*bang*” en su salida.

El mensaje “*bang*” obliga a que el mensaje concatenado a la salida de cada objeto *if*, que especifica el nombre del archivo de audio (como se encuentra

almacenado en el disco duro) del impulso correspondiente, sea enviado hacia un nuevo objeto *send* (*send abhrir*).

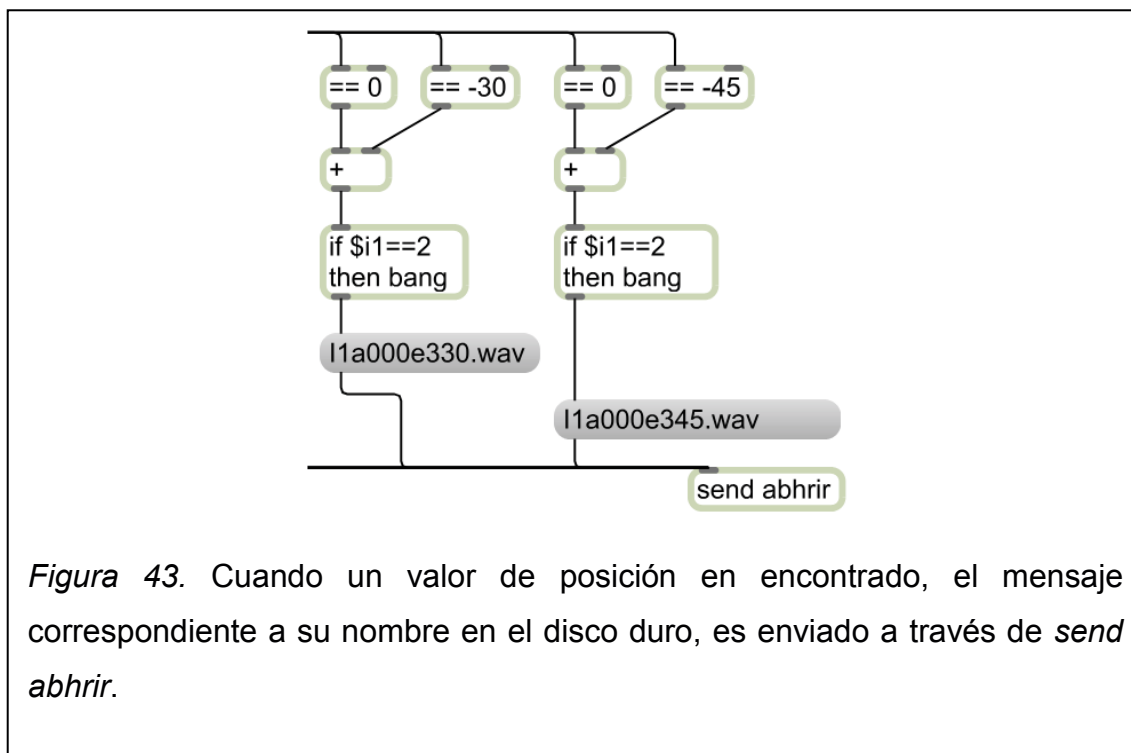


Figura 43. Cuando un valor de posición es encontrado, el mensaje correspondiente a su nombre en el disco duro, es enviado a través de *send abhrir*.

El objeto *send abhrir* envía dicho mensaje a *recive abhrir*, que lo dirige hacia la salida del *subpatch*, que en caso de estar activo y seleccionado, es enviado hacia un objeto *prepend*, para completar el mensaje que es necesario para cargar el impulso en el *buffer~ impulso* (como se explica en la Sección de Selección de Individuo).

Las modificaciones que existen entre cada *subpatch hrir* se deben a que cada *subpatch* posee un *comment box* en su parte superior que especifica el número de individuo, por ejemplo, "HRIR1" especifica el nombre del *subpatch* para el individuo 1, y así en cada caso.

Los objetos *send elevacionn*, *send acimutn*, *recive elevacionn* y *recive acimutn* también varían de acuerdo al *subpatch*. Para el individuo uno, cuyo *subpatch* es *hrir1*, los objetos se denominan *send elevacionn*, *send acimutn*, *recive elevacionn* y *recive acimutn*. Para el resto de individuos, se adjunta un número para especificar el individuo. En el caso del individuo 2, los objetos se

denominan *send elevacionn2*, *send acimutn2*, *recive elevacionn2* y *recive acimutn2*, y así sucesivamente, para el resto de *subpatch*.

2.2.3.2. Denominación de Archivos HRIR en el Software

Los cambios en cada *subpatch* también varían al variar los nombres de los archivos de audio de los impulsos. Los nombres con los que dichos archivos pueden ser descargados de la página web del proyecto LISTEN, han sido simplificados para obtener nombres más cortos y claros, únicamente con la información relevante para su control e identificación dentro del software.

El nombre genérico que poseen los archivos al descargarlos son: "*IRC_<ID del sujeto>_<estado>_<radio>_<acimut>_<elevación>.wav*". Esto para los archivos de audio en formato WAV, compensados para campo libre que son los empleados en el software. Donde IRC significa IRCAM, y los campos entre "<>", valores que cambian entre distintas posiciones e individuos.

Por ejemplo, para una posición de elevación y acimut 0°, el impulso posee el nombre de "*IRC_1002_C_R0195_T000_P000.wav*". Los argumentos IRC, C, y R (radio), son los mismos para todos los archivos, por lo que pueden ser truncados.

Para su uso dentro del software, se emplean únicamente especificaciones para número de individuo, valor de elevación y valor de acimut. El número de individuo se especifica como "1", "2", "3", "4", "5", acimut lleva la letra "a" y elevación "e", antepuesta a su valor numérico, por lo que el ejemplo anterior es renombrado como "*11a000e000.wav*".

Para valores negativos de elevación (-45°, -30°, -15°), su valor en el archivo se especifica como 345, 330, 315 respectivamente.

Los individuos de la base de datos del proyecto LISTEN que se emplearon para el software son los individuos 1002, 1003, 1004, 1005, 1006, referidos como I1, I2, I3, I4 e I5 respectivamente.

2.2.4. Sección de Convolución

La sección de convolución se encarga del procesamiento de la señal de entrada con el impulso correspondiente a la posición e individuo seleccionado. Dos *subpatch* (uno para cada canal) implementan objetos externos del paquete AHarkerExternalsv.1.0, desarrollados Alexander J. Harker que se encargan de llevar a cabo un proceso de convolución en tiempo real (y se especifican en el Anexo 4).

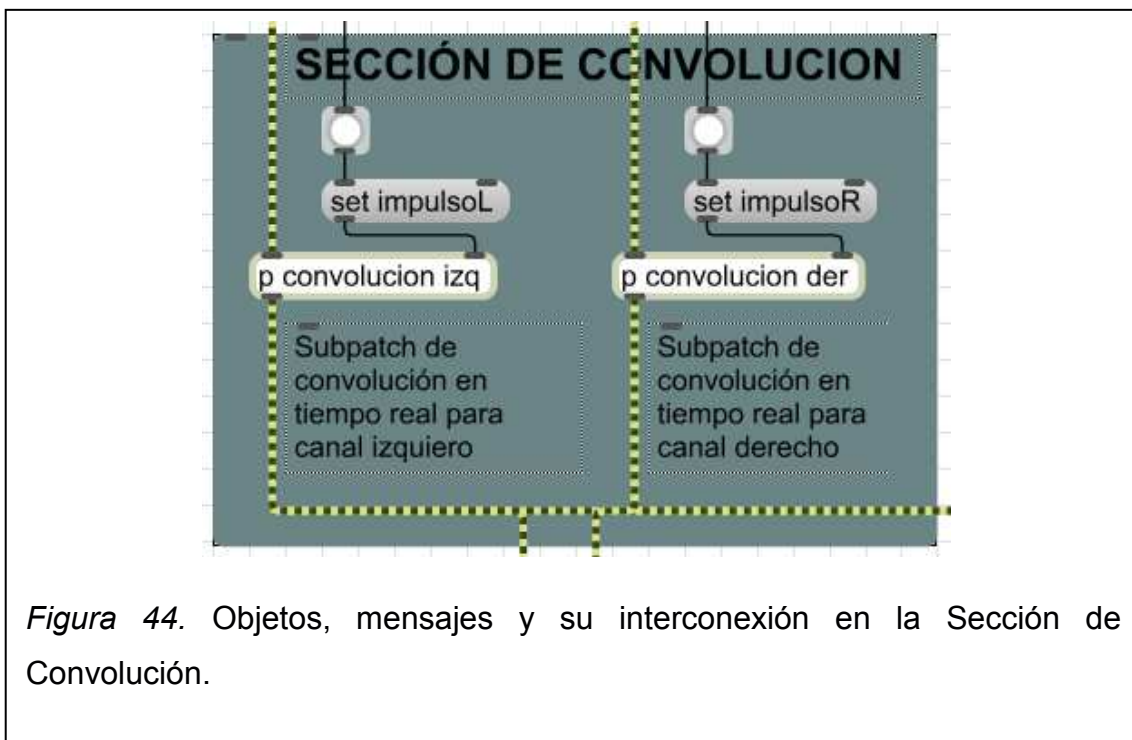


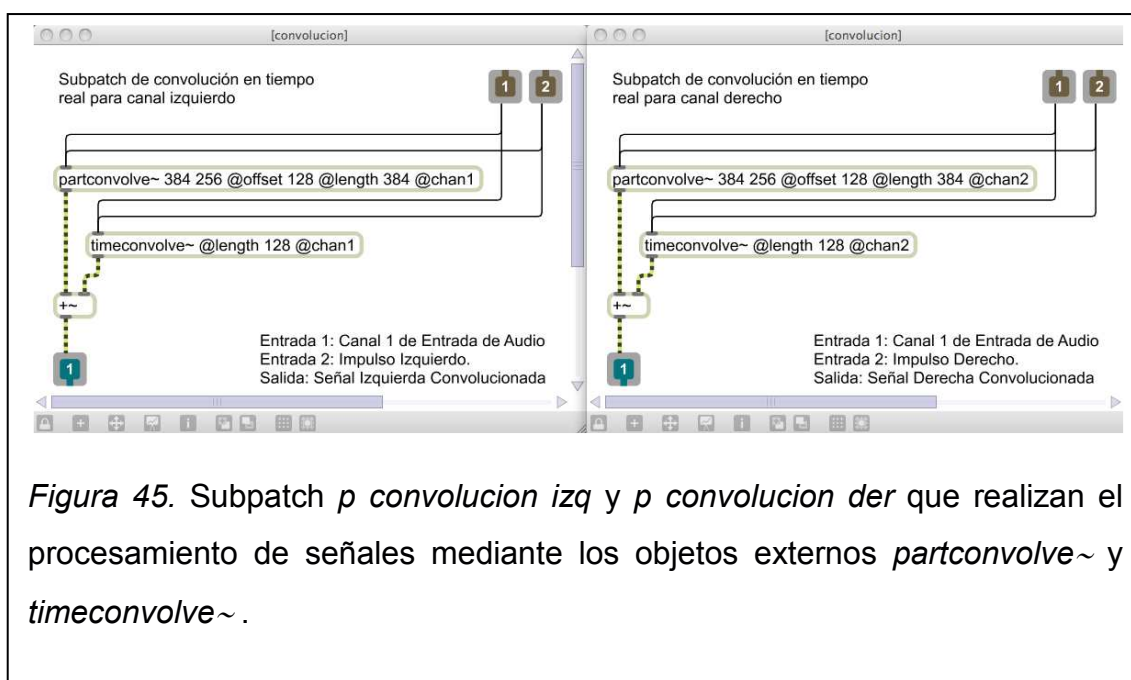
Figura 44. Objetos, mensajes y su interconexión en la Sección de Convolución.

Al ser leído y almacenado el archivo de audio correspondiente a un HRIR en el *buffer~ impulso*, dicho objeto envía un mensaje a un objeto *button* que se encarga de dividir al *buffer* estéreo en dos *buffers* monofónicos demonizados *buffer~ impulsoL* y *buffer~ impulsoR* (como se explica en la Sección de Selección del Individuo).

Paralelamente, dos objetos *button* generan un mensaje "bang" que es enviado a los mensajes "set impulsoL" y "set impulsoR", forzando a que estos se produzcan en sus salidas y se reciban en las entradas derechas de los *subpatch* que realizan la convolución.

De ésta manera, se especifica de que objeto *buffer~* se debe tomar el impulso necesario para el procesamiento, siendo *buffer~ impulsoL* necesario para *p convolucion izq* y *buffer~ impulsoR* para *p convolucion der*.

En su entrada izquierda, los *subpatch* (denominados *p convolucion izq* y *p convolucion der*) reciben la señal monofónica de la señal de entrada a través de la primera salida de audio del objeto *sfplay~* de la Sección de Entrada.



La configuración de los objetos dentro de cada *subpatch* de convolución es la misma, pero deben ser implementados por separado para cada canal de salida deseado debido a la capacidad de los objetos externos empleados.

El objeto *partconvolve~* se encarga de realizar convolución mediante FFT mientras que *timeconvolve~* lo hace en el dominio del tiempo (en tiempo real). Al implementarlos juntos, dividiendo al impulso original de 512 muestras en dos segmentos, para realizar convolución en el dominio del tiempo para el primer segmento (muestras 0 a 128) y mediante FFT en el segundo (muestras 129 a 512) y sumando sus salidas mediante un objeto *+~*, es posible obtener convolución en tiempo real y sin latencia evidente.

La división del impulso en el esquema de convolución se define en los objetos que se emplean para éste proceso, por lo que en el objeto *timeconvolve~* se especifica una longitud de 128 muestras en sus atributos como *@length 128* (seguido del canal para el *subpatch* respectivo) y deberá encargarse del primer pedazo del impulso segmentado para asegurar convolución con cero latencia.

Por lo tanto, *partconvolve~* deberá especificar en sus argumentos el número de muestras máximas (las 384 restantes), el valor máximo del tamaño de FFT (256 muestras), y como parte de sus atributos (indicado con el símbolo @) su valor de *offset* (128 muestras) como *@offset 128* y su número de canal (como *@chan*).

Las salidas de los *subpatch p convolucion izq* y *p convolucion der* que contienen la señal procesada para los canales izquierdo y derecho respectivamente son enviadas de forma simultánea a dos objetos *gain~* y dos objetos *spectroscope~* en la Sección de Salida Binaural.

2.2.5. Sección de Salida de Audio

La señal procesada es recibida por los objetos *gain~* para el control del nivel de la salida de audio, con un rango posible entre 0 y 100. Las salidas de los objetos *gain~* es enviada al objeto *ezdac~*, al cual deben ser enviados (y estar activo) para producir una señal de audio de salida.

La salida derecha del objeto *gain~* del canal izquierdo ha sido conectado a la entrada derecha del objeto *gain~* del canal derecho, de ésta forma, se puede controlar ambos objetos mediante el objeto *gain~* del canal izquierdo, formando un control único, más lógico si se busca una señal binaural en la salida de audio del software. Sin embargo, es posible seleccionar un valor independiente para la salida derecha mediante su respectivo objeto *gain~*.

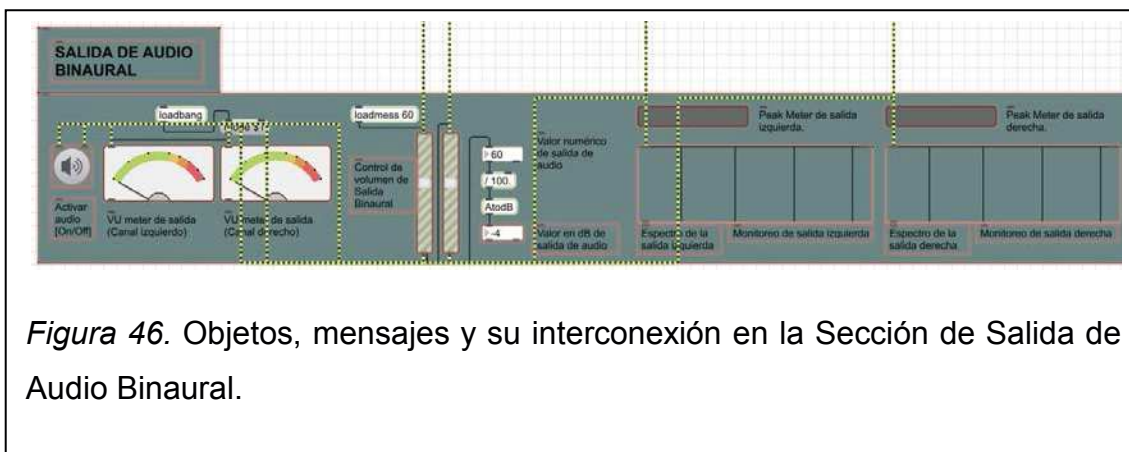


Figura 46. Objetos, mensajes y su interconexión en la Sección de Salida de Audio Binaural.

El objeto *loadmess* se emplea con el propósito de que el valor de los objetos *gain~* posea un valor escalado de salida por defecto de 60 al iniciar el programa.

El valor de la salida del objeto *gain~* izquierdo se visualiza como un valor numérico en el objeto *number* y a su vez, es enviado a un objeto */* que lo divide para "100". El resultado entra al *subpatch AtodB*, que se encarga de convertir el valor numérico en una escala de decibeles que se muestra en el último objeto *number* que recibe su valor de salida.

Las salidas de los objetos *gain~* también son enviadas a dos objetos *levelmeter~* (uno por cada canal) para el monitoreo visual de la señal binaural, donde su balística se ha configurado para monitoreo mediante indicadores tipo VU metros.

También se incluyen dos objetos *meter~*, uno por cada canal de salida, que toman la salida de los objetos *gain~* y la presentan en forma de Medidor de Picos (*Peak Meter*) donde es posible visualizar si se presenta saturación en la salida (al encenderse el último LED disponible en la escala en color rojo). De ésta manera, es posible el monitoreo mediante ambos métodos de forma simultánea, de acuerdo a la preferencia del usuario.

Por último, es posible el monitoreo del espectro de la señal procesada, empleado mediante objetos *spectroscope~*, para la señal obtenida de la salida de los *subpatch* de convolución, tanto para el canal izquierdo como derecho,

configurados en sus respectivos *inspector*, para que se presente en una escala logarítmica.

2.2.6. Configuraciones Adicionales: Presentación del Software

El *patch* del software prototipo ha sido configurado desde su inspector para ser presentado en su Vista de Presentación (*Presentation*) al ser iniciado (donde también ha sido configurado su color de fondo).

La Vista de Presentación difiere de la Vista de Edición (*Edit*) ya que muestra únicamente algunos de los objetos, sin sus interconexiones y de una manera más ordenada y agradable.

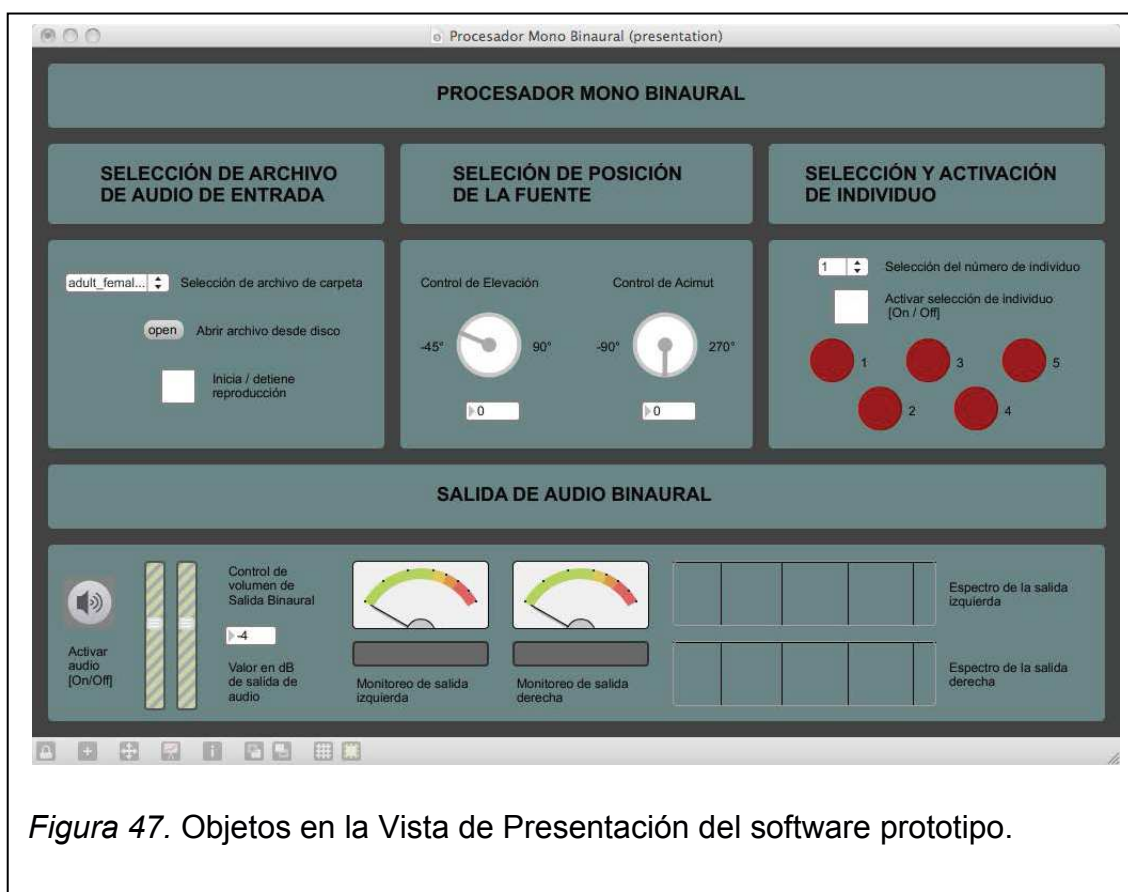


Figura 47. Objetos en la Vista de Presentación del software prototipo.

Los objetos incluidos son únicamente los considerados necesarios para que el usuario a modo de operador del programa, pueda interactuar con la interface gráfica sin ninguna otra distracción y pueda recibir la información visual de monitoreo entregada.

Todos los objetos presentes en el diseño del software siguen presentes para su correcto funcionamiento, pero únicamente los que han sido añadidos con la opción *Add to Presentation* (en el menú *Object*) están presentes en la Vista de Presentación, además de 9 objetos *panel* para los rectángulos redondeados de la interface.

2.2.7. Configuraciones y Requerimientos del Sistema

El software se presenta como un patch de MAX/MSP en un archivo único denominado “Procesador Mono Binaural” con extensión “.maxpat” por lo que solo puede ser abierto en el entorno de programación MAX/MSP.

Los objetos externos de AHarker así como los archivos de audio de los impulsos respuesta, que se encuentran en las carpetas “AHarkerExternals” y “LISTEN HRIR” respectivamente, se deben copiar en una carpeta del disco y ser añadidos al *Search Path* de MAX/MSP para su correcto funcionamiento.

Las muestras de audio de cámara anecoica se encuentran dentro de la carpeta “Muestras en Cámara Anecoica” que también deberá ser añadida en el *Search Path* de MAX/MSP y deberá copiarse dentro de la carpeta “Documentos” por lo que su ruta está determinada para dicha ubicación y es utilizada por ciertos objetos que manejan dichos archivos en el software.

Dentro del software, las configuraciones que deben establecerse en el menú DSP Status son dos salidas para el canal izquierdo y derecho respectivamente, y un *I/O Vector Size* y *Signal Vector Size* no mayor a 128 muestras.

Debido a que el software necesita de los objetos externos para el procesamiento y éstos han sido desarrollados exclusivamente para sistemas Macintosh, funciona únicamente en ésta plataforma y para sistemas operativos *Mac OS X*.

La programación se llevó a cabo en la versión 5.1.3 de MAX/MSP. Los requerimientos del desarrollador para ésta versión son un sistema PowerPC o un sistema con procesador Intel, sistema operativo *Mac OS X 10.4.11* o

superior, 1 [GB] de memoria RAM, software QuickTime 7.1, OpenGL 1.4, y una tarjeta gráfica compatible con OpenGL.

El uso de objetos *mxj*, requieren la instalación de una versión actualizada de Java Virtual Machine (JVM) en el computador.

Si bien no se ha realizado pruebas para evaluar el funcionamiento en versiones previas a MAX/MSP 5, se puede asumir que el programa no funcionará correctamente ya que ciertos objetos empleados difieren en sus características y durante en la ejecución del mismo se volverán inactivos.

2.3. Análisis del Software

Como se plantean en los objetivos, se realizan pruebas para la evaluación del software desarrollado. Dichas pruebas tienen como objetivo evaluar el procesamiento que realiza el software en transformar una señal monofónica de entrada en una señal binaural de salida.

2.3.1. Diseño de Pruebas de Análisis del Software

El tipo de análisis es cuantitativo, por lo que se compararán el espectro de la señal obtenida (en la salida) en un análisis mediante FFT. Como dicha señal de salida depende de la posición, se consideran evaluar a una señal de entrada y su convolución con distintos impulsos respuesta, que representan distintas posiciones de acimut y elevación que presentan cambios espectrales.

Se escoge a un solo individuo (Individuo 2) de los disponibles para su procesamiento en el software. La selección de los individuos en la evaluación se ha realizado de forma arbitraria y aleatoria, pudiendo ser cualquiera de los otros individuos los empleados para las pruebas debido al tipo de evaluación que se busca realizar.

Si bien el espectro de la misma señal de entrada varía al ser procesada para una misma posición para distintos individuos, no existe manera de evaluar mediante su análisis espectral las diferencias psicoacústicas en la percepción sonora, por lo que no se busca una comparación de resultados entre

individuos. Finalmente se utilizará a un segundo individuo, el Individuo 5 (seleccionado bajo el mismo criterio que el Individuo 2, con el fin de apoyar los resultados obtenidos en la evaluación).

La señal de entrada es ruido blanco monofónico, en formato WAV, generado en el software Pro Tools LE 8.0.4 de AVID, con un nivel RMS de -3 [dBFS] y una duración de 1 [s]. Dicho ruido presenta la característica de poseer un espectro constante, es decir, el mismo nivel para todas las frecuencias y a lo ancho de todo el espectro sonoro, lo que facilita la visualización del filtrado producido por las HRTF de cada posición evaluada en el análisis espectral.

La primera evaluación se realiza para la comparación de espectro en el plano vertical (acimut), por lo que el valor de posición de acimut se encuentra fija al centro de la cabeza (equidistante a ambos oídos) para cinco posiciones de elevación distintas. Se compara la misma señal procesada para las posiciones de elevación de -30° , 0° , 30° , 60° y 90° , para un mismo valor de acimut de 0° . En esta etapa se obtienen cinco pares de espectros sonoros (uno por cada canal de salida).

La segunda evaluación se realiza para la comparación de espectro en el plano horizontal (elevación), empleando ruido blanco procesado para una posición de elevación fija (al centro de la cabeza) y ocho posiciones distintas de acimut cada 45° . Es decir, para posiciones de acimut de 0° , 45° , 90° , 135° , 180° , 225° , 270° y 315° , todas para una posición de elevación de 0° . En esta etapa se obtienen ocho pares de espectros.

Para obtener las señales procesadas, se realiza una modificación al patch del software prototipo en MAX/MSP que permita la captura de las señales que se obtienen del proceso de convolución como un archivo de audio digital para su posterior análisis espectral mediante software especializado.

Se añade el objeto *sfrecord~* que se encarga de grabar las señales que se obtienen de la salida de audio del software. A dicho, objeto se le agrega un argumento "2" para establecer que existen dos canales de audio en su entrada. También se añade un mensaje "open wave" para el objeto *sfrecord~* que

permite establecer el nombre el archivo y la ubicación en el disco donde será registrado, además de especificar el formato de grabación WAV y un mensaje “*samptype int24*” para especificar la profundidad de bits de 24 en el archivo generado. El inicio y fin de la grabación es controlado por el mismo objeto *toggle* que controla el inicio y fin de la reproducción de audio en el objeto *sfplay~*, de esta manera, la grabación comenzará cuando se inicie la reproducción de la señal de ruido blanco (*toggle* envía un valor de 1 a la entrada del objeto *sfrecord~*).

Las señales obtenidas en las pruebas ingresadas al software Pro Tools LE 8.0.4 y su análisis espectral se realiza mediante el *plug-in* Ozone 5 de iZotope. El análisis de espectro se configura con un *frame* o *FFT size* es de 512 muestras, con un solapamiento (*overlap*) de 50% (cada 256 muestras) y empleando una Ventana de *Hanning* (*windowing*). Este análisis se realiza de forma sucesiva para obtener un promedio temporal durante 1 [s], equivalente a la duración total de cada muestra.

2.3.2. Presentación y Análisis de Resultados

Se presentan los resultados como un gráfico de espectro que presenta la amplitud y frecuencia para el tiempo de 1 [s] analizados para los dos canales (izquierdo y derecho) de cada posición. La amplitud se encuentra representada en el eje de las abscisas en una escala de decibeles (entre -120 y 0 [dB]). La frecuencia se encuentra en el eje de las ordenadas vertical en un rango entre 20 y 20000 [Hz].

Tanto la señal de entrada como los impulsos son archivos de audio digital que presentan una profundidad de bits o resolución de 24 [bits] y una frecuencia de muestreo de 44100 [Hz], por lo que las grabaciones obtenidas en el software han sido realizadas bajo las mismas especificaciones.

El nivel de salida es ajustado para todas las pruebas en 60 [dB] que asegura un rango dinámico suficiente donde “cualquier diferencia espectral bajo esta cantidad será probablemente o bien imperceptible, o fuera del rango total de

reproducción del sistema” (Begault, 2000, p.45), además de evitar la saturación de la señal en la salida del software.

Finalmente tanto para los resultados de las pruebas de acimut como elevación, se presenta un gráfico tridimensional (en función del tiempo) del espectro de todas las posiciones evaluadas dispuestas de forma continua (1 [s] cada una) para los canales de salida izquierdo y derecho para una comparación más directa entre los espectros de todas las muestras obtenidas. La escala de la amplitud es representada por una escala de colores entre azul y amarillo, para valores menores y mayores respectivamente. A esto le acompaña una tabla que sintetiza los resultados presentados en cada grafica de comparación de espectros para cada canal y en cada posición considerada.

Se ha optado por introducir una evaluación basada en un segundo individuo y seleccionado bajo el mismo criterio que el primer individuo. Al emplear las HRTF de un segundo individuo, en éste caso, el Individuo 5, se obtienen un segundo conjunto de pruebas con el fin de apoyar los resultados obtenidos con respecto al procesamiento del software en las pruebas del Individuo 2 inicialmente, y no como una comparación entre los mismos, que escaparía los objetivos de la evaluación.

Los resultados de la evaluación para el Individuo 5 se presentan al final de éste capítulo (en el apartado 2.2.3. Pruebas Adicionales) en forma del grafico tridimensional de comparación de espectros así como la tabla que sintetiza el análisis de los resultados obtenidos tanto para acimut como elevación, ya que las pruebas para el Individuo 5 se realizaron en las mismas posiciones y condiciones que para el Individuo 2.

Las figuras del análisis espectral de las pruebas pueden ser mejor apreciadas con mayor tamaño, por lo que han sido incluidas dentro del CD-ROM anexo a éste documento. Las imágenes presentadas a continuación se encuentran dentro de la carpeta “Resultados“, donde se presentan las subcarpetas “Individuo 2” e “Individuo 5” con los archivos adjuntos respectivos.

2.3.2.1. Señal de Entrada: Ruido Blanco Monoaural

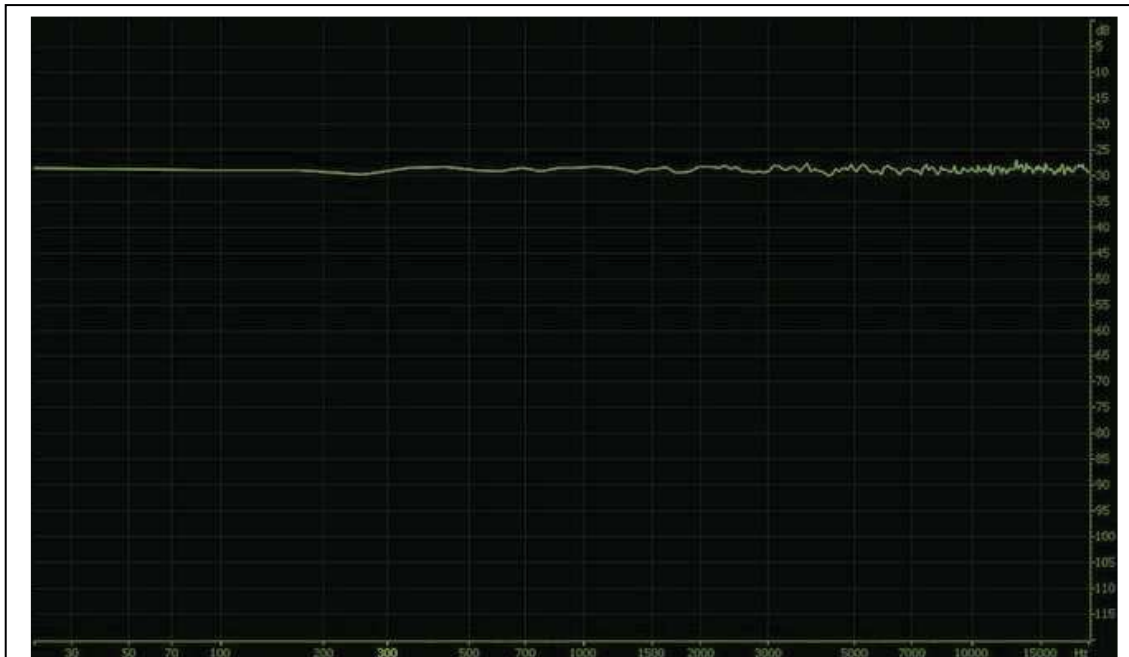


Figura 48. Espectro sonoro de la señal de entrada (ruido blanco).

La señal de entrada presenta un nivel constante alrededor de -27 y -30 [dB] para todo el espectro sonoro, a partir de los cuales se realizan las comparaciones del espectro para cada posición.

2.3.2.2. Pruebas de Elevación

Para todas las posiciones evaluadas se han encontrado que todas presentan el mismo tipo de filtrado para la zona del espectro de bajas frecuencias tanto en el canal izquierdo como derecho, por lo que su comparación no se describe en el análisis.

El filtrado es el mismo para la zona bajo los 160 [Hz], y se presenta como una especie de Filtro Pasa Bajos y es común para todas las muestras del análisis de elevación.

Un segundo filtrado es común y se presenta en la zona entre los 160 [Hz] y 400 [Hz], que si bien varía entre posiciones en la forma que modifica el espectro, es el mismo para ambos canales.

a) Elevación -30°, Acimut 0°



Figura 49. Espectro sonoro de la señal obtenida en la posición acimut 0°, elevación -30° para los canales izquierdo y derecho respectivamente.

No se presentan cambios significativos entre los canales izquierdo y derecho entre los 20 [Hz] y 1000 [Hz] aproximadamente. El valle que se presenta entre los 1300 [Hz] y 2000 [Hz] es más pronunciado (aproximadamente 3 [dB] más bajo) en el lado izquierdo y ligeramente más hacia la izquierda en el espectro, con un frecuencia más baja alrededor de los 1500 [Hz].

Se presenta un comportamiento similar para la zona entre los 2000 [Hz] y 5000 [Hz] entre ambos canales, con un valle entre los 2500 y 4000 [Hz] más pronunciado para el canal derecho (3 [dB] más aproximadamente) con un pico sobre saliente en los 3650 [Hz]. A partir de los 5000 [Hz] se pueden apreciar las diferencias notables entre los espectros.

En el canal izquierdo se presenta un valle entre los 5300 [Hz] y los 10500 [Hz] con valles notables con frecuencias centrales en 7150 [Hz] (-48,6 [dB]), 8200 [Hz] (-53 [dB]) y 9750 [Hz] (-50 [dB]). Luego se presenta un valle en los 12450 [Hz] y el pico más alto del espectro entre los 13000 [Hz] y 14000 [Hz] para posteriormente presentar un decaimiento con valles considerables en 16500 [Hz] (-50 [dB]), 18500 [Hz] (-58 [dB]) y 19500 [Hz] (-55 [dB]) para presentar un pico en los 19800 [Hz] (-40 [dB]).

En el canal derecho a partir, entre los 5000 [Hz] y 12000 [Hz], se presenta un decaimiento constante de nivel y una serie de picos y valles cada vez más

pronunciados, ocurriendo el primero en 7400 [Hz], el segundo en 8250 [Hz], el tercero en 9350 [Hz] y el último en 9900 [Hz], con una caída de -52 [dB], -55 [dB], -59 [dB] y -61 [dB] respectivamente. Inmediatamente el nivel se recupera a su mayor valor, alrededor de los -33[dB], entre 11000 [Hz] y 15300 [Hz], para presentar un valle (-52 [dB]) alrededor de los 17300 [Hz] y una subida en el fin del espectro y un último valle mínimo (de -42 [dB]) en 19500 [Hz].

b) Elevación 0°, Acimut 0°



Figura 50. Espectro sonoro de la señal obtenida en la posición acimut 0°, elevación 0° para los canales izquierdo y derecho respectivamente.

El espectro es considerablemente plano entre los 20 [Hz] y los 1200 [Hz] para ambos canales. Sin embargo, se presentan picos en ambos canales en esta zona del espectro en 425 [Hz] y 680 [Hz] (ambos de -40 [dB]) con un valles alrededor de los 600 [Hz] de -42 [dB] para el canal izquierdo y -40 [dB] para el canal derecho.

A partir de los 1200 [Hz], ambos canales presentan ligeras diferencias pero un comportamiento prácticamente similar hasta los 6000 [Hz] aproximadamente, con la excepción que para el canal derecho, el pico del espectro entre 3100 [Hz] y 3800 [Hz] presenta un nivel cerca de 3 [dB] superior con respecto al canal izquierdo.

Entre los 6300 [Hz] y los 10000 [Hz] el canal izquierdo presenta un espectro rizado seguido de una caída de nivel de aproximadamente -62 [dB] con dos valles cuyas frecuencias son 8950 [Hz] y 9250 [Hz], mientras que el canal

derecho presenta un valle en dicha zona con una caída menor (aproximadamente de -50 [dB]) con valles cuyas frecuencias están en 7150 [Hz], 8150 [Hz] y 9400 [Hz] y dos picos en 7700 [Hz] (-46 [dB]) y 9050 [Hz] (-43 [dB]) que no se aprecian para el canal izquierdo.

A partir de los 10000 [Hz] se presenta una subida de nivel en ambos canales. En el canal izquierdo entre los 10500 [Hz] y 12000 [Hz] el nivel es constante alrededor de los -38 [dB] y presenta un pico máximo entre los 13000 [Hz] y 14000 [Hz] de -33 [dB] para presentar posteriormente una caída considerable de nivel con un valor mínimo de nivel (-72 [dB]) en 16300 [Hz] que luego se recupera cerca de los 17200 [Hz] (-55 [dB]) y finalmente llega a ser -43 [dB] para el extremo del espectro (a partir de 19300 [Hz]) .

Para el canal derecho, la subida de nivel a partir de los 11000 [Hz], presenta un nivel un poco inferior (hasta los -36 [dB]) entre 10800 [Hz] y 14800 [Hz] aproximadamente, y presenta dos últimos valles considerables con frecuencias en 16500 [Hz] (de -50 [dB]) y 19500 [Hz] (-41 [dB]).

c) Elevación 30°, Acimut 0°



Los espectros presentan un comportamiento casi idéntico entre los 20 [Hz] y 700 [Hz]. A partir de éste punto, las diferencias espectrales son considerables entre ambos canales.

Para el canal izquierdo, entre los 500 [Hz] y 1650 [Hz], se produce un valle leve de con un nivel aproximado entre -42 y -44 [dB]. El nivel se recupera y se mantiene entre -40 y 38 [dB] entre los 1650 [Hz] y 5300 [Hz] con valle notorio en 2700 [Hz] de -45 [dB]. El nivel comienza a decaer presentando un nuevo valle en 7900 [Hz] de -50 [dB], para recuperar nuevamente su nivel a -43 [dB] en 8450 [Hz], nuevamente con dos valles cuyas frecuencias son 9350 [Hz] y 9950 [Hz] de -48 y -47 [dB] respectivamente.

El nivel sube a -38 [dB] hasta los 11500 [Hz] y vuelve a presentar una caída en alrededor de 12650 [Hz] de -44 [dB] para nuevamente subir a -37 [dB] entre 13500 [Hz] y 15500 [Hz]. Finalmente el nivel decae cerca de -46 [dB] con dos picos notorios en 17250 [Hz] y 17900 [Hz] (-43 [dB]) y un valle en 19500 [Hz] de -53 [dB].

Para el canal derecho, se produce un valle entre los 700 [Hz] y 1500 [Hz] con un frecuencia de 1000 [Hz] donde el nivel es el menor (-45 [dB] aproximadamente). En la zona del espectro inestable entre los 1600 y 7300 [Hz], el nivel no sobrepasa los -36 [dB] y -42 [dB] como valor de nivel máximo y mínimo respectivamente de los picos y valles que se presentan. Estos valles ocurren en las frecuencias 2230 [Hz], 2750 [Hz], 4300 [Hz], 5950 [Hz], 6550 [Hz] y los picos en 2000 [Hz], 2500 [Hz], 3200 [Hz], 3500 [Hz], 3750 [Hz], 5350 [Hz], 6050 [Hz], 6750 [Hz] y finalmente en 7300 [Hz].

Un valle de -48 [dB] en 7850 [Hz] es seguido de dos picos en 8050 [Hz] y 8650 [Hz] (de aproximadamente -40 [dB]) y dos picos menos pronunciados en 9200 [Hz] (de -42 [dB]) y 10110 [Hz] (de -43 [dB]). El nivel alcanza un valor máximo entre -33 y -34 [dB] en 11700 [Hz] y se mantiene constante y presenta un espectro rizado hasta los 16000 [Hz] para decaer hasta los -40 [dB] aproximadamente a desde los 19300 [Hz], presentando un valle con el nivel más bajo (-55 [dB]) en 18000 [Hz].

d) Elevación 60°, Acimut 0°



Figura 52. Espectro sonoro de la señal obtenida en la posición acimut 0°, elevación 60° para los canales izquierdo y derecho respectivamente.

El comportamiento es prácticamente el mismo entre los 20 [Hz] y 9000 [Hz] a pesar de que no se puede considerar como plano, se puede encontrar una similitud en ambos canales. Entre los 450 [Hz] y 1100 [Hz] se produce un valle de -45 [dB] aproximadamente con un valor mínimo de nivel en 600 [Hz] de -47 [dB], seguido de dos picos (-40 [dB]) con frecuencias centrales en 1200 [Hz] y 1550 [Hz] aproximadamente, para presentar a continuación un nuevo valle cuyas frecuencias son 1700 [Hz] (-45 [dB]) y 2050 [Hz] (-44 [dB]). Entre los 2200 [Hz] y los 9000 [Hz] el espectro del canal izquierdo es constante alrededor de los -40 [dB], sin embargo se presenta una serie de picos y valles de ± 2 [dB]. Para el espectro del canal derecho, el comportamiento es similar a pesar de su nivel se encuentre alrededor de los -38 [dB] y su rizado es más notorio debido a que los picos y valles aparecen más cercanos entre si y presentan diferencias de nivel de hasta ± 4 [dB], además de la aparición de dos valles notorios entre las zonas entre 2850 [Hz] y 3350 [Hz] y entre 4100 [Hz] y 4650 [Hz].

A partir de los 9000 [Hz], el nivel en el canal izquierdo presenta un rizado más pronunciado con una caída de -48 [dB] alrededor de los 9500 [Hz] y una subida constante a partir de dicha frecuencia hasta alcanzar un nivel máximo de -36 [dB] en 13000 [Hz]. Posteriormente se presenta un nuevo valle pronunciado de -55 [dB] en 14100 [Hz] para subir nuevamente a los -40 [dB] en 15600 [Hz] y mantenerse prácticamente constante (nuevamente presentando un rizado en el

espectro) con un valle entre los 18000 [Hz] y 19000 [Hz] de -45 [dB] aproximadamente.

Para el canal derecho, el espectro sigue presentándose como rizado a lo largo del extremo derecho con variaciones de amplitud mínimas. Entre los 9050 [Hz] y 1110 [Hz] se presenta un valle de -43 [dB] aproximadamente y el nivel se recupera a los -38 [dB] entre 11200 [Hz] y 18000 [Hz], con dos picos que alcanzan los -35 [dB] en 13000 [Hz] y 17000 [Hz]. Finalmente, se presenta un caída constante de nivel hasta los -45 [dB] con un pico notorio de -42 [dB] en 19650 [Hz].

e) Elevación 90°, Acimut 0°



Figura 53. Espectro sonoro de la señal obtenida en la posición acimut 0°, elevación 90° para los canales izquierdo y derecho respectivamente.

Los espectros son bastante similares desde los 20 [Hz] hasta aproximadamente los 10000 [Hz]. Ambos lados presentan las mismas características con ligeras variaciones en su amplitud y pequeñas diferencias en esta zona.

Ambos canales muestran un valle entre los 450 [Hz] y 850 [Hz] de -45 [dB] con un valor mínimo de nivel para una frecuencia de 770 [Hz]. Entre los 850 y 1500 [Hz] el nivel es constante en -43 [dB] aproximadamente presentando dos picos de amplitud en 1120 [Hz] (de -41 [dB]) y en 1470 [Hz] (de -42 [dB]), seguidos por un valle entre 1550 [Hz] y 1950 [Hz] de -44 [dB] para el lado izquierdo y -45 [dB] para el derecho.

Entre los 2000 [Hz] y 5500 [Hz], el nivel es inestable y se presenta una serie de picos y valles prácticamente en las mismas frecuencias en ambos canales con unas pocas excepciones pero el nivel es ligeramente superior para el canal derecho, siendo de -42 [dB] con variaciones de ± 2 [dB], mientras que para el canal izquierdo el nivel es de aproximadamente -43 [dB], con variaciones de ± 3 [dB].

A partir de los 5500 [Hz] hasta los 9000 [Hz] el nivel sube y el rizado es más notorio para ambos canales. El nivel comienza a subir llegando a los -39 [dB] aproximadamente para el canal izquierdo y -41 [dB] para el canal derecho, para decaer gradualmente a los -44 [dB] en ambos canales cerca de los 10000 [Hz].

Para el canal izquierdo, el nivel sube nuevamente hasta alcanzar su máximo de -38 [dB] en los 13000 [Hz] y presentar un caída brusca con un valle en 14300 [Hz] de -47 [dB] y nuevamente en 15500 [Hz] cayendo todavía más hasta los -54 [dB]. A partir de este punto, el nivel comienza a subir de forma constante hasta los -43 [dB] en 17900 [Hz] y presenta un valle en 19200 [Hz] de -46 [dB] y un pico pronunciado en 19930 [Hz] de -39 [dB].

Para el canal derecho, a partir de los 9000 [Hz], el nivel decae hasta los -43 [dB] aproximadamente y se mantiene hasta los 12000 [Hz]. A continuación, se presentan una caída gradual con un picos de -44 [dB] en 13000 [Hz], -46 [dB] en 14500 [Hz] y -48 [dB] en 15800 [Hz], con valles considerables en 15100 (-49 [dB]) y alcanza su valor mínimo de nivel de -51 [dB] en 16500 [Hz]. El nivel se recupera alrededor de los -45 [dB] a pesar de que se presentan numerosos valles en el fin del espectro, los más notorios 17500 [Hz] (-48 [dB]), 19050 [Hz] (-49 [dB]) y 19930 [Hz] (-50 [dB]), con un pico sobresaliente de -41 [dB] en los 19780 [dB].

2.3.2.3. Resumen y Comparación de Posiciones de Elevación

Se puede apreciar las diferencias espectrales más significativas entre las distintas posiciones al variar la elevación y visualizarlos gráficamente de forma continua como muestras de 1 [s] de duración comparándolas con el espectro del ruido blanco.

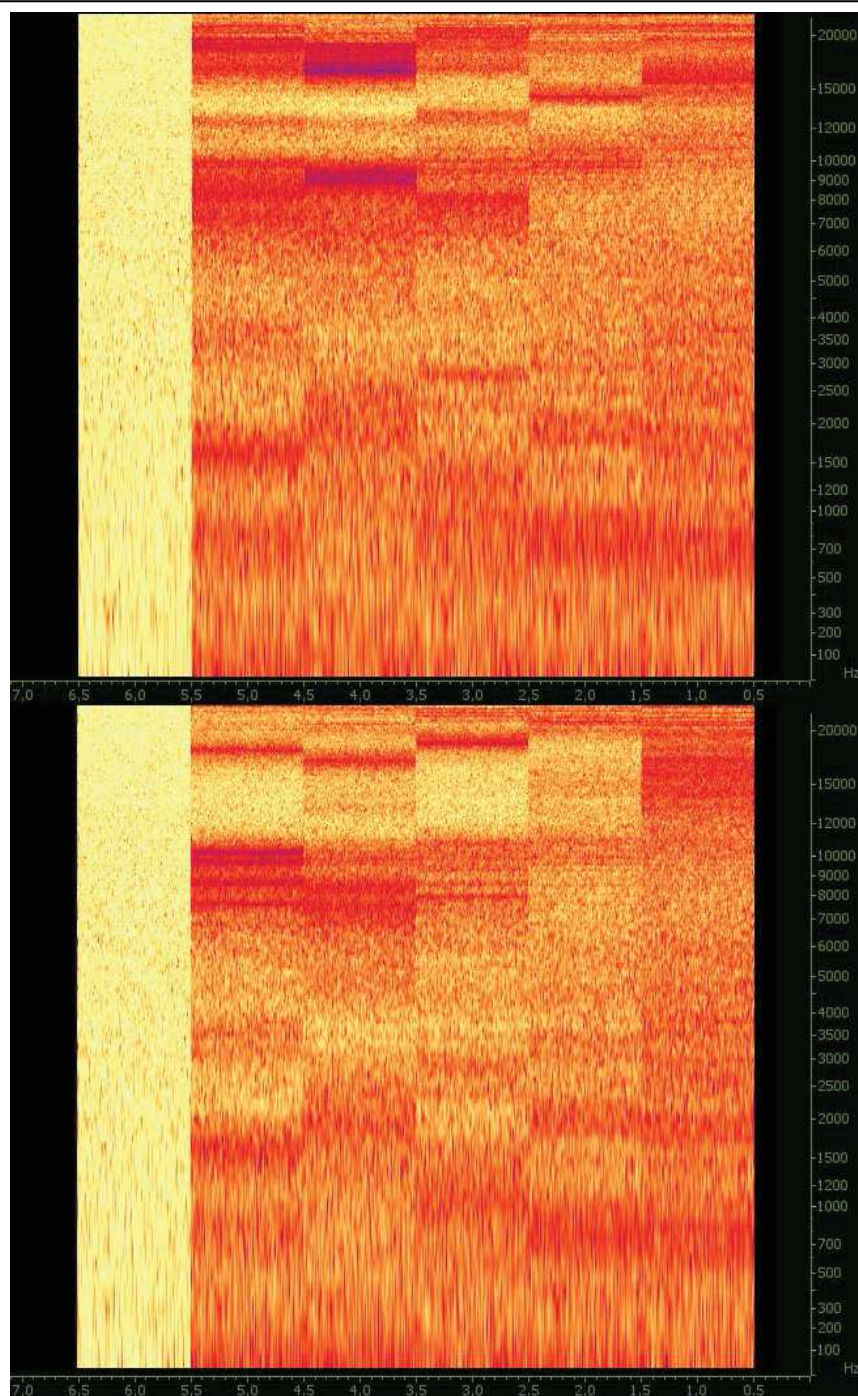


Figura 54. Comparación de espectro entre la señal de entrada seguida de todas las posiciones de elevación para los canales izquierdo (arriba) y derecho (abajo) del Individuo 2.

Se presenta también los resultados de las pruebas de elevación del Individuo 2 en forma de una tabla que permite evidenciar un resumen de la información

más relevante de los espectros obtenidos para cada posición considerada, así como realizar una comparación entre estos, sus canales y evidenciar las similitudes encontradas.

Tabla 3. Comparación de Espectros de Elevación del Individuo 2.

	Similitudes.	Canal Izquierdo.	Canal Derecho.
Elevación -30° , Acimut 0° .	Zona entre 20 y 5000 [Hz] con valle entre 1300 y 2000 [Hz] (-3 [dB] izquierda) y valle entre 2500 y 4000 [Hz] (+3 [dB] derecha).	Decaimiento de nivel entre 5300 y 10500 [Hz] con valles en 8200 [Hz] (-53 [dB]) y 9750 [Hz] (-50 [dB]) y pico entre 13000 [Hz] y 14000 [Hz]. Valles en 16500 [Hz] (-50 [dB]), 18500 [Hz] (-58 [dB]) y 19500 [Hz] (-55 [dB]).	Decaimiento de nivel entre 5000 y 12000 [Hz] con valles en 7400 [Hz] (-52 [dB]), 8250 [Hz] (-55 [dB]), 9350 [Hz] (-59 [dB]), 9900 [Hz] (-61 [dB]). Pico entre 11000 y 15300 [Hz] (33[dB]). Valle en 17300 [Hz] (-52[dB]) y 19500 [Hz] (de -42 [dB]).
Elevación 0° , Acimut 0° .	Entre 20 y 6000 [Hz] con pico entre 3100 y 3800 [Hz] (+3 [dB] derecha).	Ligero rizado entre 6300 y 10000 [Hz] y caída de nivel (-62 [dB]) con valles en 8950 [Hz] y 9250 [Hz]. Nivel constante entre 10500 y 12000 [Hz]. Subida (-38 [dB]) y pico entre	Entre 6300 y 10000 [Hz] caída (-50 [dB]) y valles en 7150 [Hz], 8150 [Hz] y 9400 [Hz] y picos en 7700 [Hz] (-46 [dB]) y 9050 [Hz] (-43 [dB]). Subida de nivel desde 11000 [Hz] hasta

		13000 [Hz] y 14000 [Hz] (-33 [dB]). Caída en 16300 [Hz] (-72 [dB]) y subida en 19300 [Hz] (-43 [dB]).	(-36 [dB]) entre 10800 y 14800 [Hz]. Valle en 16500 [Hz] (de -50 [dB]).
Elevación 30°, Acimut 0°.	Entre 20 y 700 [Hz]. Valle entre 500 y 1650 [Hz] de -42 [dB] (-3 [dB] en 1000 [Hz] en canal derecho).	Nivel se mantiene entre -39 entre 1650 y 5300 [Hz] con valle en 2700 [Hz] de -45 [dB]. Valle en 7900 [Hz] (-50 [dB]) y subida a -38 [dB] en 11500 [Hz] con valles en 9350 y 9950 [Hz] de (- 48 y -47 [dB]). Alcanza -37 [dB] entre 13500 [Hz] y 15500 [Hz] y decae a -46 [dB],	Valles en 2230 [Hz], 2750 [Hz]. Picos en 8050 [Hz] y 8650 [Hz] (-40 [dB]). Nivel máximo (-33 [dB]) en 11700 [Hz]. Rizado hasta 16000 [Hz] y decaimiento a -40 [dB] en 19300 [Hz] con un valle de -55 [dB] en 18000 [Hz].
Elevación 60°, Acimut 0°.	Entre 20 [Hz] y 9000 [Hz]. Valle entre 450 y 1100 [Hz] (-45 [dB]) y dos picos (-40 [dB]) en 1200 y 1550 [Hz]. Valle	A partir de 9000 [Hz], rizado más pronunciado con valle (-48 [dB]) entre 9500 y 13000 [Hz] (-36 [dB]). Valle de -55	Valles entre 2850 y 3350 [Hz] y entre 4100 y 4650 [Hz]. Rizado en extremo derecho con variaciones mínimas. Valle

		<p>en 1700 [Hz] (-45 [dB]) y 2050 [Hz] (-44 [dB]). Entre 2200 y 9000 [Hz] nivel constante y rizado (-40 [dB] izquierda y -38 derecha). Rizado con picos y valles más cercanos en derecha, con diferencias de nivel de ± 4 [dB], mientras que solo de ± 2 [dB] en la izquierda.</p>	<p>[dB] en 14100 [Hz] y subida a -40 [dB] en 15600 [Hz]. Valle entre 18000 y 19000 [Hz] (-45 [dB]).</p>	<p>entre 9050 y 1110 [Hz] (-43 [dB]) y alcanza -38 [dB] entre 11200 y 18000 [Hz], con dos picos de -35 [dB] en 13000 y 17000 [Hz]. Caída constante a -45 [dB].</p>
<p>Elevación 90°, Acimut 0°.</p>	<p>Entre 20 y 10000 [Hz]. Valle entre 450 y 850 [Hz] (-45 [dB]). Picos en 1120 [Hz] (-41 [dB]) y 1470 [Hz] (-42 [dB]). Nivel inestable entre 2000 y 5500 [Hz], -42 [dB] con variaciones de ± 2 [dB] a la derecha y -43 [dB] con variaciones de ± 3 [dB] a la izquierda.</p>	<p>Nivel sube a -38 [dB] en 13000 [Hz]. Cae en 14300 [Hz] a -47 [dB] y en 15500 [Hz] a -54 [dB]. El nivel sube hasta -43 [dB] en 17900 [Hz]. Valle en 19200 [Hz] (-46 [dB]) y en 19930 [Hz] (de -39 [dB]).</p>	<p>Nivel decae a partir de 9000 [Hz] hasta -43 [dB], se mantiene hasta 12000 [Hz]. Caída gradual con un picos hasta -51 [dB] en 16500 [Hz]. Subida a -45 [dB] con valles y un pico de -41 [dB].</p>	

	<p>Entre 5500 y 9000 [Hz] subida a -39 [dB] (izquierda) y -41 [dB] (derecha). Rizado mas notable. Decaimiento a -44 [dB] cerca de los 10000 [Hz].</p>		
--	---	--	--

Nota: Los valores presentados en la tabla no son exactos, han sido aproximados de la lectura obtenida del software y resumidos para una mejor presentación en la tabla.

2.3.2.4. Pruebas de Acimut

Tal como ocurre en el análisis de las muestras de comparación de posiciones de distinta elevación, se presenta un filtrado de bajas frecuencias común para todas las muestras de posiciones de acimut bajo los 160 [Hz] aproximadamente, a modo de un filtro pasa bajos.

La diferencia se encuentra que el filtrado es ligeramente más profundo entre los canales izquierdo y derecho para una misma posición de acimut de acuerdo a la posición evaluada. Esto quiere decir que para posiciones de la fuente a la izquierda (entre 0° y 180°) el nivel tiende a ser superior, alrededor de no más 3 [dB] para la zona de bajas frecuencias. Lo mismo sucede para posiciones de la fuente en el lado derecho (entre 180° y 360°).

A partir de los 160 [Hz], los espectros tienden a diferir considerablemente entre ambos canales, donde una diferencia de nivel considerable entre éstos de acuerdo a la posición evaluada es evidente.

a) Acimut 0°, Elevación 0°

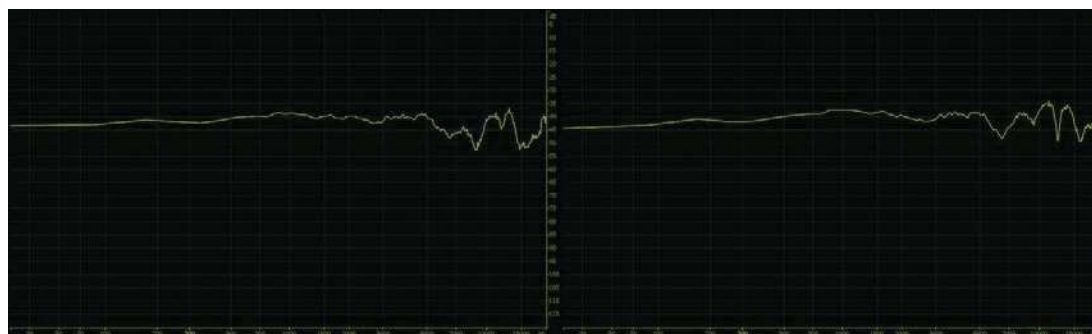


Figura 55. Espectro sonoro de la señal obtenida en la posición acimut 0°, elevación 0° para los canales izquierdo y derecho respectivamente.

Los espectros presentan un comportamiento bastante similar entre los 20 [Hz] y 6500 [Hz] con ciertas diferencias de nivel, siendo por lo general ligeramente mayor en nivel para el canal derecho. Se presenta una subida gradual de 3 [dB] desde los 350 [Hz] hasta los 600 [Hz]. Después se produce un subida a los -43 [dB] (entre 750 [Hz] y 850 [Hz]), para mantenerse alrededor de los -40 [dB] entre los 950 [Hz] y 5000 [Hz]. Sin embargo, se presentan valles alrededor de los 1350 [Hz] (-43 [dB]) y 1850 [Hz] (-42 [dB]), con la diferencia de que este último es más pronunciado y ocupa una mayor zona del espectro para el canal derecho, entre los 2000 [Hz] y 3100 [Hz], con un nivel de -43 [dB].

Otro valle se presenta para el canal izquierdo alrededor de los 2700 [Hz] (-44 [dB]) mientras el espectro comienza a mostrarse como rizado para ambos canales. Entre los 5000 [Hz] y 6500 [Hz] se presenta una caída de -9 [dB], y un nuevo pico entre 6000 [Hz] y 8850 [Hz] que alcanzan los -45 [dB] para el canal izquierdo y -40 [dB] para el canal derecho, presentando un valle en 8850 [Hz] (-54 [dB]) y 9400 [Hz] (-44 [dB]) respectivamente.

Las diferencias más claras se presentan cuando el nivel llega a su máximo para ambos canales entre 10000 [Hz] y 13500 [Hz], alcanzando los -41 [dB] aproximadamente para el canal izquierdo y los -38 [dB] para el derecho con un valle en 12100 [Hz] (-46 [dB]) y 12500 [Hz] (-50 [dB]) respectivamente. El canal

izquierdo presenta una caída brusca en los 15000 [Hz] (-55 [dB]) con una subida rizada y gradual hasta alcanzar un pico de -41 [dB] en los 19600 [Hz], mientras que el canal derecho presenta una caída similar en los 16000 [Hz] (-50 [dB]) con una subida más leve hasta los -45 [dB] con un rizado más profundo hasta el límite superior del espectro.

b) Acimut 45°, Elevación 0°



Figura 56. Espectro sonoro de la señal obtenida en la posición acimut 45°, elevación 0° para los canales izquierdo y derecho respectivamente.

Los espectros son similares únicamente en la región entre 170 [Hz] y 400 [Hz] aproximadamente, donde el espectro es plano con un nivel de -40 [dB] para el canal izquierdo y -43 [dB] para el canal derecho, a pesar de que el espectro plano se mantiene constante hasta los 1050 [Hz] en el canal izquierdo. A partir de este punto, el canal izquierdo presenta un valle entre las frecuencias 1050 [Hz] y 1450 [Hz] de -42 [dB], seguido de una subida que alcanza los -36 [dB] en 2050 [Hz]. Un segundo valle entre 2450 [Hz] y 3450 [Hz] con un mínimo de -42 [dB] en 2800 [Hz]. Entre 3800 [Hz] y 5500 [Hz] el nivel se mantiene constante con ligeros picos y valles alrededor de los -36 [dB].

El nivel comienza a decaer hasta los -47 [dB] en 10750 [Hz], con varios picos notorios en 7000 [Hz] (-38 [dB]), 7650 [Hz] (-41 [dB]), 8450 [Hz] (-40 [dB]), 9070 (-39 [dB]) y finalmente en 10350 [Hz] (-43 [dB]). Posteriormente el nivel sube hasta -35 [dB] en 11750 [Hz] y decae a -40 [dB] en 13050 [Hz] para volver a subir hasta alcanzar su máximo de -32 [dB] en 15500 [Hz]. Finalmente se

presenta un valle en 18150 [Hz] (-46 [dB]) y se recupera en un pico de -38 [dB] a partir de 19100 [Hz].

En el canal derecho, el nivel comienza a decaer desde los 350 [Hz] hasta los 850 [Hz] aproximadamente -3 [dB]. El primer valle ocurre entre 850 [Hz] y 1300 [Hz] (-49 [dB]). El nivel sube hasta los -42 [dB] en 1650 [Hz]. Se presentan un valle entre 2350 [Hz] y 2850 [Hz] que alcanza un valor mínimo de nivel de -49 [dB] en 2600 [Hz]. Dos vales mas se presentan con frecuencias centrales en 3300 [Hz] (-46 [dB]) y 3600 [Hz] (-47 [dB]).

Entre los 4000 [Hz] y 5700 [Hz] el nivel es aproximadamente -44 [dB] y comienza a caer en forma de picos y valles bastante pronunciados en la región entre 5800 y 10600 [Hz]. Las frecuencias de los valles ocurren en 5800 [Hz] (-47 [dB]), 6500 [Hz] (-50 [dB]), 7000 [Hz] (-53 [dB]), 7500 [Hz] (-55 [dB]), 8050 [Hz] (-58 [dB]), 8750 [Hz] (-56 [dB]), 9200 [Hz] (-58 [dB]), 9650 [Hz] (-54 [dB]) y 10100 [Hz] (-58 [dB]). Los picos se presentan entre dichas frecuencias con un nivel aproximado de -48 [dB], con un máximo de -44 [dB] en 8400 [Hz]. A partir de los 11100 [Hz] el nivel se encuentra alrededor de -42 [dB] hasta los 15500 [Hz], con picos y valles que no sobrepasan los 2 [dB]. Finalmente se presenta una caída de nivel hasta los -62 [dB] en 17300 [Hz] seguido de una subida a los -44 [dB] alrededor de los 19100 [Hz].

c) Acimut 90°, Elevación 0°



Figura 57. Espectro sonoro de la señal obtenida en la posición acimut 90°, elevación 0° para los canales izquierdo y derecho respectivamente.

Los espectros presentan similitud únicamente en la región entre los 20 [Hz] y los 300 [Hz] aproximadamente, a pesar que los niveles son mayores para el canal izquierdo por una diferencia cercana a los 2 [dB].

El espectro del canal izquierdo es más estable en general. Entre los 300 [Hz] y 700 [Hz] se presenta un incremento del nivel desde -40 [dB] hasta -36 [dB], seguido por un primer valle entre las frecuencias de 700 [Hz] y 1110 [Hz] (-38 [dB]), un segundo valle entre 1110 [Hz] y 1650 [Hz] con -40 [dB] (en 1350 [Hz]) y un tercer valle entre 1650 [Hz] y 2050 [Hz] con un nivel de -41 [dB] aproximadamente. El nivel sube y se mantiene alrededor de los -37 [dB] a partir de los 2150 [Hz] hasta 3100 [Hz] con un valle con una frecuencia central en 2950 [Hz] de amplitud -39 [dB] seguido de otro valle entre 3100 [Hz] y 4500 [Hz] que alcanza su nivel mínimo en de -41 [dB] en 3650 [dB].

A partir de los 4500 [Hz] hasta los 8600 [Hz] el nivel es aproximadamente de -36 [dB] presentando un rizado en el espectro con una serie de picos y valles con una variación máxima de 2 [dB] de amplitud. Entre los 8600 [Hz] y 11300 [Hz] el nivel cae a los -41 [dB] y consecuentemente presenta una caída brusca de nivel en un dos valles con frecuencias en 12000 [Hz] y 13600 [Hz] ambos de -55 [dB] (con un pico entre éstos de -44 [dB] en los 12600 [Hz]). La amplitud comienza a recuperar su valor entre los 13600 [Hz] hasta los 17000 [Hz] donde alcanza un valor máximo de -33 [dB] que se mantiene hasta los 19000 [Hz] que finalmente cae hasta llegar a los 20000 [Hz].

Para el canal derecho, el nivel se mantiene estable entre desde los 160 [Hz] hasta los 350 [Hz] en -42 [dB]. Entre los 350 [Hz] y 700 [Hz] se produce una caída de nivel a -44 [dB] seguido por un valle con una frecuencia central en 780 [Hz] (-45 [dB]). A partir de los 860 [Hz] donde el nivel es de -42 [dB], el nivel comienza a caer hasta los -57 [dB] en 2400 [Hz] presentando varios picos y valles en dicha zona. Al llegar a los 3000 [Hz] el nivel vuelve a subir a los -46 [dB], seguido por un valle con frecuencia en 3300 [Hz] (-52 [dB]) y un nuevo pico de -48 en 3550 [Hz] para volver a caer a los -55 [dB] en 4200 [Hz]. El nivel vuelve a recuperarse hasta llegar a -52 [dB] en 4500 [Hz] y cae a los -63 [dB] en 6650 [Hz].

La zona del espectro entre 6600 [Hz] y 11300 [Hz] presenta una serie de picos y valles pronunciados con significativas variaciones de nivel. Los valles más notorios se presentan en 6700 (-64 [dB]), 7500 [Hz] (-67 [dB]), 8250 [Hz] (-64 [dB]), 9650 [Hz] (-67 [dB]), 10100 [Hz] (-66 [dB]) y finalmente en 11300 [Hz] (63 [dB]). Los picos entre dichos valles poseen un nivel aproximado de -53 [dB] y ocurren en 7050 [Hz], 8000 [Hz], 9000 [Hz], 9450 [Hz], 10150 [Hz], 11000 [Hz].

El nivel se sube a -53 [dB] para una frecuencia de 12000 [Hz], y se presenta un valle entre 12580 [Hz] y 14100 [Hz]. Finalmente se presenta una caída de nivel que alcanza los -66 [dB] en 17200 [Hz] con un par de picos intermedios en 15500 [Hz] (-57 [dB]) y 16500 [Hz] (-60 [dB]) y vuelve a subir su nivel hasta el extremo superior del espectro donde el nivel es de -47 [dB] en 19700 [Hz], con un pico intermedio en 17800 [Hz] de -57 [dB].

d) Acimut 135°, Elevación 0°



Los espectros son similares únicamente en la zona entre los 20 y 160 [Hz], siendo el del canal izquierdo ligeramente superior en amplitud por 2 [dB] aproximadamente.

A partir de los 160 [Hz], el espectro izquierdo presenta una subida ligera de nivel que alcanza los -38 [dB] en 850 [Hz] para comenzar a decaer hasta llegar a los -42 [dB] en 3800 [Hz]. Algunos picos se presentan en esta zona en 1650 [Hz] (-38 [dB]), 2000 [Hz] (-39 [dB]) y 3100 [Hz] (-40 [dB]), con dos valles

significativos entre 1700 [Hz] y 2000 [Hz] y entre 2300 [Hz] y 3050 [Hz] que alcanzan niveles de -42 [dB] y -40 [dB] respectivamente.

Entre 3800 [Hz] y 4600 [Hz] se presenta una caída de nivel que llega a un nivel de -48 [dB] seguido de una subida que alcanza los -33 [dB] en 7700 [Hz] con un rizado ligero de variaciones de amplitud mínima. Un valle notorio se produce entre 8600 [Hz] y 10450 [Hz] que entrega un valor de -43 en 9300 [Hz]. Un segundo valle ocurre entre 13100 [Hz] y 14600 [Hz] que llega a los -48 [dB] en 13700 [Hz]. Para finalizar, la amplitud recupera su valor hasta llegar a los -33 [dB] nuevamente en un pico alrededor de los 18600 [Hz] y un ligero decremento (-3 [dB]) de ésta en la última región del espectro.

Para el espectro del canal derecho, el nivel decae 4 [dB] entre 160 [Hz] y 860 [Hz], seguido de una caída más pronunciada entre los 860 [Hz] y 2200 [Hz], donde se registra un valor de -59 [dB]. En esta región, se presentan picos notables en 1110 [Hz] (-47 [dB]), 1550 [Hz] (-52 [dB]), 1800 [Hz] (-55 [dB]) y valles en 1750 [Hz] (-58 [dB]) y 2200 [Hz] (-59 [dB]). El nivel se recupera hasta alcanzar los -52 [dB] en 2450 [Hz] y se mantiene prácticamente constante hasta los 4000 [Hz]. Entre los 4000 [Hz] y los 5350 [Hz] se produce un valle con un frecuencia central en 4500 [Hz] (-63 [dB]).

El nivel vuelve a recuperarse para mantenerse en -56 [dB] entre 5000 [Hz] y 6200 [Hz] y un nuevo valle se produce con una frecuencia en 6600 [Hz] (-68 [dB]), seguida por un pico que alcanza los -61 [dB] en 7600 [Hz] y vuelve a caer a -72 [dB] en 8250 [Hz]. En 9000 [Hz] aproximadamente, el nivel sube a -54 [dB], vuelve a caer a -60 [dB] en 10500 [Hz] y presenta su valle más profundo en 12000 [Hz] de -74 [dB]. Nuevamente el nivel sube hasta -54 [dB] en 14000 [Hz] y cae a -63 en 16650 [Hz] para presentar finalmente un pico entre 17000 [Hz] y 18000 [Hz] seguido de dos picos sobresalientes en el extremo de -47 [dB] en 19400 [Hz] y -49 [dB] en 19900 [Hz].

e) Acimut 180°, Elevación 0°



Figura 59. Espectro sonoro de la señal obtenida en la posición acimut 180°, elevación 0° para los canales izquierdo y derecho respectivamente.

Los espectros presentan un comportamiento similar en ambos canales entre los 20 y 260 [Hz].

En el canal izquierdo, el nivel incrementa hasta los -39 [dB] hasta los 1150 [Hz], seguido por valles en 1400 [Hz] (-42 [dB]), entre 1700 [Hz] y 1900 [Hz] (-41 [dB]) y entre 2500 [Hz] y 3050 [Hz] (-41 [dB]). Entre los 3150 [Hz] donde se registra un nivel de aproximadamente -41 [dB], comienza una caída hasta los -50 [dB] en 5850 [Hz] con varios picos y valles intermedios de ± 1 [dB].

Entre 6000 [Hz] y 9000 [Hz] el nivel se mantiene alrededor de -49 [dB] con picos notorios en 6150 [Hz], 7750 [Hz], 8500 [Hz] que alcanzan los -47 [dB]. A continuación se presentan dos valles profundos que alcanzan sus mínimos en 9500 [Hz] (-57 [dB]) y 9800 [Hz] (-58 [dB]). El nivel se recupera y alcanza sus máximos en dos picos 11600 [Hz] y 13400 [Hz] (de -42 [dB]). Finalmente se presenta el valle de mayor profundidad en 14500 [Hz] (-60 [dB]) que sube gradualmente a los 16100 [Hz] (-47 [dB]) y por último, alcanza los -41 [dB] en 17800 [Hz], presentando una caída final a los -45 [dB] cerca de los 19400 [Hz].

Para el canal derecho, entre los 260 [Hz] y 850 [Hz] se presenta un valle donde el nivel disminuye en un máximo de 2 [dB]. Entre los 1200 [Hz] y 3100 [Hz], el espectro presenta una serie de picos y valles, entre los que más destacan los valles que ocurren en las frecuencias de 1400 [Hz], 1700 [Hz], 1900 [Hz] y 2300

[Hz] (todos a un nivel de -44 [dB]) mientras que los picos más prominentes ocurren en 1550 [Hz] (-41 [dB]), 2100 [Hz] (-42 [dB]). En 3650 [Hz] se presenta un valle de -47 [dB] que termina en un pico de -45 [dB] en 3800 [Hz]. A partir de esta frecuencia comienza a producirse una caída del nivel progresiva hasta alcanza los -62 [dB] en 7700 [Hz], con una serie de picos y valles en su trayecto. El nivel se recupera hasta los -49 [dB] en 9150 [Hz] y nuevamente a -46 [dB] en 10100 [Hz] donde se mantiene constante hasta los 12000 [Hz].

Un nuevo valle se presenta entre 12000 [Hz] y 15500 [Hz], donde el nivel es de -52 en 13400 [Hz]. Por último, el nivel se mantiene constante alrededor de los -47 [dB] a partir de los 16600 [Hz] con la excepción de dos valles de -54 [dB], en los 19500 [Hz] y en 19950 [Hz].

f) Acimut 225°, Elevación 0



Figura 60. Espectro sonoro de la señal obtenida en la posición acimut 225°, elevación 0° para los canales izquierdo y derecho respectivamente.

Los espectros son similares únicamente en la región entre los 20 [Hz] y 160 [Hz] con la diferencia que para el canal derecho es ligeramente más alto en su nivel (aproximadamente 1 [dB]).

El canal izquierdo presenta una caída de nivel entre los 170 [Hz] y 1400 [Hz] de 3 [dB] con un valle notorio entre 520 [Hz] y 880 [Hz] con un nivel de 47 [dB] en 700 [Hz]. A esto le siguen un pico cuya frecuencia central está en 1550 [Hz] (-43 [dB]). A continuación se presenta una caída de nivel que llega a los -53 [dB] en 2500 [Hz] que recupera su valor de amplitud a -47 en 3500 [Hz]. Vuelve a

presentar una caída que registra -52 [dB] en 4400 [Hz] y una subida a -49 [dB] en 5300 [Hz]., seguido de otro decremento de nivel a -55 [dB] en 5850 [Hz] con un incremento más estable que alcanza los -51 [dB] en 8600 [Hz] y un último decremento a -54 [dB] en 9350 [Hz] y un incremento de nivel en 11300 [Hz] (-48 [dB]).

Entre los 11500 [Hz] y 19700 [Hz] aproximadamente, el espectro presenta un comportamiento bastante irregular en forma de un valle profundo que abarca el ancho de la región con varios picos y valles con significativa variación de amplitud. Los picos ocurren en 12600 (-57 [dB]), 13800 [Hz] (-62 [dB]), 14800 (-63 [dB]), 15650 [Hz] (-62 [dB]), 17700 [Hz] (-62 [dB]), 19000 [Hz] (-60 [dB]) y finalmente en 19700 (-51 [dB]). Los valles aparecen en 12400 (-58 [dB]), 13400 (-65 [dB]), 14350 [Hz] (-72 [dB]), 16700 [Hz] (-72 [dB]) y 18 450 (-68 [dB]).

Para el canal derecho, a partir de los 170 [Hz], el nivel incrementa 3 [dB] hasta los 870 [Hz] y comienza a decaer hasta los -40 [dB] en 2300 [Hz], con un valles notorio cuya frecuencia central se encuentra cerca de los 1840 [Hz] (-43 [dB]). Entre los 2300 [Hz] y 3500 [Hz] el nivel es prácticamente constante presenta una caída hasta los -48 [dB] en 4650 [Hz] para comenzar a subir hasta -38 [dB] en 7600 [Hz].

Entre los 7500 [Hz] y 20000 [Hz] aproximadamente el espectro se conforma de zonas de frecuencia conformadas por picos y valles de una banda de frecuencia considerable. Los picos presentan sus máximos en 7800 [Hz] (-38 [dB]), 10800 [Hz] (-39 [dB]), 14000 [Hz] (-40 [dB]) y 19600 [Hz] (-43 [dB]), mientras que los valles se encuentran entre estos y sus valores mínimos de nivel ocurren en sus frecuencias centrales que son 9550 [Hz] (-60 [dB]), 12300 [Hz] (-50 [dB]) y 16750 [Hz] (-65 [dB]).

g) Acimut 270°, Elevación 0°

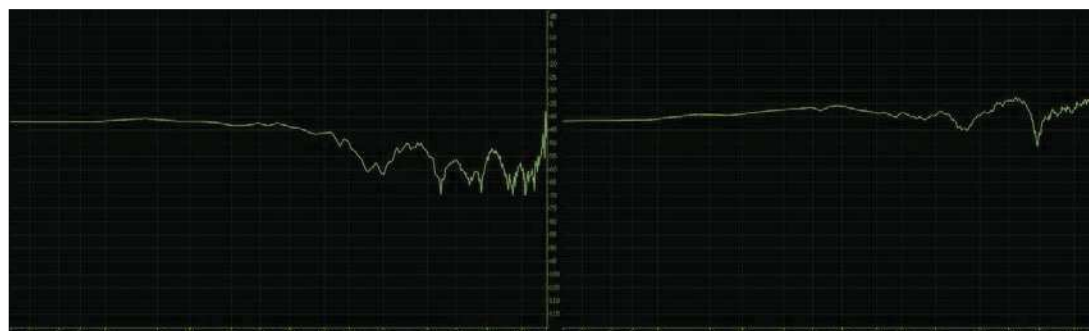


Figura 61. Espectro sonoro de la señal obtenida en la posición acimut 270°, elevación 0° para los canales izquierdo y derecho respectivamente.

Las semejanzas en los espectros para los canales izquierdo y derecho ocurren únicamente entre los 20 [Hz] y 160 [Hz] aproximadamente, siendo levemente superior en nivel para el canal derecho.

Para el canal izquierdo, el nivel decae cerca de 2 [dB] entre 160 [Hz] y 850 [Hz] y presenta una caída de -4 [dB] hasta alcanzar los -47 [dB] alrededor de los 1600 [Hz], seguido de un valle en 1800 [Hz] de -53 [dB]. Entre los 1900 [Hz] y 4500 [Hz] se presenta un valle considerable que alcanza los -63 [dB] en 2500 [Hz] y nuevamente en los 3000 [Hz] (con un pico intermedio de -59 [dB] en 2750 [Hz]). El nivel vuelve a caer hasta un mínimo de -71 [dB] en 6000 [Hz] en un segundo valle entre 5300 [Hz] y 6600 [Hz]. Un tercer valle ocurre entre 7300 [Hz] y 9900 [Hz] con un pico de -61 [dB] en 8850 [Hz] entre 8200 [Hz] (-67 [dB]) y 9450 [Hz] (-70 [dB]). El nivel se recupera y alcanza los -53 [dB] en la región entre los 10300 [Hz] y 11300 [Hz].

Una nueva caída ocurre y la amplitud registra un valor de -69 [dB] en 13000 [Hz], seguido por un pico en 13100 (-63 [dB]) para presentar un incremento hasta los -58 [dB] en 15000 [Hz]. Una vez más, el nivel decae a -70 [dB] en 15700 [Hz] y presenta un incremento gradual en forma de picos pronunciados hasta alcanzar los -39 [dB] en 19950 [Hz].

En el espectro derecho, a partir de los 160 [Hz] y hasta alcanzar los 1000 [Hz] aproximadamente, el nivel incrementa en 4 [dB], para proceder a caer 10 [dB] (hasta los -47 [dB]) en 4400 [Hz], con la aparición de un pico considerable en 3250 [Hz] de -39 [dB].

Entre los 4200 [Hz] y 8400 [Hz] el nivel incrementa hasta los -34 [dB] de forma gradual hasta que ocurre un valle profundo con un mínimo de nivel de -53 [dB] en 9850 [Hz]. Una vez más el nivel comienza a recuperarse hasta el extremo derecho del espectro de forma gradual presentando un rizado con varios picos y valles de amplitud relativamente pequeña.

h) Acimut 315°, Elevación 0°



Figura 62. Espectro sonoro de la señal obtenida en la posición acimut 315°, elevación 0° para los canales izquierdo y derecho respectivamente.

La respuesta entre 20 [Hz] y 160 [Hz] es similar entre ambos canales pero la diferencia de nivel entre estos es de 2 [dB] siendo mayor para el canal derecho.

En el canal izquierdo, el nivel decae desde los -41 [dB] en 160 [Hz] hasta los -46 [dB] en 1150 [Hz]. Se presenta un valle en una región amplia del espectro entre 1150 y 3050 [Hz] con mínimos en 1400 [Hz] y 1800 [Hz] de -58 [dB] con un pico intermedio de -51 [dB] en 1630 [Hz]. El nivel se recupera y alcanza los -45 [dB] en 3100 [Hz] y vuelve a caer hasta los -64 [dB] en 6700 [Hz], con varios picos intermedios en 3700 [Hz] (-47 [dB]), 4750 [Hz] (-50 [dB]) y 5350 [Hz] (-51 [dB]). Entre los 6700 [Hz] y 12400 [Hz] se evidencia una zona de numerosos picos y valles con variaciones de nivel en el orden entre los -55 [dB] y -65 [dB]

cada 400 [Hz] aproximadamente, presentando un valle final en 12400 de -67 [dB].

El nivel se recupera a un valor de -55 [dB] en 13300 [Hz] y se mantiene hasta los 15100 [Hz] y vuelve a decaer hasta alcanzar su mínimo de -70 en 16000 [Hz]. A continuación se presenta una subida de nivel en forma de picos notorios en 17200 [Hz] (-58 [dB]), 18200 [Hz] (-56 [dB]) y un máximo de -41 [dB] en 19920 [Hz].

El espectro del canal derecho es más estable. A partir de los 160 [Hz] donde se registra un nivel de -39 [dB], el nivel incrementa 1 [dB] hasta los 870 [Hz]. Un valle ocurre entre los 1300 [Hz] y 1650 [Hz] con un nivel de -42 [dB] en 1450 [Hz] y alcanza los -35 [dB] en 2550 [Hz]. Otro valle aparece entre los 2550 [Hz] y 4600 [Hz] con una disminución de 3 [dB]. En la región comprendida entre 4700 [Hz] y 9000 [Hz] el nivel es prácticamente constante alrededor de los -35 [dB].

Se presentan los valles, el primero que alcanza un mínimo de -51 [dB] en 9950 [Hz] y el segundo de -49 [dB] en 12650 con un pico intermedio entre dichas frecuencias con un nivel de -44 [dB] aproximadamente. Finalmente, el nivel se recupera gradualmente a partir de los 12700 [Hz] hasta los 15600 [Hz] donde se registran -36 [dB] y un ligero decremento de 3 [dB] al llegar a los 19900 [Hz].

2.3.2.5. Resumen y Comparación de Posiciones de Acimut

Se puede apreciar las diferencias espectrales más significativas entre las distintas posiciones al variar la acimut y visualizarlos gráficamente de forma continua como muestras de 1 [s] de duración comparándolas con el espectro del ruido blanco.

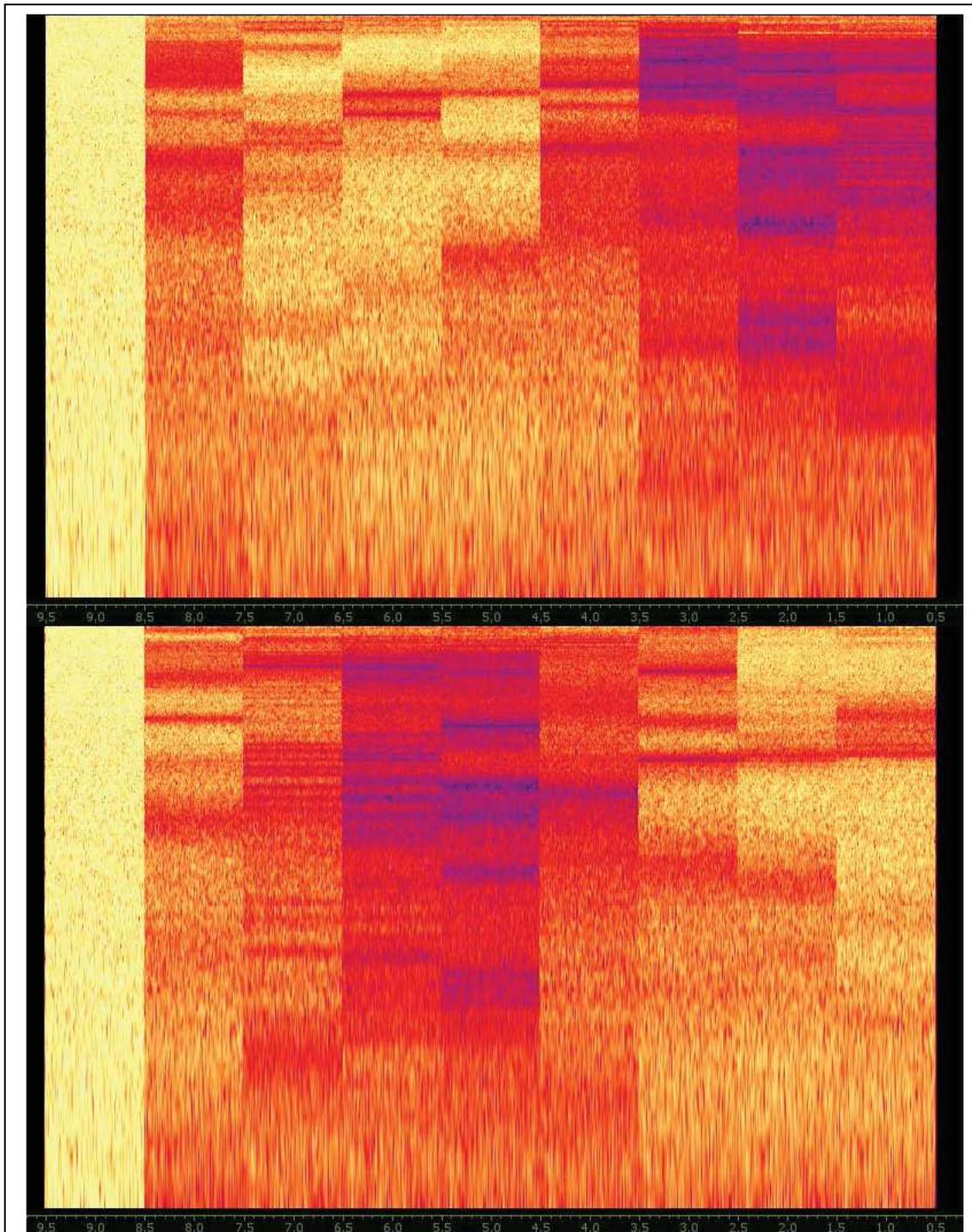


Figura 63. Comparación de espectro entre la señal de entrada seguida de todas las posiciones de acimut para los canales izquierdo (arriba) y derecho (abajo) del Individuo 2.

Se presenta también los resultados de las pruebas de acimut del Individuo 2 en forma de una tabla que permite evidenciar un resumen de la información más relevante de los espectros obtenidos para cada posición considerada, así como realizar una comparación entre estos, sus canales y evidenciar las similitudes encontradas.

Tabla 4. Comparación de Espectros de Acimut del Individuo 2.

	Similitudes.	Canal Izquierdo.	Canal Derecho.
Acimut 0°, Elevación 0°.	Entre 20 y 6500 [Hz]. Subida de 3 [dB] desde 350 hasta 600 [Hz]. Subida a -43 [dB] entre 750 y 850 [Hz]. Estable en -40 [dB] entre 950 y 5000 [Hz]. Valles en 1350 [Hz] (-43 [dB]) y 1850 [Hz] (-42 [dB]). Caída entre 5000 y 6500 [Hz] (-9 [dB]). Valle en 9400 [Hz] (-44 [dB]).	Valle en 2700 [Hz] (-44 [dB]). Pico entre 6000 y 8850 [Hz] (-45 [dB]). Máximo entre 10000 y 13500 [Hz] (-41 [dB]). Valle en 12100 [Hz] (-46 [dB]). Caída en 15000 [Hz] (-55 [dB]), subida rizada hasta -41 [dB] en 19600 [Hz].	Nivel mayor para el canal derecho. Valle entre 2000 y 3100 [Hz] (-43 [dB]). Pico entre 6000 y 8850 [Hz] (-40 [dB]). Máximo entre 10000 y 13500 [Hz] (-38 [dB]). Valle en 12500 [Hz] (-50 [dB]). Caída en 16000 [Hz] (-50 [dB]), subida hasta -45 [dB] y rizado más profundo.
Acimut 45°, Elevación 0°.	Entre 170 y 400 [Hz] (-40 [dB] izquierdo y -43 [dB] derecho).	Constante hasta los 1050 [Hz]. Valle entre 1050 y 1450 [Hz] (-42 [dB]), subida a -36 [dB] en 2050 [Hz]. Valle entre 2450 y 3450 [Hz] (-42	Decaimiento entre 350 y 850 [Hz]. Valle entre 850 y 1300 [Hz] (-49 [dB]). Subida a -42 [dB] en 1650 [Hz]. Valle entre 2350 y 2850 [Hz], mínimo

		<p>[dB]). Constante entre 3800 y 5500 [Hz] en -36 [dB]. Decaimiento a -47 [dB] en 10750 [Hz] con picos en 7000 [Hz], 9070 [Hz] (-39 [dB]) y finalmente en 10350 [Hz] (-43 [dB]). Subida a -35 [dB] en 11750 [Hz], caída a -40 [dB] en 13050 [Hz] con máximo de -32 [dB] en 15500 [Hz]. Valle en 18150 [Hz] (-46 [dB]) y pico de -38 [dB] en 19100 [Hz].</p>	<p>en -49 [dB] en 2600 [Hz], valles en 3300 [Hz] (-46 [dB]) y 3600 [Hz] (-47 [dB]). Nivel de -44 [dB] entre 4000 y 5700 [Hz], caída como picos y valles entre 5800 y 10600 [Hz]. Valles con nivel desde -50 [dB] hasta -58 [dB] y picos con un nivel de -48 [dB], con máximo de -44 [dB] en 8400 [Hz]. Nivel de -42 [dB] entre 11100 y 15500 [Hz]. Caída a -62 [dB] en 17300 [Hz].</p>
<p>Acimut 90°, Elevación 0°.</p>	<p>Entre los 20 [Hz] y los 300 [Hz] (+2 [dB] izquierda).</p>	<p>Entre 300 y 700 [Hz] subida de -40 a -36 [dB], valle entre 1110 y 1650 [Hz] (-40 [dB]) y 1650 y 2050 [Hz] (-41 [dB]). Subida a -37 [dB] desde 2150 hasta 3100 [Hz] con</p>	<p>Nivel estable entre 160 y 350 [Hz] (-42 [dB]). Valle en 780 [Hz] (-45 [dB]). Caída desde 860 [Hz] (-42 [dB]) hasta -57 [dB] en 2400 [Hz]. Subida en 3000 [Hz] (-46 [dB]), pico de -48</p>

		<p>valle en 2950 [Hz] (-39 [dB]) y entre 3100 y 4500 [Hz] (-41 [dB]). Rizado entre 4500 y 8600 [Hz] (-36 [dB]). Caída entre 8600 y 11300 [Hz] a -41 [dB]. Valles en 12000 y 13600 [Hz] (ambos -55 [dB]). Subida entre 13600 y 17000 [Hz] hasta -33 [dB].</p>	<p>[dB] en 3550 [Hz] y caída a -55 [dB] en 4200 [Hz], -52 [dB] (4500 [Hz]) y -63 [dB] (6650 [Hz]). Entre 6600 y 11300 [Hz], serie de picos y valles con variaciones de nivel entre -63 [dB] y -67 [dB] y picos de -53 [dB]. Subida a -53 [dB] en 12000 [Hz] y un valle de -66 [dB] en 17200 [Hz].</p>
<p>Acimut 135°, Elevación 0°.</p>	<p>20 y 160 [Hz] (+2 [dB] izquierda).</p>	<p>Subida hasta -38 [dB] entre 160 y 850 [Hz], caída a -42 [dB] en 3800 [Hz] con picos en 1650 [Hz] (-38 [Hz]), 2000 [Hz] (-39 [dB]) y 3100 [Hz] (-40 [dB]). Caída entre 3800 y 4600 [Hz] hasta -48 [dB], subida hasta los -33 [dB] en 7700 [Hz]. Valle entre 8600 y 10450 [Hz] (-43</p>	<p>Entre 160 y 860 [Hz] caída de 4 [dB], caída entre 860 y 2200 [Hz] hasta -59 [dB]. Pico en 1550 [Hz] (-52 [dB]). Subida a -52 [dB] en 2450 [Hz], constante hasta 4000 [Hz]. Valle entre 4000 y 5350 [Hz] (-63 [dB]), subida hasta -56 [dB] entre 5000 y 6200 [Hz] y valle en 6600 [Hz]</p>

		[dB]), entre 13100 y 14600 [Hz] (-48 [dB]). Subida a -33 [dB] en 18600 [Hz].	(-68 [dB]) y 8250 [Hz] (-72 [dB]). Subida a -54 [dB] en 9000 [Hz] y caída a -60 [dB] en 10500 [Hz], valle en 12000 [Hz] (-74 [dB]). Subida hasta -54 [dB] en 14000 [Hz] que decae a -63 en 16650 [Hz].
Acimut 180°, Elevación 0°.	Entre 20 y 260 [Hz].	Subida a -39 [dB] hasta 1150 [Hz]. Valle entre 1700 y 1900 [Hz] (-41 [dB]) y entre 2500 y 3050 [Hz] (-41 [dB]). Caída desde 3150 [Hz] (-41 [dB]) hasta -50 [dB] en 5850 [Hz]. Entre 6000 y 9000 [Hz] nivel de -49 [dB]. Valles en 9500 [Hz] (-57 [dB]) y 9800 [Hz] (-58 [dB]). Picos en 11600 [Hz] y 13400 [Hz] (-42 [dB]). Valle en 14500 [Hz] (-60	Valle entre 260 y 850 [Hz]. Entre los 1200 [Hz] y 3100 [Hz], valles (-44 [dB]) y picos (-41 [dB]). Caída a partir de 3800 [Hz] (-45 [dB]) hasta -62 [dB] en 7700 [Hz], con picos y valles. Subida hasta -49 [dB] en 9150 [Hz] y a -46 [dB] en 10100 [Hz], constante hasta 12000 [Hz]. Valle entre 12000 y 15500 [Hz] (-52 [dB]). Nivel estable en -47 [dB] a partir

		[dB]) y subida desde 16100 [Hz] (-47 [dB]) hasta -41 [dB] en 17800 [Hz].	de los 16600 [Hz].
Acimut 225°, Elevación 0°.	Entre 20 y 160 [Hz].	Serie de caídas y subidas de nivel. Caída entre 170 y 1400 [Hz]. Valle entre 520 y 880 [Hz] (-47 [dB]), pico en 1550 [Hz] (-43 [dB]). Caída a -53 [dB] (2500 [Hz]) y subida a -47 [dB] (3500 [Hz]). Caída a -52 [dB] (4400 [Hz]) y subida a -49 [dB] (5300 [Hz]), caída a -55 [dB] (5850 [Hz]) y subida a -51 [dB] (8600 [Hz]) y caída a -54 [dB] (9350 [Hz]) y subida a -48 [dB] (11300 [Hz]). Valle entre 11500 y 19700 [Hz] con picos en 15650 [Hz] (-62 [dB]), 17700 [Hz] (-62	Subida de 3 [dB] desde 170 hasta 870 [Hz]. Caída a -40 [dB] en 2300 [Hz], con valle en 1840 [Hz] (-43 [dB]). Constante entre 2300 y 3500 [Hz] y caída a -48 [dB] en 4650 [Hz], subida a -38 [dB] en 7600 [Hz]. Entre 7500 y 20000 [Hz] existen picos en 7800 [Hz] (-38 [dB]), 10800 [Hz] (-39 [dB]), 14000 [Hz] (-40 [dB]) y 19600 [Hz] (-43 [dB]) y valles en 9550 [Hz] (-60 [dB]), 12300 [Hz] (-50 [dB]) y 16750 [Hz] (-65 [dB]).

		[dB]), 19000 [Hz] (-60 [dB]).	
Acimut 270°, Elevación 0°.	Entre 20 y 160 [Hz].	Caída de 2 [dB] entre 160 y 850 [Hz] hasta -47 [dB] en 1600 [Hz], con valle en 1800 [Hz] (-53 [dB]) y entre 1900 y 4500 [Hz] (-63 [dB]) con pico de -59 [dB] (2750 [Hz]). Caída a -71 [dB] (6000 [Hz]) y valle entre 7300 y 9900 [Hz] con pico de -61 [dB] (8850 [Hz]). Subida a -53 [dB] entre 10300 y 11300 [Hz] y caída a -69 [dB] en 13000 [Hz] con pico en 13100 (-63 [dB]) y subida a -58 [dB] en 15000 [Hz]. Caída a -70 [dB] en 15700 [Hz], incremento gradual con picos hasta -39 [dB] en 19950 [Hz].	Subida de 4 [dB] desde 160 hasta 1000 [Hz] y caída a -47 [dB] en 4400 [Hz], con pico 3250 [Hz] (-39 [dB]). Subida a -34 [dB] entre 4200 y 8400 [Hz] con valle de -53 [dB] en 9850 [Hz]. Incremento rizado hasta el extremo derecho del espectro.
Acimut 315°,	Entre 20 y 160	Caída desde -41	Estable en -39

Elevación 0°.	[Hz] (+2 [dB] en canal derecho).	[dB] (160 [Hz]) hasta -46 [dB] (1150 [Hz]). Valle entre 1150 y 3050 [Hz] (-58 [dB]) y pico de -51 [dB] (1630 [Hz]). Caída a -64 [dB] (6700 [Hz]) con picos en 3700 [Hz] (-47 [dB]), 4750 [Hz] (-50 [dB]), y 5350 [Hz] (-51 [dB]). Entre 6700 y 12400 [Hz], picos y valles con entre -55 y -65 [dB] cada 400 [Hz], con valle en 12400 de -67 [dB]. Subida a -55 [dB] (13300 [Hz]), constante hasta 15100 [Hz] y caída a -70 [dB] (16000 [Hz]). Subida con picos en 17200 [Hz] (-58 [dB]), 18200 [Hz] (-56 [dB]) y 19920 [Hz] (-41 [dB]).	[dB] entre 160 y 870 [Hz]. Valle entre 1300 y 1650 [Hz] (-42 [dB]) y subida a -35 [dB] (2550 [Hz]). Valle entre 2550 y 4600 [Hz] (-3 [dB]). Constante entre 4700 [Hz] y 9000 [Hz] (-35 [dB]). Valles de -51 [dB] (9950 [Hz]) y -49 [dB] (12650 [Hz]) con pico intermedio de -44 [dB]. Subida gradual -a 36 [dB] desde 12700 hasta 15600 [Hz].
---------------	----------------------------------	---	---

Nota. Los valores presentados en la tabla no son exactos, han sido aproximados de la lectura obtenida del software y resumidos para una mejor presentación en la tabla.

2.3.3. Pruebas Adicionales

Una vez presentados todos los resultados para el Individuo 2, se presentan los obtenidos para el Individuo 5, en forma del gráfico tridimensional de comparación de espectros y tablas de comparación del análisis por canal y posición tanto para elevación y acimut, con el objetivo de presentar un nuevo conjunto de datos que apoyan los resultados buscados y obtenidos con las pruebas del Individuo 2.

2.3.3.1. Resumen y Comparación de Posiciones de Elevación

Se puede apreciar las diferencias espectrales más significativas entre las distintas posiciones al variar la elevación y visualizarlos gráficamente de forma continua como muestras de 1 [s] de duración comparándolas con el espectro del ruido blanco.

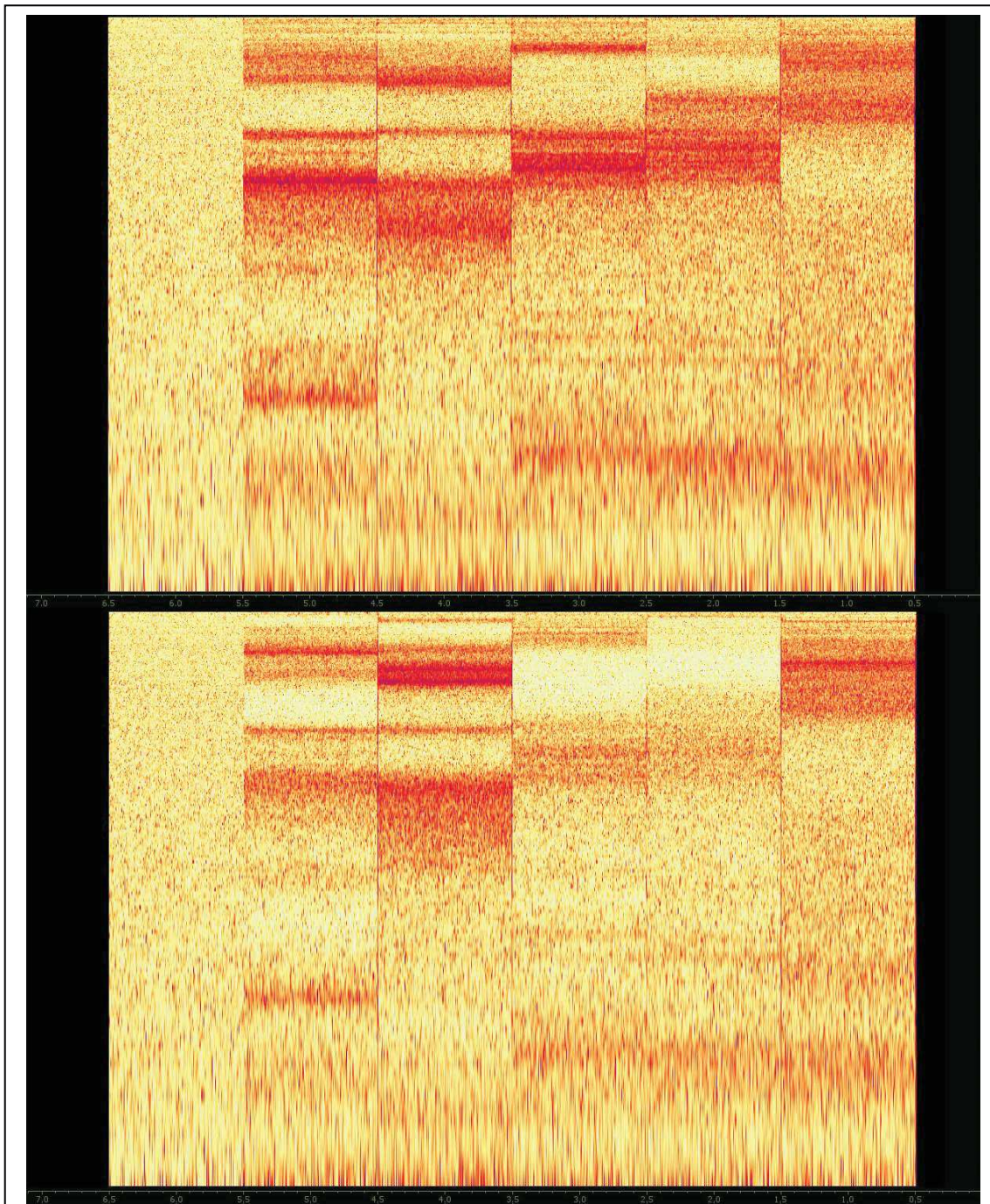


Figura 64. Comparación de espectro entre la señal de entrada seguida de todas las posiciones de elevación para los canales izquierdo (arriba) y derecho (abajo) del Individuo 5.

Se presenta también de las pruebas de elevación del Individuo 5 en forma de una tabla que permite evidenciar un resumen de la información más relevante

de los espectros obtenidos para cada posición considerada, así como realizar una comparación entre estos, sus canales y evidenciar las similitudes encontradas.

Tabla 5. Comparación de Espectros de Elevación del Individuo 5.

	Similitudes.	Canal Izquierdo.	Canal Derecho.
Elevación -30°, Acimut 0°.	Subida de 20 a 430 [Hz] (alcanza -32 [dB] izquierda y -27 [dB] derecha). Valle entre 430 y 1300 [Hz] (-35 [dB] izquierda y -32 [dB] derecha).	Caída a -42 [dB] en 1550 [Hz] y subida a -29 [dB] en 2650 [Hz]. Decaimiento rizado con valle en 4200 [Hz] (-35 [dB]) hasta -55 [dB] en 7740 [Hz]. Subida a -29 [dB] en 11500 [Hz] con picos en 9000 [Hz] y 9600 [Hz] (-34 [dB]).	Caída a -36 [dB] en 1550 [Hz] y subida a -22 [dB] en 2900 [Hz]. Valle entre 3400 y 4300 [Hz] (-30 [dB]). Caída rizada a -38 [dB] en 7300 y subida a -28 [dB] en 8500, valle en 10500 [Hz] (-39 [dB]) y subida a -24 [dB] entre 10900 y 13600 [Hz]. Caída a -44 [dB] en 17500 con pico de -32 [dB] en 16050 [Hz]
Elevación 0°, Acimut 0°.	Subida desde 20 hasta 1300 [Hz] (-29 [dB] izquierda y -24 [dB] derecha). Rizado con variaciones de ± 1 [dB] hasta 3300	Caída a -45 [dB] en 5400 [Hz] y pico de -40 [dB] en 6900 [Hz]. Subida hasta -32 [dB] y constante hasta el fin con	Caída rizada a -43 [dB] (7200 [Hz]). Subida a -26 [dB] entre 8500 y 12500 [Hz]. Caída a -53 [dB] en 14100 [Hz] y

		[Hz] (izquierda) y 3800 [Hz] (derecha).	valle entre 10300 y 11400 [Hz] (-42 [dB]).	subida gradual hasta -23 [dB] con picos en 14800 [Hz] y -35 [dB] en 16800 [Hz].
Elevación 30°, Acimut 0°.		Subida desde 20 hasta 260 [Hz] (-30 [dB] izquierda y -26 [dB] derecha), caída gradual a -38 [dB] izquierda y -34 [dB] derecha (850 [Hz]) y subida a -30 [dB] (izquierda) y -25 [dB] (derecha) (2000 [Hz]).	Picos y valles de ± 4 [dB] hasta 3400 [Hz]. Rizado y caída a -53 [dB] en 8400 [Hz] y subida gradual a -30 [dB] en extremo izquierdo, con pico de -39 [dB] en 9500 [Hz].	Caída a -28 [dB] en 7000 [Hz] con valles en 2100 [Hz] (-29 [dB]) y 2550 [Hz] (-31 [dB]) y pico en 3250 [Hz] (-24 [dB]). Valle entre 7000 y 11500 [Hz] (-36 [dB]) con picos en 8000 y 8500 [Hz] (-32 [dB]). Subida rizada hasta -20 [dB] en 14100 [Hz] y caída a -34 [dB] en 18500 [Hz].
Elevación 60°, Acimut 0°.		Subida desde 20 hasta 260 [Hz] (-30 [dB] izquierda y -27 [dB] derecha).	Caída a -40 [dB] en 860 [Hz] y subida a -31 [dB] en 1600 [Hz]. Picos y valles de ± 3 [dB] hasta 3100 [Hz]. Nivel constante y rizado en -32 [dB]. Valle	Valle entre 430 y 1300 [Hz] (-33 [dB]). Rizado con variaciones mínimas hasta 6300 [Hz] y valle en 2150 [Hz] (-31 [dB]). Valle rizado entre 6300 y

		entre 7500 y 12000 [Hz] (-48 [dB]) con picos y valles de ± 2 [dB].	13000 [Hz] (-34 [dB]) y subida a -20 [dB] en 15800 [Hz].
Elevación 90°, Acimut 0°.	Subida desde 20 hasta 270 [Hz] (-27 [dB]). Valle entre 450 y 1100 [Hz] (-34 [dB]).	Constante y rizado en -30 [dB] entre 1150 y 6800 [Hz]. Subida de 1 [dB] entre 6800 y 9600 [Hz]. Valles entre 10900 y 14400 [Hz] (-40 [dB]) y entre 15500 y 19000 [Hz] (-40 [dB]).	Valle entre 1300 y 2300 [Hz] (-32 [dB]) y nivel constante hasta 6000 [Hz] (-29 [dB]). Subida de 1 [dB] entre 6000 y 10600 [Hz]. Caída a -38 [dB] en 11850 [Hz], subida de 2 [dB] hasta 18200 [Hz] y valle en 16200 [Hz] (-46 [dB]).

Nota. Los valores presentados en la tabla no son exactos, han sido aproximados de la lectura obtenida del software y resumidos para una mejor presentación en la tabla.

2.3.3.2. Resumen y Comparación de Posiciones de Acimut

Se puede apreciar las diferencias espectrales más significativas entre las distintas posiciones al variar la elevación y visualizarlos gráficamente de forma continua como muestras de 1 [s] de duración comparándolas con el espectro del ruido blanco.

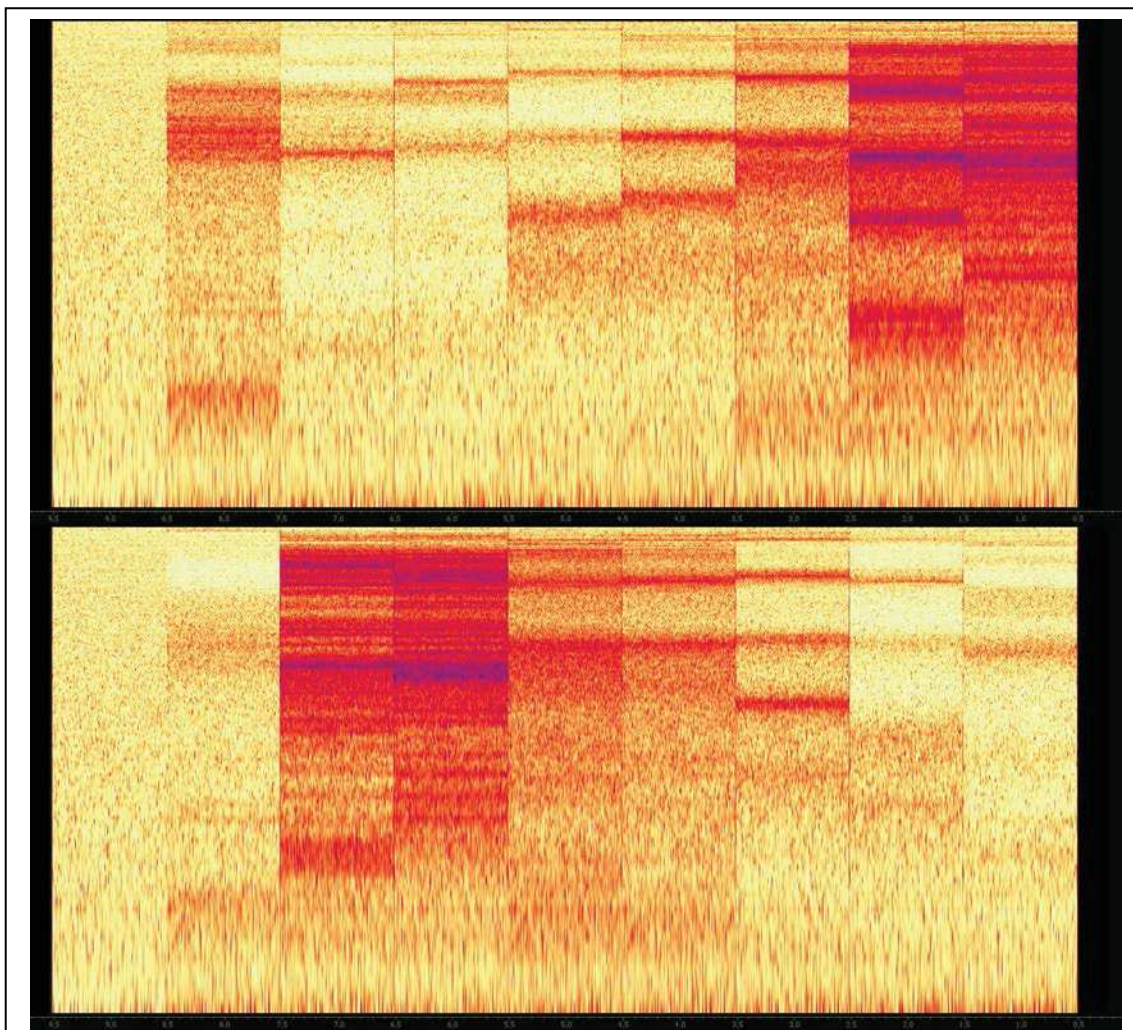


Figura 65. Comparación de espectro entre la señal de entrada seguida de todas las posiciones de acimut para los canales izquierdo (arriba) y derecho (abajo) del Individuo 5.

Se presenta también los resultados de las pruebas de acimut del Individuo 5 en forma de una tabla que permite evidenciar un resumen de la información más relevante de los espectros obtenidos para cada posición considerada, así como realizar una comparación entre estos, sus canales y evidenciar las similitudes encontradas.

Tabla 6. Comparación de Espectros de Acimut del Individuo 5.

	Similitudes.	Canal Izquierdo.	Canal Derecho.
Acimut 0°, Elevación 0°.	Subida desde 20 hasta 260 [Hz] (-27 [dB] izquierda y -26 [dB] derecha). Valle entre 430 y 1300 [Hz] (-38 [dB] izquierda y -34 [dB] derecha). Nivel constante rizado en -30 [dB] (izquierda) y -27 [dB] (derecha) hasta 5200 [Hz] (izquierda) y 6800 [Hz] (derecha) con picos y valles de ± 3 [dB].	Caída de 2 [dB] entre 5800 y 7300 [Hz]. Valle entre 7400 y 12000 [Hz] (-45 [dB]), con picos con variaciones de ± 3 [dB]. Valle entre 12000 y 13900 [Hz] (-40 [dB]) y subida a -25 [dB] en 16600 [Hz].	Valle rizado entre 6800 y 13100 [Hz] (-34 [dB]) y subida hasta -19 [dB] en 16000 [Hz].
Acimut 45°, Elevación 0°.		Subida desde 20 hasta 2000 [Hz] (-25 [dB]) con valle en 1500 [Hz] (-30 [dB]). Subida a -23 [dB] y constante, rizado hasta 5750 [Hz] con valle en 4100 [Hz] (-26 [dB]), caída a -42 [dB] en 8200 [Hz] y subida a -20 [dB]	Subida desde 20 hasta 270 [Hz] (-28 [dB]), caída a -35 [dB] en 1150 [Hz]. Valle entre 1150 y 1800 [Hz] (-42 [dB]). Picos y valles de ± 5 [dB] hasta 3800 [Hz], caída a -49 [dB] en 8200 [Hz] con picos y valle de -

		en 15250 [Hz] con picos de ± 3 [dB] y un valle entre 11300 y 14000 [Hz] (-35 [dB]). Valle de -27 [dB] entre 15700 y 17750 [Hz].	60 [dB] en 7800 [Hz]. Entre 8300 y 18000 [Hz], nivel altamente inestable, rizado con variaciones de hasta ± 15 [dB].
Acimut 90°, Elevación 0°.		Subida desde 20 hasta 760 [Hz] (-25 [dB]). Leve rizado pero constante hasta 7500 [Hz], con variaciones de ± 3 [dB]. Valles entre 7500 y 10300 [Hz] (-35 [dB]) y entre 11000 y 15600 [Hz] (-38 [dB]) con pico en 13300 (-30 [dB]). Valle entre 15800 y 18500 [Hz] (-28 [dB]).	Subida desde 20 hasta 270 [Hz] (-27 [dB]), caída a -37 [dB] en 1800 [Hz]. Picos y valles de ± 6 [dB] hasta 4700 [Hz]. Caída rizada hasta -60 [dB] en 7100 [Hz]. Picos y valles de ± 12 [dB] (entre -43 y -55 [dB]) hasta 14000 [Hz]. Subida rizada a -30 [dB] en 19500 [Hz].
Acimut 135°, Elevación 0°.		Subida desde 20 a 1300 [Hz] (-25 [dB]). Caída rizada hasta -39 [dB] en 4960 [Hz] con subida entre 3700 y 4300 [Hz] (-31	Subida desde 20 a 270 [Hz] (-28 [dB]). Valle entre 270 y 1850 [Hz] (-35 [dB]) con variaciones de ± 2 [dB]. Caída suave

		[dB]). Subida a -25 [dB] en 7050 [Hz], caída a -34 [dB] en 9330 [Hz]. Subida a -20 [dB] en 10860 [Hz] y caída a -36,5 [dB] en 15150 [Hz]. Subida a -25 [dB] en 16200 [Hz].	y rizada hasta -47 [dB] en 9000 [Hz]. Subida a -33 [dB] en 10850 [Hz] y valle en 14700 [Hz] (-44 [dB]).
Acimut 180°, Elevación 0°.		Subida desde 20 a 1300 [Hz] (-25 [dB]). Caída rizada hasta -30 [dB] en 4550 [Hz]. Valle entre 4550 y 6750 [Hz] (-41 [dB]), subida a -29 en 7090 [Hz], valle entre 7400 y 10100 [Hz] (-46 [dB]). Subida a -26 [dB] (10600 [Hz]) y constante hasta valle entre 13700 y 16900 [Hz] (-40 [dB]).	Subida desde 20 a 260 [Hz] (-28 [dB]). Valle entre 270 y 1200 [Hz] (-33 [dB]). Caída rizada con variaciones de ± 2 [dB] hasta -38 [dB] en 7900 [Hz]. Valle entre 8050 y 9850 [Hz] (-46 [dB]) y subida a -32 [dB] en 10500 [Hz]. Valle en 15050 [Hz] (-47 [dB]).
Acimut 225°, Elevación 0°.		Subida desde 20 a 260 [Hz] (-29 [dB]). Valle entre	Subida desde 20 a 1300 [Hz] (-25 [dB]). Caída rizada

		260 y 1200 [Hz] (-34 [dB]). Caída a -34 [dB] en 5500 [Hz] con picos y valles de ± 3 [dB]. Valle rizado entre 5500 y 10050 [Hz] (-48 [dB]) con pico de -39 [dB] en 8080 [Hz]. Nivel constante y rizado en -31 [dB] entre 10200 y 17500 [Hz] con valle en 14800 [Hz] (-51 [dB]).	hasta -35 [dB] en 3100 [Hz]. Subida a -30 [dB] en 4200 [Hz], valle entre 4800 y 6750 [Hz] (-45 [dB]), subida a -29 en 7040 [Hz], valle entre 7500 y 10400 [Hz] (-39 [dB]). Subida a -25 [dB] (10700 [Hz]) y constante hasta valle entre 13600 y 16800 [Hz] (-43 [dB]).
Acimut 270°, Elevación 0°.		Subida desde 20 a 260 [Hz] (-29 [dB]). Caída a -47 [dB] en 2100 [Hz] con picos en 680 [Hz] (-31 [dB]), 1130 [Hz] (-33 [dB]) y 1880 [Hz] (-43 [dB]). Subida hasta -35 [dB] en 3200 [Hz] y caída a -54 [dB] en 4700 [Hz], sube a -43 [dB] en 5600 [Hz] y cae a -65 [dB] en	Subida desde 20 a 1300 [Hz] (-26 [dB]). Caída rizada hasta -32 [dB] en 4200 [Hz]. Subida a -22 en 6100 [Hz], caída a -31 [dB] en 940 [Hz]. Subida a -19 [dB] en 10860 [Hz] y caída a -38 [dB] en 14900 [Hz]. Subida a -21 [dB] en 17000 [Hz].

		8080 [Hz], sube con rizado a -38 [dB] en 10100 [Hz] y cae a -60 [dB] en 13100 [Hz]. Subida rizada hasta -31 [dB] con valle en 18000 [Hz] (-54 [dB]).	
Acimut 315°, Elevación 0°.		Subida desde 20 a 260 [Hz] (-27 [dB]). Caída a -40 [dB] en 2800 [Hz] con pico en 1980 [Hz] (-35 [dB]). Entre 2800 y 18800 [Hz] el nivel es demasiado inestable, con picos y valles de ± 3 [dB] (hasta 6300 [Hz]) seguido de variaciones de hasta ± 7 [dB].	Subida desde 20 hasta 2300 [Hz] (-24 [dB]) con valle en 1010 [Hz] (-28 [dB]) y 1450 [Hz] (-30 [dB]). Subida a -23 [dB] nivel y constante, rizado hasta 6800 [Hz], con valle en 3800 [Hz] (-28 [dB]). Caída a -36 [dB] en 8700 [Hz] y subida a -23 [dB] en 10850 [Hz]. Valle entre 11000 y 14400 [Hz] (-29 [dB]). Subida a -20 [dB] en 15500 y valle de -27 [dB] entre 16500 y 18400 [Hz].

Nota. Los valores presentados en la tabla no son exactos, han sido aproximados de la lectura obtenida del software y resumidos para una mejor presentación en la tabla.

3. Análisis Costo-Beneficio

3.1. Análisis Costo-Beneficio

Para el análisis del costo total del proyecto se tienen en cuenta únicamente a los recursos que se consideran necesarios para la etapa de investigación, el desarrollo del software y su posterior evaluación, como son el computador y software utilizados.

Tabla 7. Análisis de Costos.

Detalle	Cantidad	Valor Unitario	Valor Total
Computador iMac 11,2.	1	700.00	700.00
Software MAX/MSP 5.1.3.	1	250.00	250.00
Software Pro Tools LE 8.0.4.	1	250.00	250.00
Software Ozone 5 Advanced.	1	999.00	999.00
Servicio de Internet.	7	20.00	140.00
TOTAL			2399.00

El computador es un Macintosh iMac modelo 11,2 con un procesador Intel Core i3 de 3,06 [GHz], memoria RAM DDR3 de 4 [GB] de 1333 [MHz] y sistema operativo *Mac OS X 10.6.8*.

El software empleado para el diseño del prototipo es MAX/MSP desarrollado por la empresa Cycling '74 cuya versión utilizada es la 5.1.3. Para la etapa de análisis de resultados, se utilizó el software Pro Tools LE 8.0.4 desarrollado por

AVID y Ozone 5 Advanced desarrollado por iZotope. La interface de audio utilizada durante el diseño y las pruebas es la predeterminada del computador.

Como recurso adicional se considera el uso de servicio de Internet durante los primeros 7 meses del proceso, empleado para consultas y referencias, así como descargas de archivos necesarios para el software.

El costo puede ser todavía menor al emplear herramientas disponibles de forma gratuita para la etapa del análisis, además de considerar la ventaja de optar por recursos que no representan costos como son los objetos externos para MAX/MSP, los archivos de audio de los impulsos respuesta del proyecto LISTEN y los archivos de audio de cámara anecoica de la librería OpenAIR.

Éstos, al ser empleados mediante licencias que permiten su uso de forma libre en aplicaciones no comerciales, no se consideran en el análisis de costo.

4. Proyecciones

4.1. Capacidad de Expansión y Potencial del Proyecto

La aplicación principal para la cual se realizó éste programa, es la de una herramienta de producción de audio. Se considera el necesario empleo de audífonos estereofónicos para su funcionamiento como una ventaja al aprovechar la creciente y cada vez más presente tecnología móvil para consumo de producciones de material donde se puede emplear la herramienta.

La compatibilidad de la reproducción de material binaural en un sistema estereofónico tradicional mediante altavoces (a pesar de la diafonía y las reflexiones del ambiente de escucha), se puede considerar como una ventaja, donde, si bien la técnica y percepción binaural se pierden (pues la entrega de las señales no es ideal), no se vuelve necesario la realización de una segunda mezcla estéreo ni el empleo de equipamiento adicional para la misma (los mismos dos canales de la cadena digital-analógica así como electroacústica pueden ser empleados).

Existe una cantidad de mejoras que pueden ser implementadas y han sido consideradas como parte de éste trabajo, pero bien o estas están fuera de los objetivos del mismo y el tiempo asignado para llevarlo a cabo no lo ha permitido. Sin embargo, se las enlistan a continuación como sugerencias en caso de que éste proyecto sea tomado por otro investigador con el fin de avanzarlas.

- Implementación de un sistema de *crossfade* (cambio de nivel gradual) para evitar ruidos causados por la discontinuidad en la amplitud de la señal de salida, producidos al cambiar de HRIR en medio de la reproducción.
- Ampliación de la base de datos de HRIR, ya sea empleando otra base de datos libre, o mediante interpolación de la base de datos proporcionada.
- Incorporación de los 46 de individuos restantes en la base de datos de HRIR del proyecto LISTEN.

- Implementación de un módulo de reverberación realista para expandir la capacidad de localización a tres dimensiones.
- Especificar valor de distancia a señales de entrada de acuerdo a un análisis de su nivel o en base a su distancia de captación, con el fin de emplear simulación de distancia absoluta.
- Implementar un sistema hardware para análisis de movimiento de la cabeza del usuario (*head-tracking*) para simulación más realista.
- Implementación de automatización vía MIDI.
- Implementación de señales dinámicas y su automatización, considerando por ejemplo, simulación del efecto *Doppler*.

La posibilidad de expandir el proyecto una mezcladora es un tema de sumo interés, capaz de manejar varias señales monofónicas de entrada que puedan ser ubicadas en un campo tridimensional y se propone como continuación una vez resueltas algunas de las sugerencias anteriores (en especial, la inclusión del modelo y algoritmo de reverberación).

En general la aplicación de sonido binaural en grabación y mezcla no solo puede beneficiar a producciones audiovisuales para entretenimiento. Moller (1992, p.206) plantea la posibilidad de implementarlo para obtener y evaluar, desde comportamiento de altavoces en ambientes, acústica de salas, teleconferencias e inclusive evaluación de ruido, todo esto debido una captación con un acercamiento más realista, que suelen regirse por estándares que no necesariamente reflejan la realidad de la situación acústica que se evalúa o simula, o considera insuficientes.

El software también tiene el potencial de emplearse como una herramienta para la realización de pruebas para la evaluación de la percepción y localización de fuentes sonoras, haciendo posible una cantidad de investigaciones distintas cuyos resultados pueden traer nuevos conocimientos que contribuyan al campo de la psicoacústica. El problema radica en la cantidad de variables, el entorno visual y la interface gráfica a la que es sometido el participante, que pueden influir y parcializar los resultados, por lo

que un diseño correcto de las mismas deberá considerar todas las variables presentes durante las pruebas.

5. Conclusiones y Recomendaciones

5.1. Conclusiones

- Se llega a la conclusión de que el diseño y la implementación de un software capaz de convertir señales monofónicas de entrada en señales binaurales de salida permite el procesamiento de señales de audio digital considerando los fenómenos psicoacústicos de la audición y representa una potencial herramienta de producción de audio por su capacidad de simular la ubicación de fuentes sonoras en un campo tridimensional alrededor del usuario.
- El procesamiento de señales monofónicas considera los distintos fenómenos psicoacústicos presentes en la audición a pesar de no haber sido diseñado como un simulador de una cabeza humana como ha sido planteado en un principio.
- La psicoacústica y en especial, la capacidad de localización de fuentes sonoras es un campo amplio por explorar a pesar de haber sido motivo de estudio por décadas, por lo que se puede concluir que difícilmente será definido o establecerlos en su totalidad de forma teórica ya que trata con procesos subjetivos como la percepción, y funciones complementarias como lo son el sentido de la visión y procesos cognitivos y de memoria para la localización de fuentes, imposible de medir y cuantificarlos.
- Las claves de psicoacústicas que determinan la localización de sonidos en los planos de acimut, elevación y distancia han sido revisadas durante el desarrollo del proyecto, llegando a la conclusión que parte de éstas, como lo son las contribuciones del pabellón auditivo y la reverberación, se tratan de propuestas de distintos autores que no tienen manera de ser establecidas al no existir una forma de cuantificar su influencia en la percepción ni manera de generalizar los resultados.
- Si bien algunas claves han llegado a ser aceptadas dentro del campo al coincidir varias investigaciones y varios autores como el caso las diferencias interaurales y sombra acústica de la cabeza para la

localización en el plano horizontal, las claves para el plano vertical entregadas como variaciones de espectro producidas por el pabellón auditivo y en especial, las claves del entorno acústico donde se encuentra y percibe la fuente para la localización de distancia, son todavía menos claras.

- Considerando la variedad de métodos estructurales propuestos por distintos autores para su implementación en el software, se optó que la implementación mediante HRTF debido a la ventajas de obtener todas las claves psicoacústicas mediante una única etapa de procesamiento de señales, agilitando tiempo y recursos en el proyecto.
- El emplear un proceso de convolución de una señal de entrada con las HRIR no se considera únicamente la influencia de la cabeza pero la de todo el cuerpo del individuo del cual se obtienen las funciones de transferencia.
- Los parámetros que definen a la señal de audio monofónica son sometidos a un único proceso de convolución con el fin de generar dos señales discretas de salida, una para cada oído, que posean la información espectral y temporal necesaria para la localización de fuentes en la escucha binaural.
- Con la convolución de señales de entrada y HRTF, el espectro de las señales resultantes tienen un espectro particular que considere las claves psicoacústicas afectando la percepción del mismo como timbre, sin bien no hay manera de evaluar dicha propiedad por su nivel de subjetividad asociada.
- No ha sido posible el establecer un algoritmo para el software desarrollado, como se plantea en uno de los objetivos, debido al tiempo necesario para expresar el software desarrollado en un lenguaje de programación orientado a objetos en forma de una ecuación matemática.
- La falla en la determinación del algoritmo del software escapa el nivel de conocimientos y su búsqueda no influye en el desarrollo del proyecto,

pues no impide alcanzar el tipo de procesamiento buscado desde un principio.

- Por medio de la comparación de los espectros, el análisis es estrictamente objetivo por lo que no se pretende evaluar la eficiencia del software ni su precisión de la localización de las mismas con respecto al usuario, al ser una propiedad subjetiva difícilmente cuantificable.
- Las capacidades de evaluación consideradas para el proyecto son demasiado amplias y pueden en sí ser el motivo de nuevas investigaciones, por lo que se ve necesaria la limitación de resultados a un cierto número de posiciones que se puedan agrupar bajo el criterio de su variación de posición en una dimensión.
- Debido a las limitaciones de tiempo y condiciones para éste proyecto, no han sido posibles de realizar pruebas de evaluación del software para la percepción, para lo que es necesario una evaluación psicoacústica que considere la experiencia del usuario al usar el software.
- Una evaluación psicoacústica puede ser una herramienta de investigación potente donde el análisis de resultados puede derivar en descubrimientos significativas dentro del campo, con el problema de la inversión económica, de tiempo y personal que por lo general requieren de años para el diseño y ejecución de las mismas, así como un equipo de trabajo capacitado y laboratorios equipados que cumplan con condiciones de trabajo necesarias.
- Debido a la implementación de HRTF no individualizadas se emplea una amplia base de datos para cinco individuos distintos, que permite la opción de una prueba auditiva donde se determine cuál de éstos se adapta mejor a la localización binaural del usuario, en lugar de la solución e impráctica de la medición y generación de los impulsos de cada uno.
- Un esquema de convolución que lleve a cabo dicho proceso tanto en el dominio del tiempo y dominio de la frecuencia de los impulsos, permite el procesamiento en tiempo real de las señales y el cambio dinámico de la posición de la fuente durante la reproducción de la señal de entrada.

- El único control de la posición de la fuente en la dimensión de la distancia existe por medio del control de nivel de salida de audio a pesar de que se consideró emplear un algoritmo de generación de reflexiones o reverberación artificial.
- El no incluir generación de reverberación artificial representa una ventaja al no perjudicar la calidad espectral y temporal que aportan las HRTF, por lo que funciona de manera óptima para señales de entrada que han sido registradas bajo condiciones anecoicas.
- Si bien las pruebas realizadas presentan un análisis cuantitativo en un software basado en claves psicoacústicas, se concluye que los resultados permiten visualizar algunas tendencias en los espectros resultantes.
- El empleo un ruido de espectro constante como el ruido blanco permite apreciar fácilmente y de forma visible el filtrado producido por las HRTF correspondientes a las posiciones evaluadas.
- El empleo de un único individuo en las pruebas es suficiente para el objetivo de la evaluación aunque diferencias interpersonales existirán, dando lugar a un análisis espectral distinto para cualquier otro individuo.
- Los espectros comparados tanto en acimut como en elevación son indicadores de los cambios producidos y demuestran el procesamiento realizado, a pesar de que sea imposible determinar mediante las variaciones de los mismos, los fenómenos de percepción y su influencia en la audición.
- El empleo de HRIR considera las ITD entre canales en el procesamiento a pesar de que no sea considerado en su análisis.
- Es común una respuesta de frecuencia similar a la de un filtro pasa bajos para el extremo de bajas frecuencias, bajo los 160 [Hz] aproximadamente y para todas las posiciones evaluadas por lo que probablemente se debe a la convolución de señales, defecto común en dicho proceso o bien, debido a la falta de bajas frecuencias de las HRIR empleadas. Si bien el filtrado varía con la posición, la variación es leve y

es prácticamente idéntica para los dos canales de la salida binaural evaluada.

- En términos prácticos, se puede considerar que las señales son bastante similares en un rango del espectro comprendido entre los 20 [Hz] y los 3000 [Hz]. A partir de ésta frecuencia, las diferencias entre ambos canales son más notables, lo que coincide de cierta forma con las propuestas de varios autores que plantean la necesidad de contenido en altas frecuencias del sonido emitido por la fuente, para la aparición de claves espectrales fundamentales en la localización en el plano vertical.
- A pesar de que existen estudios que demuestran la influencia de bajas frecuencias en la percepción en la localización vertical, éstas no son evidentes ya sea por el proceso de convolución o la respuesta de las HRTF empleadas.
- Al aumentar la posición de elevación, los espectros comienzan a volverse más constantes, sin llegar a ser planos totalmente, pero perdiendo las deficiencias notorias que se evidencian en ciertas posiciones de elevación en forma de picos y valles pronunciados.
- La diferencia más evidente entre cada par de espectros izquierdo y derecho es la de la caída de nivel que se presenta entre ambos en posiciones extremas, donde un oído es claramente opacado por la cabeza, evidenciando la sombra acústica producida por la misma.
- Esta diferencia de nivel comienza a ser considerable inclusive inmediatamente luego del filtrado común para todas las señales, por lo que es difícil establecer un rango común entre pares de espectros para cada posición, lo que evidencia la presencia de ILD.
- Al no ser necesariamente similares en sus espectros y mostrar considerables diferencias espectrales entre canales izquierdo y derecho, es evidente que no solo las diferencias interaurales están presentes en la localización de fuentes en el plano horizontal. La mayoría de autores sostiene que la localización sonora en acimut es conseguida y suficiente con el empleo de diferencias interaurales de tiempo y nivel.

- Al realizar una comparación de pares de espectros de posiciones equivalentes en lados opuestos (donde el oído ipsilateral en la primera posición es el oído contralateral en el lado opuesto), no se puede observar una similitud clara y esperada (en ningún individuo), debido a que las similitudes son mínimas en comparación a las diferencias presentadas, dificultando el definir un patrón entre sí.
- Un segundo individuo (Individuo 5) en la evaluación del software presenta resultados similares a los obtenidos para el primero, con cambios espectrales esperados de acuerdo al cambio de posición de acimut y elevación (diferencias interaurales y diferencias espectrales en alta frecuencia respectivamente), así como diferencias esperadas debidas a empleo de funciones específicas a la antropometría (desconocidas) de cada individuo y cuya razón que no se puede deducir con exactitud. Estos cambios serán comunes para cualquier conjunto de HRTF empleados.
- El comparar los resultados entre ambos individuos no ha sido un objetivo de las pruebas debido a que a pesar de poder evaluar sus diferencias espectrales, no se podrían obtener resultados que indiquen definitivamente la influencia de dichas diferencias en la capacidad de localización binaural de los individuos, pues dependerá de la percepción de cada uno.
- Se puede observar las tendencias apreciadas para el Individuo 2 en los espectros sonoros del Individuo 5, lo que se esperaba al introducir un segundo individuo en la evaluación del procesamiento software. Las conclusiones previas en base al análisis de las pruebas del Individuo 2, se verifican y apoyan con los resultados del Individuo 5.
- El filtrado tipo Filtro Pasa Bajos también se evidencia en el Individuo 5 alrededor de 160 [Hz] y la respuesta tiende a estabilizarse en el extremo derecho del espectro al incrementar la elevación, sin llegar a ser plana.
- De las pruebas de acimut se evidencia la sombra acústica y las diferencias de nivel interaural así como la caída de nivel en el oído contralateral en posiciones opuestas a la fuente. También se presenta

un espectro más complejo, rizado y generalmente inestable en altas frecuencias, difícilmente mostrando similitudes entre canales para una misma posición.

5.2. Recomendaciones

- Un análisis cualitativo del procesamiento del software mediante pruebas a individuos puede entregar resultados que evalúen la capacidad del software y su efectividad en la localización de fuentes en la percepción psicoacústica de las mismas. Sin embargo, para éste tipo de pruebas, las condiciones en las que se realizan deberán ser controladas, limitando las variables presentes y los resultados que se pueden obtener, además de la necesidad de considerar un número de individuos suficientemente grande que represente una muestra válida de la población.
- Es posible el análisis cuantitativo de las diferencias interaurales presentes de las muestras obtenidas para las pruebas realizadas. Si bien existen diferencias entre individuos, investigaciones previas presentan resultados definitivos y ampliamente aceptados para éstas diferencias, por lo que no se recomienda un estudio de las mismas.
- Se recomienda la experimentación con otros esquemas de convolución (división de los impulsos y técnicas empleadas) del procesamiento ya que éste no solo afecta al timbre del sonido pero puede llegar a ser un tema de investigación independiente y derivado de éste proyecto, que estudie la eficiencia e influencia en el espectro sonoro de distintos tipos de convolución de señales.
- Se recomienda buscar dispositivos de procesamiento externos que añadan la reverberación necesaria para una etapa posterior y se adapte el procesamiento que realiza el software con el fin de obtener resultados convincentes en todas las dimensiones.
- Se recomienda el software en MAX/MSP como plataforma para implementación de aplicaciones de audio para estudiantes sin conocimiento previo en áreas de programación en lenguajes

tradicionales donde su entorno modular mediante objetos y la opción de implementar objetos de terceros (libres y de fácil acceso) son empleados en redes para el procesamiento de las señales de audio y un diseño de una interface gráfica clara y amigable para el usuario.

Referencias

- Algazi, R., Avendano, C. y Duda, R. (1999). *An Adaptable Ellipsoidal Head Model For The Interaural Time Difference*. California, Estados Unidos: CIPIC, University of California Davis / Department of Electrical Engineering, San Jose State University.
- Algazi, R., Avendano, C. y Duda, R. (2001). *Estimation Of A Spherical-Head Model From Anthropometry*. Estados Unidos: Department of Electrical Engineering, San Jose State University / CIPIC, University of California Davis / Creative Advanced Technology Center, Scotts Valley.
- Araneda, M., Ramos, O. y Tommasini, F. (septiembre, 2010). *Auralización En Tiempo Real: Implementación Del Modelo Del Oyente*. Caseros, Argentina: 2^{do} Congreso Internacional de Acústica UNTREF.
- Araneda, M., Ramos, O. y Tommasini, F. (septiembre, 2010). *Contribución De La Cabeza, El Torso Y El Oído Externo En Las Funciones De Transferencia Relativas A La Cabeza*. Caseros, Argentina: 2^{do} Congreso Internacional de Acústica UNTREF.
- Audiolab, University of York. (2009). *Acoustics And Psychoacoustics Book*. Recuperado el 15 de febrero de 2013 de <http://www.openairlib.net/anechoicdb/content/acoustics-and-psychoacoustics-book>
- Begault, D. (2000). *3-D Sound For Virtual Reality And Multimedia*. Moffett Field, Estados Unidos: NASA/TM.
- Brice, B., David, D., Duncan, B., Hood, J., Patronis, E., Self, D., Sinclair, I., Singmin, A. y Watkinson, J. (2009). *Audio Engineering Know It All*. Burlington, Estados Unidos: Newnes / Elsevier.
- Brown, P. y Duda, R. (septiembre, 1998). A Structural Model For Binaural Sound Synthesis. *IEEE Transactions On Speech And Audio Processing*, 6 (5), 476-488.

- Carrión, A. (1998). *Diseño Acústico De Espacios Arquitectónicos*. Barcelona, España: Edicions UPS.
- Casey, M. (1996). *Implementation Of Binaural Spatializer*. Recuperado el 30 de abril de 2013 de <http://xenia.media.mit.edu/~mkc/micArray/node13.html>
- De La Fraga, L. (2001). *La Transformada Discreta de Fourier y la Transformada Rápida de Fourier*. Recuperado de 28 de septiembre de 2013 de <http://cs.cinvestav.mx/~fraga/Cursos/PDI/tdf.pdf>
- Duda, R. (2011). *3-D Audio For HCI*. Recuperado el 8 de julio de 2012 de <http://interface.cipic.ucdavis.edu/sound/tutorial/>
- Duda, R. (2011). *Head-Related Transfer Functions*. Recuperado el 8 de julio de 2012 de <http://interface.cipic.ucdavis.edu/sound/tutorial/hrtf.html>
- Duda, R. (2011). *HRTF-Based Systems*. Recuperado el 8 de julio de 2012 de <http://interface.cipic.ucdavis.edu/sound/tutorial/hrtfsys.html>
- Duda, R. (2011). *Psychoacoustics Of Spatial Hearing*. Recuperado el 8 de julio de 2012 de <http://interface.cipic.ucdavis.edu/sound/tutorial/psych.html>
- Duda, R. (2011). *Simple Spatial Audio Systems*. Recuperado el 8 de julio de 2012 de <http://interface.cipic.ucdavis.edu/sound/tutorial/simplesys.html>
- Duda, R. (2011). *Spatial Sound: An Introduction*. Recuperado el 8 de julio de 2012. <http://interface.idav.ucdavis.edu/sound/>
- Duda, R. (2011). *The Physics Of Sound*. Recuperado el 8 de julio de 2012 de <http://interface.cipic.ucdavis.edu/sound/tutorial/physics.html>
- Everest, A. (2007). *Critical Listening Skills For Audio Professionals*. Boston, Estados Unidos: Thomson Course Technology PTR.
- Everest, A. (2001). *The Master Handbook of Acoustics*. (4^a. ed.). Estados Unidos: McGraw-Hill.

- Fletcher, H. (septiembre, 1953). Stereophonic Recording and Reproducing System. *SMPTE Motion Imaging Journal*, 61 (3), 355-363.
- Harker, A. (2012). *Software*. Recuperado el 20 de febrero de 2013 de <http://alexanderjharker.co.uk/Software.html>
- ISOVER. (2000-2013). *Manual de Aislamiento en la Industria*. Recuperado el 2 de mayo de 2012 de <http://www.isover.es/content/download/568/4520/version/1/file/manual-aislamiento-industria.pdf>
- Lahoz, A. (2009). *Sonido 3D*. Barcelona, España: SAE Institute / University of Middlesex.
- LDS. (2003). *Understanding FFT Windows*. Recuperado el 28 de septiembre de 2013 de <http://www.physik.uni-wuerzburg.de/~praktiku/Anleitung/Fremde/ANO14.pdf>
- Long, M. (2006). *Architectural Acoustics*. Burlington, Estados Unidos: Elsevier Academic Press.
- Miles, D. y Runstein, R. (2005). *Modern Recording Techniques*. (6^a. ed.). Burlington, Estados Unidos: Focal Press / Elsevier.
- Miyara, F. (1999). *Control de Ruido*. Rosario, Argentina: Universidad Nacional de Rosario (UNR).
- Moller, H. (marzo, 1992). Fundamentals of Binaural Technology. *Applied Acoustics*, 36 (3-4), 171-218.
- Monte, E., Ramírez, N. y Valles, M. (diciembre, 2003). Sonido Virtual o Cómo Mover Una Fuente Monofónica Por El Espacio. *Burán*, 20, 39-44.
- National Instruments Corporation. (2013). *Windowing: Optimizing FFTs Using Window Functions*. Recuperado el 28 de septiembre de 2013 de <http://www.ni.com/white-paper/4844/en/>
- Penha, R. (2013). *Spatium · Max Objects*. Recuperado el 16 de febrero de 2013 de <http://spatium.ruipenha.pt/max/>

- Satarzadeh, P. (2006). *A Study Of Physical And Circuit Models Of The Human Pinnae*. California, Estados Unidos: University of California Davis.
- Thebinauralsource. (s.f.). *Binaural For Beginners (FAQs)*. Recuperado el 12 de julio de 2012 de <http://www.binaural.com/binfaq.html>.
- UC Davis College Of Engineering. (2009). *Binaural Tools*. Recuperado el 15 de febrero de 2013 de http://www.ece.ucdavis.edu/binaural/binaural_tools.html
- Warusfel, O. (s.f.). *Listen HRTF Database*. Recuperado el 2 de abril de 2013 de <http://recherche.ircam.fr/equipes/salles/listen/index.html>
- Warusfel, O. (s.f.). *Glossary*. Recuperado el 2 de abril de 2013 de <http://recherche.ircam.fr/equipes/salles/listen/glossary.html>.
- Warusfel, O. (s.f.). *System Protocol*. Recuperado el 2 de abril de 2013 de http://recherche.ircam.fr/equipes/salles/listen/system_protocol.html.
- Weisstein, E. (s.f.). *Convolution*. Recuperado el 1 de octubre de 2013 de <http://mathworld.wolfram.com/Convolution.html>
- White, G. (2010). *Introducción al Análisis de Vibraciones*. Woburn, Estados Unidos: Azima / DLI.

ANEXOS

Anexo 1. Glosario

- °: símbolo utilizado para indicar grados.
- λ : símbolo utilizado para indicar longitud de onda.
- *: símbolo utilizado para indicar convolución de funciones.
- Σ : Suma de términos de la ecuación.
- μ : micro. Prefijo de unidades que indica un factor de 10^{-6} .
- Ambiphonics: sistema de reproducción de señales binaurales mediante altavoces y procesamiento digital de señales.
- Apparent Source Width: término en inglés para Ancho Aparente de La Fuente, que hace referencia al tamaño o ancho del campo que ocupa la fuente en un Sistema de Sonido 3D.
- Auditory Spaciousness: término en inglés para Espacio Auditivo, que hace referencia al campo o imagen sonora que se genera en un Sistema de Sonido 3D.
- Bias: término en inglés que indica un prejuicio o parcialidad.
- Buffer: término en inglés que hace referencia dentro de un computador, a un espacio de memoria para el almacenamiento temporal de información necesario para la transferencia de la misma.
- cm: abreviación de centímetros, la centésima parte de la unidad de distancia, el metro [m].
- Comb: término en inglés para Peine, que se refiere a un tipo de filtro cuya función de transferencia presenta dicha forma gráficamente.
- Cos: abreviación de función Coseno.
- CPU: siglas en inglés de Central Processing Unit, elemento de un computador que se encarga del control y ejecución de todo procesamiento que se lleve a cabo, determinando la capacidad del mismo.
- Crossfade: término en inglés de Fundido Cruzado que hace referencia a la transición entre dos sonidos de forma suave mediante un control de amplitud de éstos.

- Crosstalk: término en inglés para Diafonía, la aparición de una señal transmitida por un canal en otro.
- Crosstalk Cancelled Stereo: término en inglés para Cancelación de Diafonía Estereofónica, que hace referencia a la decodificación de señales binaurales para su reproducción mediante altavoces evitando el cruce de señales entre canales.
- Cue. terminó en inglés que hace referencia a Indicador o Clave.
- dB: abreviación de Decibel, unidad empleada para la cuantificación de la presión sonora como Nivel de Presión Sonora.
- dBFS: abreviación en inglés de Decibel relative to Full Scale o Decibelio relativo a Escala Completa, unidad empleada en la cuantificación de señales de audio digital, cuyo rango dinámico depende del sistema y se encuentra entre un valor negativo y 0, donde el sistema se sobrecarga y distorsiona.
- Delay: terminó en inglés para Tiempo de Retardo.
- DFT: siglas en inglés de Discrete Fourier Transform, que hace referencia al algoritmo de la Transformada Discreta de Fourier para señales digitales con un número de muestras finitas.
- Diffuse Field Equalization: término en inglés para Ecuación en Campo Difuso, tipo de ecualización para compensación en la reproducción de señales en campo difuso.
- Doppler: fenómeno del mismo nombre, donde la frecuencia del sonido emitido por una fuente varía al variar la posición de ésta con respecto al oyente.
- DSP: siglas en inglés de Digital Signal Processing, término que hace referencia al campo encargado del estudio, diseño e implementación de procesamiento para señales digitales.
- Dummy Head: cabeza artificial equipada con micrófonos en la posición de los oídos empleada para grabación de sonido binaural.
- Eco: reflexión sonora que se recibe en un tiempo mayor al tiempo de integración del oído, por lo que es percibida como una repetición del sonido directo.

- FFT: siglas en inglés de Fast Fourier Transform, que hace referencia al algoritmo de la Transformada Rápida de Fourier para el análisis espectral de señales.
- FTF: siglas en inglés de Free Field Transfer Function, grupo de Funciones de Transferencia de Campo Libre de las cuales se derivan las HRTF.
- Filtro Pasa Bajo: tipo de filtro de audio que permite el paso del espectro sonoro de bajas frecuencias determinado por una frecuencia denominada frecuencia de corte.
- FIR: siglas en inglés de Finite Impulse Response o Impulso Respuesta Finita, tipo de filtro digital con un número finito de términos distintos a cero, para una señal impulso de entrada.
- Frame / FFT size: términos en inglés que especifica el número de muestras en un fragmento analizado en un algoritmo FFT.
- Free Field Equalization: término en inglés para Ecuación en Campo Libre, tipo de ecualización para compensación en la reproducción de señales en campo libre.
- Fuente sonora: cualquier mecanismo capaz de la generación y emisión de sonido.
- G: giga: Prefijo de unidades que indica un factor de 10^9 .
- Group Delay: término en inglés para Retardo de Grupo que se refiere al tiempo de retardo que existe entre componentes de distinta frecuencia de una misma señal en su propagación entre ambos oídos, conocidos como Retardo de Fase en función de la frecuencia.
- Head-Tracking: término en inglés que hace referencia a un sistema de seguimiento y detección de movimientos de la cabeza.
- HRIR: siglas en inglés de Head Related Impulse Response o Impulsos Respuesta Relativos a la Cabeza, equivalentes a los HRTF en el dominio del tiempo.
- HRTF: siglas en inglés de Head Related Transfer Function o Funciones de Transferencia Relativas a la Cabeza, que se obtienen por una

relación de presiones y presentan los cambios espectrales y temporales para la localización binaural.

- Hz: abreviación de Hertz, unidad de la frecuencia.
- IIR: siglas en inglés de Infinite Impulse Response o Impulso Respuesta Infinita, tipo de filtro digital con un número infinito de términos distintos a cero, para una señal impulso de entrada.
- ILD: siglas en inglés de Interaural Level Difference, que indica la Diferencia de Nivel Interaural entre dos oídos.
- In Head Location: término en inglés para Localización Dentro de la Cabeza, fenómeno que se presenta en la percepción de una fuente aparentemente dentro de la cabeza al emplear audífonos en su reproducción.
- Intermodulación: tipo de distorsión generado por interacción de dos ondas sonoras que producen componentes armónicos como resultado.
- Interpolación: proceso matemático de análisis numérico a partir del cual es posible obtener valores desconocidos para ciertos puntos en base a puntos donde el valor es conocido.
- ITD: siglas en inglés de Interaural Time Difference, que indica la Diferencia de Tiempo Interaural entre dos oídos.
- Java: lenguaje de programación orientado a objetos empleado para la creación de aplicaciones multiplataforma.
- K: kilo. Prefijo de unidades que indica un factor de 10^3 .
- Loudness: término en inglés para Sonoridad, la percepción subjetiva de la magnitud de un campo sonoro sonido, que depende de la amplitud y la frecuencia del sonido.
- Leakage: término en inglés que hace referencia al filtrado de energía en el espectro sonoro, como en el caso de un análisis DFT de señales no periódicas.
- M: mega. Prefijo de unidades que indica un factor de 10^6 .
- M: abreviación de metro [m], unidad de distancia.
- Microfonía: captación de ondas sonoras mediante transductores acústico-eléctrico denominados micrófonos.

- MIDI: siglas en inglés para Musical Instrument Digital Interface o Interfaz Digital para Instrumentos Musicales y representa un protocolo de comunicación estandarizado para la comunicación entre distintos dispositivos.
- mm²: abreviación de milímetro Cuadrado, la milésima parte de la unidad de área, el metro Cuadrado [m²].
- ms: abreviación de milisegundo, la milésima parte de la unidad de tiempo, el segundo [s].
- NPS: siglas de Nivel de Presión Sonora.
- Notch: término en inglés que hace referencia a un valle o caña de nivel en cierta zona del espectro.
- Omnidireccional: patrón de captación y reproducción de transductores que especifican una misma sensibilidad o misma intensidad respectivamente para todas las direcciones.
- Onda Estacionaria: onda que se produce a partir de la suma de amplitudes de dos ondas sinusoidales que se propagan con sentido opuesto, creando un desfase entre sí y generando puntos donde la amplitud es nula (nodos) y máxima (antinodos).
- Oscilación: variación de magnitud que se manifiesta alrededor de un punto de forma repetitiva.
- Overlapping: término en inglés que hace referencia a la sobre posición de frames sucesivos en un análisis FFT.
- Overlap-add: término en inglés para Solapamiento-Suma, técnica de procesamiento de señales digitales donde la señal de entrada es dividida para ser procesada por separado, y recombinada en su salida.
- P₀: presión de referencia en la ecuación de Nivel de Presión Sonora, igual a 2×10^{-6} [Pa].
- Pa: abreviación de Pascal, unidad de presión.
- Phon: término en inglés para Fon, unidad de sonoridad basadas en las Curvas Isofónicas.
- Pinna: término en inglés para Pabellón Auditivo, uno de los componentes del oído externo.

- Pinna Notch: término en inglés que indica un valle en el espectro del sonido que se produce debido a la interacción de la onda sonora y sus reflexiones en el pabellón auditivo.
- Pitch: término en inglés para Altura Tonal, la percepción subjetiva de la frecuencia de un sonido.
- Playback: término en inglés que hace referencia a la reproducción de audio almacenado previamente registrado.
- Point Of View: término en inglés para Punto de Vista que hace referencia a técnica de narrativa visual donde lo que es captado por la cámara es presentado en primera persona como lo que esta visualizando un individuo.
- Pole/Zero: término en inglés para Polo/Cero, indica en una función de transferencia, las frecuencias donde la ganancia es infinita (resonancia) o cero respectivamente.
- Post-producción: última etapa en la producción de material audiovisual, donde se trabaja y procesa el material obtenido en la etapa de producción (grabación).
- Predelay. término en inglés para el parámetro en un procesador de reverberación artificial que controla el tiempo entre el sonido directo y las reflexiones y reverberación.
- Raw: término en inglés para un tipo de formato de archivo de audio digital, que al no poseer ninguna indicación, sus características no son conocidas y generalmente se emplea para archivos que no han sido codificados.
- Resonancia: propiedad de todo cuerpo que posea elasticidad, que al ser estimulado por una fuerza externa de igual frecuencia que su frecuencia natural, presentará una vibración de amplitud máxima.
- Reversal: término en inglés que hace referencia a la confusión en la localización de la fuente en la escucha que crea una imagen invertida sobre el eje interaural de la posición correcta de la misma.
- s: abreviación de segundo [s], unidad de tiempo.

- Señal Binaural: señal que contiene información sonora o de audio en dos canales independientes, para ser reproducida en los oídos izquierdo y derecho de forma discreta.
- Señal Estereofónica: señal que contiene información sonora o de audio en dos canales independientes.
- Señal Monofónica: señal que contiene información sonora o de audio en un solo canal.
- Shoulder Reflection: término en inglés para Reflexión de Hombro que puede producirse para ciertas ondas sonoras como parte de la escucha binaural.
- Sine Sweep: término en inglés para una señal de audio con forma de onda sinusoidal y frecuencia variable de forma constante entre un rango definido, conocido como Barrido de Frecuencia
- Side Lobe: término en inglés que hace referencia a los lóbulos de frecuencia de menor nivel que se presentan a los costados de un pico de amplitud máximo visibles en el espectro sonoro de una señal.
- Síntesis de Sonido 3D / Sistemas de Sonido 3D: hace referencia a cualquier tipo de procesador software o hardware que tenga como objetivo producir señales binaurales.
- Snowman: modelo estructural de HRTF que consiste en una cabeza esférica sobre un torso esférico.
- Sone: término en inglés para Son, unidad de sonoridad de un sonido con respecto a un sonido de 40 [fon].
- Split Image: término en inglés para Imagen Dividida que se refiere a la percepción de dos señales como resultado de ciertos modelos estructurales de HRTF.
- Subwoofer: término en inglés para altavoces que se encargan de reproducir contenido de bajas frecuencias.
- Surround: término en inglés que hace referencia a Sonido Envoltivo mediante altavoces dispuestos alrededor del oyente.

- Threshold of reflection detection: término en inglés para Umbral de Detección de Reflexión que especifica un nivel y tiempo de retardo y sobre el cual las reflexiones son percibidas como eco.
- Tono Puro: sonido de forma de onda sinusoidal con una frecuencia única.
- Torso Diffraction: término en inglés para Difracción en el Torso, que se presenta como un obstáculo para ondas sonoras como parte de la escucha binaural.
- WAV: término en inglés para Onda, que hace referencia a un formato de archivos de audio digital con extensión .wav o .wave desarrollado por Microsoft e IBM capaz de almacenar información sin compresión en varias frecuencias de muestreo y profundidad de bits.
- Windowing: término en inglés que indica el suavizado de frames en un análisis FFT mediante una envolvente de amplitud.

Anexo 2. Lista y Función de Objetos de MAX/MSP Utilizados

Se presenta una lista de todos los objetos del entorno de programación MAX/MSP en orden alfabético según el tipo de archivo y una breve explicación de los mismos.

Objetos MAX

- **==**: realiza comparaciones de dos valores numéricos que recibe en sus entradas, en caso de ser iguales, produce un valor de “1” en su salida, en caso contrario, producen un valor de “0”.
- **/:** operador que realiza la división del valor que recibe en su entrada con el valor especificado en su argumento.
- **button**: objeto en forma de botón que genera un mensaje “bang” en su salida al ser pulsado o recibir cualquier mensaje en su entrada.
- **comment**: permite introducir comentarios en el patch y no afectan el comportamiento del resto de objetos.
- **dial**: controles rotatorios para generar valores numéricos en sus salidas.
- **folder**: especifica una carpeta en el disco. Entre paréntesis, se establece la ruta de dicha carpeta, donde se emplean el símbolo “/” para especificar la jerarquía de subcarpeta.
- **gate**: el valor que recibe en su entrada derecha (mientras sea es diferente de “0”) determina la salida correspondiente que recibe el mensaje que ingresa por la entrada derecha. Al recibir un valor de “0” en su entrada derecha, no permite el paso para cualquiera de sus salidas.
- **gswitch2**: puerta que posee una entrada derecha cuyo mensaje puede pasar mediante dos salidas. La entrada izquierda actúa como la entrada de control que alterna entre sus salidas al recibir un mensaje “bang”.
- **if**: realiza una condición especificada en su argumento para una o varias entradas, denominadas como “\$i”. Es posible establecer la salida deseada que produce el objeto empleando un modelo de condición “if-then” (en caso de obtener el resultado deseado) o “if-else” (en caso de no obtenerla) como es empleado en lenguajes de programación tradicionales.

- *inlet/outlet*: representan entradas y salidas de *subpatch*, necesarias para el recibir y enviar de datos entre éstos.
- *led*: indicador en forma de LED que se ilumina al recibir un valor de “1” y se apaga al recibir un valor de “0”, generando los mismos valores en su salida.
- *loadbang*: genera un mensaje “*bang*” en su salida al iniciarse el *patch*.
- *loadmess*: genera el mensaje especificado en su argumento al iniciarse el *patch*.
- *mxj buf.Op*: se encarga de realizar operaciones adicionales para *buffers~* que no se pueden realizar mediante mensajes o argumentos en los mismos. El prefijo “*mxj*” en el nombre del objeto indica que se trata de un objeto escrito en Java, que interactúa e intercambia información con MAX/MSP.
- *number*: permite seleccionar un valor o mostrar un valor numérico que ingresa en su entrada en su salida.
- *pak*: recibe valores en sus entradas y genera una lista de los mismos en una salida única cada vez que un nuevo valor de entrada es recibido.
- *panel*: genera un panel rectangular con múltiples opciones de configuración mediante mensajes o inspector para el diseño de la interface visual del *patch*.
- *prepend*: antepone en su salida lo que se especifique como su argumento, es decir, lo que se especifique seguido de la palabra “*prepend*”.
- *send/recive*: se emplean en lugar de *patchcords* como conexiones inalámbricas. Los mensajes enviados mediante objeto *send* son recibidos (y únicamente entendidos) por objetos *recive* con su mismo nombre (establecido en su argumento). Facilitan el control y visualización de las conexiones entre grupos de objetos donde existen un gran número de éstas.
- *toggle*: switch *on/off* que se activa mediante un mensaje “1”, seleccionándolo mediante un *clic* (y representado por una X en el ícono

del objeto) y se desactiva con un mensaje “0”, generando los mismos valores en su salida.

- *umenu*: presenta una lista de ítems como son denominados en el inspector en forma de un menú desplegable.
- *unpack*: recibe una lista en su entrada y la descompone en sus elementos, enviándolos de forma individual por cada una de sus salidas.

Objetos MSP

- *+~*: realiza la suma de los valores recibidos en sus entradas, generando su resultado en su salida.
- **~*: realiza la multiplicación de los valores recibidos en sus entradas, generando su resultado en su salida.
- *buffer~*: almacena muestras de audio en el disco en memoria que es procesada por otros objetos, por lo que debe poseer un nombre único y un tiempo máximo de almacenamiento en su argumento.
- *ezdac~*: realiza la función de activar y desactivar la salida de audio de MAX/MSP y la de un conversor digital/analógico. Para poder escuchar cualquier señal, este deberá activarse mediante un *click* que cambiará su color a azul y puede desactivarse al ser presionado nuevamente (recuperando su color original).
- *gain~*: permite realiza un control de nivel escalando una señal de audio
- que recibe en su entrada izquierda y la envía por su salida izquierda.
- *levelmeter~*: indicadores tipo VU metros del nivel en valor eficaz de una señal de audio. Presentan varios modos de operación establecidos mediante mensajes externos. El mensaje “*mode \$1*” (denominado “VU”) establece un tiempo de ataque y relevo (*attack* y *release* respectivamente) de 330 [ms].
- *meter~*: indicador tipo medidor de pico o *peak meter* de una señal de audio. Cada led representa 3 [dB].
- *sfplay~*: lleva a cabo la reproducción de señales de audio que recibe como mensajes “*open*” seguido del nombre del archivo, o seleccionándolo directamente desde el disco duro (mediante un click en

un mensaje “*open*”). Para iniciar la reproducción deberá recibir un mensaje de valor “1”. Si el mensaje que recibe es “0”, la reproducción se detendrá.

- *sfrecord~*: se encarga de la grabación de una señal de audio como un archivo en el disco. Mediante un mensaje “*open*” seguido del nombre del formato, permite establecer el formato del archivo de audio de la grabación. Para iniciar el proceso, deberá recibir un mensaje de valor “1” y especificar un nombre para el archivo y su lugar de almacenamiento en el disco. Si el mensaje que recibe es “0”, la grabación se detendrá.
- *spectroscope~*: indicador tipo espectrograma en un gráfico de amplitud versus frecuencia para una señal de audio.

Objetos Externos

- *partconvolve~*: realizar convolución mediante un algoritmo FFT. Toma las muestras de un archivo de audio almacenado en un objeto *buffer~* y lo emplea como un impulso respuesta en la convolución con una señal de audio, recibiendo a ambas señales en su única entrada. El valor de los “*Vector Size*” no puede ser menor a 4 para su funcionamiento.
- *timeconvolve~*: realiza convolución en el dominio del tiempo (en tiempo real). Toma las muestras de un archivo de audio almacenado en un objeto *buffer~* y lo emplea como un impulso respuesta en la convolución con una señal de audio, recibiendo a ambas señales en su única entrada.

Anexo 3. Información Técnica del Proceso de Medición de HRIR del Proyecto Listen

Los archivos de audio correspondientes a las HRIR de los cinco individuos empleados en la Sección de Selección de Individuo y adjuntos en el CD (dentro de la carpeta "LISTEN HRIR") fueron obtenidos del proyecto LISTEN, por *IRCAM* y *AKG* como parte del programa *Information Society Technologies* (IST), parte del Quinto Programa Marco (*Fifth Framework Program*) de la Unión Europea.

Especificaciones e información técnica sobre el procedimiento de medición son descritos como parte de Sistema y Protocolo del proyecto LISTEN.

Control de posición de altavoces:

- Grúa motorizada.
- Estructura metálica cubierta con paneles de melanina.

Control de posición de individuo:

- Silla común con altura e inclinación ajustable.
- Tornamesa B&K 9640 y controlador B&K 5960.
- Interface de control IEEE-488.

Cámara Anecoica.

- Dimensiones: 8.1 × 6.2 × 6.45 [m].

Cadena Electroacústica.

- **Medición de impulsos de los altavoces:**
 - Micrófonos B&K type 4149.
 - Preamplificador B&K 2801.

- **Medición de impulsos de los micrófonos:**
 - B&K artificial mouth.
- **Medición de impulsos de individuos:**
 - Micrófonos: Knowles FG3329.
 - Preamplificador: personalizado de 40 [dB] de ganancia.
- **Altavoces:**
 - TANNOY system 600 con amplificador Yamaha.

Sistema de Control:

- Computador Macintosh: procesador G4 de 500 [MHz] memoria RAM de 256 [Mb] y sistema operativo Mac OS 9.2.
- Tarjeta de Sonido: RME Multiface con capacidad de cuantización en 24 [bits] y frecuencia de muestreo de 96 [kHz].
- Software Max/MSP: medición de impulsos respuesta y control de todos los dispositivos.
- Señal de prueba: barrido logarítmico de 8192 muestras para dos canales con frecuencia de muestreo de 44100 [Hz].

Anexo 4. Archivos Externos Utilizados en el Software

Muestras en Cámara Anecoica

Los archivos de audio de grabaciones en cámara anecoica empleados en la Sección de Entrada del software desarrollado fueron obtenidos de la librería Open Acoustic Impulse Response (OpenAIR). Las siete grabaciones, adjuntas en el CD (dentro de la carpeta “Muestras en Cámara Anecoica”) forman parte del libro “*Acoustics and Psychoacoustics*” de David Howard y Jamie Angus. (Audiolab, University of York, 2009).

Tabla 8. Denominación de los archivos adjuntos y sus nombres originales en la librería OpenAIR.

Nombre en Carpeta	Nombre Original
<i>adult_female_speech</i>	<i>Adult Female Speech: Read passage.</i>
<i>bagpipe_music</i>	<i>Bagpipe music: Brief music snippet on bagpipes.</i>
<i>bagpipe_steady_chord</i>	<i>Bagpipe steady chord: Drone and steady note on bagpipes.</i>
<i>flute_arpeggio</i>	<i>Flute Arpeggio: G major arpeggio (D4 to B6) on a flute.</i>
<i>flute_music</i>	<i>Flute music: Brief music snippet on flute.</i>
<i>tuba_arpeggio</i>	<i>Tuba arpeggio: G major arpeggio (G1 to G3) on a tuba.</i>
<i>tuba_music</i>	<i>Tuba music: Brief music snippet on a tuba.</i>

Todas las grabaciones fueron realizadas por los ingenieros Jez Wells y David Howard y presentan una cuantización de 16 [bits], frecuencia de muestreo de 44100 [Hz] en formato WAV, estéreo.

La atribución de los derechos se la hace a OpenAIR y Audiolab de la Universidad de York.

Objetos Externos AHarker.

Los objetos *partconvolve~* y *timeconvolve~* empleados en el software desarrollado son parte del paquete de objetos externos AHarker External v.1.0 desarrollados por Alexander J. Harker en 2011, quien es portador de los derechos de los mismos.

La licencia de uso, especifica que los objetos pueden utilizarse “forma libre para propósitos no comerciales” (Harker, 2011) y es adjuntada en el CD (en la carpeta “AHarker External”) como parte de las condiciones de redistribución del software.

Anexo 5. Contenidos del CD Anexo

Con la entrega del trabajo se adjunta un CD-ROM anexo que presenta los siguientes contenidos:

- “Procesador Mono Binaural”: carpeta que contiene al software desarrollado así como todos los archivos necesarios para su funcionamiento, como se indica en el apartado 2.2.7.
- “Código Fuente del Software”: archivo de procesador de texto que contiene el código fuente del software desarrollado.
- “Desarrollo De Un Software De Procesamiento Para Obtener Señales De Audio Binaurales A Partir De Señales De Audio Monofónicas”: todo el contenido presentado en un archivo único en formato PDF.
- Carpeta “Resultados” con las imágenes del análisis de las pruebas presentadas en mayor tamaño y resolución.