DISEÑO Y PRODUCCIÓN DE UN BANCO DE AMBIENTES 3D IMPLEMENTANDO LA HRTF *(HEAD RELATED TRANSFER FUNCTION)*

JUAN CAMILO HEREDIA CARVAJAL

UNIVERSIDAD DE SAN BUENAVENTURA

FACULTAD DE INGENIERÍA

INGENIERÍA DE SONIDO

BOGOTÁ D.C.

2010

DISEÑO Y PRODUCCIÓN DE UN BANCO DE AMBIENTES 3D IMPLEMENTANDO LA HRTF *(HEAD RELATED TRANSFER FUNCTION)*

JUAN CAMILO HEREDIA CARVAJAL

PROYECTO PARA OPTAR AL TITULO DE INGENIERO DE SONIDO

Tutor de proyecto
CARLOS ANDRES CABAS
Ing. De Sonido

UNIVERSIDAD DE SAN BUENAVENTURA
FACULTAD DE INGENIERÍA
INGENIERÍA DE SONIDO
BOGOTÁ D.C.
2010

CONTENIDO

pág.

INTRODUCCIÓN ***8***

1. PLANTEAMIENTO DEL PROBLEMA ***9***

1.1 ANTECEDENTES **9**

1.2 DESCRIPCIÓN Y FORMULACIÓN DEL PROBLEMA **10**

1.3 JUSTIFICACIÓN **11**

1.4 OBJETIVOS DE LA INVESTIGACIÓN **12**

1.4.1 Objetivo General 12

1.4.2 Objetivos Específicos 12

1.5. ALCANCES Y LIMITACIONES **13**

1.5.1 Alcances 13

1.5.2 Limitaciones 13

2. MARCO TEÓRICO-CONCEPTUAL ***14***

2.1 OÍDO HUMANO **14**

2.1.1 Fisiología 14

2.1.2 Psicoacústica 17

2.2 HRTF (*HEAD RELATED TRANSFER FUNCTION*) **21**

2.2.1 Medición de la HRTF 22

2.2.2 Extracción de Características significativas 23

2.2.3 Sistemas Basados en HRTF 26

2.3 SISTEMAS COMERCIALES PARA LA REPRODUCCIÓN DE AUDIO 3D **27**

2.3.1 Auriculares 27

2.3.2 Altavoces Estéreo 28

2.3.3 Sistemas Multicanal (*Surround 5.1*) 29

2.4 CANCELACIÓN DE LA INTERFERENCIA EN LOS ALTAVOCES (*CROSSTALK CANCELLATION*) **29**

3. METODOLOGÍA ***33***

3.1 ENFOQUE DE LA INVESTIGACIÓN **33**

3.2 LÍNEA DE INVESTIGACIÓN DE USB / SUB-LÍNEA DE FACULTAD / CAMPO TEMÁTICO DEL PROGRAMA **33**

3.3 TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN **34**

3.3.1 Análisis de las principales locaciones utilizadas y de mayor interés en las producciones locales. 34

3.3.2 Captura de los ambientes 34

3.3.3 Extracción de características de la Función de transferencia relacionada a la cabeza. 35

3.3.4 Cancelación de la interferencia en altavoces (*Crosstalk Cancellation*) 35

3.4 POBLACIÓN Y MUESTRA **36**
3.5 HIPÓTESIS **36**
3.6 VARIABLES **36**
3.6.1 Variables Independientes 36
3.6.2 Variables Dependientes 36
4. DESARROLLO INGENIERIL ***37***
4.1 BASE DE DATOS **37**
4.1.1 Listado de Locaciones 37
4.1.2 Selección de Locaciones 40
4.2 CAPTURA DE AMBIENTES **41**
4.2.1 Selección de Equipos 41
4.2.2 Configuración de Captura 43
4.2.3 Detalles de cada captura 44
4.3 EDICIÓN DE LOS AMBIENTES **52**
5. PRESENTACIÓN Y ANÁLISIS DE RESULTADOS ***54***
5.1 EXTRACCIÓN DE DIT (*Diferencia Interaural de Tiempo)* **54**
5.2 DISEÑO DEL AUDIO TRANSAURAL PARA UN SISTEMA ESTEREO 2.0 ..... **63**
5.3 DISEÑO DEL AUDIO TRANSAURAL PARA UN SISTEMA SURROUND 5.1 ..... **67**
5.4 EVALUACIÓN DEL AUDIO 3D **68**
5.4.1 Evaluación de parámetros significativos 68
5.4.2 Evaluación de los sistemas de reproducción 71
6. CONCLUSIONES ***73***
7. RECOMENDACIONES ***75***
8. BIBLIOGRAFÍA ***76***
ANEXOS ***78***

LISTA DE FIGURAS

pág.

Figura 1: Fisiología del Oído humano 14
Figura 2: Ubicación del oído Externo 15
Figura 3: Ubicación del oído medio 15
Figura 4: Curvas de igual sonoridad 17
Figura 5: Grafica de la relación entre Pitch y Frecuencia a 60dB 18
Figura 6: Frente de onda incidiendo sobre la Pinna 19
Figura 7: Proceso de localización mediante DIT y DII. 20
Figura 8: Planos de localización sonora. a) Plano frontal. b) Plano Horizontal. c) Plano Medio. 20
Figura 9: Respuesta en frecuencia para sonidos incidentes sobre la cabeza y de frente. 21
Figura 10: Planos formados por las 3 dimensiones de la HRTF. 22
Figura 11: Expresiones matemáticas para la HRTF. 26
Figura 12: Esquema de reproducción binaural por audífonos. 27
Figura 13: Señales de Interferencia producidas por los altavoces (Crosstalk) ... 28
Figura 14: Configuración de altavoces a 30º 28
Figura 15: Configuración de un sistema 5.1 29
Figura 16: Propagación de la señal en una reproducción estereofonica. 30
Figura 17: Diagrama de Conexión para la captura. 43
Figura 18: Respuesta al impulso L30. 56
Figura 19: Zoom de la respuesta L30. 56
Figura 20: Respuesta al impulso R30. 56
Figura 21: Zoom de la respuesta R30. 56
Figura 22: Respuesta al impulso L330. 57
Figura 23: Zoom de la respuesta L330. 57
Figura 24: Respuesta al impulso R330. 57
Figura 25: Zoom de la respuesta R330. 57
Figura 26: Respuesta al impulso L110. 60
Figura 27: Zoom de la respuesta L110. 60
Figura 28: Respuesta al impulso R110. 60
Figura 29: Zoom de la respuesta R110. 60
Figura 30: Respuesta al impulso L250. 61
Figura 31: Zoom de la respuesta L250. 61
Figura 32: Respuesta al impulso R250. 61
Figura 33: Zoom de la respuesta R250. 61
Figura 34: Flujo de señal para los filtros de cancelación. 64
Figura 35: Calificación de parámetros por persona. 70
Figura 36: Calificación de parámetros general. 70
Figura 37: Áreas de preferencia por persona. 72
Figura 38: Preferencia de sistema general. 72

LISTA DE IMÁGENES

pág.

Imagen 1: Interferencia que impide la reproducción del audio 3D 8
Imagen 2: Wavesurround. 9
Imagen 3: Fisiología del oído interno 16
Imagen 4: Medición de la HRTF con Dummy Head. 22
Imagen 5: Medición de la HRTF con una persona real. 22
Imagen 6: Medición de la HRTF en Cámara Anecóica. 23
Imagen 7: Diferencia interaural de tiempo extraída de la HRIR. 23
Imagen 8: Gráfica de las DIT para 30º de elevación y variaciones en el plano horizontal. 24
Imagen 9: Diferencia interaural de intensidad extraída de la HRIR 24
Imagen 10: Respuesta en frecuencia de la HRTF para diferentes elevaciones. 25
Imagen 11: Micrófonos In-Ear SP-TBF-2 de Sound Professionals 34
Imagen 12: Grabadora Digital Microtrack II de m-audio 34
Imagen 13: Extracción de DIT 35
Imagen 14: Respuesta en frecuencia para los 2 oídos con 15º de elevación. 35
Imagen 15: Respuesta en Frecuencia de los Micrófonos SP-TFB-2. 42
Imagen 16: Biblioteca pública el tintal. 44
Imagen 17: Centro Comercial Gran Estación. 45
Imagen 18: Billar Londres. 45
Imagen 19: Estadio Nemesio Camacho el Campín. 46
Imagen 20: Puente de la 68 con 63. 46
Imagen 21: Terminal de Transporte de Bogotá. 47
Imagen 22: Monserrate. 47
Imagen 23: Salitre Mágico. 48
Imagen 24: Parque Simon Bolívar. 48
Imagen 25: Plaza de Banderas. 49
Imagen 26: Plaza Parque de Lourdes. 49
Imagen 27: Parque de los Novios. 50
Imagen 28: Plaza de Bolívar. 50
Imagen 29: Museo Iglesia de Santa Clara 51
Imagen 30: Complejo acuático de Bogotá. 51
Imagen 31: Corte de región en proceso de edición. 52
Imagen 32: Crossfade en proceso de edición. 52
Imagen 33: Filtro pasa bajas en proceso de edición. 53
Imagen 34: Propiedades de las respuestas a 30 y 330 grados. 55
Imagen 35: Ajustes para los filtros que se dirigen al altavoz izquierdo. 65
Imagen 36: Ajustes para los filtros que se dirigen al altavoz derecho. 65
Imagen 37: Flujo de señal y nivelación en Pro Tools. 66
Imagen 38: Ajustes para los filtros de los altavoces LS y RS. 67

LISTA DE TABLAS

pág.

TABLA 1: 4º ENCUENTRO DISTRITAL AUDIOVISUAL (NOVIEMBRE DE 2009). 37
TABLA 2: LISTADO INICIAL DE LOCACIONES 1ª SECCIÓN. 38
TABLA 3: LISTADO INICIAL DE LOCACIONES 2ª SECCIÓN. 39
TABLA 4: LISTADO DE LOCACIONES ESCOGIDAS PARA GRABACIÓN. 40
TABLA 5: ESPECIFICACIONES DE LOS MICRÓFONOS SP-TFB-2. 41
TABLA 6: ESPECIFICACIONES DE LA MICROTRACK II PARA LA ENTRADA ESTÉREO DE 1/8". 42
TABLA 7: AJUSTES DE GRABACIÓN EN LA MICROTRACK II 44
TABLA 8: RESULTADOS EN LA CALIFICACIÓN DE PARÁMETROS. 69
TABLA 9: EVALUACIÓN DE LOS SISTEMAS DE REPRODUCCIÓN. 71

INTRODUCCIÓN

Con una captura binaural podemos obtener la señal generada por una fuente, junto con las componentes de la función de transferencia relacionada a la cabeza. Estas componentes son producto de las reflexiones y difracciones del sonido incidente sobre el hombro, la cabeza y la pinna. Teniendo en cuenta estos factores que afectan a la señal original, podemos generar la sensación de tridimensionalidad en el espacio sonoro recreado.

Las capturas que se realizaran en este proyecto serán de ambientes y locaciones significativos en la ciudad de Bogotá, teniendo en cuenta que sean usables en producciones locales.

Existen varias formas de obtener audio en 3D, las mas exitosas han resultado en el uso un *Dummy Head* (maniquí) para la captura y en procesamiento sobre una señal estereofónica. Los *Dummy Head* son construidos teniendo en cuenta las características físicas del ser humano, algunos solo la cabeza y otros incluyendo el torso, poseen dos micrófonos omnidireccionales uno a cada lado emulando la posición de los oídos, y en las mejores construcciones se emula la fisiología del pabellón auditivo externo. El procesamiento de la señal estereofónica se realiza por medio de aproximaciones matemáticas a la función de transferencia relacionada a la cabeza.

Para propósitos de este proyecto se usará una función de transferencia individual, realizando la captura con micrófonos in-ear sobre una persona real.

Por otra parte, uno de los grandes limitantes que ha tenido la captura binaural, es su reproducción ligada al uso de audífonos, esto se debe a que cada oído requiere de una señal individual respectiva para izquierda y derecha, lo que no ocurre en una reproducción por altavoces.

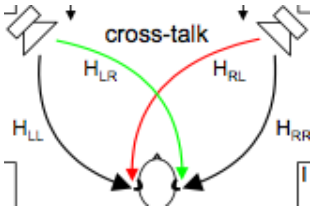


Imagen 1: Interferencia que impide la reproducción del audio 3D

En este proyecto se usara el análisis de la HRTF para generar una cancelación de la interferencia y producir la tridimensionalidad en formatos 2.0 y 5.1 para reproducción por altavoces.

1. PLANTEAMIENTO DEL PROBLEMA

1.1 ANTECEDENTES

Desde la década de los 70 con la aparición del *Soundfield Microphone* y configuraciones de grabación micrófono-parlante en diferentes salas, el ser humano ha buscado la recreación de los espacios sonoros de una forma muy cercana a la realidad, lo que hoy en día ha generado un campo de estudio llamado *Realidad Acústica Virtual*.

En 1983 se realizó la primera aplicación de la binauralidad para su difusión comercial, el argentino Zucarelli participó en la ingeniería del álbum *The Final Cut* de Pink Floyd. Utilizando para este capturas con *Dummy Head* y su algoritmo que emula las características del pabellón auditivo.

Para 1989 como avance para el procesamiento de señales, Richter y Persterer crearon el primer procesador de audio binaural, capaz de emular las características de la HRTF para cualquier señal monofónica.

En el año 2000 Pearl Jam lanza al mercado su álbum *Binaural* llevando como titulo la técnica implementada en las grabaciones del disco. De este año hasta hoy se han escrito varios documentos acerca de la implementación de la HRTF en el sonido tridimensional, realizando mejoras en cuanto a la reproducción y la comodidad del oyente en el ambiente.

En el 2004 el ingles Steve Marshal saca al mercado su producción *Bilocation* en donde utiliza técnicas que simulan una captura binaural sobre ambientes y música reproducida en espacios de la India. Posteriormente crea el audio transaural para que pueda ser reproducido por altavoces en formatos 2.0 y 5.1. En los últimos años se han desarrollado herramientas de reproducción y edición de audio implementando la HRTF. Ejemplos de esto son el *Yamaha Virtual Cinema DSP* que genera fuentes virtuales atrás del oyente para reproducciones a 2 altavoces, y el *WaveSurround* de *Wave Arts* que permite el posicionamiento de una señal en un espacio tridimensional.

Imagen 2: Wavesurround.[1]

[1] Imagen extraída de <http://wavearts.com/products/plugins/wavesurround-4/>

La Universidad de Virginia en Estados Unidos, tiene publicada en su pagina Web el trabajo de Michael Rasbury titulado *Earth Recordings* en donde implementa la HRTF sobre ambientes para reproducción por audífonos.

En Colombia la implementación de la binauralidad se encuentra aún en proyectos de investigación. En la Universidad Javeriana, Jorge Torres realizó una tesis que incluye la grabación binaural para elementos de un producto publicitario. En la Universidad de San Buenaventura existen trabajos de grado terminados y en proceso que utilizan capturas binaurales para el diseño sonoro, o que muestran la aplicabilidad de la binauralidad en los medios.

1.2 DESCRIPCIÓN Y FORMULACIÓN DEL PROBLEMA

Algunas producciones colombianas, implementan en su diseño sonoro librerías extranjeras que incluyen ambientes, efectos y *foley*. Por supuesto una librería extranjera tendrá un contenido de productos ajenos a la cultura propia de nuestro país y específicamente de la cultura y ambientes de nuestra ciudad.

La falta de elementos autóctonos en las producciones locales, solo nos convierte en imitadores y seguidores de los pioneros en el área del diseño sonoro. Es importante que se puedan producir herramientas propias de una forma ingenieril e innovadora que permitan competir a las producciones locales con las producciones extranjeras.

¿Cómo aplicar las características de la HRTF sobre ambientes para uso en el diseño sonoro de producciones audiovisuales locales?

1.3 JUSTIFICACIÓN

El sonido siempre ha jugado un papel importante a la hora de transmitir sensaciones e interactuar con el consumidor dentro de entornos virtuales. La tridimensionalidad generada haciendo uso de la HRTF se ha convertido en una herramienta muy útil, permitiendo sumergir al oyente en el espacio recreado.

Para poder competir con producciones de otros lugares del mundo, hay que producir con herramientas ingenieriles y de calidad, es importante demostrar la aplicabilidad de la acústica y la psicoacústica en áreas donde el trabajo se maneja aún de una forma muy artesanal, con esto incrementar el interés de las producciones audiovisuales en poseer un ingeniero de sonido propio y que brinde herramientas con conocimiento.

Este proyecto usará las características significativas de localización sonora, para generar sensaciones de espacialidad cercanas a la realidad y aplicables al diseño sonoro.

Se diseñara el producto de ambientes 3D para su reproducción por 3 medios: Audífonos, 2.0 por altavoces y 5.1. Estos 3 formatos amplían la aplicabilidad del banco de ambientes y permiten su uso en el diseño sonoro: desde radio y videojuegos hasta cortometrajes y cine.

1.4 OBJETIVOS DE LA INVESTIGACIÓN

1.4.1 Objetivo General

- Diseñar y producir un banco de ambientes[2] bogotanos en 3D implementando las características de localización de la HRTF.

1.4.2 Objetivos Específicos

- Realizar el proceso de selección de 15 espacios y locaciones, sobresalientes en la ciudad y aplicables al diseño sonoro.
- Realizar la captura con una HRTF individual de los ambientes seleccionados.
- Editar las capturas obtenidas para su reproducción por audífonos.
- Crear el audio transaural de los ambientes para su reproducción por altavoces en un sistema 2.0.
- Crear el audio transaural de los ambientes para su reproducción en un sistema 5.1

[2] Componentes sonoros de lugares cerrados y abiertos.

1.5. ALCANCES Y LIMITACIONES

1.5.1 Alcances

Hacer uso del banco de ambientes en producciones locales.

Realizar la difusión del banco en las principales productoras locales.

Implementar las características de la HRTF en musicalización.

Realizar la convolución de señales con respuestas al impulso obtenidas sobre el canal auditivo (HRIR[3]).

Los anteriores puntos son aplicaciones y usos que podría llegar a tener el sistema de captura y edición, esto con el fin de ampliar y aumentar la necesidad del ingeniero de sonido en producciones que requieran manejo de audio y creación con alta calidad y conocimiento.

1.5.2 Limitaciones

El resultado de la tridimensionalidad en el audio transaural, está limitado a una configuración de los altavoces específica.

No se cuenta por el momento con un estudio profesional para monitoreo en 5.1.

Individualizando la HRTF es posible que en algunas capturas sea necesario retirar efectos de pinna.

[3] Head Related Impulse Response

2. MARCO TEÓRICO-CONCEPTUAL

2.1 OÍDO HUMANO

Oído es un término generalmente usado para expresar tanto la fisiología del sistema auditivo como la sensación de escucha. Físicamente funciona como una serie de transductores de energía, empezando con una transducción de energía sonora a energía mecánica, seguido de una transducción de energía mecánica a energía eléctrica (Impulsos eléctricos en el cerebro). La sensación de escucha es generada en el cerebro cuando este interpreta la información que llega por medio de las señales neuronales.

2.1.1 Fisiología

El sistema auditivo se encuentra dividido en 3 sectores: El oído externo que comprende desde el pabellón auditivo hasta el tímpano, el oído medio que comprende la zona entre el tímpano y la ventana oval y por ultimo el oído interno que se encuentra dividido en 2 zonas, el laberinto anterior y el laberinto posterior.[4]

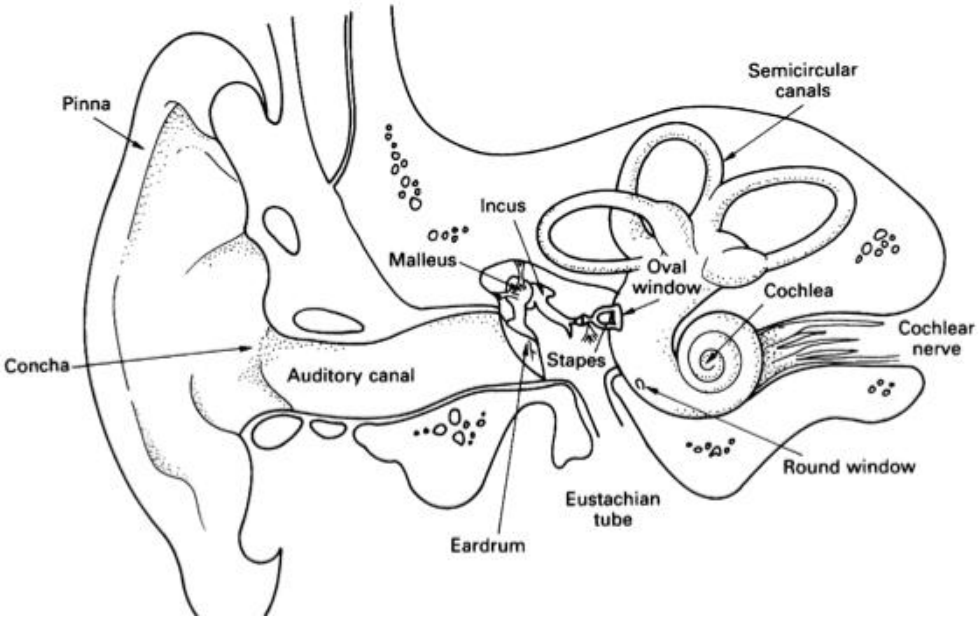


Figura 1: Fisiología del Oído humano

[4] Ing. Francisco Ruffa. Capitulo Segundo, Diseño de sistemas de sonido 2. Diapositiva 3.

2.1.1.1 Oído Externo

Inicia con el pabellón auditivo, una estructura cartilaginosa que capta las vibraciones y ayuda a la localización de la dirección de la fuente generadora de sonido. Seguido se encuentra el canal auditivo externo, que se extiende desde el pabellón hasta la membrana del tímpano con un diámetro promedio de 5mm a 7mm y una longitud promedio de 27mm, se comporta como un tubo cerrado, transformando las ondas sonoras esféricas en planas.

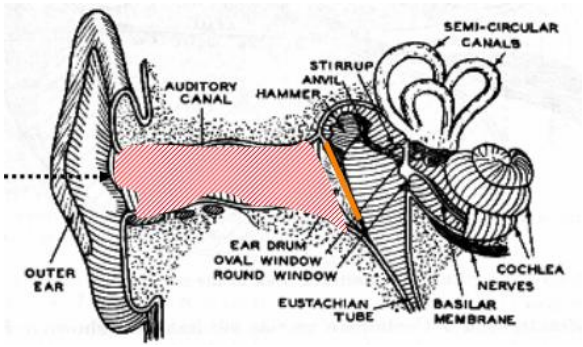


Figura 2: Ubicación del oído Externo

2.1.1.2 Oído Medio

Lo primero que se encuentra en esta zona es la caja timpánica, una cavidad llena de aire donde se encuentra la cadena de huesecillos (Martillo Yunque y Estribo). El tímpano ejerce un efecto de palanca trasladando el movimiento de las partículas de aire hacia la cadena de pequeños huesos convirtiéndolo en un movimiento mecánico. El área del tímpano es de unos 60mm² y la superficie de la platina del estribo es de unos 3mm², resultando en una energía 20 veces mas intensa que la energía incidente.

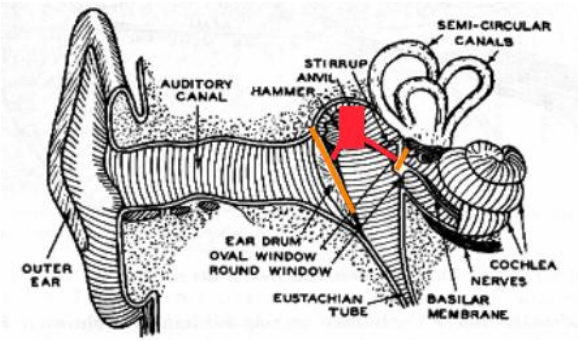


Figura 3: Ubicación del oído medio

El oído medio también funciona como mecanismo de protección a sonidos de muy alto nivel. Cuando inciden sonidos con una presión sonora mayor a 90dB, los músculos y el estribo se contraen, disminuyendo la cantidad de energía entregada al oído interno, este movimiento recibe el nombre de Reflejo Timpánico.

Para que el oído medio pueda transmitir adecuadamente la energía, es necesario que la presión al interior de la caja y al exterior sean iguales; la Trompa de Eustaquio es la encargada de esto.

2.1.1.3 Oído Interno

Consta de una serie de cavidades óseas que contienen en su interior el Caracol Membranoso ó Cóclea. En la cóclea se convierten las señales mecánicas en señales nerviosas, se discriminan los sonidos por frecuencia y se codifican los estímulos en el tiempo.

La función coclear está dividida en 3 periodos. El primero es la mecánica coclear originada por los movimientos de los líquidos y las membranas. El segundo periodo es de micromecánica coclear en el que se desplaza el órgano de Corti. En el tercer periodo se produce la transducción de la energía.

Imagen 3: Fisiología del oído interno

2.1.2 Psicoacústica

La psicoacústica trata la relación que existe entre el estimulo, y la sensación generada a partir de este. La forma en que la sensación es interpretada, percibida o alterada en los procesos cerebrales.

2.1.2.1 Sonoridad (Volumen)

Teóricamente el oído humano es sensible a sonidos arriba de 20μPa. Este valor esta expresado solo para un tono de 1KHz ya que la sensibilidad es mayor en frecuencias medias (especialmente el rango del habla) que en frecuencias bajas y altas. La relación entre sonoridad y frecuencia se encuentra en las curvas de igual sonoridad, realizadas por Fletcher y Munson.

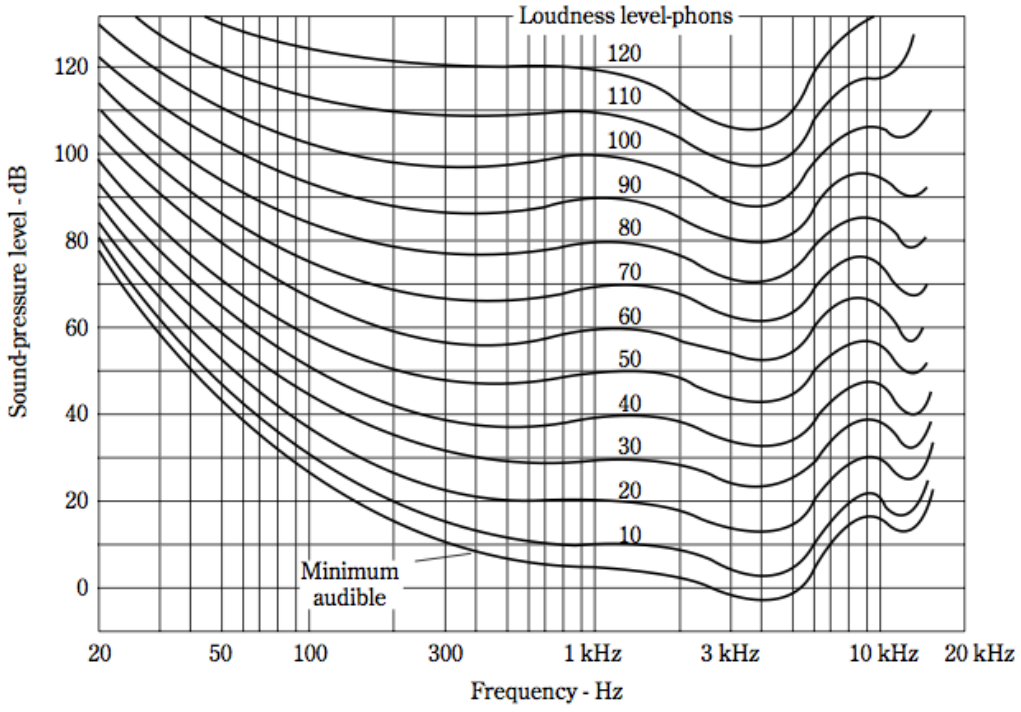


Figura 4: Curvas de igual sonoridad

2.1.2.2 Altura Tonal (Pitch)

La altura tonal es un término subjetivo que depende directamente de la frecuencia, pero estos no guardan una relación lineal. Por ejemplo un tono de 1KHz reproducido con dos diferentes amplitudes, seguirá siendo el mismo tono de 1KHz, pero la altura tonal percibida cambiara. Los experimentos realizados para la altura tonal en función de la frecuencia, han culminado en graficas muy similares al movimiento de la membrana basilar en función de la misma, esto ha llevado a pensar hasta el día de hoy que la altura tonal depende del movimiento de la membrana basilar. Así, para bajas frecuencias al aumentar la amplitud, disminuye la altura tonal percibida y para altas frecuencias al aumentar la amplitud, aumenta la altura tonal percibida.

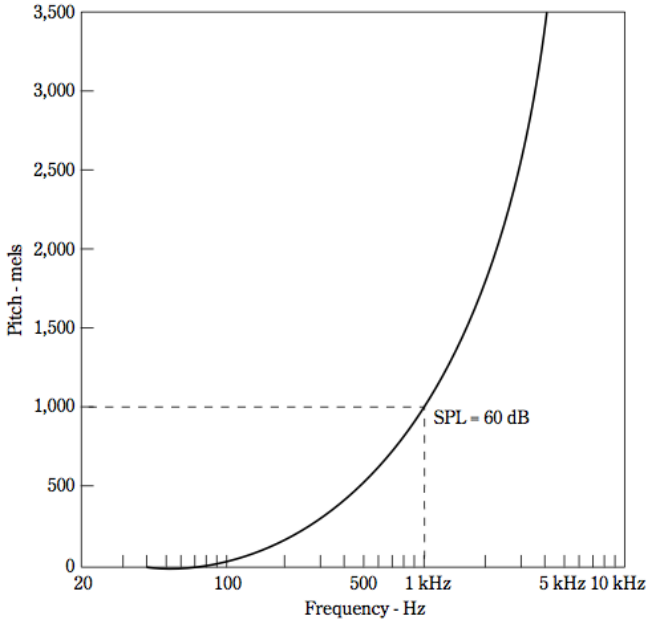


Figura 5: Grafica de la relación entre Pitch y Frecuencia a 60dB

2.1.2.3 Timbre

Este término es utilizado específicamente en la percepción de sonidos complejos, es decir, con un contenido espectral de diferentes frecuencias con diferente amplitud.

Es por esta razón que podemos identificar el sonido de un instrumento musical y reconocerlo al instante. Por ejemplo una guitarra tiene un timbre diferente a una flauta y podemos diferenciarlos al escucharlos. Así, el timbre es el término subjetivo utilizado para calificar la percepción del espectro de una señal, teniendo en cuenta que lo que escuchemos va a depender del nivel y puede ser diferente del espectro medido.

2.1.2.4 Localización de la fuente sonora

La localización sonora es el resultado del impresionante proceso de codificación del oído externo, principalmente los efectos generados por la *pinna*. Con las reflexiones y difracciones producidas en la superficie de la pinna y las modificaciones que ocurren antes de llegar al canal auditivo, el sonido generado por la fuente pasa al cerebro para que la información sea interpretada.

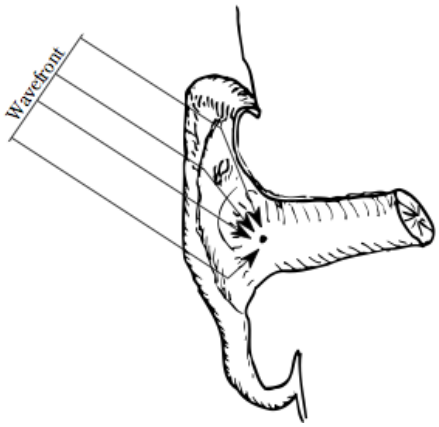


Figura 6: Frente de onda incidiendo sobre la Pinna

Al momento de localizar una fuente, las primeras preguntas que surgen son ¿De qué dirección viene el sonido? Y ¿Qué tan lejos esta?

Para determinar la dirección de la fuente emisora, el cerebro aprovecha la fisiología, y teniendo en cuenta que poseemos 2 oídos entra a evaluar los siguientes factores:

- A que oído llega primero el sonido
- En que oído el sonido es más fuerte
- Cual es el timbre del sonido

Las diferencias interaurales son dependientes de la frecuencia. Para frecuencias altas las diferencias de amplitud van a ser notorias debido a la pequeña longitud de onda, en donde la cabeza se convierte en un obstáculo para el sonido incidente. Para frecuencias bajas, cuando la longitud de onda se hace grande en comparación al tamaño de la cabeza, predominan las diferencias de tiempo.

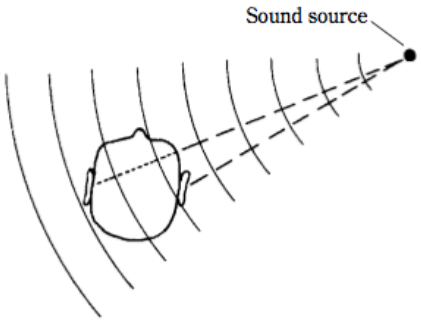
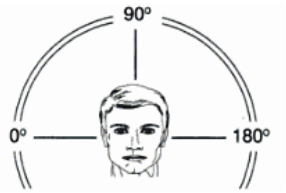


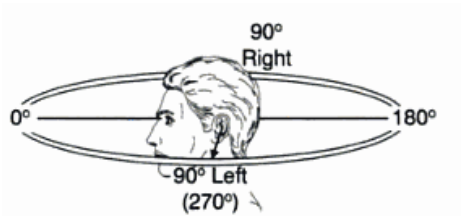
Figura 7: Proceso de localización mediante DIT y DII.[5]

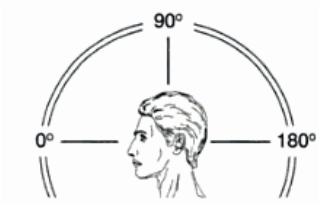
La evaluación del timbre, que depende principalmente de los efectos producidos por la *pinna*, es usada para reconocer si un sonido esta adelante o atrás, arriba o abajo. Así, según experimentos y mediciones, para un sonido proveniente de atrás el contenido en altas frecuencia va a ser menor que uno proveniente del frente. Los sonidos provenientes de arriba o abajo también tienen un cambio espectral.

La percepción de distancia de la fuente, será interpretada analizando los siguientes factores:

- Patrón de reflexiones
- Relación Señal directa a señal reverberada
- Nivel de presión sonora
- Contenido en altas frecuencias







(a) *(b)* *(c)*

Figura 8: Planos de localización sonora. a) Plano frontal. b) Plano Horizontal. c) Plano Medio.

[5] Diferencia Interaural de Tiempo y Diferencia Interaural de intensidad.

2.2 HRTF (*HEAD RELATED TRANSFER FUNCTION*)

En el ambiente natural, el ser humano percibe los sonidos en 3 dimensiones y es capaz de localizarlos en cualquier momento. El cerebro se entrena con el pasar del tiempo y reconoce una serie de parámetros que ya se han nombraron en el punto 2.1.2.

El estéreo tradicional proporciona información espacial, pero no es suficiente para recrear espacios reales en entornos virtuales, la sensación que produce es la de estar de frente a la fuente mas no en un lugar del espacio.

La clave del desarrollo del sonido tridimensional esta en el uso de funciones de transferencia dependientes de la dirección y la frecuencia[6]. Anteriormente el procesado de señal por medios analógicos era muy complicado, lo que retrasó el avance en el campo de la realidad acústica virtual, hoy en día con medios de edición y manipulación de señales digitales, este campo puede tener una gran exploración y aplicación en actividades relacionadas con el audio.

La HRTF teóricamente es la transformada de Fourier de la respuesta al impulso sobre el canal auditivo HRIR (*Head Related Impulse Response*). De la respuesta al impulso podemos extraer las diferencias interaurales de tiempo e intensidad. Realizando la transformada de Fourier, podemos obtener la respuesta en frecuencia y los cambios espectrales dependientes de la dirección de incidencia del sonido.

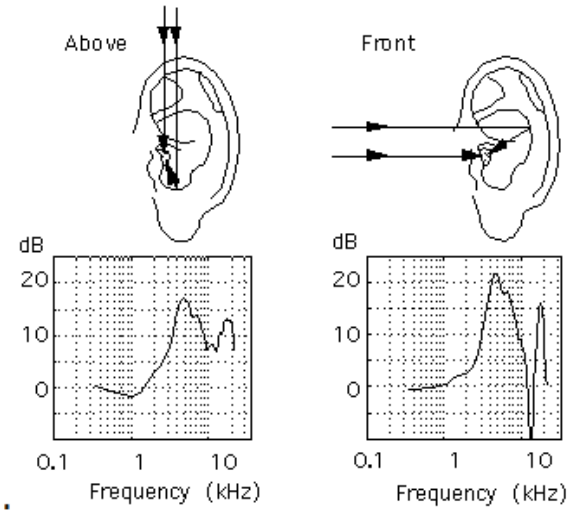


Figura 9: Respuesta en frecuencia para sonidos incidentes sobre la cabeza y de frente.

[6]Universidad de Valladolid 4.4 HRTF
<http://www.lpi.tel.uva.es/~nacho/docencia/ing_ond_1/trabajos_03_04/Csound/44.htm>

La HRTF captura muchas cualidades que el ser humano utiliza para la localización de fuentes sonoras. Una vez conocidos estos parámetros para ambos oídos, se pueden sintetizar señales binaurales a partir de una señal monoaural.

Esta función depende de 4 variables: 3 espaciales X, Y, Elevación, que forman los tres planos de localización (Frontal, Horizontal y Medio). La cuarta variable es la frecuencia. La mayoría de las HRTF son medidas en campo lejano (a 1 metro de la fuente) para que caiga con la ley del inverso al cuadrado y las únicas variables de las que dependa sean las nombradas anteriormente.

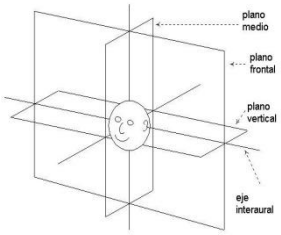


Figura 10: Planos formados por las 3 dimensiones de la HRTF.

2.2.1 Medición de la HRTF

Existen 2 formas de realizar la medición de la HRTF, la primera y más común es por medio de un maniquí que emule las características fisiológicas del ser humano, la segunda es realizar la medición directamente sobre personas.

Imagen 4: Medición de la HRTF con Dummy Head.[7]

Imagen 5: Medición de la HRTF con una persona real.[8]

[7] Imagen extraída de <http://www.lpi.tel.uva.es/~nacho/docencia/ing_ond_1/trabajos_03_04/Csound/44.htm>

Estas mediciones deben ser realizadas en un espacio libre de reflexiones, preferiblemente en una cámara anecóica.

Imagen 6: Medición de la HRTF en Cámara Anecóica.[9]

Para obtener la HRIR se generan secuencias LSS *(Long Sine Sweep)* y se captura con los transductores exactamente donde deberían ir los oídos para el caso del maniquí, y a la entrada del canal auditivo para el caso de una persona real. Para obtener una respuesta en frecuencia directa, se pueden realizar mediciones con ruido rosa también sobre el canal auditivo. Cualquier otra técnica para obtener respuestas al impulso puede ser utilizada, secuencias MLS *(Maximum Length Sequences),* impactos, pulsos sonoros, ruido interrumpido, etc.

2.2.2 Extracción de Características significativas

- Diferencias interaurales de tiempo

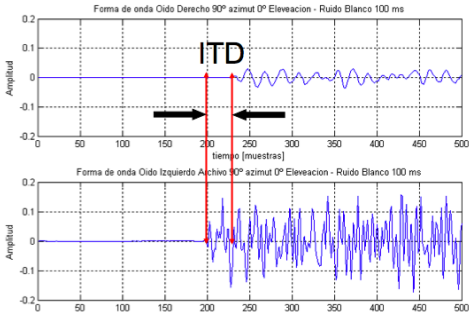


Imagen 7: Diferencia interaural de tiempo extraída de la HRIR.[10]

---

[8] Imagen extraída de < http://www.ausim3d.com/products/images/HeadZap_800.jpg>
[9] Imagen extraída de "Medición de HRTF y extracción de características". Universidad Pérez Rosales.

Analizando las diferencias de tiempo entre las respuestas al impulso de los dos oídos, se extrae el ITD (*Interaural Time Delay*), para diferentes posiciones de fuente.

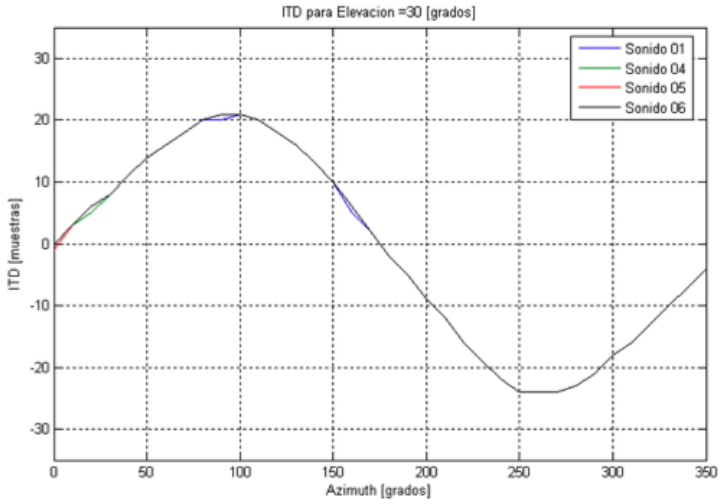


Imagen 8: Gráfica de las DIT para 30º de elevación y variaciones en el plano horizontal.[11]

- Diferencias Interaurales de Intensidad

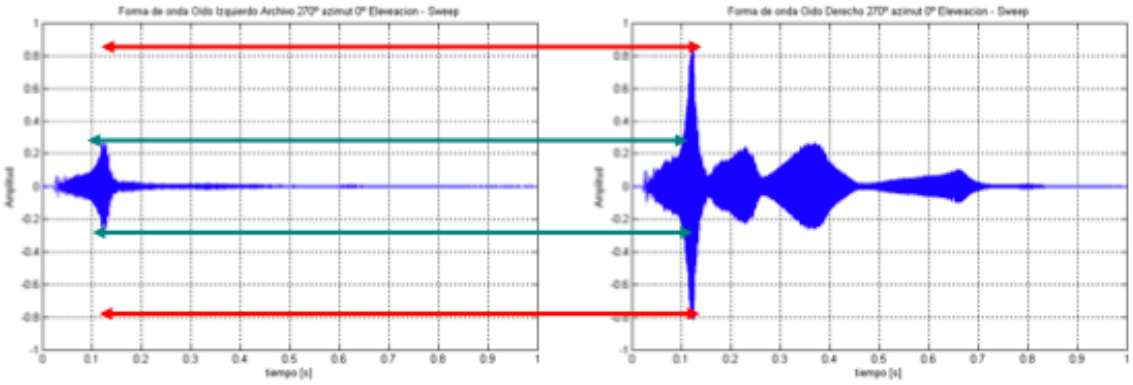


Imagen 9: Diferencia interaural de intensidad extraída de la HRIR[12]

[10] Imagen extraída de "Medición de HRTF y extracción de características". Universidad Pérez Rosales.
[11] Imagen extraída de "Medición de HRTF y extracción de características". Universidad Pérez Rosales.
[12] Imagen extraída de "Medición de HRTF y extracción de características". Universidad Pérez Rosales.

El ILD (*Interaural Level Diference*) junto con el ITD son denominados pistas binaurales, de donde se extrae la información para la localización en el plano Azimuth.

- Diferencias espectrales

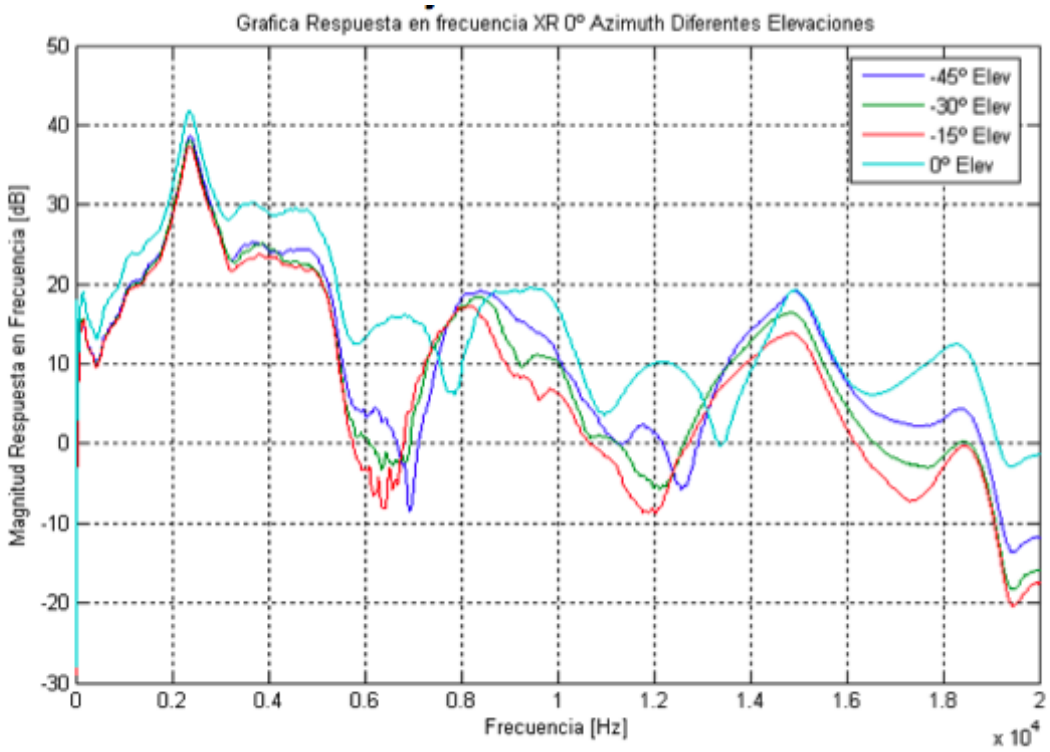


Imagen 10: Respuesta en frecuencia de la HRTF para diferentes elevaciones.[13]

La respuesta en frecuencia varía según la posición de la fuente en el espacio tridimensional. El pabellón auditivo actúa como una cavidad de resonancia, donde amplifica algunas frecuencias, y su geometría genera ciertos efectos de interferencia que atenúan otras frecuencias.

[13] Imagen extraída de "Medición de HRTF y extracción de características". Universidad Pérez Rosales.

Cuando el ITD y el ILD se hacen iguales la función de transferencia se convierte en una herramienta fundamental para la ubicación de la fuente, la localización se lleva a cabo con la convolución de la señal y los efectos de la fisiología humana, incluyendo torso, cabeza, cuello y oídos. Esta información es indispensable para determinar la elevación y la ubicación trasera o delantera.

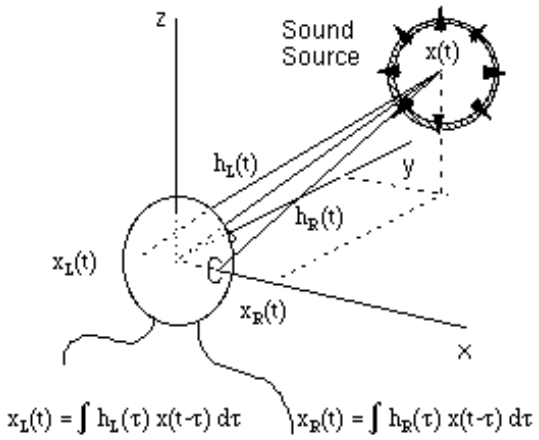


Figura 11: Expresiones matemáticas para la HRTF.

2.2.3 Sistemas Basados en HRTF

Los sistemas basados en HRTF son capaces de simular todas las características de localización que se han nombrado a lo largo de este marco teórico, dando la sensación de tridimensionalidad al oyente. Aunque los resultados pueden variar con cada persona, estos sistemas se han convertido hoy en día en un estándar para la aplicación de la Realidad Acústica Virtual.

El primer sistema basado en HRTF fue creado por la NASA y manufacturado por Cristal River Engineering llevando el nombre de *"convolvotrón"*. Su funcionamiento está basado en realizar la convolución de la respuesta al impulso para cada oído con la señal a reproducir, el resultado es llevado después a una reproducción por audífonos. El problema de este sistema es que su funcionamiento depende de la función de transferencia propia del oyente, ya que si difiere considerablemente de la usada en la convolución, los resultados de tridimensionalidad no van a ser los esperados.

Se han creado modelos computacionales de las características de la HRTF para la simulación de espacios, muy utilizados en las auralizaciones. En el campo de la acústica esto se ha implementado en el diseño de recintos, permitiendo simular el estar allí sin haber construido el recinto.

Empresas como Yamaha, AKG y *Wave Arts*, han lanzado al mercado productos que, implementando la HRTF, son capaces de generar fuentes virtuales y producir resultados de tridimensionalidad.

2.3 SISTEMAS COMERCIALES PARA LA REPRODUCCIÓN DE AUDIO 3D

Dependiendo de las aplicaciones, se dividirán las formas y medios para recrear el campo tridimensional en un entorno virtual en 3 grupos, siendo estos los más usados en el mercado.

2.3.1 Auriculares

Las reproducciones binaurales por auriculares son las más simples de adaptar en la realidad, pues de esta forma se tiene una señal individual para cada oído. Su aplicación en la realidad virtual es muy grande pero siempre va a estar limitada al uso de estos elementos.

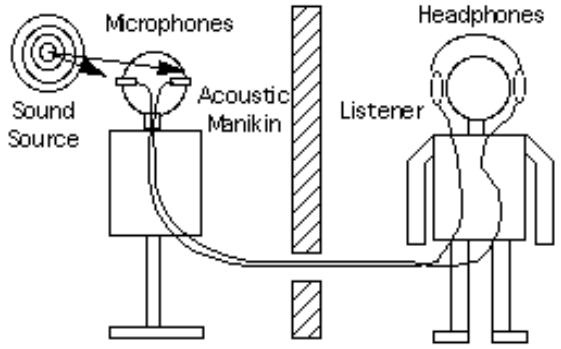


Figura 12: Esquema de reproducción binaural por audífonos.

Los auriculares poseen ciertos problemas a la hora de la reproducción del campo, principalmente la respuesta en frecuencia que poseen. Esta respuesta puede generar alteraciones en frecuencias altas relacionadas a los efectos de la *Pinna* sobre la señal capturada. En caso de encontrar este problema, deberá ser compensado para el funcionamiento óptimo del audio 3D.

2.3.2 Altavoces Estéreo

Los altavoces resuelven la mayoría de los problemas generados por los audífonos, pero generan una interferencia sobre cada oído opuesto que hay que eliminar para obtener la señal individual para cada oído.

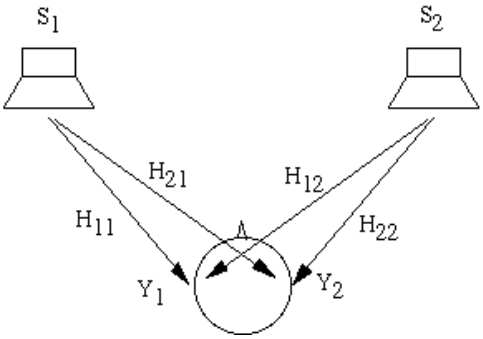


Figura 13: Señales de Interferencia producidas por los altavoces (Crosstalk)

Este sistema de reproducción para audio 3D esta limitado a configuraciones especificas de los altavoces en donde se pueda realizar la cancelación de la interferencia. Las más comunes son a 30º a 20º y a 10º.

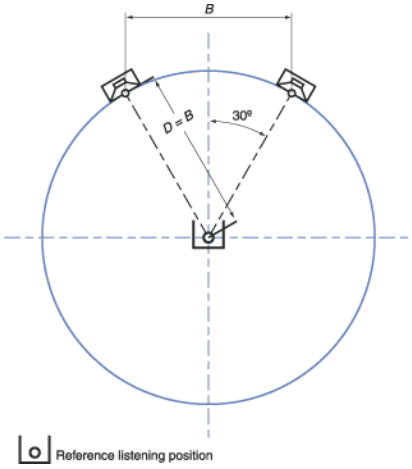


Figura 14: Configuración de altavoces a 30º

2.3.3 Sistemas Multicanal (*Surround 5.1*)

El sonido envolvente 5.1 es considerado como una reproducción tridimensional del audio, sin embargo harían falta altavoces en la parte superior e inferior para poder simular la elevación.

Es posible también, realizar edición y procesos en el audio para recrear estas sensaciones, tratando 2 interferencias diferentes: La interferencia entre L-R y la interferencia entre LS-RS.

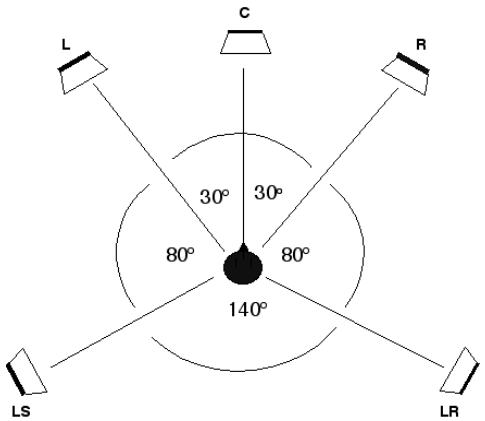


Figura 15: Configuración de un sistema 5.1

2.4 CANCELACIÓN DE LA INTERFERENCIA EN LOS ALTAVOCES (*CROSSTALK CANCELLATION*)

La reproducción del audio 3D por medio de altavoces requiere la cancelación de la interferencia generada, de tal manera que el oído izquierdo reciba la señal L y el odio derecho reciba la señal R de forma individual. Esto se puede realizar mediante procesamiento de señal.

El primer filtro fue formulado por Atal y Schroeder como un filtro de substracción. Se realiza la compensación de las señales de interferencia radiando una señal invertida y con un tiempo de llegada idéntico al de la señal de la contraparte. Esta compensación también genera interferencia, por lo que es necesario realizar varios filtros. 5 Filtros en cada canal son suficientes para obtener una cancelación con atenuación de 20dB a 25dB.

Otra forma de realizar la cancelación de la interferencia, es con la inversión y reproducción por bandas de frecuencia, teniendo en cuenta que los retardos para compensar el tiempo de arribo dependerán de la longitud de onda y de las diferencias interaurales de tiempo.

Para la cancelación de interferencia en altavoces ubicados atrás del oyente estos retardos cambiaran y deberán ser ajustados según el audio que se este trabajando.

La generación del audio para ser reproducido por altavoces y crear la sensación de tridimensionalidad es lo que se conoce hoy en día como audio transaural.

La base matemática completa de todo el trabajo transaural puede encontrarse en los aportes del trabajo realizado por Schroeder y Atal.

Una de las ventajas que tiene el audio transaural, sobre otras técnicas estereofónicas, es la reproducción de la imagen sonora mucho más fiel y amplia. En una reproducción estereofónica normal, la propagación de la señal se da de la siguiente forma.

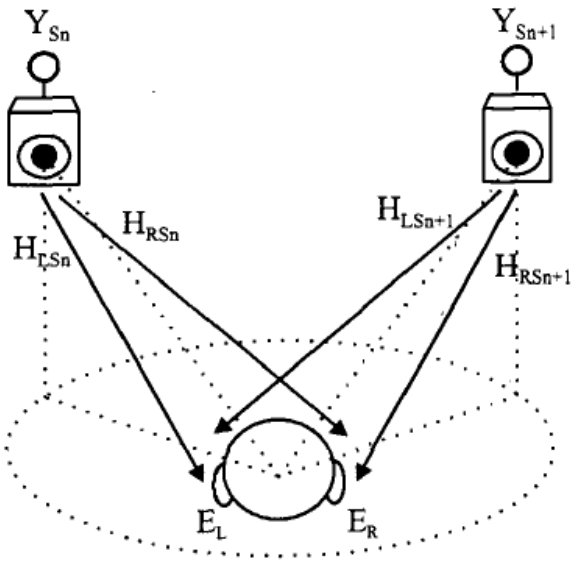


Figura 16: Propagación de la señal en una reproducción estereofonica.

El concepto teórico se desarrolla de la siguiente manera.

$$\begin{bmatrix} E_L \\ E_R \end{bmatrix} = \begin{bmatrix} H_{S_n L}(\theta_n, \phi_n) & H_{S_{n+1} L}(-\theta_{n+1}, \phi_{n+1}) \\ H_{S_n R}(-\theta_n, \phi_n) & H_{S_{n+1} R}(\theta_{n+1}, \phi_{n+1}) \end{bmatrix} \cdot \begin{bmatrix} Y_{S_n} \\ Y_{S_{n+1}} \end{bmatrix}$$

EL y ER son las señales binaurales sobre los oídos.

HSIX es la función de transferencia del altavoz Si al oído X, esto significa que HSNL es la función de transferencia del altavoz n al oído izquierdo.

YSi es la señal que se transmite al altavoz.

θn y ϕn son los ángulos de posición de los altavoces con respecto al oyente.

En este caso especifico de reproducción n es el altavoz izquierdo y n+1 es el altavoz derecho, ubicados en θ= 30º y ϕ= 0º.

La cancelación de la interferencia estará dada entonces por

$$\begin{bmatrix} Y_{S_n} \\ Y_{S_{n+1}} \end{bmatrix} = \begin{bmatrix} H_{S_n L}(\theta_n, \phi_n) & H_{S_{n+1} L}(-\theta_{n+1}, \phi_{n+1}) \\ H_{S_n R}(-\theta_n, \phi_n) & H_{S_{n+1} R}(\theta_{n+1}, \phi_{n+1}) \end{bmatrix}^{-1} \cdot \begin{bmatrix} E_L \\ E_R \end{bmatrix}$$

Manipulando la ecuación se obtiene una general

$$\begin{bmatrix} Y_{S_n} \\ Y_{S_{n+1}} \end{bmatrix} = \frac{1}{\Delta} \cdot \begin{bmatrix} H_{S_{n+1} R}(\theta_{n+1}, \phi_{n+1}) & -H_{S_{n+1} L}(-\theta_{n+1}, \phi_{n+1}) \\ -H_{S_n R}(-\theta_n, \phi_n) & H_{S_n L}(\theta_n, \phi_n) \end{bmatrix} \cdot \begin{bmatrix} E_L \\ E_R \end{bmatrix}$$

En donde

$$\Delta = H_{S_n L}(\theta_n, \phi_n) \cdot H_{S_{n+1} R}(\theta_{n+1}, \phi_{n+1}) - H_{S_n R}(-\theta_n, \phi_n) \cdot H_{S_{n+1} L}(-\theta_{n+1}, \phi_{n+1})$$

En el desarrollo práctico del diseño del audio transaural surgen las limitaciones grandes del proceso, la dependencia de la posición y orientación del oyente, y la dependencia del recinto de reproducción.

La posición de los altavoces determina el camino que seguirá la señal al oyente, teniendo en cuenta esto, se diseñan los filtros para actuar en el *sweet spot*. De esta forma, movimientos significativos del oyente afectaran la llegada de la señal a los oídos y los filtros de cancelación no funcionaran apropiadamente.

3. METODOLOGÍA

3.1 ENFOQUE DE LA INVESTIGACIÓN

Este proyecto tiene un enfoque EMPÍRICO-ANALÍTICO, ya que pretende realizar un producto de forma experimental recreando un campo tridimensional en un entorno virtual, teniendo en cuenta la base teórica para poder ser comprobado y relacionado con las características de la función de transferencia de la cabeza, específicamente con las alteraciones del sonido antes de incidir y al incidir sobre el canal auditivo.

3.2 LÍNEA DE INVESTIGACIÓN DE USB / SUB-LÍNEA DE FACULTAD / CAMPO TEMÁTICO DEL PROGRAMA

El trabajo a realizar pertenece a la línea de investigación Tecnologías Actuales y Sociedad, ya que se están implementando nuevos métodos de recreación de la realidad en entornos virtuales para ser usados posteriormente en producciones que mejoren la sensación y la interacción del consumidor.

La Sub-Línea de la facultad es Instrumentación y control de procesos ya que a partir de instrumentos de captura enfocados sobre la función de transferencia de la cabeza, se realizara la edición y modificación del audio obtenido para su apropiado funcionamiento en la recreación del campo tridimensional.

El campo temático del programa es el de Grabación y producción, ya que a partir de las grabaciones obtenidas se realizara la producción del banco de ambientes, teniendo en cuenta su aplicabilidad al diseño sonoro de las producciones locales.

3.3 TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN

3.3.1 Análisis de las principales locaciones utilizadas y de mayor interés en las producciones locales.

Este estudio se realizará teniendo en cuenta las muestras de audiovisuales locales, los concursos y convocatorias recientes para producciones; obteniendo de estas las sinopsis y extrayendo los ambientes requeridos en el diseño sonoro.

Esta información puede ser obtenida en el área de cinematografía del Ministerio de Cultura de la Republica de Colombia[14]; y de las demás entidades encargadas de fomentar las producciones locales[15].

3.3.2 Captura de los ambientes

Para la captura de los ambientes se utilizaran los micrófonos *SP-TFB-2 in ear* de *Sound Professionals*.

Imagen 11: Micrófonos In-Ear SP-TBF-2 de Sound Professionals[16]

Se utilizará la grabadora digital *Microtrack II* de *m-audio*, que reúne las características requeridas para este trabajo, principalmente su sencilla portabilidad.

Imagen 12: Grabadora Digital Microtrack II de m-audio[17]

[14] http://www.mincultura.gov.co/?idcategoria=1154
[15] http://www.proimagenescolombia.com/secciones/fdc/intro.php
[16] Imagen extraída de < http://www.soundprofessionals.com/cgi-bin/gold/item/SP-TFB-2>
[17] Imagen extraída de < http://www.m-audio.com/products/en_us/MicroTrackII.html>

3.3.3 Extracción de características de la Función de transferencia relacionada a la cabeza.

Para la extracción de efectos de *pinna* y diferencias interaurales de tiempo se usará la base de datos que tiene disponible *Mit Media Lab* en Internet de su trabajo *HRTF Measurements of a Kemar Dummy-Head Microphone*. Mediciones realizadas de manera profesional en una cámara anecóica y con los equipos requeridos.

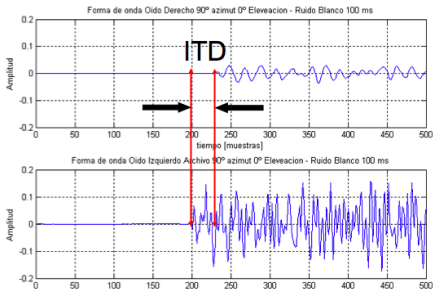


Imagen 13: Extracción de DIT

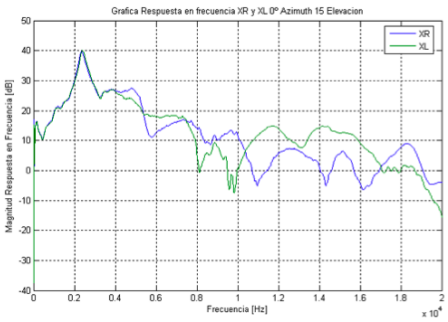


Imagen 14: Respuesta en frecuencia para los 2 oídos con 15º de elevación.

3.3.4 Cancelación de la interferencia en altavoces (*Crosstalk Cancellation*)

La cancelación de la interferencia y la generación del audio 3D para una reproducción por altavoces se hará teniendo en cuenta algunos aportes del trabajo de Steve Marshall *Bilocation* y la extracción de características del punto anterior para ser relacionada con *papers* de la AES sobre *Crosstalk Cancellation*.

3.4 POBLACIÓN Y MUESTRA

Los equipos utilizados para la captura de ambientes deben cumplir con parámetros de calidad en el muestreo de la onda sonora y su conversión a audio.

Las locaciones y espacios seleccionados para su recreación virtual deberán describir comportamientos culturales propios de la ciudad y/o deberán tener un uso constante en las producciones locales.

3.5 HIPÓTESIS

Al implementar las características de la HRTF en la captura y edición de ambientes sonoros, se puede obtener un espacio tridimensional recreado, una realidad virtual aplicable al diseño sonoro. Generando los 3 formatos de reproducción (Audífonos, Altavoces 2.0, Surround 5.1) se hará más amplia la aplicabilidad del banco de sonidos para producciones de radio, videojuegos, largometrajes, cine, etc.

3.6 VARIABLES

3.6.1 Variables Independientes

- Las condiciones climáticas a la hora de realizar la captura.
- Los cambios en el ambiente a capturar en las diferentes horas del día o situaciones.
- La HRTF individual empleada en la captura.

3.6.2 Variables Dependientes

- Los micrófonos usados en la captura.
- La grabadora digital empleada para la conversión análogo/digital.
- El software utilizado en la edición del audio.
- El sistema de monitoreo empleado para la reproducción del audio 3D

4. DESARROLLO INGENIERIL

4.1 BASE DE DATOS

4.1.1 Listado de Locaciones

En el momento del planteamiento del proyecto, se diseñó una lista con base en las producciones audiovisuales del medio local reconocidas en los dos últimos años (2009 – 2010). Para esto, la búsqueda de lugares inició en los registros pertenecientes a festivales, encuentros y demás eventos desarrollados en la temporada. Se partió buscando en las entidades distritales[18] que fomentan el desarrollo de los medios audiovisuales.

NOMBRE	DIRECTOR	AÑO	DURACION (MINS)	LOCACIONES
ALGUIEN MATÓ ALGO	Jorge Navas	1998	27	No encontrado
LA VUELTA DE HOJA	Carlos Hernández	2001	11	Carrera 7 frente teatro Jorge Eliécer
CUANDO VUELVAS DE TUS MUERTES	Carlos Mario Urrea	2000	19	Chapinero, Hospital San Juan de Dios
LÚDICA MACÁBRICA	Carlos Mogollón	2003	14	No encontrado
LOS CICLOS	Juan Manuel Acuña	2007	20	No encontrado
INSTRUCCIONES PARA ROBAR UNA MOTOCICLETA	Felipe Solarte	1998	15	No encontrado
LA CERCA	Rubén Mendoza	2004	18	No encontrado
MARIPOSA NOCTURNA	Diana Sánchez	2008	8	No encontrado

Tabla 1: 4º Encuentro Distrital Audiovisual (Noviembre de 2009).

En este proceso surgieron dos factores poco convenientes para el desarrollo del proyecto que se pueden evidenciar en la tabla 1. Primero, la falta de información sobre las producciones y segundo, la generalización de las locaciones en donde se hacia referencia a las calles y avenidas en donde se encuentran. Esto dificultaba la selección de lugares debido a la poca exactitud en la lista.

[18] Secretaría de cultura, recreación y deporte <http://www.culturarecreacionydeporte.gov.co/>
Instituto distrital de recreación y deporte <http://www.idrd.gov.co/>
Ministerio de Cultura <http://www.mincultura.gov.co/>

La solución se encontró en el Fondo para el desarrollo cinematográfico en Colombia[19], que financia cada año proyectos cinematográficos desde la escritura de guiones, producción y postproducción hasta la distribución y exhibición.

PROIMAGENES en movimiento es el encargado de la comisión fílmica colombiana, y su objetivo es desarrollar una relación comercial con el exterior mediante la promoción del territorio colombiano como destino para la filmación de películas.[20]

Contactando con Proimágenes vía email, se preguntó por un listado de locaciones para producciones audiovisuales en Bogotá. La respuesta, dirigía al portal Web que maneja la comisión fílmica colombiana para lograr su objetivo[21]. Allí se encuentra una completa galería de locaciones en Colombia divididas por regiones y organizadas por códigos. Para el interés de este proyecto se encontró una división llamada Bogotá Región, de allí se partió para el primer listado de locaciones que cumplieran con las condiciones iniciales del proyecto, descripción de comportamientos culturales y uso en producciones audiovisuales.

NOMBRE	DIRECCION	CÓDIGO
Biblioteca Pública el Tintal Manuel Zapata Olivella	Avenida Ciudad de Cali # 6C-09	1
Centro Comercial Gran Estación	Avenida Calle 26 # 62-47	2
Complejo Acuático Simón Bolívar	Calle 63 # 42-00	3
Centro de Entrenamiento Familiar Compensar	Avenida 68 # 49A-47	5
Biblioteca Publica Virgilio Barco	Avenida carrera 60 # 57-60	13
Hospital de Fontibon	Carrera 104 # 20C-21	27
Billar Londres	Carrera 7 # 21-91	30
Estadio Nemesio Camacho el Campin	Avenida Norte Quito Sur con calle 57	31
Puente de la 68 con 63	Calle 63 con Av. 68	38
Parroquia San Gregorio Magno	Carrera 32 No. 19 15	39
Terminal de transporte	Diagonal 23 No 69 - 60	40

Tabla 2: Listado Inicial de Locaciones 1ª sección.

[19] http://www.proimagenescolombia.com/
[20] http://www.proimagenescolombia.com/secciones/comision_filmica/intro.htm
[21] http://www.locationcolombia.com/

NOMBRE	DIRECCIÓN	CÓDIGO
Iglesia de Nuestra Señora del Carmen (Santa Teresita)	Carrera 18A No. 43A-59	41
Monserrate	Paseo Bolívar No. 2-48 Este	42
Salitre Mágico	Cll 63 No. 48-96	46
Mundo Aventura	Transversal 71D 6-30 Sur	47
Plaza El chorro de Quevedo	calle 13 con carrera 2ª	52
Puente Cll 116 con carrera 7	Calle116 con Carrera 7	63
Parque Simón Bolívar	Calle 63 y 53 entre carreras 48 y 68	77
Edificios empresariales avenida El Dorado	Avenida El Dorado N68C – 61	78
Plazoleta de la Rebeca	Carrera 13 calle 25, San Diego	84
Plaza de Banderas	Avenida de Las Américas Cr76	85
Parque de Los Periodistas	entre Eje Ambiental y la carrera 4	86
Plaza Parque de Lourdes	Carrera 13 Nº 63-2	87
Plazoleta del Rosario	Calle 16 # 6 - 25	88
Jardín Botánico José Celestino Mutis	Calle 57 No. 61-13	89
Parque de los novios - El Lago	Calle 63 Carrera 36A	90
Mirador de Hilos	Avenida circunvalar con 34	93
Corferias	Entrada Principal Carrera 37 No 24 – 67	129
Centro Cultural Gabriel García Márquez	Calle de la enseñanza (11) 5-60	149
Museo Botero, Banco de la República	Calle 11 No. 4-21	150
Plaza de Bolívar	Calle 10 a 11 Carreras 7 y 8	164
Estación de la Sabana	Calle 13 No. 18 – 24.	174
Biblioteca Nacional de Colombia	Calle 24 N° 5- 60	175
Museo de la Independencia, Casa del Florero	Calle 11 No 6-94.	178
Teatro de Cristóbal Colón	Calle 10 No 5-32	181
Intercambio Vial Calle 92 con autopista	Calle 92 con autopista	184
Museo Iglesia de Santa Clara	Carrera 8 No 8-91	187
Maloka	Cra 68 D No. 24 A 51	196
Planetario	Carrera 7 entre 26 y 30	197
Parque de la Independencia	Carrera 5 y la Carrera 7 Avenida 26	198
Las torres del parque	Cra. 5 No. 26 A - 47	201
Torre Colpatria	Calle 26 con Carrera 7	218
Iglesia de San Agustín	Carrera 15 # 10 - 43	262
Parque de Diversiones Cafam	Autopista Norte Calle 215 Kilometro 14	280
Parque Jaime Duque	Tocancipa Km 34 Autopista Norte	281

Tabla 3: Listado Inicial de Locaciones 2ª sección.

4.1.2 Selección de Locaciones

Para la selección de los 15 espacios y locaciones de grabación, se tuvieron en cuenta los siguientes factores.

- Ambiente sonoro diferente entre cada lugar
- Facilidad de acceso
- Seguridad
- Ajuste al tiempo y día destinado a grabación.

No	NOMBRE	DIRECCION	CODIGO
1	Biblioteca Pública el Tintal Manuel Zapata Olivella	Avenida Ciudad de Cali # 6C-09	1
2	Centro Comercial Gran Estación	Avenida calle 26 # 62-47	2
3	Complejo Acuático Simón Bolívar	Calle 63 # 42-00	3
4	Billar Londres	Carrera 7 # 21-91	30
5	Estadio Nemesio Camacho el Campin	Avenida Norte Quito Sur con calle 57	31
6	Puente de la 68 con 63	Calle 63 con Av. 68	38
7	Terminal de transporte	Diagonal 23 No 69 - 60	40
8	Monserrate	Paseo Bolívar No. 2-48 Este	42
9	Salitre Mágico	Calle 63 No. 48-96	46
10	Parque Simón Bolívar	Calle 63 y 53 entre carreras 48 y 68	77
11	Plaza de Banderas	Avenida de Las Américas Cr76	85
12	Plaza Parque de Lourdes	Carrera 13 Nº 63-2	87
13	Parque de los novios - El Lago	Calle 63 Carrera 36A	90
14	Plaza de Bolívar	Calle 10 a 11 Carreras 7 y 8	164
15	Museo Iglesia de Santa Clara	Carrera 8 No 8-91	187

Tabla 4: Listado de Locaciones escogidas para grabación.

4.2 CAPTURA DE AMBIENTES

4.2.1 Selección de Equipos

El desarrollo del proyecto y su finalidad, requerían de sistemas de grabación de fácil transporte y configuración. Teniendo en cuenta esto, la primera conclusión a la que se llegó, es que se debía implementar un sistema de captura para una HRTF individual, es decir, que no se haría uso de un *Dummy Head*, se usarían micrófonos *in-ear*.

En primer lugar, se contempló la posibilidad de su diseño y construcción pero esto se salía un poco de la línea general y objetivo del proyecto. La mayoría de los micrófonos in-ear en el mercado para su momento, que eran pocos, debían ser insertados en el canal auditivo; el único par de micrófonos que no obstruían el canal eran los SP-TFB-2[22] (Imagen 11).

Marca	Sound Professionals
Respuesta en Frecuencia	20 – 20000 Hz
Relación Señal a Ruido	Sensibilidad Estándar 60dB/ Alta Sensibilidad 62dB
Sensibilidad en Circuito Abierto	Sensibilidad Estándar -42dB / Alta Sensibilidad -35dB
Máximo Nivel de Entrada	105dB / 120dB
Rango Dinámico	81dB / 96dB
Patrón Polar	Omnidireccional
Requerimientos de fuente de poder	2 – 12 vdc

Tabla 5: Especificaciones de los micrófonos SP-TFB-2.

[22] http://www.soundprofessionals.com/cgi-bin/gold/category.cgi?item=SP-TFB-2&type=store&template=specifications

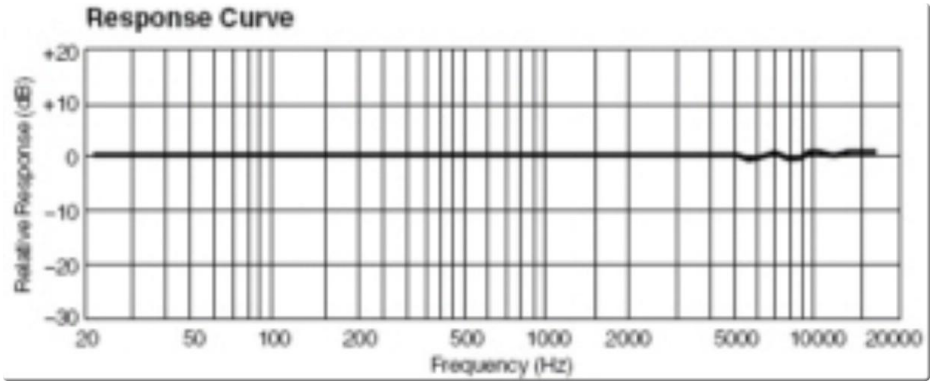


Imagen 15: Respuesta en Frecuencia de los Micrófonos SP-TFB-2.[23]

El sistema de almacenamiento, debía tener una entrada de 1/8" para micrófono estéreo y fuente de poder para los micrófonos escogidos. La mejor opción fue la Microtrack II[24] de m-audio (Imagen 12).

Impedancia de Entrada	2.2 K Ohms
Crosstalk entre canales	<-95dB
Relación Señal a Ruido	-101 dBA
Rango Dinámico	101 dBA
THD+N	0.003% (-90 dB)
Respuesta en Frecuencia	20 Hz – 20KHz (+/- 0.5 dB)
Ganancia del Preamplificador	>52 dB
Fuente de poder para micrófono electret/condensador estéreo	

Tabla 6: Especificaciones de la Microtrack II para la entrada estéreo de 1/8".

[23] Imagen extraída de < http://www.soundprofessionals.com/cgi-bin/gold/item/SP-TFB-2 >
[24] http://www.m-audio.com/products/en_us/MicroTrackII.html

4.2.2 Configuración de Captura

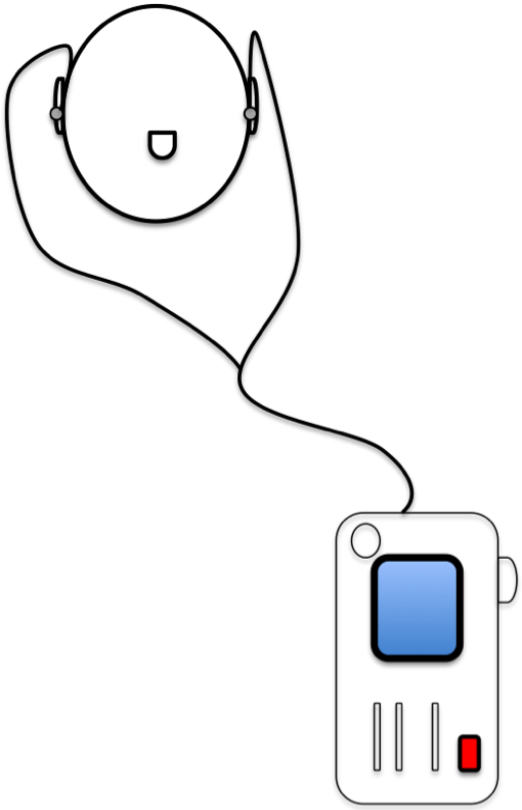


Figura 17: Diagrama de Conexión para la captura.

Otra de las ventajas que ofrece esta grabadora digital portátil, es su conversión análogo – digital, con una frecuencia de muestreo de hasta 96KHz y una resolución en bits de 16 y 24.

Almacena los datos en una tarjeta Compact Flash y luego estos pueden ser llevados a un computador por medio de un puerto USB.

Fuente de Entrada	1/8”
Monitorización de Entrada	Encendido
Codificación	WAV
Frecuencia de Muestreo	48 KHz
Resolución	24 Bits
Canales	Estéreo
Conexión L+R	Encendido
Limitador	Encendido

Tabla 7: Ajustes de Grabación en la Microtrack II

4.2.3 Detalles de cada captura

- Biblioteca Pública el Tintal Manuel Zapata Olivella

  Fecha de grabación: Miercoles15 de septiembre de 2010.

  Componentes generales del ambiente sonoro: Voces de adultos, voces de niños, movimiento de sillas, hojas de libros.

Imagen 16: Biblioteca pública el tintal.

- Centro Comercial Gran Estación

  Fecha de Grabación: Sábado 24 de Julio de 2010

  Componentes generales del ambiente sonoro: Voces de adultos, voces de niños, detector antirrobo, tren, sala de juegos, pasos.

Imagen 17: Centro Comercial Gran Estación.

- Billar Londres

  Fecha de Grabación: Jueves 5 de Agosto de 2010

  Componentes generales del ambiente sonoro: Música, voces de adultos, choque entre bolas.

Imagen 18: Billar Londres.

- Estadio Nemesio Camacho el Campín

  Fecha de Grabación: Domingo 8 de Agosto de 2010

  Componentes generales del ambiente sonoro: Cornetas, aplausos, canto de la barra, voces de adultos, voces de niños, silbidos, percusión.

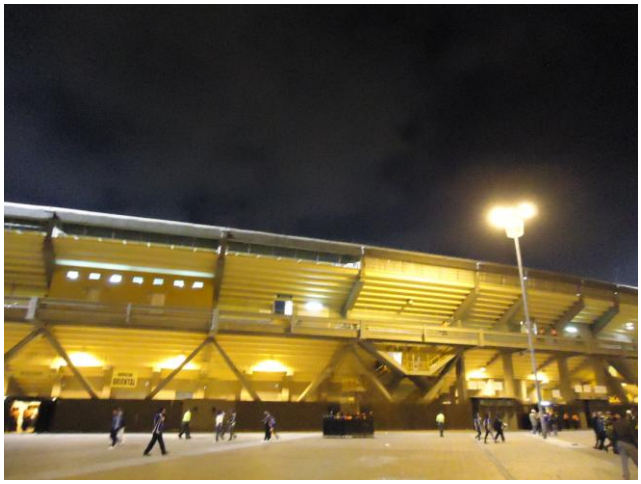


Imagen 19: Estadio Nemesio Camacho el Campín.

- Puente de la 68 con 63

  Fecha de Grabación: Miércoles 21 de Julio de 2010

  Componentes generales del ambiente sonoro: Viento, automóviles, buses, motos, paso de aviones, voces de adultos, voces de niños.

Imagen 20: Puente de la 68 con 63.

- Terminal de Transporte de Bogotá

  Fecha de Grabación: Viernes 23 de Julio de 2010

  Componentes generales del ambiente sonoro: Voces de adultos, pasos, pito, nombre de destinos según modulo, voces de niños.

Imagen 21: Terminal de Transporte de Bogotá.

- Monserrate

  Fecha de Grabación: Domingo 8 de Agosto de 2010

  Componentes generales del ambiente sonoro: Viento, voces de adultos, música, rieles en funicular, voces de niños.

Imagen 22: Monserrate.

- Salitre Mágico

  Fecha de Grabación: Sábado 31 de Julio de 2010

  Componentes generales del ambiente sonoro: Gritos, viento, música, ruido de las atracciones mecánicas, voces de adultos, voces de niños, paso de aviones.

Imagen 23: Salitre Mágico.

- Parque Simón Bolívar

  Fecha de Grabación: Domingo 26 de Septiembre de 2010.

  Componentes generales del ambiente sonoro: Viento, aves, voces de adultos, voces de niños.

Imagen 24: Parque Simon Bolívar.

- Plaza de Banderas

  Fecha de Grabación: Martes 3 de Agosto de 2010

  Componentes generales del ambiente sonoro: Automóviles, motos, buses, aves, viento.

Imagen 25: Plaza de Banderas.

- Plaza Parque de Lourdes

  Fecha de Grabación: Lunes 2 de Agosto de 2010

  Componentes generales del ambiente sonoro: Automóviles, buses, motos, palomas, voces de adultos.

Imagen 26: Plaza Parque de Lourdes.

- Parque de los Novios – El Lago

  Fecha de Grabación: Lunes 2 de Agosto de 2010

  Componentes generales del ambiente sonoro: Viento, aves, caída de agua.

Imagen 27: Parque de los Novios.

- Plaza de Bolívar

  Fecha de Grabación: Sábado 31 de Julio de 2010

  Componentes generales del ambiente sonoro: Palomas, voces de adultos, voces de niños, viento, buses, automóviles, campana de helados.

Imagen 28: Plaza de Bolívar.

- Museo Iglesia de Santa Clara

  Fecha de Grabación: Sábado 31 de Julio de 2010.

  Componentes generales del ambiente sonoro: Automóviles, buses, coro, violín, susurros.

Imagen 29: Museo Iglesia de Santa Clara

- Complejo Acuático de Bogotá

  Fecha de grabación: Martes 21 de septiembre de 2010.

  Componentes generales del ambiente sonoro: Ruido de la maquinaria, voces de niños, voces de adultos, silbidos.

Imagen 30: Complejo acuático de Bogotá.

4.3 EDICIÓN DE LOS AMBIENTES

El proceso de edición se realizó usando el software Pro Tools LE 8, que ofrece herramientas y facilidad para esto.

Cada ambiente posee en promedio una duración de 5 minutos, en algunos casos un poco más cortos debido a los errores en captura. Ruido interno generado por equipamiento o ruido de alta frecuencia debido al viento son factores que se tuvieron en cuenta como errores al momento de la edición.

Las voces demasiado cercanas al sistema de captura, fueron consideradas con mayor énfasis, como error en la grabación al momento de editar.

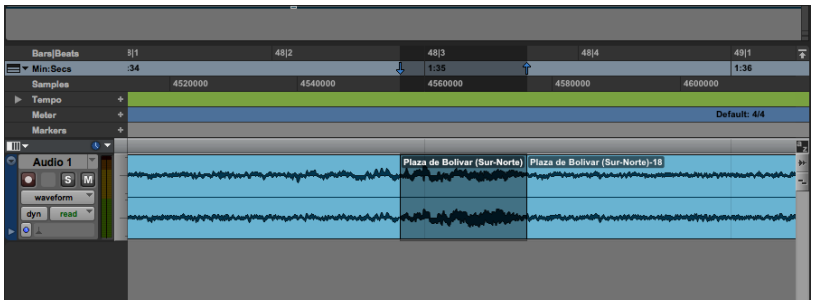


Imagen 31: Corte de región en proceso de edición.

Se retiraron los segmentos con las voces de las personas que pasaban demasiado cerca, luego se unieron las regiones que quedaban aparte.

La ventaja encontrada al momento de edición, fue la continuidad en la sonoridad de los ambientes, lo que permitía la unión de las regiones sin notar cambios bruscos, diferencias espectrales, silencios o niveles pico en el audio. (Cambios en dinámica demasiado notorios).

En algunas grabaciones fue necesario realizar *crossfades* en la unión de las regiones separadas para evitar los efectos nombrados en el párrafo anterior.

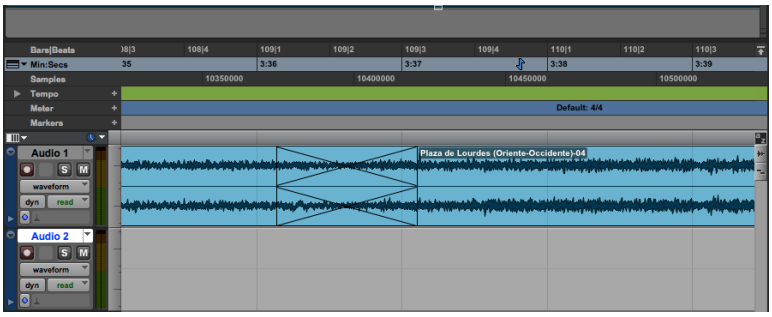


Imagen 32: Crossfade en proceso de edición.

Para la solución de los errores por ruido de alta frecuencia, primero se optó por el uso de *windscreen* sobre los micrófonos, un filtro en el momento de la captura.

En los casos más drásticos, se filtró la señal ya capturada con el *Kill EQ* de AIR[25]. Se diseñó un filtro pasa bajas con su frecuencia de corte igual a la utilizada en la técnica para realizar el *crosstalk cancellation* y generar el audio transaural, 8000 Hz.

Este procedimiento solo pudo ser realizado en ambientes que no tuvieran mucho contenido en alta frecuencia, especialmente interiores. Para los ambientes con contenido significante en alta frecuencia y en la banda de aire, fue necesario repetir la captura.

Imagen 33: Filtro pasa bajas en proceso de edición.

[25] Advanced Instrument Research

5. PRESENTACIÓN Y ANÁLISIS DE RESULTADOS

5.1 EXTRACCIÓN DE DIT (*Diferencia Interaural de Tiempo)*

Para la implementación de la HRTF en el diseño del audio transaural, se usaron las respuestas al impulso sobre el canal auditivo que ofrece *MIT Media Lab* como paquetes de libre descarga[26].

El planteamiento y análisis del diseño sobre un sistema estéreo se realiza sobre las respuestas con 0 grados de elevación, 30[27] grados y 330[28] grados en azimuth, para cada oído *(ver figura 14)*. Los archivos están registrados a 44100 Hz y 16 Bits, se encuentran nombrados siguiendo el patrón: *Oído-Elevación-Azimuth,* de esta forma se encuentra el siguiente archivo: *L0e030a.wav* que indica que es la respuesta al impulso sobre el canal auditivo del oído izquierdo con 0 grados de elevación y 30 grados en azimuth.

El análisis de las respuestas se realizó con el software Matlab, en su versión educativa.

El primer paso es asignar los archivos a vectores de datos en matlab con los siguientes comandos.

```
>> [l30,y]=wavread('L0e030a');

>> [r30,y]=wavread('R0e030a');

>> [l330,y]=wavread('L0e330a');

>> [r330,y]=wavread('R0e330a');
```

En l30, r30, l330 y r330 se encuentra el valor de cada muestra de audio y en la variable 'y' se encuentra la frecuencia de muestreo 44100 Hz. En las propiedades del vector se evidencia una longitud de 512 muestras, es decir que cada respuesta al impulso tiene una duración de 11,6 milisegundos.

En el informe de la medición de *MIT Media Lab,* se explica que en el archivo original existía un retraso de aproximadamente 230 muestras antes de que la respuesta de la cabeza ocurriera, este retraso corresponde a 180 muestras por la distancia fuente-oído (1,4 metros) y 50 muestras del sistema playback/grabación. Teniendo en cuenta esto, decidieron publicar los archivos de las respuestas al impulso, descartando las primeras 200 muestras de 712 en total.

---

[26] http://sound.media.mit.edu/resources/KEMAR.html
[27] Posición del altavoz derecho con respecto al oyente en un sistema estéreo.
[28] Posición del altavoz izquierdo con respecto al oyente en un sistema estéreo.

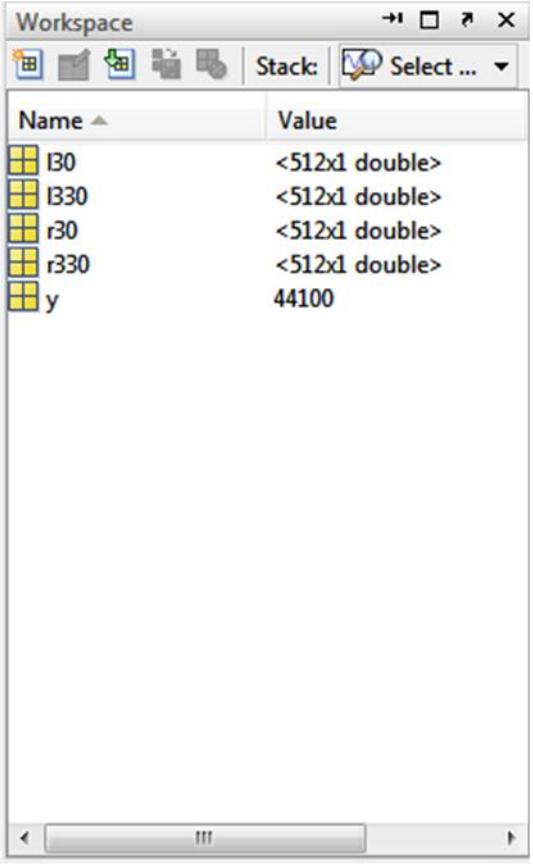


Imagen 34: Propiedades de las respuestas a 30 y 330 grados.

En la imagen 34 se observa la frecuencia de muestreo *(y)* y la longitud en muestras de cada respuesta al impulso, para cada oído *(l y r)* y para cada ángulo en azimuth *(30 y 330)*; ambos datos son iguales para todos los archivos.

Es necesario realizar una estimación visual del punto de inicio de cada respuesta al impulso, antes de analizarlas matemáticamente con el fin de comprobar la teoría de las diferencias interaurales de tiempo e intensidad. Este proceso se realiza con los siguientes comandos y la herramienta zoom.

```
>>plot(l30)

>>plot(r30)

>>plot(l330)

>>plot(r330)
```

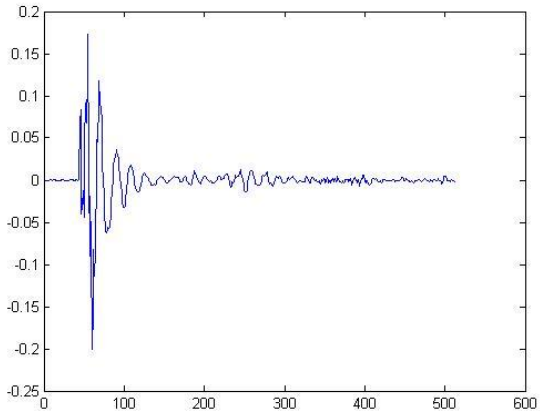


Figura 18: Respuesta al impulso l30.

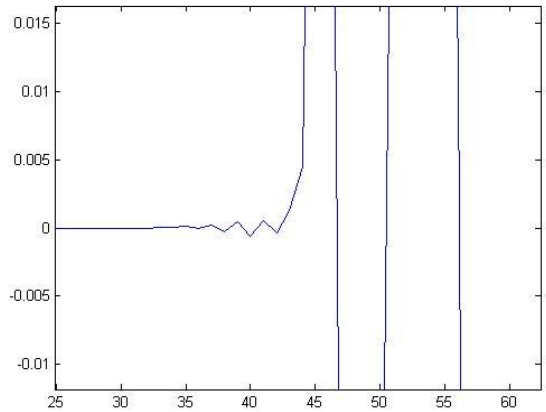


Figura 19: Zoom de la respuesta l30.

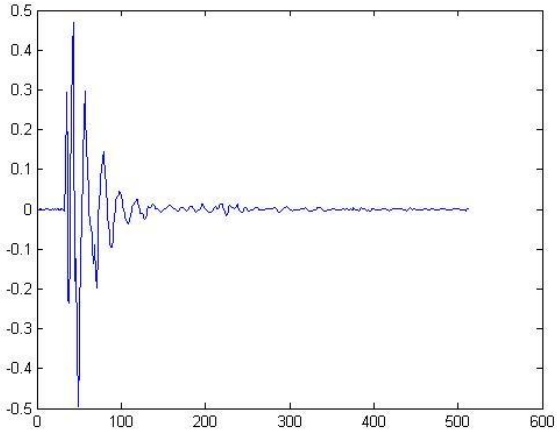


Figura 20: Respuesta al impulso r30.

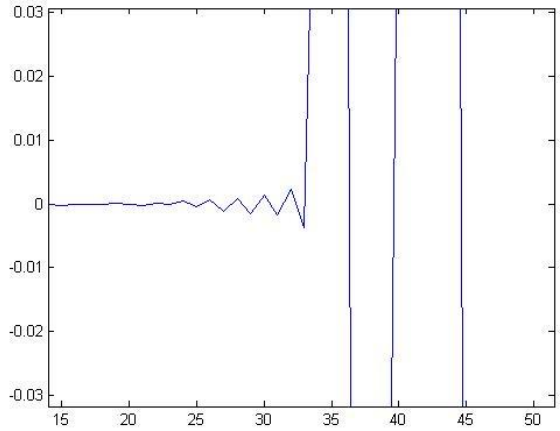


Figura 21: Zoom de la respuesta r30.

Entre las figuras 17 y 19 se observa la diferencia en la amplitud de las respuestas al impulso, siendo mayor esta en el oído derecho.

En las figuras 18 y 20 se encuentra el punto de inicio de cada respuesta. Para el oído izquierdo entre las muestras 40 y 45, y para el oído derecho entre las muestras 30 y 35. Esto demuestra el retraso que existe en el oído izquierdo con respecto al derecho para una fuente a 30 grados.

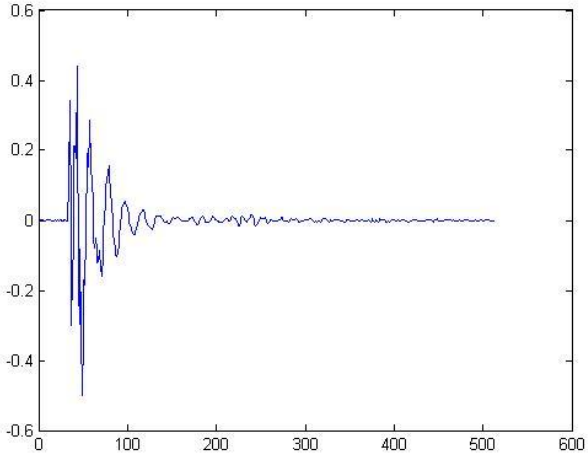


Figura 22: Respuesta al impulso l330.

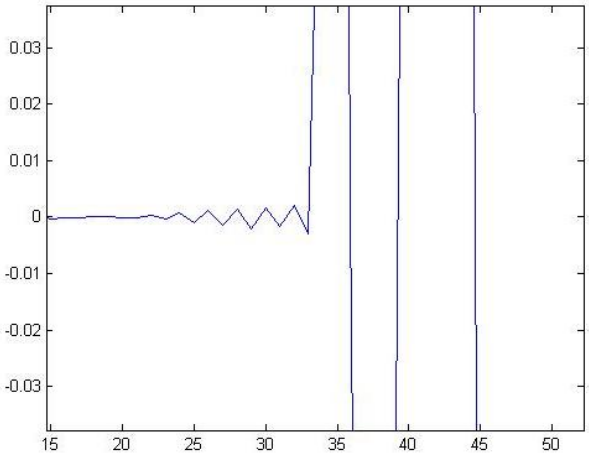


Figura 23: Zoom de la respuesta l330.

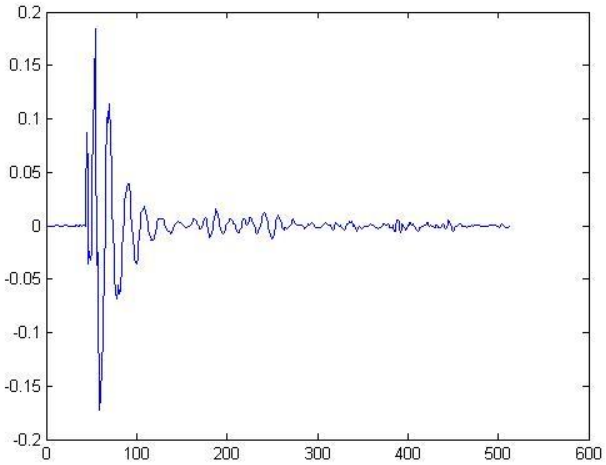


Figura 24: Respuesta al impulso r330.

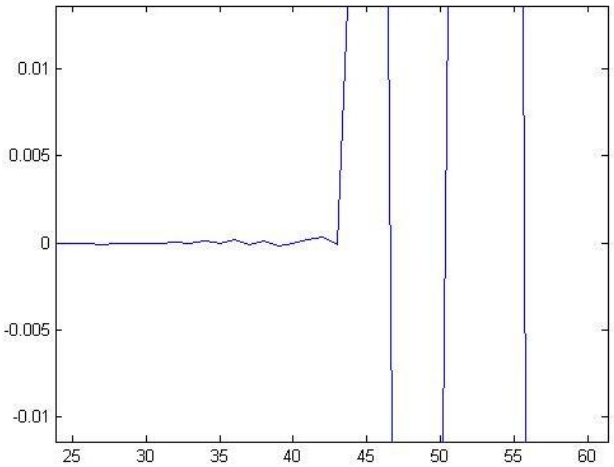


Figura 25: Zoom de la respuesta r330.

Entre las figuras 21 y 23 se observa que la amplitud es mayor para el oído izquierdo.

En las figuras 22 y 24, el punto de inicio se encuentra para el oído izquierdo entre las muestras 30 y 35, y para el oído derecho entre las muestras 40 y 45. Esta vez el retraso se encuentra en el oído derecho con respecto al izquierdo para la fuente a 330 grados.

Luego de tener las aproximaciones visuales, y de comprobar que las respuestas al impulso demuestran la teoría expuesta para la localización de fuentes sonoras, se realiza el análisis matemático para extracción de las diferencias interaurales de tiempo.

Primero se analizan las respuestas al impulso para una fuente a 30 grados.

Se asigna a una variable el punto máximo de las respuestas con los siguientes comandos.

```
>> a=max(l30)

a =

    0.1727

>> b=max(r30)

b =

    0.4700
```

Esto con el fin de encontrar su posición en el vector, es decir la muestra que posee el valor máximo de la respuesta. Se ubica esta muestra con los siguientes comandos.

```
>> find(l30==a)

ans =

    55

>> find(r30==b)

ans =

    43
```

Teniendo la ubicación de las muestras con el punto máximo en cada señal se puede calcular la diferencia de tiempo entre los dos oídos de la siguiente forma.

```
>> DIT30=(55-43)/y

DIT30 =

  2.7211e-004
```

Para una fuente a 30 grados en azimuth y 0 grados de elevación, existe una diferencia interaural de tiempo de 0,27 milisegundos.

El análisis de las respuestas al impulso para una fuente a 330 grados en azimuth, se realiza con el mismo procedimiento.

```
>> c=max(l330)

c =

   0.4404

>> d=max(r330)

d =

   0.1838

>> find(l330==c)

ans =

   43

>> find(r330==d)

ans =

   54

>> DIT330=(54-43)/y

DIT330 =

  2.4943e-004
```

Para una fuente a 330 grados en azimuth y 0 grados de elevación existe una diferencia interaural de tiempo de 0,25 milisegundos.

En este procedimiento aparece una diferencia inusual entre las DIT para 30 y 330 grados de 0,02 milisegundos, que de igual forma se tuvo en cuenta para el diseño del audio transaural.

El planteamiento y análisis del diseño transaural para los altavoces surround, se obtiene a partir de las respuestas para 110[29] y 250[30] grados en azimuth, siguiendo el mismo procedimiento.

[29] Posición del altavoz derecho surround con respecto al oyente en un sistema 5.1.
[30] Posición del altavoz izquierdo surround con respecto al oyente en un sistema 5.1.

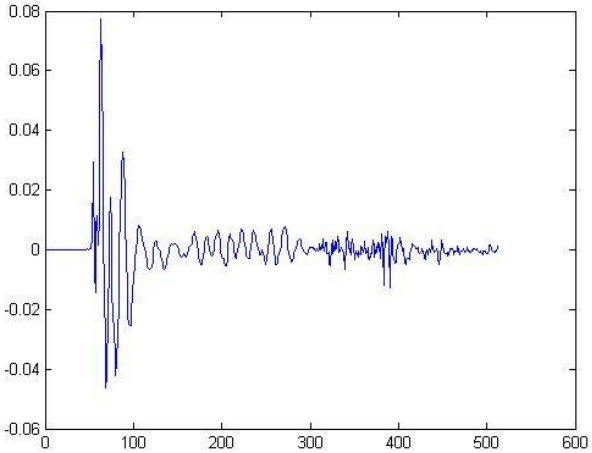


Figura 26: Respuesta al impulso l110.

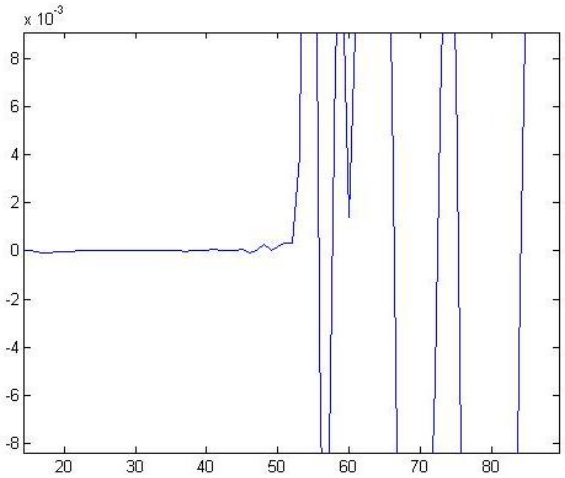


Figura 27: Zoom de la respuesta l110.

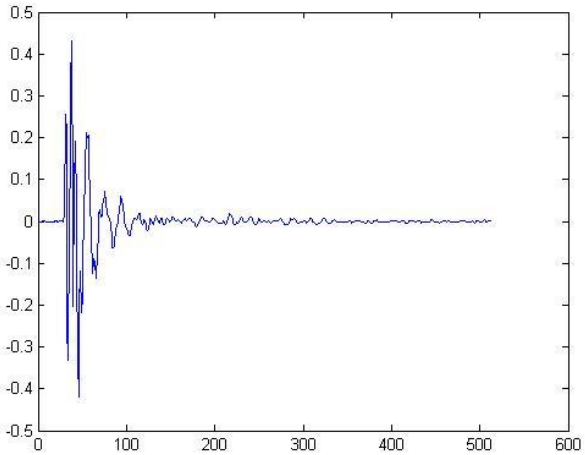


Figura 28: Respuesta al impulso r110.

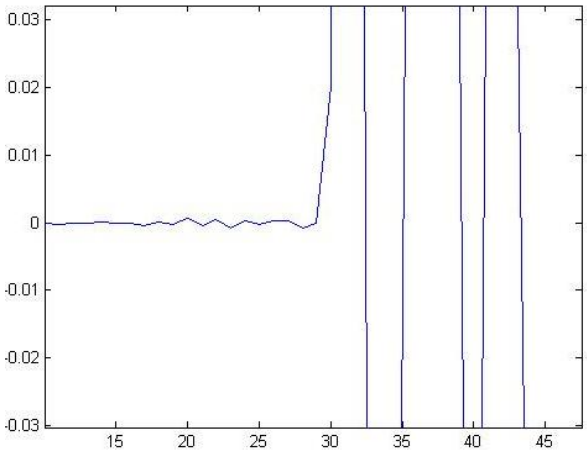


Figura 29: Zoom de la respuesta r110.

Entre las figuras 25 y 27 se observa la diferencia en la amplitud de las respuestas al impulso, siendo mayor esta en el oído derecho.

En las figuras 26 y 28, el punto de inicio se encuentra para el oído izquierdo entre las muestras 50 y 60, y para el oído derecho entre las muestras 25 y 30. Esto demuestra el retraso que existe en el oído izquierdo con respecto al derecho para una fuente a 110 grados.

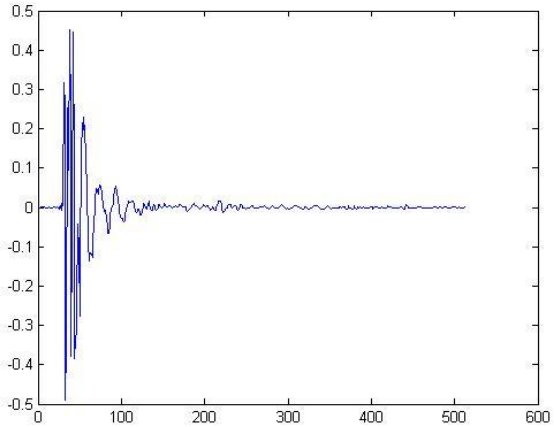


Figura 30: Respuesta al impulso l250.

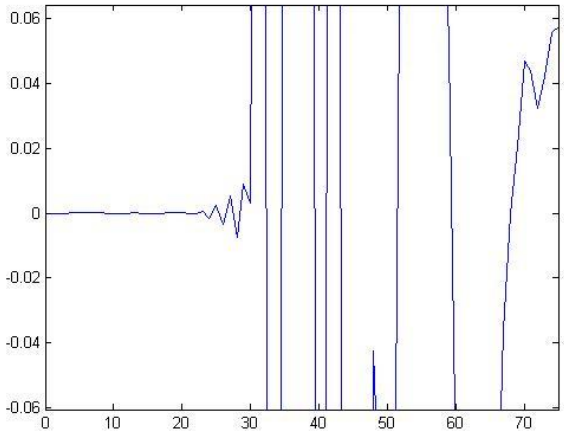


Figura 31: Zoom de la respuesta l250.

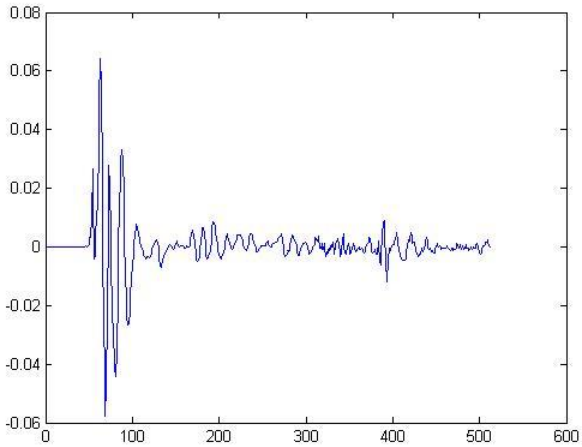


Figura 32: Respuesta al impulso r250.

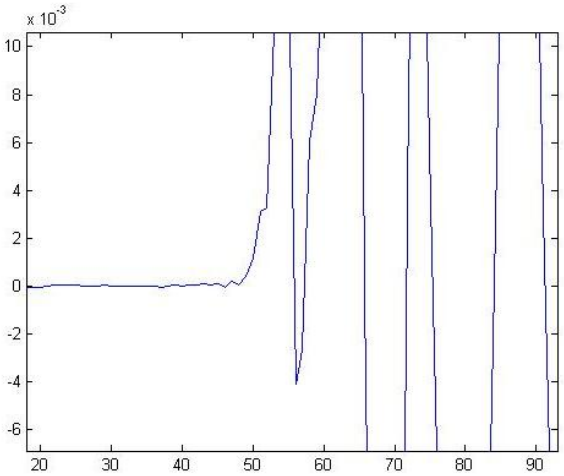


Figura 33: Zoom de la respuesta r250.

Entre las figuras 29 y 31 se observa que la amplitud es mayor para el oído izquierdo.

En las figuras 30 y 32, el punto de inicio se encuentra para el oído izquierdo entre las muestras 20 y 30, y para el oído derecho entre las muestras 40 y 50. Esta vez el retraso se encuentra en el oído derecho con respecto al izquierdo para la fuente a 250 grados.

Nuevamente la aproximación visual guarda relación con la teoría. De esta forma se procede con el análisis matemático.

```
>> e=max(l110)

e =

    0.0772

>> f=max(r110)

f =

    0.4326

>> find(l110==e)

ans =

    63

>> find(r110==f)

ans =

    38

>> DIT110=(63-38)/y

DIT110 =

  5.6689e-004
```

Para una fuente a 110 grados en azimuth y 0 grados de elevación existe una diferencia interaural de tiempo de 0,57 milisegundos.

```
>> g=max(l250)

g =

    0.4531

>> h=max(r250)

h =

    0.0642
```

```
>> find(l250==g)

ans =

    38

>> find(r250==h)

ans =

    63

>> DIT250=(63-38)/y

DIT250 =

  5.6689e-004
```

Para una fuente a 250 grados en azimuth y 0 grados de elevación existe una diferencia interaural de tiempo de 0,57 milisegundos.

5.2 DISEÑO DEL AUDIO TRANSAURAL PARA UN SISTEMA ESTEREO 2.0

La teoría propone que con 5 filtros por canal, se obtiene una cancelación de la interferencia de 20dB a 25dB. Este proceso resultó en un cambio drástico del contenido espectral de la señal, así que se realizó el procesamiento con 2 filtros por canal, con el fin de mantener las características en frecuencia de la señal grabada.

Con el primer filtro se obtiene una excelente localización sonora en el espacio tridimensional a través de los altavoces, pero se crea una cancelación total en baja frecuencia. Esto concuerda con la longitud de onda y las características direccionales de las frecuencias bajas.

Con el fin de mantener la calidad de las grabaciones se realizó un segundo filtro que efectivamente recupera el contenido en baja frecuencia. Con la suma de estos dos filtros, se obtiene una pérdida en la localización de fuentes en el espacio virtual.

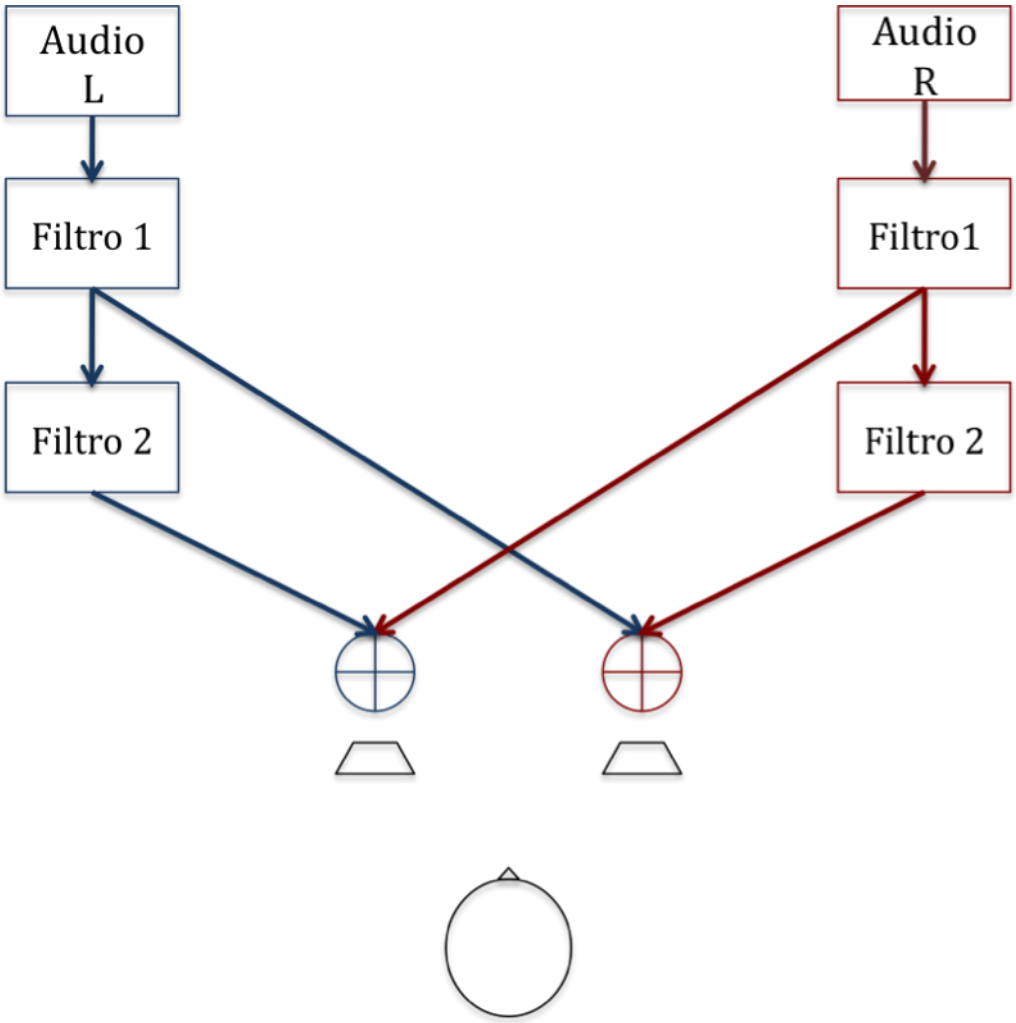


Figura 34: Flujo de señal para los filtros de cancelación.

La solución a los problemas de pérdida en baja frecuencia y pérdida en localización se soluciona mediante un ajuste de nivel en cada filtro.

Los 4 filtros tienen la fase invertida, y en este caso, por los resultados obtenidos de las DIT, varia el retraso dependiendo de si se dirige al canal izquierdo (0.27ms) o al canal derecho (0.25ms). El delay es utilizado como un *Time Adjuster* con 0% de feedback 100% en la mezcla y profundidad, esto permite que el audio sea escuchado sin repeticiones en la posición temporal deseada.

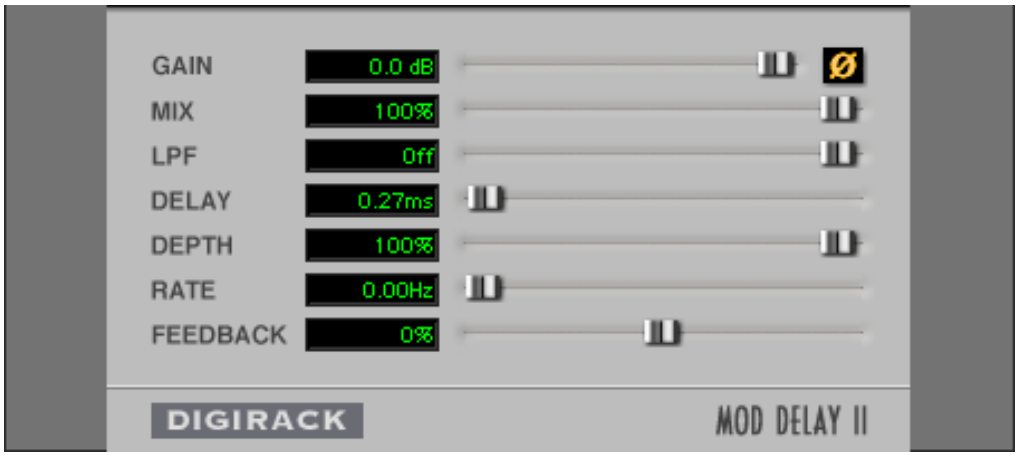


Imagen 35: Ajustes para los filtros que se dirigen al altavoz izquierdo.

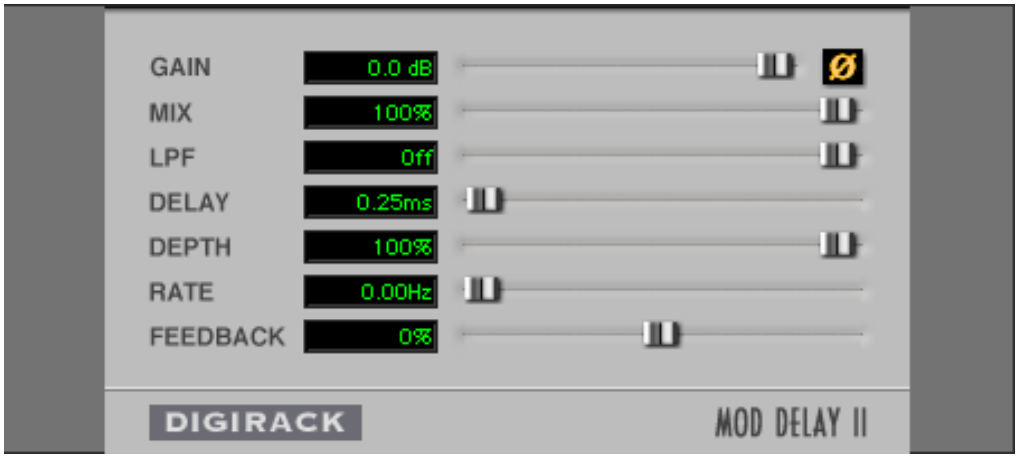


Imagen 36: Ajustes para los filtros que se dirigen al altavoz derecho.

La nivelación se realiza de la siguiente forma para obtener el mejor balance entre contenido espectral y localización de fuentes sonoras.

- Audio: 0 dBfs
- Filtro 1: -2.5 dBfs
- Filtro 2: -4.9 dBfs

Después de monitorear los resultados y comprobar que el audio 3D se generara en todos los ambientes, se grabó todo el proceso en un *audio track* estéreo para su exportación.

Imagen 37: Flujo de señal y nivelación en Pro Tools.

Con el balance entre contenido en frecuencia y localización de fuentes, se llegó al mejor resultado en donde la generación de fuentes virtuales llega máximo a los 110 grados en azimuth, con respecto al eje central del oyente. Lugar en donde irán los altavoces surround, que tendrán un mejor desempeño como se evidencia en el capitulo 5.4.

5.3 DISEÑO DEL AUDIO TRANSAURAL PARA UN SISTEMA SURROUND 5.1

Debido a la imposibilidad de trabajar directamente en un sistema 5.1, se siguió el mismo procedimiento para crear la cancelación del crosstalk en los altavoces surround.

De esta manera se tiene un audio estéreo para el frente, un audio estéreo para atrás y una señal monofónica para los canales Centro y Sub.

En este caso el retraso es el mismo para las señales independientemente del canal al que se dirige. De igual forma el delay se trabaja como un *Time Adjuster.*

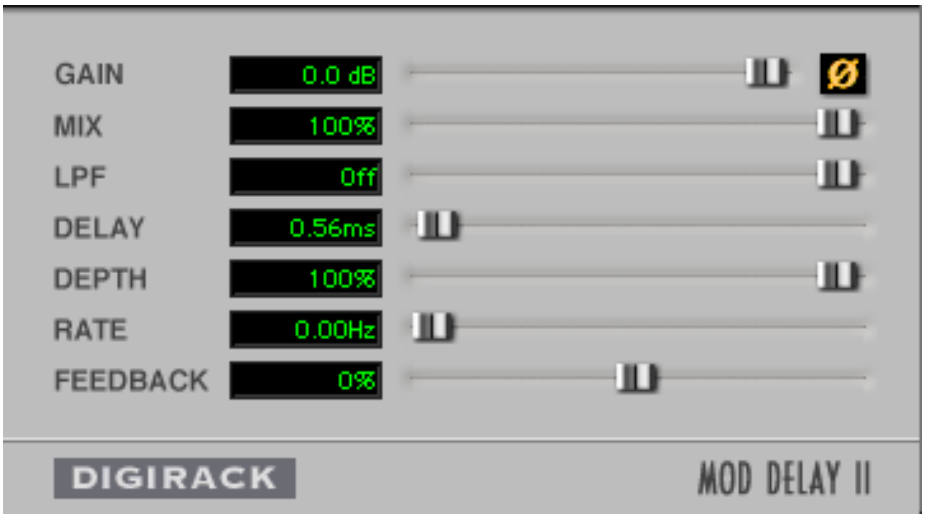


Imagen 38: Ajustes para los filtros de los altavoces LS y RS.

Para monitorear su funcionamiento antes de la codificación, es necesario un sistema de reproducción que permita enviar una señal monofónica a cada uno de los 6 canales del 5.1. El software utilizado para este fin fue Logic 9. Se utilizó la interfaz *Firewave* de Griffin, que proporciona 6 salidas por pares estéreo y soporte para Dolby Digital.

La configuración I/O del software se realizó de la siguiente forma: 1-L, 2-R, 3-C, 4-LFE, 5-LS, 6-RS[31]. La conexión se realizó a un sistema de parlantes 5.1 para casa.

Con el audio transaural en 5.1 se llega a una mejor generación de fuentes virtuales, gracias al refuerzo posterior. La ubicación en el espacio tridimensional se hace mas clara y los movimientos de la cabeza del oyente, se hacen menos significativos debido a la amplitud del *Sweet Spot* en este tipo de sistemas de reproducción.

Al comprobar que el audio transaural funcionara en el sistema, se procede a la codificación de las 6 señales para obtener un solo archivo. Esta codificación se realizó con la aplicación Compressor. Se codificó un archivo Dolby digital AC-3 para cada ambiente transaural en 5.1.

5.4 EVALUACIÓN DEL AUDIO 3D

Se realizó un test[32] de escucha para evaluar el desempeño del audio 3D y de cada sistema de reproducción. Se dejó a libre selección el ambiente a escuchar, teniendo en cuenta que el audio 3D y los filtros de cancelación deben funcionar en todos los ambientes y no deben depender de las características[33] del lugar. El test lo realizaron 25 personas, de las cuales 10 poseen un entrenamiento auditivo[34].

5.4.1 Evaluación de parámetros significativos

Los parámetros a evaluar son los siguientes:

a) Profundidad del ambiente sonoro.

b) Realidad del ambiente sonoro.

c) Generación de fuentes virtuales.

d) Distinción de fuentes sonoras en el espacio.

e) Claridad en los elementos sonoros.

---

[31] Left, Rigth, Center, Low Frecuency Effects(Subwoofer), Left Surround, Right Surround.
[32] Ver anexo 1.
[33] Lugar: Cerrado, abierto, ruidoso, reverberante, etc.
[34] Ingenieros de sonido y músicos.

CALIFICACIÓN DE PARÁMETROS					
#	A	B	C	D	E
1	5	4	4	4	5
2	5	4	4	5	5
3	4	3	4	4	4
4	2	4	5	4	4
5	5	4	4	5	4
6	3	3	4	4	5
7	5	5	3	5	5
8	4	4	3	4	5
9	5	5	5	4	4
10	5	5	5	5	5
11	3	5	5	3	4
12	3	5	3	5	5
13	4	5	5	5	4
14	2	4	4	5	5
15	3	4	4	2	4
16	5	4	5	3	5
17	2	3	3	4	4
18	3	5	5	4	3
19	4	5	5	4	4
20	4	5	4	4	5
21	4	5	3	5	5
22	5	5	5	5	5
23	5	3	4	5	5
24	4	5	4	5	5
25	4	5	4	5	5
PROMEDIO	3,92	4,36	4,16	4,32	4,56

Tabla 8: Resultados en la calificación de parámetros.

El parámetro con menos calificación fue la profundidad del ambiente. En lugares con mucha afluencia de gente, es complicado retirar todas las voces. En la aplicación del test se notó la inquietud de las personas por las voces sobre los ambientes sonoros de salitre mágico, estadio Nemesio Camacho el campín, y monserrate, en donde estas voces son un elemento distractor sobre todo el entorno virtual. Sin embargo se presenta una claridad en los elementos sonoros del ambiente, siendo este factor el de mayor calificación. Esto quiere decir que la grabación y el procesamiento son adecuadas y no cambian los componentes propios del sonido en el lugar. El timbre de cada elemento sonoro se mantiene concluyendo que no hubo modificación en el contenido espectral de la señal, ni en el ataque de su envolvente.

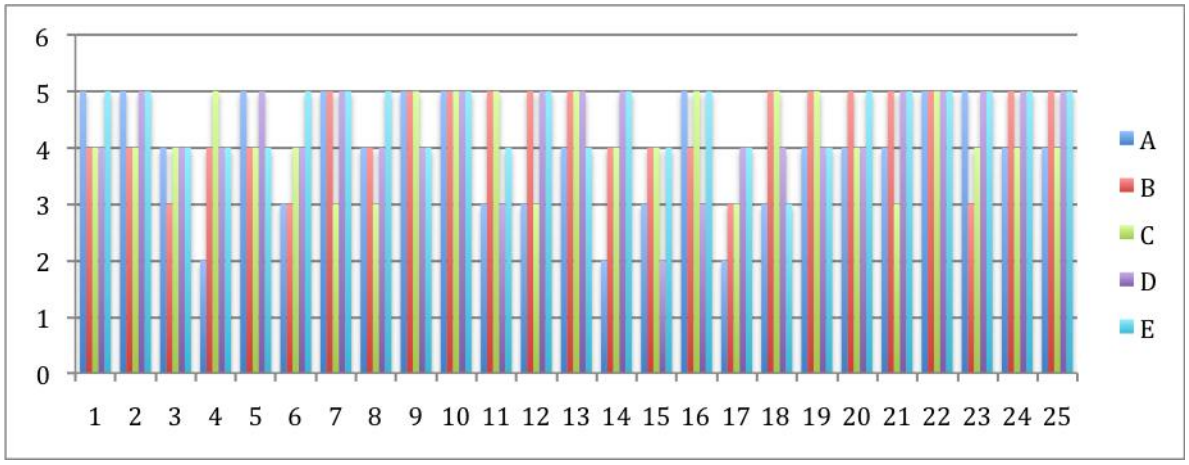


Figura 35: Calificación de parámetros por persona.

Las calificaciones sobre 2 que se observan en la figura 35, pertenecen al ambiente sonoro del billar londres. En este lugar la calificación recibida fue baja en los parámetros de profundidad y generación de fuentes virtuales, esto debido a un contenido musical constante en el lugar al momento de la grabación. Este contenido hace la imagen sonora confusa y reduce la capacidad del oyente para la localización de fuentes sonoras.

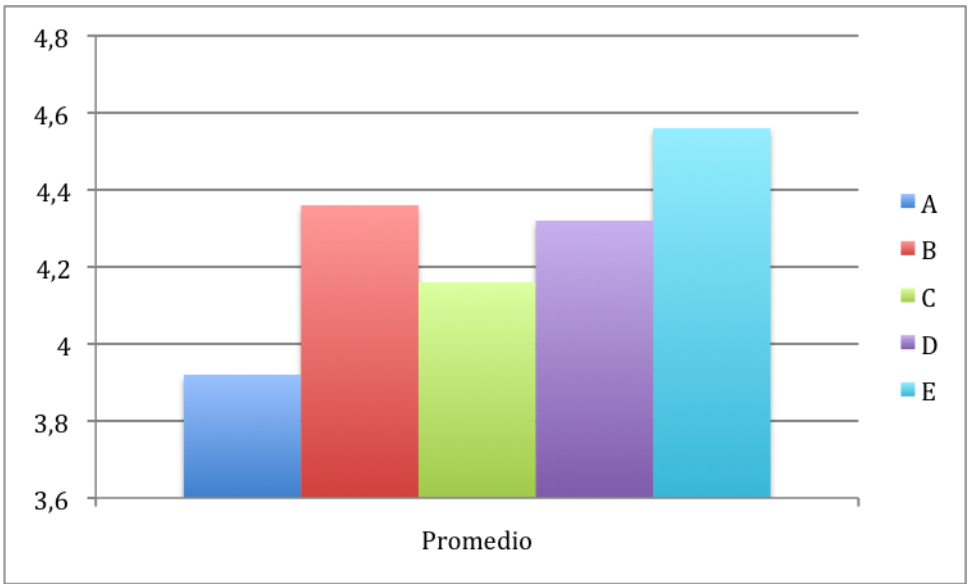


Figura 36: Calificación de parámetros general.

Los parámetros de realidad, generación de fuentes virtuales, distinción y claridad están por encima de 4 en la calificación general. Es claro para el oyente identificar fuentes 360 grados alrededor de la cabeza y arriba, solamente la imagen se hace confusa al discriminar fuentes que provienen de una elevación negativa[35].

La realidad del ambiente sonoro lograda es óptima, demostrando un proceso adecuado en toda la cadena de producción (Captura-Edición-Filtrado), se evidencia una localización de fuentes acertada por parte del oyente, al igual que la distinción en los elementos sonoros que componen el ambiente.

La profundidad del ambiente, es el factor mas dependiente del entorno de escucha. Está limitado por el tipo de sistema que se este utilizando[36], y la sala en la que se este reproduciendo para el caso de los altavoces.

5.4.2 Evaluación de los sistemas de reproducción

Con pocas muestras fue posible evidenciar el resultado concluido en el capitulo 5.3. El *sweet spot* del sistema 5.1, es una gran ventaja en el diseño del audio transaural.

EVALUACIÓN DE LOS SISTEMAS DE REPRODUCCIÓN					
Audifonos	2	2	3	2	2
Altavoces Estéreo	1	1	1	1	1
Altavoces Surround	3	3	2	3	3

Tabla 9: Evaluación de los sistemas de reproducción.

En la tabla 9, se asigna un número a cada sistema, siendo 3 el de mejor desempeño y fidelidad en la reproducción de los parámetros calificados anteriormente.

Se evidencia que el audio transaural en estéreo es el de menor preferencia, esto debido a la influencia significativa del movimiento del oyente en el limitado *Sweet Spot* propio del sistema.

[35] Fuentes provenientes de abajo del eje interaural.
[36] Tipo de audífonos, Tipo de altavoces.

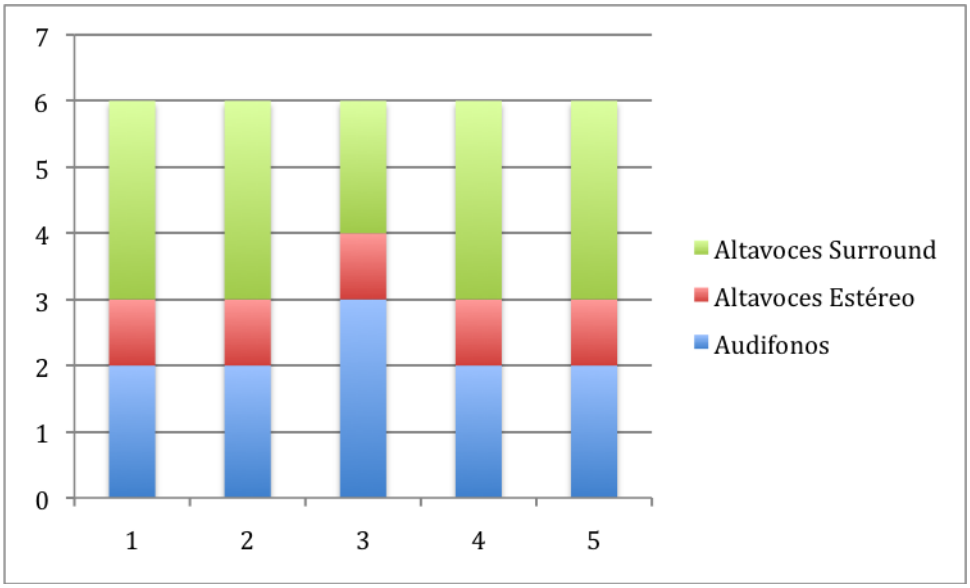


Figura 37: Áreas de preferencia por persona.

En audífonos, se manifiesta una mayor inmersión en el espacio virtual debido al asilamiento del ruido externo, sin embargo, la mejor sensación tridimensional en todos los aspectos se evidencia en el sistema 5.1.

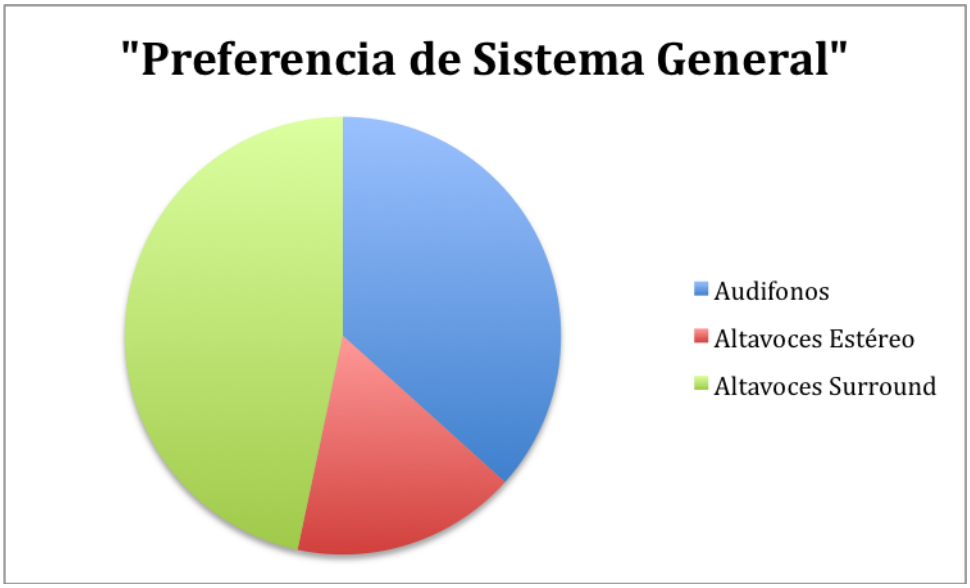


Figura 38: Preferencia de sistema general.

6. CONCLUSIONES

- La selección de locaciones en Colombia para la grabación de producciones audiovisuales, se puede hacer de una forma organizada y legal, con las herramientas que brindan las entidades que apoyan este trabajo artístico.

- La base de datos de locaciones que posee el fondo para el desarrollo cinematográfico en Colombia, es una propuesta con una estructura excelente que deja al país como un destino atractivo para productoras internacionales.

- La captura de los ambientes, implementando una HRTF individual, concuerda con lo expuesto teóricamente sobre espacialidad y localización de fuentes. Esta localización sonora se mantiene para diferentes oyentes y en diferentes condiciones de reproducción, siempre que estén dentro de los limites que se plantean para el proyecto.

- Con los equipos utilizados para la captura, se obtiene un ambiente sonoro que conserva el contenido espectral y transiente de los elementos que componen el lugar. Es posible localizar y discriminar cada sonido en el espacio tridimensional generado. Aún siendo contenido digital, se llega a una aproximación a la realidad, sin manifestación de elementos sintéticos en el ambiente.

- La grabación de ambientes con bastante afluencia de público, generan una imagen confusa debido a elementos muy cercanos al sistema de captura. Algunos de estos elementos, como las voces, se pueden sustraer del ambiente sin generar cambios significativos en dinámica, teniendo en cuenta la continuidad en los componentes de fondo.

- La reproducción por audífonos genera una mejor inmersión en el espacio tridimensional debido al aislamiento del ruido externo. Sin embargo en una sala aislada del ruido externo se obtiene este resultado con altavoces. Este sistema de reproducción esta limitado por el tipo de audífonos que se usen, siendo el factor mas importante su respuesta en frecuencia.

- El análisis de resultados para las DIT guarda relación con la teoría expuesta sobre la HRTF, exceptuando la diferencia de los tiempos para 30 grados y 330 grados, que puede deberse a las muestras que fueron retiradas del archivo de análisis o una variación en la posición de medición. Las diferencias de tiempo interaural para las fuentes a 110 y 250 grados coinciden en 0,57ms. Siendo 270 grados la mayor diferencia de tiempo en este caso, los resultados concuerdan con el máximo expuesto en la teoría de 0,6ms.

- El sistema de reproducción por altavoces en estéreo, es el más limitado para la generación de la tridimensionalidad. Depende del tipo de altavoces que se estén usando, su respuesta en frecuencia, la imagen estéreo que generan, la definición de centro y el sweet spot. Depende significativamente de la sala que se use para su reproducción y de los movimientos del oyente con respecto al eje central de escucha. Su mejor desempeño se da en altavoces profesionales.

- Los filtros canceladores para el audio transaural en 2.0, generan fuentes virtuales a máximo 110 grados y 250 grados en azimuth, lugar en donde van situados los altavoces surround para el refuerzo posterior. La localización de fuentes en elevación positiva[37] se percibe al 100%. Con la nivelación de cada filtro se mantienen las características sonoras de la captura y la calidad del ambiente para obtener la realidad virtual deseada.

- El sistema de reproducción por altavoces en 5.1, tiene un sweet spot amplio que permite el movimiento del oyente sin cambios drásticos en la llegada de las señales a los oídos, permitiendo una mejor percepción del 3D. La localización de fuentes es clara en el sistema 5.1 y permite el uso de sistemas no profesionales para su funcionamiento (Teatro en Casa).

- Los filtros canceladores para 5.1 generar una simulación de fuentes 360 grados alrededor del oyente. Se perciben 100% las fuentes con elevación positiva y se mantiene el contenido espectral y transiente del ambiente capturado.

- La profundidad del ambiente sonoro, es un factor que depende notablemente de los elementos en la reproducción (Sistema-Sala-Posición). La realidad del ambiente, la generación de fuentes virtuales, la localización y claridad de los elementos sonoros en el espacio tridimensional, están arriba del punto medio de calificación, los oyentes demostraron el funcionamiento de la captura y de los filtros con una preferencia hacia un sistema5.1.

[37] Fuentes sonoras arriba del eje interaural.

7. RECOMENDACIONES

- Actualmente se está trabajando en otras partes del mundo el concepto de audio en alta definición. Con el paso del tiempo y el posicionamiento del Blue Ray en Colombia, se puede pensar en un banco de ambientes grabado a 96KHz y 24 Bits.
- Un proyecto de investigación, podría enfocarse en solucionar los limites del audio transaural en 2.0, teniendo en cuenta el movimiento del oyente y el sweet spot de los altavoces.
- Con la instalación del sistema 5.1 en los estudios de la Universidad de San Buenaventura Bogotá, se puede explorar mucho mas el campo del audio en 3D y proponer experiencias auditivas que acompañen la tecnología desarrollada hoy en día en el campo visual.

8. BIBLIOGRAFÍA

- Borwik, Jhon. *Loudspeaker and headphone handbook.* Gran Bretaña: Focal Press, 2001.

- Everest, Aton: *The master Handbook of acoustic*. USA: McGraw-Hill, 2001.

- Foo, Kelvin; Hawksford, Malcolm; Hollier, Michael. "Optimization of virtual sound reproduced using two loudspeakers". Centre for Audio Research and Engineering, Department of Electronic Systems Engineering, University of Essex, Wivenhoe Park, Colchester, UK ; BT Laboratories, Martlesham Heath, Ipswich, UK: Marzo, 1999.

- Gradner, Bill; Martin, Keith. "HRTF Measurements of a KEMAR Dummy-Head microphone". MIT Media Lab: Mayo, 1994.

- Holman, Tomlinson. *Sound for film and television*. USA: Focal Press, 1997.

- Kinsler, Lawrence. *Fundamentals of acoustics*. USA:2000

- Lacouture, Yesenia. "Analysis of design parameters for crosstalk cancellation filters applied to different loudspeaker configurations". Aalborg University, Dinamarca: Octubre de 2008

- Martin, Geoff. *Introduction to sound recording*. 2004.

- Pico, Rubén; Ramis, Jaime; Redondo, Javier; Alba, Jesús. "Medida de la función de transferencia de un maniquí acústico". Escuela Politécnica superior de Gandia. España.

- Rose, Jay. *Audio Postproduction for film and video*. USA: Focal Press, 2009.

- Ruffa, Francisco. "Capitulo segundo: El mecanismo de la percepción y sus características subjetivas". Diseño de sistemas de sonido 2. Universidad de San Buenaventura. Colombia: 2009

- Shumaker, Randall. *Virtual Reality*. Segunda conferencia internacional ICVR. China: Julio, 2007

- Talbot, Michael. *Audio engineer's reference book*. Gran Bretaña: Focal Press, 1999.

- Venegas, R; Lara, M; Correa, R; Floody, S. "Medición de HRTF y extracción de características significativas para localización sonora". Departamento de Acústica. Universidad Pérez Rosales. Chile: 2005.

- Vorländer, Michael. *Auralization*. Alemania: Springer, 2008

- http://www.soundonsound.com/sos/Oct04/articles/bilocation.htm

- http://faculty.virginia.edu/fieldrecording/

- http://www.lpi.tel.uva.es/~nacho/docencia/ing_ond_1/trabajos_03_04/Csound/44.htm

ANEXOS

Encuesta de calificación de audio 3D

Nombre:

Ambiente Seleccionado:

1. Califique los siguientes factores de 1 a 5, siendo 1 la menor calificación y 5 la mayor.

a. Profundidad del ambiente sonoro.(Distancia de los elementos sonoros con respecto al oyente)

1	2	3	4	5

b. Realidad del ambiente sonoro.

1	2	3	4	5

c. Generación de fuentes virtuales (Sonidos que no parecen venir del sistema de reproducción, audifonos-altavoces).

1	2	3	4	5

d. Distinción de fuentes sonoras en el espacio. (Localización)

1	2	3	4	5

e. Claridad en los elementos sonoros (Puede distinguir a que pertenece cada sonido).

1	2	3	4	5

2. Teniendo en cuenta los factores que calificó en el punto anterior, ordene según su desempeño los sistemas de reproducción, siendo 3 el mejor.

a. Audífonos.

b. Altavoces Estéreo.

c. Altavoces Surround.