UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN

Medidas objetivas de localización de un evento sonoro para sistemas de teleconferencia

TESIS DOCTORAL

ELENA BLANCO MARTÍN
Ingeniera de Telecomunicación

2009

Departamento de
Señales, Sistemas y Radiocomunicaciones

ESCUELA TÉCNICA SUPERIOR DE
INGENIEROS DE TELECOMUNICACIÓN

Medidas objetivas de localización de un evento sonoro para sistemas de teleconferencia

ELENA BLANCO MARTÍN
Ingeniera de Telecomunicación

Director:
Francisco Javier Casajús Quirós
Doctor Ingeniero de Telecomunicación

2009

TESIS DOCTORAL

Medidas objetivas de localización de un evento sonoro para sistemas de teleconferencia

AUTOR

Elena Blanco Martín

DIRECTOR

Francisco Javier Casajús Quirós

El Tribunal nombrado para juzgar la Tesis Doctoral arriba citada, compuesto por:

Presidente:

D. Ramón García Gómez

Vocales:

Dª. Soledad Torres Guijarro

D. Lino García Morales

D. José Javier López Monfort

Secretario:

D. Luis Alfonso Hernández Gómez

Acuerda otorgar la calificación de:

En Madrid, a ……. de ……de 2009

AGRADECIMIENTOS

Gracias a todos los que me han ayudado de forma directa e indirecta a la realización de esta Tesis Doctoral.

A mi Director de Tesis, Francisco Javier Casajús Quirós, por creer que podría salir de mi trabajo una Tesis Doctoral.

A mis amigos y compañeros del DIAC Juan José Gómez Alfageme y Luis Ignacio Ortiz Berenguer por no dejar que abandonara en los momentos complicados; a José Luis Sánchez Bote por sus ayudas científicas y técnicas.

A Soledad Torres Guijarro y Jon Ander Beracoechea Álava, antiguos miembros del GAPS, por ayudarme en lo que he necesitado.

A Luis Carlos Seco y Antonio Rodríguez, que me han ayudado a montar y desmontar todo el equipamiento en repetidas ocasiones.

Y sobre todo a mi familia, a mis padres por encargarse mientras que yo no estaba de los pequeños y permitirme dedicar tiempo a esta Tesis.

Gracias a todos.

ÍNDICE

RESUMEN

El objetivo de esta Tesis Doctoral es el desarrollo de una herramienta que permita medir, de forma objetiva, sobre el sonido recibido en los oídos (señal biaural) la sensación subjetiva de ubicación en el espacio del evento sonoro que genera dicha señal. Es importante diferenciar entre la posición de las fuentes sonoras, ubicación física real, y la posición del evento sonoro que representan dichas fuentes, ubicación subjetiva.

Esta herramienta se aplicará al caso de sistemas de teleconferencia, sistemas que intentan reproducir virtualmente la información sonora y visual que procede de otra sala, representándola en una pantalla con altavoces. El sistema de reproducción de audio que se va a estudiar está basado en el concepto de ventana acústica virtual. La ventana acústica virtual es un sistema de comunicaciones que recrea el Principio de Huygens. Lo que se pretende es hacer que entre dos salas (emisora y receptora, que pueden estar a cientos de kilómetros) aparezca una ventana virtual entre ellas de forma que la sensación subjetiva sea que están comunicadas por una ventana física en una de las paredes. Para ello se capta el frente de onda sonora que llega a la pared donde se sitúa la ventana en la sala emisora y se reproduce en la pared de la ventana acústica virtual en la sala receptora. Así pues, a diferencia de otros sistemas de reproducción multicanal, como el estéreo o los sistemas surround, donde la "espacialidad" de las fuentes de sonido sólo se reproduce correctamente en una zona limitada de la sala (llamada "sweet spot"); la ventana acústica virtual permite reconstruir el campo sonoro que se introduciría por la ventana virtual dejando que se propague por la sala el sonido que viene de la sala emisora como si hubiera una apertura física real. Por lo tanto las aplicaciones basadas en el concepto de ventana acústica virtual proporcionan mejoras evidentes sobre un conjunto de atributos a menudo relacionados con el concepto de calidad espacial: localización, profundidad y extensión. Concretamente uno de esos atributos subjetivos, la localización del evento sonoro reproducido, es el que se investiga en esta Tesis, poniendo de manifiesto las limitaciones de la ventana acústica virtual a la hora de reproducir la información de localización del sonido.

El concepto de ventana acústica virtual está basado en el uso de un array de micrófonos capaz de capturar el frente de onda en una pared de la sala emisora, más un array de altavoces encargado de reconstruir dicho frente de onda en otra pared de la sala de recepción. En el caso hipotético de que el número de transductores fuera infinito y la pared posterior totalmente absorbente, la reproducción del campo sonoro sería perfecta. Cómo dicho número está limitado, la reproducción es imperfecta y lo que se busca es valorar como influye la distorsión introducida en la calidad del sonido reproducido, y en concreto sobre los parámetros de localización subjetiva.

Se ha dividido el estudio de la localización en los dos planos principales: el plano horizontal y el plano medio. Es en el plano horizontal donde se desarrollan más investigaciones, ya que el sistema auditivo humano está desarrollado principalmente para ubicar los "peligros" en este plano, de ahí la ubicación de dos sensores (oídos) localizados

en diferentes puntos de este plano. En lo que respecta al plano medio, existen escasos estudios, ya que la localización en este plano no parte de dos señales sino de una (señal monoaural) y además la resolución subjetiva de localización del sistema auditivo humano es peor.

A continuación se presenta la Tesis Doctoral titulada "Medidas objetivas de localización de un evento sonoro para sistemas de teleconferencia". Dicha Tesis está estructurada en una serie de capítulos que se presentan a continuación.

En el primer capítulo, de introducción, se ofrece una justificación de la Tesis y su ámbito de aplicación, además de una enumeración de los objetivos previstos.

En el segundo capítulo se presenta una visión general del marco teórico, así como del estado del arte sobre los aspectos que fundamentan la localización espacial de eventos sonoros. Estos son los tópicos en los que se centra el trabajo de investigación de esta Tesis.

En los dos siguientes capítulos, se presentan el desarrollo y los medios materiales con los que se ha contado para la realización de la Tesis Doctoral, así como los resultados obtenidos. Basándose en el estudio de los modelos psicoacústicos que analizan el sistema de audición, se ha desarrollado un Localizador de Eventos Sonoros y se ha aplicado a escenarios acústicos simulados y a escenarios acústicos reales, analizando las diferencias y la fiabilidad de la simulación. Hay que destacar que la temática con la que se relaciona este trabajo tiene carácter multidisciplinar, debido a que abarca temas generales de procesado de señal, de acústica y electroacústica y de psicoacústica. Por este motivo resulta extremadamente difícil estudiar y exponer de forma completa el marco teórico y estado del arte sobre la investigación de la Tesis Doctoral, por lo que se ha preferido exponer solamente los temas clave que aplican a la localización espacial en el plano horizontal y en el plano medio. Como existe una clara diferencia entre los mecanismos que utiliza el sistema auditivo en el plano horizontal y en el plano medio para determinar la dirección del sonido, el tercer capítulo se dedica al proceso de localización en el plano horizontal y el cuarto capítulo al proceso de localización en el plano medio.

En el capítulo quinto se refieren las conclusiones y aportaciones destacables de la investigación y las líneas propuestas de trabajo para investigaciones futuras.

El último capítulo está dedicado a la bibliografía y a las publicaciones ya realizadas.

Para terminar, como producto final de esta Tesis, queda una herramienta práctica que se puede utilizar para estudiar, mejorar y desarrollar sistemas de reproducción de audio en los que se quiera analizar y valorar la localización de los eventos sonoros reproducidos.

ABSTRACT

The aim of this Doctoral Thesis is to implement a tool for evaluating the direction of sound events (on the horizontal plane and on the median plane), named LSE (Localization of Sound Events). LSE mimics the auditory system that a person uses for localizing sound. LSE algorithm is able to localize sound event on the horizontal plane using the binaural signal and on the median plane using the monaural signal. Moreover, LSE can simulate some acoustic configurations such as the virtual acoustic opening.

Mainly, LSE is a Matlab application developed for simulating a virtual acoustic opening configuration, which is used for teleconference. The virtual acoustic opening is implemented by using the principles of Wave Field Synthesis. The quality of reproduced sound event is very important in multichannel systems.

The subjective quality of the sound reproduced by a virtual acoustic opening must be evaluated. Performing tests using listeners have a high cost. Therefore, there is an increasing need to formulate objective measurements that model the human auditory perception. Such measurements should combine the best of both approaches: the relevance of listening tests and the efficiency as well as repeatability of the objective measurements. Such auditory perception or psychoacoustic models are becoming an important tool for audio quality evaluations. The direction of the sound perceived by a listener is an important cue when a multichannel audio system is evaluated. Accurate localization of the sound sources depends on many cues that are related to the nature of the sound, the anthropometry and hearing characteristics of the listener, the voluntary or involuntary motion of source or of listener, and the physical environment in which the listener is immersed. While some of these cues are valuables for accurate localization, others, such as echoes and reverberation, are often detrimental.

In a virtual acoustic opening application there are two acoustic spaces connected by a multichannel audio communication system. Inside the emitting room, the sound source position is variable and there is a microphone array that receives the sound field. This field is synthesized according to Huygens Principle into the receiving room by a loudspeaker array (WFS technique). Both arrays have the same number of transducers placed at the same position. Into the receiving room, the listener can be located at any place. The system tries to replicate the sound as if there was a physical opening in the wall between the two rooms. Listener will localize a sound event reproduced by the loudspeaker array. The direction of the sound perceived must be the same as if sound comes through a window made in the wall. It is important to differentiate between the position from the sound sources, real physical location, and the position of the sound event that reproduces such sources, subjective localization.

Other systems of multichannel reproduction, as the stereo or surround systems, can only reproduced a correct sound image in a limited zone of the room, called “sweet spot”.

Therefore the applications based on the concept of virtual acoustic opening provide clear improvements on a set of attributes often related to the concept of space quality: location, depth and extension. Specifically, one of those subjective attributes, the location of sound event, is the one that is investigated in this Thesis. Moreover the limitations of the virtual acoustic opening for reproducing the localization cues are presented.

The localization has been divided into two subjects: the horizontal plane and the median plane. On horizontal plane there are many investigations based on the binaural signal from the ears. On median plane few studies has been made, since the location in this plane only analyzes one signal (monaural signal).

Next the Doctoral Thesis titled “Objective measurements of sound event localization event for teleconference systems” is presented. This Thesis is organized in the following chapters. In the first chapter, Introduction, the reason of the Thesis and its appliance scope are presented. In addition, the aims are enumerated. In the second chapter, general view of theoretical subjects, as well as recent researches on spatial localization of sound events, are studied.

In third and fourth chapters, the own research and results are presented. Moreover, the methods and systems of measurements are described. “Localization of Sound Event” has been developed in based to psychoacoustics models of auditory system. This tool has been applied to virtual acoustic scenes and to real acoustics scenes for testing the accurate of the models and simulations. This Thesis has a multidisciplinary scope: psychoacoustics, electro-acoustical and processing signal aspects. Each subject could be deeply studied, but only the key factors on spatial localization are put forward. Third chapter is dedicated to localization on horizontal plane and fourth chapter to localization on median plane.

In the fifth chapter, the main conclusions and contributions are presented, moreover the futures works.

Last chapter is dedicated to the referred bibliography and to the publications made by the author.

Relación de acrónimos y símbolos utilizados

ACRÓNIMOS

AI	Articulation Index.
DTS	Digital Theatre Sound.
ECM	Error Cuadrático Medio.
ERB	Equivalent Rectangular Band.
fft	fast Fourier transform.
GTFB	Gammatone Filter Bank.
HATS	Head and Torso Simulator
hrir	head-related impulse response.
HRTF	Head Related Transfer Function.
IACC	Interaural Cross Correlation.
IC	Interaural Coherency
ifft	inverse fast Fourier transform.
ILD	Interaural Level Difference.
IPD	Interaural Phase Difference.
ITD	Interaural Time Difference.
ITU	International Telecommunication Union.
JAES	Journal of Audio Engineering Society.
KEMAR	Knowles Electronics Manikin for Auditory Research.
MLS	Maximum Length Sequences.
MOS	Mean Opinion Score.
OEM	Oído externo y medio.
PCA	Principle Component Analysis.
PRTF	Pinna Related Transfer Function.
LES	Localizador de Eventos Sonoros.
STFT	Short Time Fourier Transform.
STI	Speech Transmission Index
WFS	Wave Field Synthesis.

SÍMBOLOS

θ	Azimut, ángulo de llegada del sonido en el plano horizontal.
φ	Elevación, ángulo de llegada del sonido en el plano medio.
c	Velocidad del sonido en el aire, 343 m/s.
d_{alt}	Distancia entre altavoces del array.
$\theta_{max,M}$	Componente angular máxima del campo sonoro recibido por el array de micrófonos.
$\theta_{max,L}$	Componente angular máxima radiada por el array de altavoces.
a	Radio de la cabeza.
h_L, h_R	Respuesta impulsiva referida a la cabeza (hrir) del oído izquierdo y derecho.
H_L, H_R	Función de transferencia de la cabeza (HRTF), del oído izquierdo y derecho.
x_L, x_R	Señal biaural: señal del oído izquierdo y derecho.
f_c	Frecuencia central de los filtros.
D	Diámetro de la cabeza: 0.152 m.
r	Distancia entre la fuente y el oyente.
T_g	Retardo de grupo.
τ_i	Retardo de la señal de cada micrófono del array.
θ_{max}	Apertura acústica del array de altavoces.

Capítulo 1: INTRODUCCIÓN Y OBJETIVOS

1.1 INTRODUCCIÓN

En este capítulo se ofrece una justificación de la Tesis y su ámbito de aplicación.

Con este trabajo se pretende implementar una herramienta de medida objetiva que proporcione una valoración de la calidad de reproducción de los sistemas de audio multicanal en uno de los aspectos más importantes: la localización espacial del evento sonoro (que es una apreciación subjetiva). En concreto, para sistemas como los de teleconferencia, es importante que el oyente en la sala receptora localice adecuadamente el evento sonoro cuando en la pantalla se proyecta la imagen de diversos locutores y se reproduce el sonido con múltiples fuentes sonoras. La aplicación principal de la herramienta serán los sistemas de teleconferencia aunque una vez implementada y validada podrá ser utilizada en cualquier sistema de audio multicanal. Por lo tanto, el trabajo se divide en dos partes: la primera será la creación de un entorno simulado donde se podrá probar la herramienta para diferentes configuraciones de sistemas de teleconferencia, y la segunda el estudio, implementación y mejora de la propia herramienta de estimación del ángulo de llegada del evento sonoro que simule el comportamiento del sistema auditivo humano. Para realizar esto último, es necesario estudiar el sistema de percepción humana desde un punto de vista físico, pasando por los niveles fisiológico, neuronal y también cognitivo.

Es importante diferenciar entre localización de la fuente sonora, lugar físico donde está el foco emisor de las ondas acústicas; y localización del evento sonoro, lugar donde nuestro sistema auditivo “siente” que está la fuente sonora. Estos dos lugares no tienen porque coincidir, como se verá más adelante. Esta diferencia entre la fuente real y la fuente virtual es muy importante cuando se trata de temas como audio tridimensional, técnicas biaurales y “auralización”.

1.2 SISTEMAS DE TELECONFERENCIA O TELEPRESENCIA

La mayoría de los sistemas de telecomunicación están pensados para establecer comunicaciones entre dos personas, cada una en uno de los extremos del canal establecido. Actualmente, y también como tendencia en el futuro, existe la necesidad de contar con sistemas de comunicación de buena calidad que permitan establecer canales entre varias personas de forma simultánea. La teleconferencia o videoconferencia es el conjunto de técnicas y sistemas que proporcionan una comunicación entre dos grupos de personas situados en dos salas diferentes. La videoconferencia, principalmente utilizada en organizaciones grandes, consiste en un sistema donde el vídeo de una cámara y el audio de

un micrófono situados en una sala, se reproducen en una pantalla y un altavoz en la sala situada en el otro extremo del canal de comunicación. En la actualidad se está investigando en la mejora de estos sistemas de forma que dicha reproducción suponga una inmersión total de las personas en una sala virtual común, en la cual la separación sea una ventana acústica virtual [Harma 02].

1.2.1 SISTEMAS DE SONIDO ESPACIAL

La historia de los sistemas de reproducción espacial de audio se remonta prácticamente a las primeras décadas del siglo XX. Desde los días del audio monocanal hasta los sistemas de reproducción de audio en los cines actuales se ha recorrido un largo camino. A principio de los años 30, se describió los cimientos del sistema estéreo, que puede considerarse como el primer sistema de audio espacial. En aquella época la posibilidad de crear una fuente fantasma, es decir la capacidad de situar los sonidos subjetivamente en posiciones donde no existía ningún altavoz, suponía todo un avance sobre los sistemas monofónicos existentes. Parecía evidente que la calidad espacial estaba de alguna forma relacionada con el número de canales o altavoces del sistema y se sucedieron numerosos ensayos que intentaron mejorar la estructura del sistema estéreo. Desafortunadamente la solución no era tan sencilla, y en los años 50 se demostró que el hecho de añadir más altavoces no mejoraba significativamente la calidad y desde luego no justificaba el coste asociado a emplear un mayor número de canales. En realidad aquellos pioneros tropezaron con el problema fundamental asociado al audio multicanal y es que no parece sencillo capturar un campo sonoro tridimensional empleando transductores de dimensiones puntuales (generalmente los micrófonos y altavoces se consideran puntos en el espacio). Los años 70 vieron el nacimiento de los sistemas de cuadrafonía que añadían un par de canales, además del estéreo convencional, y que debido a los pobres resultados obtenidos fueron rápidamente abandonados. Así pues, el estéreo se convirtió en el sistema de reproducción más utilizado en el mundo hasta nuestros días.

Hasta hace pocos años no aparecieron los nuevos sistemas de reproducción que mejoran la calidad espacial. Dentro de estos sistemas, los conocidos como esquemas 5.1 (puesto que se emplean 5 altavoces más uno de graves) son ampliamente utilizados por la industria y el público en general como base en los sistemas de reproducción de cine domésticos, pero no son los únicos. La reproducción biaural, el sistema Ambisonic o la síntesis de campos sonoros son algunos de los sistemas que intentan ir un paso más allá y despiertan el interés de la comunidad científica. A continuación se presentan brevemente algunos de estos sistemas, sus características, sus ventajas y sus inconvenientes.

- **Sistema estéreo**: El sistema estéreo es el primer sistema comercial de sonido espacial que tuvo un éxito masivo en la industria de consumo. Aún hoy en día es el sistema de reproducción más empleado debido a su sencillez y efectividad. Se utilizan dos altavoces, uno a cada lado del oyente (L: canal izquierdo y R: canal de derecho) formando un triángulo equilátero en su disposición ideal, tal y como se ve en la Figura 1.

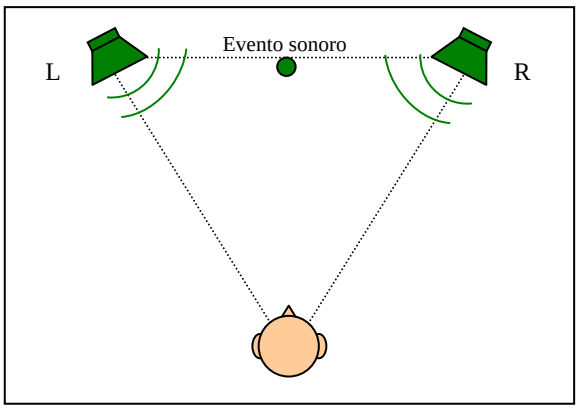


Figura 1. Sistema de reproducción estéreo.

- **Sistema 5.1**: Los sistemas tipo 5.1, entre ellos los sistemas Dolby y DTS (Digital Theater Sound), en un primer momento fueron introducidos sólo en las salas cinematográficas aunque poco a poco, en los últimos años han llegado al mercado de consumo. Este tipo de sistemas constituyen una auténtica familia, puesto que el sistema original ha ido evolucionando con el tiempo dando lugar a esquemas mejorados (Dolby Pro Logic, Dolby Digital) aunque básicamente todos comparten el mismo funcionamiento. Se suelen emplear seis altavoces (Figura 2: L: canal izquierdo, R: canal derecho, C: canal central, Ls: canal trasero izquierdo, Rs: canal trasero derecho y S: canal de bajos) para proporcionar al oyente una sensación envolvente de sonido [ITU Rec. BS.775-1]. Así, por ejemplo, empleando uno de los canales traseros es posible hacer creer al oyente que un helicóptero proviene de uno de los laterales de la pantalla si se emplea simultáneamente el canal frontal del mismo lado. Se busca mejorar las sensaciones percibidas por los oyentes proporcionándoles una mayor calidad espacial. Los resultados obtenidos van desde muy buenos a muy malos dependiendo del tipo de grabación realizada y de la forma de reproducción empleada. Esto está directamente relacionado con las propiedades del campo acústico y los efectos psicoacústicos reproducibles con un número limitado de altavoces. Además, al igual que sucedía en estéreo, la posición del evento sonoro es totalmente dependiente de la posición relativa oyente-altavoces. Este efecto negativo se acentúa al simultanear señales sonoras con señales visuales ya que la posición visual del hablante puede no coincidir con la posición del evento sonoro [De Bruijn 03]. Por ello se suele situar un canal central encargado de reproducir los diálogos. Con posterioridad se han diseñado esquemas que emplean un mayor número de altavoces (como por ejemplo los sistemas 7.1) pero que funcionan con la misma filosofía y si bien se obtienen algunas mejoras (mayor direccionalidad y una mejor sensación del movimiento), lo cierto es que no aportan ventajas realmente significativas.

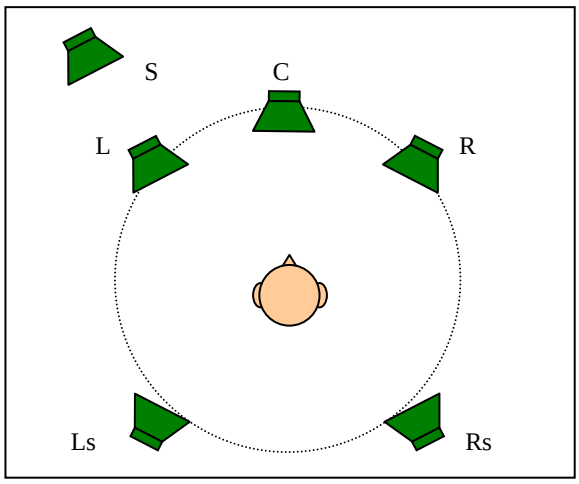


Figura 2. Sistema de reproducción 5.1.

- **Sistemas biaurales**: La reproducción biaural es un concepto antiguo pensada inicialmente para una reproducción empleando auriculares. El sistema consta de dos fases principales. En la primera se realiza una grabación mediante un maniquí acústico (Figura 3) que intenta simular la cabeza de una persona, sus orejas, el conducto auditivo, etc. Este maniquí incorpora dos pequeños micrófonos en sus oídos que registran la señal. En una segunda fase se reproduce esta grabación mediante auriculares en los oídos. De esta forma se volverán a reproducir en el oyente las mismas señales sonoras que se grabaron en el maniquí.

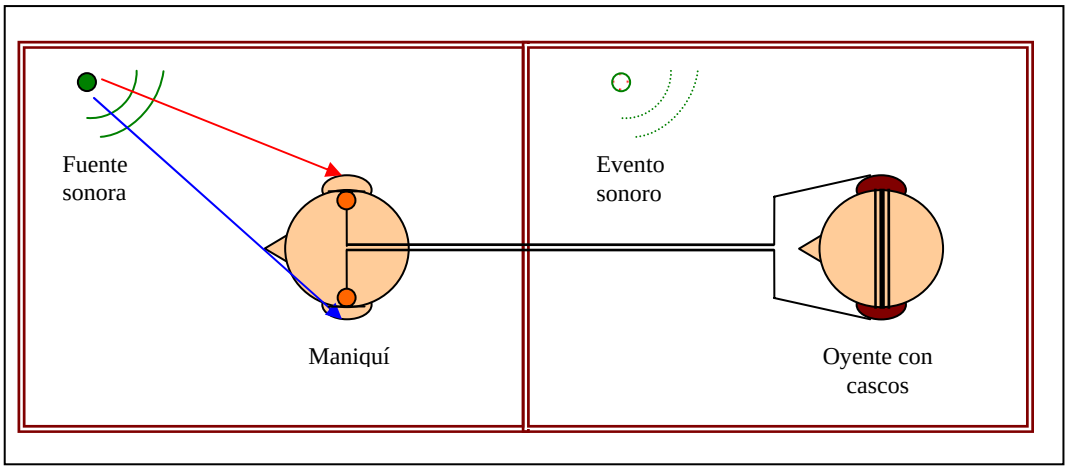


Figura 3. Sistema de reproducción biaural.

La ventaja frente a los sistemas basados en altavoces radica en que no existe el "sweet spot" donde se debe colocar siempre al oyente para recrear con fidelidad la sensación espacial, sino que el evento sonoro se mueve con el oyente. A pesar de este hecho, este sistema no disfruta de mucha popularidad debido a que el proceso de grabación previo está restringido a una sala en concreto y a un maniquí que en muchas ocasiones posee un pabellón auditivo que difiere del de la persona que va a escuchar (hecho que afecta a la capacidad de ubicación de los eventos sonoros). El segundo inconveniente está relacionado con la rigidez inherente al sistema, puesto que a diferencia de lo que sucede en la realidad, al emplear un sistema biaural si se realiza un movimiento de la cabeza la escena sonora no se ve alterada.

Dejando a un lado la reproducción biaural, lo cierto es que tanto los sistemas estéreo como los sistemas de sonido envolvente tipo 5.1 adolecen de un conjunto de problemas importantes, entre otros:

- La posición de los altavoces es muy estricta y cualquier modificación de la misma altera considerablemente el campo sonoro generado.
- Sólo se obtiene una buena sensación espacial en una pequeña porción de la sala (lo que se denomina “sweet spot”).
- Es posible situar eventos sonoros en los altavoces o en cualquier punto de la línea que los une, pero no en un punto intermedio entre los altavoces y la posición de escucha.
- Por supuesto no existe una tercera dimensión, es decir, los eventos sonoros se sitúan siempre en el plano horizontal que forman los altavoces y la cabeza del oyente.

1.2.1.1 WAVE FIELD SYNTHESIS

Este sistema de reproducción de audio pertenece a un conjunto conocido como sistemas holofónicos que tratan justamente de reconstruir un campo sonoro en un volumen a base de altavoces situados en su periferia. Dentro de estos sistemas holofónicos la síntesis de campos sonoros o Wave Field Synthesis (WFS) está suscitando en los últimos tiempos una enorme expectación e interés dentro de la comunidad científica debido a su enorme potencial.

La síntesis de campos sonoros está basada en la reconstrucción casi perfecta de un frente de onda mediante un conjunto de fuentes secundarias colocadas en la superficie donde estaría un frente de onda de la fuente primaria que se intenta reproducir. Este sistema se basa en el Principio de Huygens y Fresnel [Blauert 97] por el cual cualquier frente de onda primario puede ser sintetizado a partir de un número elevado de fuentes secundarias. La reconstrucción del campo sonoro elimina la necesidad de colocar al oyente en un punto determinado de la sala como sucede en los sistemas estéreo o 5.1. Otra ventaja frente al audio biaural, que generalmente se realiza con cascos, es que los movimientos de la cabeza instintivos que ayudan a la mejora de la localización se pueden realizar, ya que el evento sonoro se recrea en la sala y no en nuestra mente.

Desde el punto de vista práctico, en lugar de emplear superficies de altavoces se emplean arrays que rodean una superficie plana. Así pues, la sensación sonora no se regenera dentro de un volumen sino sólo en un plano (horizontal). Yendo más allá, es posible emplear un único array lineal, lo que simplifica notablemente el sistema con la contraprestación de que aparecen ciertos efectos de borde en los límites del array. Aún así las ventajas de este esquema de funcionamiento son notables frente a los sistemas tradicionales diseñados hasta la fecha. Cabe señalar además que el evento sonoro recreado no está limitado a la distancia donde está colocado el array de altavoces sino que se pueden recrear eventos sonoros que se sitúan a mayor distancia e incluso a menor distancia mediante procesado de la señal.

Existe un efecto de aliasing espacial [Boone 95] que aparece porque el número de altavoces empleados es finito y están separados una cierta distancia. La frecuencia de Nyquist espacial a partir de la cual aparece el efecto aliasing, y por tanto la reconstrucción del campo sonoro no es tan buena, viene dada por [De Brujin 03]:

$$f_{max} = \frac{c}{d_{alt}\left(\sin\theta_{max,M} + \sin\theta_{max,L}\right)} \qquad \text{Ec. 1}$$

Donde c es la velocidad del sonido en el aire, d_{alt} es la distancia entre altavoces, $\theta_{max,M}$ es la componente angular máxima del campo sonoro recibido por el array de micrófonos y $\theta_{max,L}$ es la componente angular máxima radiada por el array de altavoces. Por ejemplo, en el array construido para esta Tesis, la frecuencia de Nyquist está entre 2027 Hz y 6524 Hz [Anexo 1-1]. Por debajo de esta frecuencia el campo sonoro reproducido es idéntico al muestreado por el array de micrófonos. Por encima de la frecuencia de Nyquist la radiación de cada altavoz se puede identificar como una contribución independiente en lugar de una interferencia constructiva y por lo tanto el campo acústico no se reproduce perfectamente, aunque este hecho no parece que suponga un decremento drástico de la calidad para la mayor parte de las aplicaciones [Boone 95].

1.2.2 VENTANA ACÚSTICA VIRTUAL (VIRTUAL ACOUSTIC OPENING)

La ventana acústica virtual es un sistema de comunicación de audio multicanal. Se compone de un array de micrófonos en la sala emisora y otro array de altavoces en la sala receptora (Figura 4). El objetivo es proporcionar a los oyentes en la sala receptora la impresión de que hay una apertura o ventana en la pared que separa las dos salas. Por ello, este sistema es capaz de proporcionar una extraordinaria calidad espacial. Esto se lleva a cabo con un elevado número de canales y usando las técnicas de síntesis del campo sonoro. La señal que capta cada micrófono es la señal de la fuente sonora, retardada y atenuada según la ley de divergencia esférica [Blauert 97] suponiendo un entorno anecoico. El problema está en como transmitir y codificar todos estos canales tan correlacionados pero diferentes con un régimen binario no muy alto [Torres 03].

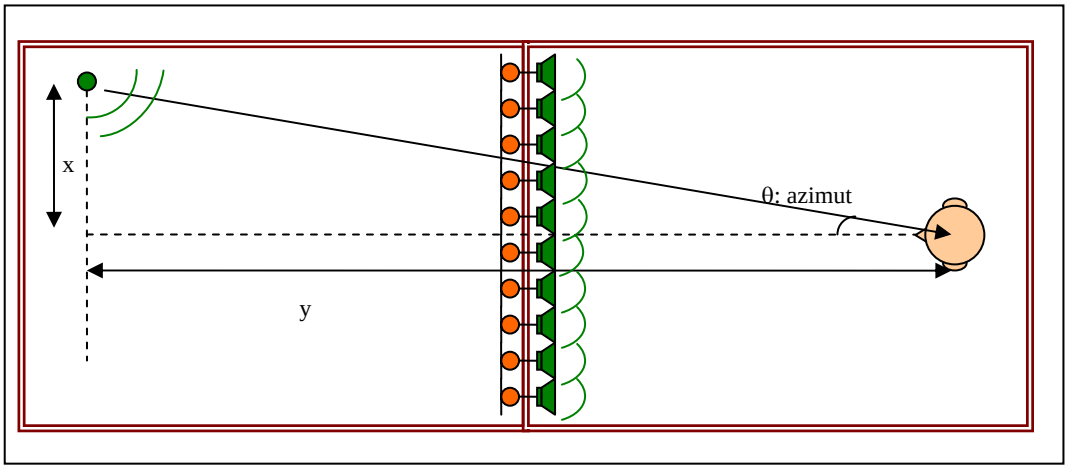


Figura 4. Ventana acústica virtual: a pesar de existir un array de fuentes sonoras emitiendo simultáneamente, la localización del evento sonoro se produce en la dirección que tendría la fuente sonora original si no hubiera separación entre las dos salas.

Las técnicas de codificación buscan una disminución en el régimen binario sin disminuir la calidad subjetiva del oyente. Para evaluar diferentes tipos de codificaciones o en general, diferentes tipos de sistemas de captura, transmisión y reproducción se pueden utilizar medidas objetivas o subjetivas.

En resumen, a través del WFS se intentará recrear el campo sonoro de una sala emisora en una sala receptora manteniendo la calidad del sonido percibido de forma que parezca que hay una ventana física de separación entre ambas salas. Este hecho se basa en la propiedad de localización sumatoria del sistema auditivo. La localización sumatoria se produce cuando las señales radiadas por diferentes altavoces difieren ligeramente en nivel o en tiempo (menor de 1 ms), efecto por el cual sólo existe un evento sonoro, localizado según las diferencias de nivel y tiempo de las señales que llegan a los oídos [Blauert 97], siempre que hablemos del plano horizontal.

A menudo se denomina a la ventana acústica virtual como un sistema WFS de conexión dura, o hard-wired, puesto que no se crean eventos sonoros virtuales ni se procesa la señal, sino que el campo sonoro capturado en la sala de emisión tal cual es reproducido directamente en la sala de recepción. Esto implica que las señales capturadas por el array de micrófonos atacan directamente al array de altavoces en una conexión, virtual o real, de uno a uno.

1.3 CALIDAD SONORA

La calidad del sonido, y en concreto la calidad espacial, es un término ciertamente ambiguo puesto que hace referencia a un conjunto de atributos que en muchas ocasiones son difíciles de medir y evaluar.

Las medidas objetivas vienen cualificando la reproducción del sonido desde el punto de vista eléctrico de la señal: distorsión, relación señal-ruido,... Para evaluar la calidad subjetiva se pueden analizar numerosos parámetros mediante estudios con sujetos reales y técnicas de análisis estadísticos. Las medidas subjetivas cualifican la señal percibida por un grupo de oyentes, valorando aspectos como la localización o difusión, claridad, inteligibilidad,... [Berg 03] [Becker 01] [ITU Recommendation P.800 y P.800.1]. Este tipo de medidas es muy caro y costoso en tiempo porque supone utilizar a personas a las que hay que entrenar para cualificar los sistemas. Por lo tanto, se hace necesario desarrollar técnicas objetivas que evalúen la calidad percibida cuyos resultados estén verificados a través de estudios subjetivos paralelos.

En la actualidad se han desarrollado métodos de medida objetivos cuyos resultados tienen equivalencia directa con resultados de medidas subjetivas, pudiéndose calificar los sistemas de audio desde el punto de vista de la percepción humana con medidas eléctricas realizadas sobre la señal de audio. Entre ellos está la [ITU Recommendation BS.1387]: “Method for objective measurements of perceived audio quality”, donde se valora la calidad percibida de la señal de audio según un índice que tiene equivalencia directa con parámetros de calidad subjetiva. También se realizan medidas de inteligibilidad de la palabra mediante equipos de medida comerciales que proporcionan, por ejemplo, el Índice de Articulación (AI, Articulation Index) o el Índice de Transmisión de Voz (STI, Speech Transmission Index).

En concreto, el audio multicanal intenta recrear la sensación de “espacialidad”, de sonido envolvente. Por ello posee un conjunto de propiedades que entran dentro del ámbito de lo subjetivo como son: inteligibilidad, presencia de una composición, localización espacial de los eventos sonoros o profundidad. La estimación de estas propiedades es ciertamente difícil y por tanto el desarrollo de herramientas para evaluarlas se complica y abre todo un campo por explorar. A pesar de ser un tema de gran actualidad, lo cierto es que aún no existe dentro de la comunidad científica un consenso sobre el significado que conlleva el término calidad espacial.

No obstante, bajo el contexto de una ventana acústica virtual, el atributo más importante y sobre el que se ha centrado el trabajo de esta Tesis es el de localización espacial de los eventos sonoros reproducidos en los planos horizontal y medio. Es decir, la capacidad del sistema de reproducir virtualmente las fuentes en la dirección que se percibirían si la ventana acústica fuera real. Dado que la principal dirección de localización de las fuentes en este sistema es horizontal, y que es necesario localizar los interlocutores en este eje, se van a estudiar los parámetros de localización en el plano horizontal, estimándose el azimut de la fuente virtual reproducida. Este es el primer y principal objetivo de la Tesis, partiendo de la teoría de localización muy conocida pasamos a la implementación de una herramienta que hace uso de esa teoría y simula algo tan desconocido como el proceso neurológico del sistema auditivo.

Una vez conseguido el anterior objetivo, se ha añadido un segundo objetivo que es estudiar los parámetros de localización en el plano medio para poder estimar el ángulo de elevación de los eventos sonoros. Este plano menos estudiado, y a veces olvidado, permitiría recrear la "espacialidad" sonora de una forma más completa, consiguiendo por ejemplo despegar el sonido de la horizontal y colocarlo en un punto cenital.

1.3.1 EVALUACIÓN SUBJETIVA DE LA CALIDAD

Existen dos grandes filosofías a la hora de realizar pruebas para determinar la calidad de un sistema. Generalmente, se suele optar por métodos subjetivos los cuales emplean un conjunto de personas, con una sensibilidad auditiva especialmente entrenada, a las cuales se les realiza las preguntas apropiadas tras oír un conjunto de sonidos. Este método es largo y costoso y difícilmente repetible ya que precisa de un número de personas relativamente elevado y por ello, muchas veces, se suele optar por otros sistemas que no necesitan el uso de voluntarios y que emplean algún tipo de medida objetiva.

Las condiciones para este tipo de experimentos están perfectamente regladas ya que existen recomendaciones que especifican estrictas condiciones a seguir para realizar este tipo de experimentos y aun a día de hoy se sigue trabajando en la normalización de este tipo de medidas [Miyasaka 91]. Así pues la recomendación [ITU-T Recommendation P.800]: "Methods for subjective determination of transmission quality" presenta los métodos para la determinación subjetiva de la calidad en sistemas de transmisión de señales de voz. Para determinar la calidad de audio en general, incluyendo los sistemas multicanal, la recomendación [ITU-R Recommendation BS.1116]: "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems" establece las condiciones a seguir. En general este tipo de recomendaciones devuelven un índice conocido como MOS (Mean Opinion Score, [ITU Recommendation P.800.1]: "Mean Opinion Score (MOS) terminology"). La evaluación subjetiva, siempre que siga las recomendaciones, es una herramienta muy potente para determinar el buen comportamiento de los sistemas desarrollados pero tiene el gran inconveniente de que su puesta en marcha es tremendamente costosa ya que necesita de un gran número de voluntarios (se suele considerar que a partir de los 50 individuos los resultados estadísticos pueden ser considerados como correctos), que además tienen que cumplir unas condiciones (sensibilidad auditiva) especiales. Por ello se suele recurrir a medidas objetivas y únicamente se emplean los métodos subjetivos cuando el desarrollo del sistema está muy avanzado y se pretende su validación final.

1.3.2 EVALUACIÓN OBJETIVA DE LA CALIDAD

Las medidas de tipo objetivo no emplean voluntarios humanos sino que tratan de encontrar algún parámetro, que esté relacionado con la calidad, y medirlo o calcularlo. Hay que tener en cuenta que el concepto de calidad es subjetivo, y por tanto la relación entre los parámetros medibles y la calidad subjetiva no es directa.

Se están investigando diferentes parámetros de la calidad subjetiva que pueden ser medidos objetivamente a partir de las señales de audio. Como ejemplo de este esfuerzo investigador se desarrolló la [ITU Recommendation BS.1387]: "Method for objective measurements of perceived audio quality". Esta norma evalúa señales mono-canal comparando el grado de similitud entre una señal de referencia y la misma señal ya procesada. También sirve para la evaluación de señales estéreo pero a través de la media aritmética de los resultados de las señales mono. Además, la valoración de la calidad que proporciona está correlacionada con la valoración subjetiva que tendría un grupo de oyentes, quedando por lo tanto validado el método.

En esta Tesis se utilizarán métodos de medida objetivos que evalúan la posición de la fuente sonora virtual o evento sonoro, partiendo de los parámetros utilizados por el cerebro en la localización espacial. La relación subjetiva entre estos parámetros y la localización se ha estudiado profundamente durante más de 50 años, como muestra el libro de [Blauert 97] "Spatial hearing". El cerebro analiza la señal recibida en cada uno de los oídos (señal biaural) y determina la posición, no sólo con parámetros que se pueden medir en estas señales, sino con parámetros psicológicos, totalmente subjetivos, que están dentro del campo cognitivo. Por lo tanto, quedando patente la profunda complejidad de intentar evaluar una sensación subjetiva, se procesará la señal biaural para estimar el ángulo en el plano horizontal de una fuente sonora virtual reproducida, o de forma general, de una fuente sonora. Además se procesará la señal monoaural para estimar el ángulo en el plano medio. Puesto que en este plano la señal biaural es idéntica para los dos oídos, la información de elevación sólo se puede obtener de una única señal que es la señal monoaural.

En resumen, se determinará la calidad espacial del audio reproducido por un sistema multicanal, comparando la estimación del ángulo "percibido" y el ángulo que se quiere reproducir, así como evaluando la capacidad de reproducir los parámetros de localización que utiliza el sistema auditivo en la estimación de la dirección del sonido.

1.4 APLICACIONES

A continuación se va a diferenciar entre las aplicaciones de la ventana acústica virtual y las aplicaciones de la herramienta de localización de eventos sonoros.

Las particulares propiedades de la ventana acústica virtual pueden ser utilizadas en un gran número de escenarios acústicos entre los cuales cabe destacar los siguientes:

- **Salas de conciertos**: los sistemas empleados en la actualidad para amplificar las actuaciones en vivo destruyen la escena sonora creando una disparidad entre lo que los oyentes ven y por donde les está llegando el sonido. La ventana acústica virtual permitiría un refuerzo de la señal mucho menos agresivo para la "espacialidad" del sistema.
- **Eventos al aire libre**: uno de los mayores problemas a los que se enfrentan los eventos al aire libre es la necesidad de crear una presión sonora lo suficientemente

elevada para todo el estadio o recinto sin que ello suponga que los espectadores de las primeras filas tengan que soportar niveles peligrosos para la salud. Evidentemente es posible colocar altavoces a medio camino dentro del recinto pero, a menudo, ello conlleva una pérdida de la coherencia espacial. De nuevo la ventana acústica virtual podría aliviar estos problemas de forma significativa.

- **Teleconferencia**: quizás la más interesante y directa de las aplicaciones posibles para la ventana acústica virtual. Hasta la fecha la mayor parte de los sistemas de teleconferencia emplean esquemas mono, o a lo sumo estéreo, para la captación y reproducción del audio. Existen situaciones donde el número de participantes es elevado y donde puede darse el caso de tener a varias personas hablando simultáneamente. Los sistemas mono y estéreo sufren en estas situaciones puesto que se produce el conocido efecto gallinero, que reduce significativamente la inteligibilidad global del sistema. Se ha demostrado que el aumento de "espacialidad" conlleva un aumento de los niveles de inteligibilidad, reduciéndose de forma notable dicho efecto gallinero [Beracoechea 07].

Las aplicaciones anteriores son de la ventana acústica virtual, en las cuales valorar la capacidad del sistema de reproducir la "espacialidad" del campo sonoro reproducido es muy importante.

Por otra parte las aplicaciones de la herramienta desarrollada en la Tesis serían todas aquellas donde se necesite analizar y evaluar la localización de los eventos sonoros representados por sistemas multicanal o directamente como herramienta de localización de eventos sonoros de fuentes únicas. Además de estimar el ángulo del evento sonoro, se puede hacer estudios de los diferentes parámetros de localización en frecuencia, analizando por ejemplo como afectan diferentes tipos de procesado o diferentes escenarios acústicos, configuraciones de altavoces, etc.

1.5 OBJETIVOS

En este apartado se exponen los objetivos previstos de la Tesis exponiendo sus alcances y las tareas que se han realizado para lograrlos.

1.5.1 OBJETIVOS PREVISTOS

El trabajo de esta Tesis se puede dividir en dos partes. Una primera que consiste en investigar y desarrollar una herramienta de medida objetiva que estime el ángulo de azimut percibido por un oyente en el plano horizontal y el ángulo de elevación en el plano medio. Ello permitirá comprobar si los sistemas de reproducción de audio preservan la información de localización espacial, o de una forma más general, comprobar cuál sería la dirección percibida por un sujeto en presencia de un campo sonoro partiendo de la señal biaural captada por un maniquí acústico. La gran ventaja es poder medir una sensación subjetiva a través de un maniquí acústico, al cual se le puede someter a medidas repetitivas sin cansancio, sin entrenamiento,… Será necesario por lo tanto comprobar el grado de acierto de las estimaciones con relación a las posiciones reales de la fuente, así como comprobar que esta capacidad de resolución coincide con la capacidad subjetiva de los oyentes establecidas en los numerosos estudios. A esta herramienta la vamos a llamar LES (Localizador de Eventos Sonoros). El alcance del localizador se limita al plano horizontal y al plano medio.

La segunda parte de la Tesis consiste en desarrollar una aplicación donde se pueda simular un entorno de ventana acústica virtual con un array de altavoces y un array de micrófonos, ambos configurables, que sirva para evaluar la capacidad de estos sistemas multicanal de preservar la información de localización espacial. Esta información puede depender de parámetros de transmisión como por ejemplo: tipo de codificación, régimen binario, número de canales, etc. Dicha evaluación se realizará con el LES. Dentro de esta simulación hay que sintetizar la señal recibida por un oyente, señal biaural. La señal biaural se calcula a partir de las respuestas en frecuencia de cada uno de los oídos para un ángulo de incidencia del sonido determinado. Existen bases de datos disponibles para el mundo científico, p.e. [Gardner 00a] [Algazi 01], con la función de transferencia de la cabeza (HRTF, Head Related Transfer Function) de diversos maniquís y personas. La creación de estas bases de datos es muy costosa en tiempo e infraestructura, sin embargo, se ha generado una nueva base de datos para el maniquí utilizado en esta Tesis de forma que se pueda comprobar la influencia de las bases de datos en los resultados. Para poder medir las HRTF se ha desarrollado un nuevo método de medida, diferente a los utilizados en otras bases, y que ha venido determinado por el equipamiento de medida, consiguiendo una total independencia de la HRTF que se quiere medir del resto de las funciones de transferencia involucradas en la medida.

También es necesario que el entorno simulado se implemente en la realidad para comprobar la fiabilidad del simulador. Por lo tanto es imprescindible disponer de un maniquí acústico que se coloque en un escenario acústico y se grabe la señal biaural de sus oídos. Posteriormente se aplicará la herramienta LES para comparar los resultados entre un entorno simulado y un entorno real.

En resumen, los objetivos que se ha pretendido cubrir en la Tesis son los siguientes:

- Parte I: Localizador de Eventos Sonoros, LES:
    - Plasmación del estado del arte de localización espacial.
    - Estudio de los métodos para localización biaural existentes y su implementación.
    - Estudio de los métodos para localización monoaural existentes y su implementación.
    - Mejora de los métodos existentes mediante la conjunción de la información obtenida a través de los diferentes parámetros de localización y mediante la revisión profunda del modelo de percepción auditiva.

Todo lo anterior lleva a la creación de la herramienta que estima el azimut en el plano horizontal y la elevación en el plano medio: LES.

Una vez creada la herramienta, el siguiente paso es realizar pruebas de validación con simulación y sin simulación.

- Parte II: Simulación de ventana acústica virtual:
    - Creación de una base de datos de Head Related Transfer Function (HRTF) para el maniquí acústico HATS de Head Acoustics.
    - Creación de una aplicación en Matlab para la simulación de sistemas de teleconferencia configurables basados en la ventana acústica virtual.
    - Implementación práctica de un array de altavoces para la validación de la simulación de los entornos acústicos.

- Pruebas de validación con simulación y sin simulación del LES para un sistema de teleconferencia.

Esta Tesis se ha desarrollado en paralelo con la investigación sobre codificación de audio multicanal realizada por Torres-Guijarro y Beracoechea-Alava. Como ya se ha justificado en apartados anteriores, el sistema de audio basado en una ventana acústica virtual presenta indudables ventajas, ya que no requiere la colocación del oyente en un punto determinado de la sala para recrear eventos sonoros correctamente y además permite el movimiento de la cabeza para mejorar la localización.

1.5.2 PRECEDENTES Y MARCO DE TRABAJO

Es bien sabida la conveniencia de que los trabajos de investigación se integren como parte de una tarea más amplia. El desarrollo de sistemas de teleconferencia, WFS y codificación de señales multicanal es una de las líneas de trabajo del Grupo de Aplicaciones de Procesado de Señal (GAPS, www.gaps.ssr.upm.es) del Departamento de Señales, Sistemas y Radiocomunicaciones (SSR) de la E.T.S.I. de Telecomunicación de la Universidad Politécnica de Madrid.

Como trabajo posterior a los objetivos de esta Tesis quedaría ver la influencia de la codificación de transmisión de las señales en la localización espacial en un entorno de teleconferencia. También se podría evaluar la capacidad de recreación de eventos sonoros en cuanto a su localización en sistemas multicanal tipo 5.1, 7.1, estéreo, etc. Otra línea de progreso se centra en la ampliación del alcance de la herramienta LES fuera del plano horizontal y del plano medio a cualquier punto en el espacio tridimensional.

Capítulo 2: ASPECTOS TEÓRICOS

2.1 INTRODUCCIÓN

En este capítulo se presenta una visión general del marco teórico así como el estado del arte sobre los aspectos que fundamentan la localización espacial. Hay que tener en cuenta que para realizar una correcta localización de las fuentes, el sistema auditivo se basa en un conjunto de parámetros o atributos relacionados con la naturaleza del sonido, la antropometría, las características auditivas del oyente, el movimiento de la fuente y del oyente y las propiedades acústicas del entorno. Algunos de estos atributos favorecen una correcta localización mientras que otros, particularmente los ecos y la reverberación, son perjudiciales. De hecho, las estrategias empleadas hasta la fecha que intentan modelar el sistema de percepción auditivo humano o al menos aquellos atributos del mismo relacionados con la localización de fuentes están basados en sencillos modelos biaurales y a pesar de los avances obtenidos, aún queda un largo camino por recorrer.

A continuación se exponen los aspectos teóricos utilizados en el desarrollo de la Tesis, tanto para el proceso de simulación de entornos acústicos como para el proceso de estimación del ángulo de llegada que realiza el LES. Existen dos libros que cubren bastante bien el marco teórico de la audición espacial [Begault 94] [Blauert 97], así como un informe técnico de la Universidad de San José [DUDA 00]. Para empezar se analizarán los parámetros fundamentales de la psicoacústica, para continuar con las funciones de transferencia de la cabeza, descritas muy bien en [YANG 01].

2.2 PSICOACÚSTICA DE LA LOCALIZACIÓN ESPACIAL

Los seres vivos han desarrollado diversos mecanismos para extraer del sonido información sobre su localización. Aunque todavía existen misterios en la percepción auditiva, los mecanismos principales son conocidos desde hace mucho tiempo. Estudios psicológicos extensos y profundos han establecido con cuanta precisión podemos determinar la localización de las fuentes [Blauert 97]. Cualquiera que quiera generar o analizar el sonido, desde un punto de vista espacial, tiene que conocer la influencia del sistema de auditivo humano. Un oyente es capaz de localizar la dirección de la fuente del sonido porque la onda sonora sufre un “filtrado espacial” debido a la dispersión, reflexión y difracción de las ondas en su propio torso, cabeza y orejas. Este apartado resume los principales factores que influyen en la audición espacial.

La posición de una fuente sonora o de un evento sonoro es representada mediante un sistema de coordenadas polares centrado en la cabeza del oyente. Se define el plano horizontal como el plano paralelo al suelo que pasa por los dos oídos. En este plano se define el ángulo de azimut, θ, como el desplazamiento de la fuente hacia el oído derecho

siendo θ = 0º para un fuente colocada delante, θ = 90º, colocada a la derecha y θ = 180º colocada detrás. Se define el plano medio como el plano perpendicular al suelo que pasa por la nariz y el centro de la cabeza. En este plano se define el ángulo de elevación, φ, como el desplazamiento de la fuente hacia la parte superior de la cabeza siendo φ = 0º para una fuente colocada delante, φ = 90º colocada encima de la cabeza y φ = 180º colocada detrás.

2.2.1 PARÁMETROS DE ACIMUT

Lord Rayleigh desarrolló la Teoría Dúplex [Blauert 97]. De acuerdo con esta teoría, hay dos parámetros primarios para la localización en el plano horizontal: la diferencia de tiempo interaural (ITD, Interaural Time Difference) y la diferencia de nivel interaural (ILD, Interaural Level Difference).
Lord Rayleigh dio una explicación sencilla para la ITD. Si se considera una onda sonora procedente de una fuente lejana que impacta con una cabeza esférica de radio a según una dirección determinada por el ángulo de azimut θ (ver Figura 1), claramente, el sonido llega al oído derecho antes que al izquierdo ya que debe viajar una distancia extra a(θ + senθ) hasta alcanzar este último. Dividiendo esta distancia por la velocidad del sonido, c, se obtiene una sencilla ecuación para la diferencia de tiempo interaural:

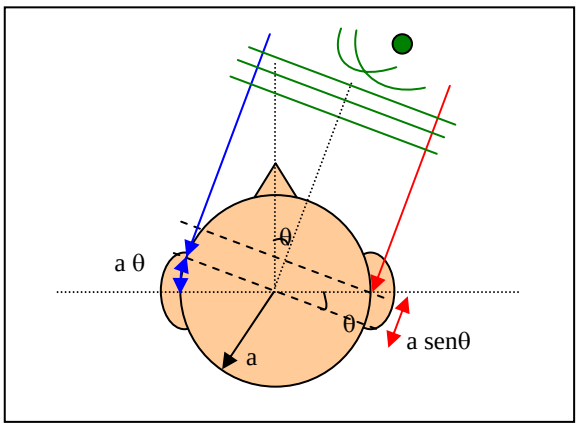


Figura 1. Cálculo de la diferencia de tiempo interaural en el caso de una cabeza esférica y propagación de ondas plana (fuente alejada).

$$ITD = \frac{a}{c}(\theta + \text{sen}\theta) \qquad -90^o \le \theta \le 90^o \qquad \text{Ec. 1}$$

Esto es, la ITD es cero cuando la fuente está directamente enfrente (θ = 0º), y es máxima, $\frac{a}{c}\left(\frac{\pi}{2}+1\right)$, cuando la fuente está justo en un lateral (θ = 90º). Esto supone una diferencia de llegada de aproximadamente 0.7 ms (para un tamaño típico de cabeza humana 17.5 cm de diámetro) que es fácilmente apreciada por el cerebro.

Obviamente, se puede invertir esta ecuación (Ec. 1) y obtener el acimut θ a partir de la ITD. El sistema auditivo realiza, aproximadamente, la función equivalente de recuperar el acimut a partir de la información dada por la ITD.

La precisión con la que se puede realizar esto depende de las circunstancias. Para señales de voz en salas con reverberación normal, la precisión humana típica es del orden de 10º a 20º. Aunque, bajo condiciones óptimas, se pueden conseguir precisiones mayores (del

orden de 1º) cuando la cuestión es decidir solamente si la fuente de sonido se mueve o no. Esto es importante, ya que significa que un cambio en el tiempo de llegada de tan sólo 10 μs es perceptible.

Finalmente, hay que observar que la ITD por si sola restringe únicamente la posición de la fuente a algún punto de acimut constante. En concreto, esto no es suficiente para determinar si la fuente está delante o detrás.

Lord Rayleigh también observó que las ondas incidentes de sonido son difractadas por la cabeza. Realmente, él resolvió la ecuación de onda para demostrar como una onda plana es difractada por una esfera rígida. Su solución demostró que además de la diferencia de tiempo existe una diferencia significativa entre los niveles de señal en los dos oídos, la ILD.

Como cabría esperar, la ILD depende mucho de la frecuencia. A frecuencias bajas, donde la longitud de onda del sonido es del orden del diámetro de la cabeza, prácticamente no hay diferencia de presión sonora en los dos oídos. Aunque, a frecuencias altas, donde la longitud de onda es pequeña puede haber una diferencia de 20 dB o más. Este hecho se denomina el efecto sombra de la cabeza, ya que un oído está en la sombra de la cabeza.

La Teoría Dúplex asevera que la ILD y la ITD son complementarias, a la hora de determinar el ángulo de incidencia del sonido. A frecuencias bajas (por debajo de 1.6 kHz), la ILD proporciona poca información, pero la ITD desplaza la señal una fracción del periodo, lo cual es fácilmente detectable. A frecuencias altas (por encima de 1.6 kHz) hay ambigüedad en la ITD, ya que supone varios periodos de desplazamiento, pero la ILD sirve para solventar esta ambigüedad en las posibles direcciones marcadas por la ITD. La Teoría Dúplex de Rayleigh dice que la ILD y la ITD juntas proporcionan información de localización para todo el rango de frecuencias. Sin embargo estos parámetros interaurales no proporcionan información para determinar la elevación de la fuente de sonido [Blauert 97].

Por lo tanto, se puede decir que el sistema auditivo analiza estos parámetros de forma dependiente con la frecuencia. Se ha comprobado con ensayos subjetivos que la ITD es analizada como el desfase de las señales portadoras por debajo de 1.6 kHz y como el desfase de las envolventes por encima de 1.6 kHz.

2.2.2 PARÁMETROS DE ELEVACIÓN

El mecanismo de localización espacial empleado por el cerebro en el plano medio es diferente al mecanismo descrito en el apartado anterior, empleado para la localización en el plano horizontal. Cuando la fuente de sonido está en el plano medio, las señales en los dos oídos son prácticamente idénticas, luego las diferencias interaurales no proporcionan información de localización. En este plano la resolución del sistema auditivo es menor y el mínimo incremento de ángulo que se aprecia (localization blur o umbral de percepción de localización), para $\varphi = 0º$ va desde 17º a 4º dependiendo del tipo de señal [Blauert 97]. Este umbral de percepción además varía con el ángulo de elevación según muestra la Figura 2.

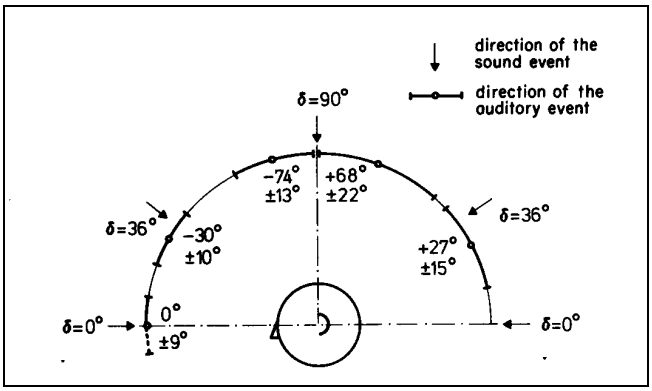


Figura 2. Localización y umbral de percepción de localización en el plano medio para voz continua de una persona familiar (Damaske 1969). [Blauert 97].

Existe una tendencia de los sonidos cortos que contienen impulsos a desplazarse a la parte posterior del plano medio. Pero cuando estas señales ya han sido presentadas con anterioridad a los oyentes, este efecto no se produce. Por lo tanto la familiaridad de la señal juega un papel importante en la localización en el plano medio. Para señales de ancho de bandas menores a 2/3 de octava no es posible su localización. No hay reglas que relacionen la dirección de la fuente sonora y la del evento sonoro, como se puede ver en la Figura 3.

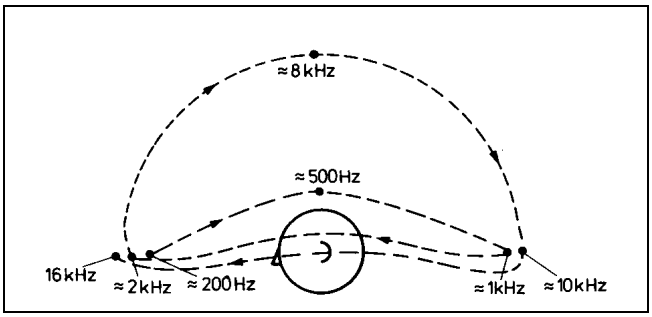


Figura 3. Camino de la dirección del evento sonoro en función de la frecuencia central de un ruido de banda estrecha para cualquier dirección de la fuente en el plano medio. [Blauert 97].

Se sabe que el ángulo de elevación influye en la respuesta en frecuencia de la HRTF [Toledo 08] [Iida 07] [Rodríguez 05a], por ejemplo existe un nodo pronunciado alrededor de los 8 kHz cuya profundidad depende del ángulo de elevación y que es debido a la influencia de la oreja. Este fenómeno se debe a que las resonancias del oído externo e interno se excitan más o menos dependiendo del ángulo de elevación. La coincidencia entre las direcciones del evento sonoro y de la fuente sonora es mayor si el sonido es de banda ancha (sobre todo si tiene componentes mayores de 4-5 kHz) y de larga duración [Ferguson 05]. Además los sonidos con mayor nivel tienden a localizarse delante y los de menor nivel detrás.

Se ha estudiado [Blauert 97] [Itoh 07] que en el dominio espectral existen unas bandas llamadas direccionales (Figura 4): cuando se excita con sonidos de 1/3 de octava y se va variando la frecuencia central del estímulo la aparición del evento sonoro va variando entre delante, detrás y arriba.

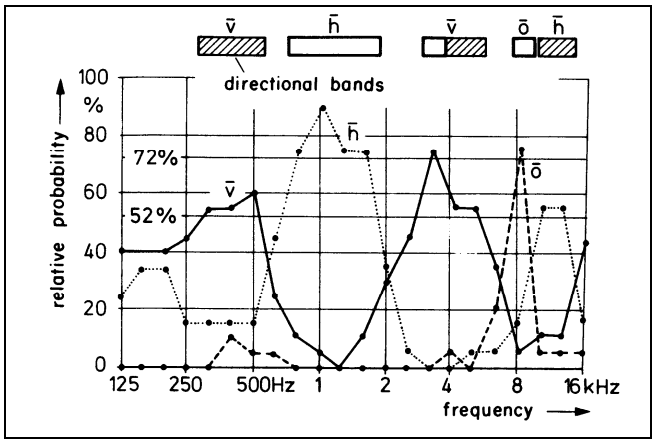


Figura 4. Probabilidad de respuesta (delante v, detrás h, arriba o) cuando se excita en tercios de octava. La zona rallada es para probabilidades relativas del 50 ± 10%. [Blauert 97].

Además, cuando se mide la diferencia de nivel para φ = 0º y φ = 180º si se utiliza una señal de nivel constante para todas las frecuencias, se ve claramente que hay unas bandas donde el nivel en los oídos es mayor para φ = 0º y otras donde el nivel es mayor para φ = 180º, estas bandas se llaman bandas de realce (Figura 5).

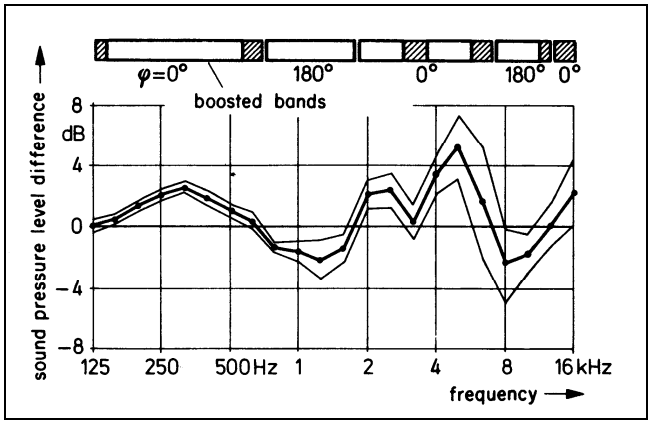


Figura 5. Diferencia de nivel entre φ = 0º y φ = 180º, y división de las bandas de realce. [Blauert 97].

Comparando las bandas direccionales y las de realce se ve que coinciden ampliamente (Figura 6), luego dependiendo de si el sonido está detrás o delante la respuesta en frecuencia del sistema auditivo realza unas bandas u otras, y el cerebro valora en que bandas direccionales llega más nivel y localiza el evento sonoro en esa dirección. Este modelo da explicación al hecho de que variando la frecuencia de un sonido de banda estrecha se pueda modificar su dirección en el plano medio.

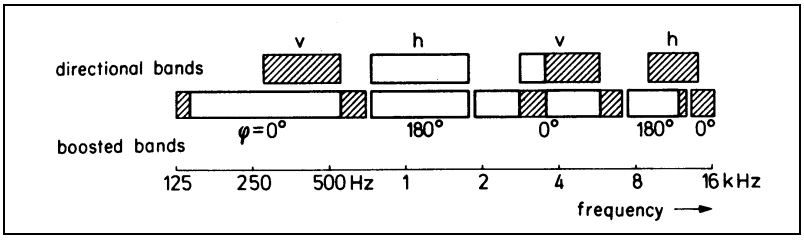


Figura 6. Coincidencia en frecuencia entre las bandas direccionales y las bandas de realce. [Blauert 97].

Se puede concluir según Blauert y Mellert [Blauert 97] que los nodos en el espectro de las señales de entrada a los oídos, así como los picos, son parámetros importantes en la estimación de la dirección del evento sonoro. En particular, el nodo que varía entre 6 y 10 kHz según el ángulo de elevación de la fuente sonora.

La influencia de la respuesta espectral de la oreja, que influye mucho en la localización vertical, es predominante en las frecuencias altas del rango de audición; por lo tanto si las señales que alcanzan a los oídos no tienen componentes altas de frecuencia, la precisión de la localización en elevación disminuye mucho. Además existe también información de elevación en baja frecuencia (alrededor de los 700 Hz) debido a la reflexión en el torso.

Esta variabilidad en la localización del evento sonoro en función de la frecuencia, cuando la fuente sonora no se mueve ha sido estudiada subjetivamente y demuestra que diferentes pasajes musicales se localizan a diferentes alturas dependiendo de su contenido espectral [Ferguson 05].

Por lo tanto, como parámetros de localización en elevación se podría hablar de la función de transferencia de la oreja (Pinna-Related Transfer Function, PRTF). Si se extrae esta información de las HRTF y se la caracteriza según algún parámetro en función del ángulo de elevación, se podría estimar el ángulo de llegada.

2.2.3 PARÁMETROS DE AZIMUT VS. PARÁMETROS DE ELEVACIÓN

Para ver como influyen unos parámetros en los otros pongamos dos ejemplos.

La fuente de sonido está en $\theta = 0^o$ y $\varphi = 0^o$ (delante). La cabeza gira hasta que la fuente se coloca a $\theta = 90^o$ y $\varphi = 0^o$ (en el lado derecho). Las diferencias interaurales pasan desde el valor mínimo al máximo.

La fuente de sonido está en $\theta = 0^o$ y $\varphi = 90^o$ (arriba). La cabeza gira otra vez 90^o en el plano horizontal. En este caso las diferencias interaurales se mantienen a cero durante el giro ya que la fuente se mantiene en el plano medio.

Entre ambos casos existe una variación continua entre máximo cambio de las diferencias interaurales para $\varphi = 0^o$ y ningún cambio para $\varphi = 90^o$.

Por lo tanto, el valor que determina las diferencias de la señal biaural es el ángulo γ entre la dirección del sonido incidente y el plano medio. Este ángulo depende de la elevación y del azimut de la dirección de la fuente (Figura 7):

$$\operatorname{sen}\gamma = \cos\theta \operatorname{sen}\varphi \qquad \text{Ec. 2}$$

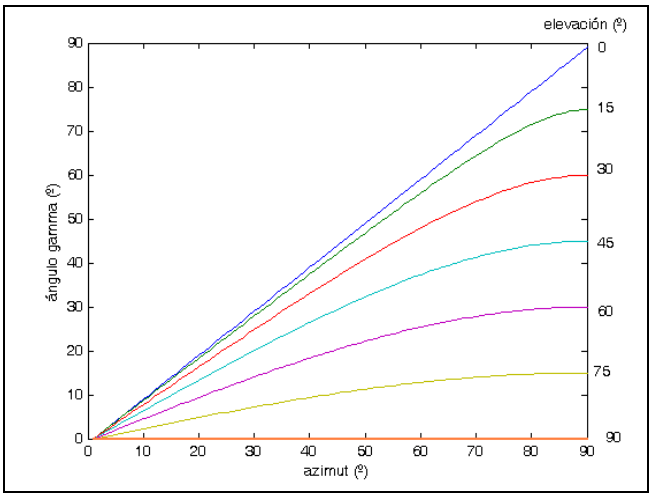


Figura 7. Variación del ángulo γ en función de la elevación y del azimut.

Las curvas con la elevación φ = 0º y φ = 90º se distinguen fácilmente y representan los dos casos extremos mencionados anteriormente. La variación de γ es mayor para valores de elevación más pequeños.

Este fenómeno permite llegar a la siguiente conclusión avalada en múltiples estudios subjetivos [Blauert 97]: el movimiento de la cabeza permite determinar la dirección de la fuente sonora con bastante precisión. Se ha evidenciado que después de dejar mover la cabeza libremente, la localización del evento sonoro y la localización de la fuente sonora coinciden (la cabeza busca de forma inconsciente ángulos γ más grandes).

2.2.4 LA SEÑAL BIAURAL

Es de sobra conocido que no es necesario tener múltiples canales para crear sonido envolvente convincentemente; dos canales es suficiente. El método consiste en recrear la presión sonora en el oído derecho e izquierdo que existiría si el oyente realmente estuviera en presencia del sonido.

Una aproximación conceptualmente sencilla es poner dos micrófonos en los canales auditivos de un maniquí acústico y grabar lo que recogen (Figura 8). Cuando estas señales derecha e izquierda alimentan la unidad derecha e izquierda de unos auriculares, es como si el oyente estuviera presente en el campo sonoro original. En concreto, si la cabeza del maniquí y el oyente tienen la misma forma y tamaño presentan la misma información de ITD y de ILD. Similarmente, si el maniquí y el oyente tienen las orejas y tronco con la misma forma y tamaño, presentarán los mismos parámetros de elevación. Las grabaciones hechas de esta forma se llaman grabaciones biaurales.

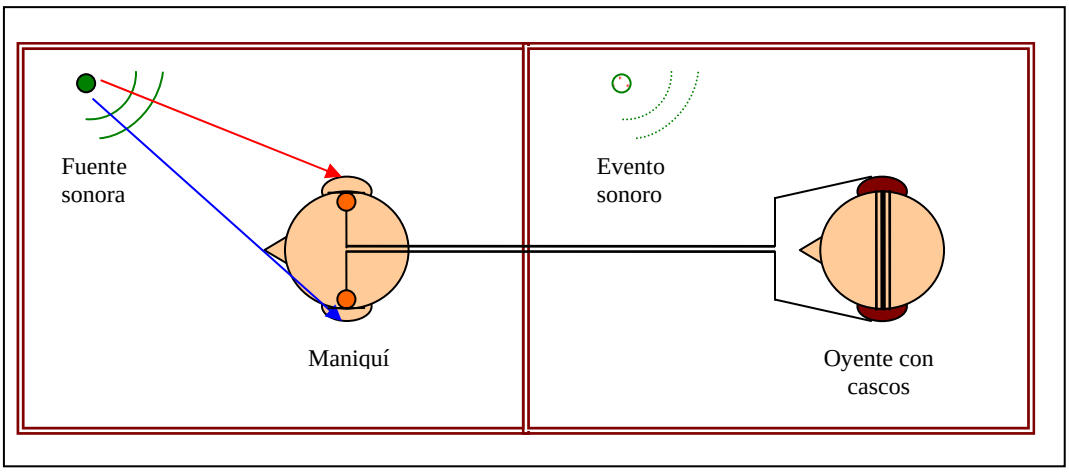


Figura 8. Representación biaural ideal: el evento sonoro aparece en la posición que la fuente sonora tiene respecto al maniquí.

Una alternativa a la grabación biaural consiste en sintetizar las señales para el oído derecho y el oído izquierdo (señal biaural), convolucionando la señal de sonido de la fuente con las respuestas impulsivas de la transmisión del sonido desde la fuente a cada oído. Por lo tanto, la señal biaural puede obtenerse con un maniquí acústico o con procesado de señal según muestra la Figura 9.

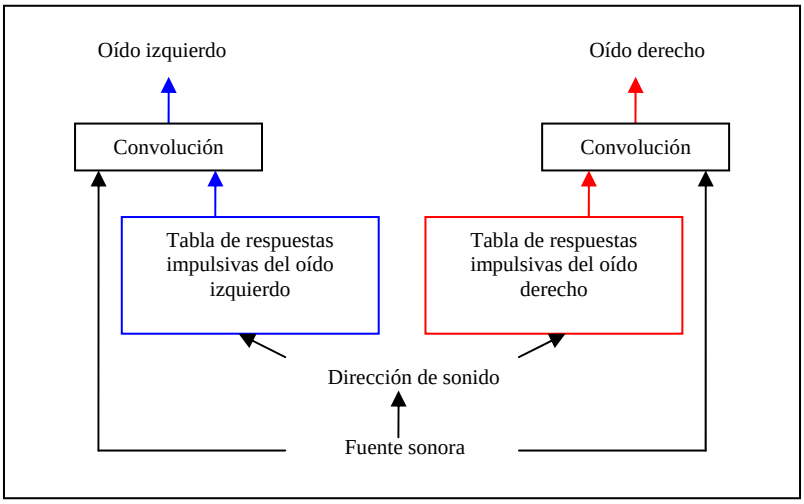


Figura 9. Representación biaural basada en las respuestas impulsivas de los oídos.

El número de respuestas impulsivas almacenadas en la tabla se puede reducir usando muestreo espacial grueso e interpolando adecuadamente entre puntos de dirección contiguos.

Para conseguir las representaciones biaurales es necesario entender las funciones de transferencia de la cabeza (HRTF, Head-Related Transfer Function).

Para poder simular y comprobar el correcto funcionamiento de los modelos de estimación de azimut y elevación desarrollados en esta Tesis, es necesario obtener la señal biaural a partir de las señales emitidas por los altavoces o fuentes sonoras. Para ello utilizaremos las funciones de transferencia de la cabeza. Se utilizará la base de datos de HRTF obtenidas para el maniquí acústico KEMAR (Knowles Electronics Manikin for Auditory Research) [Gardner 00] [Gardner 00a]. Y también, utilizaremos la base de datos creada por nosotros del maniquí de Head Acoustics, HATS (Head and Torso Simulator). Como la localización

de los altavoces puede ser cualquiera, y las bases de datos tienen una resolución de 5º en azimut y 10º en elevación, será necesario interpolar las respuestas para ángulos distintos a los de la base de datos.

2.3 FUNCIONES DE TRANSFERENCIA DE LA CABEZA, HRTF

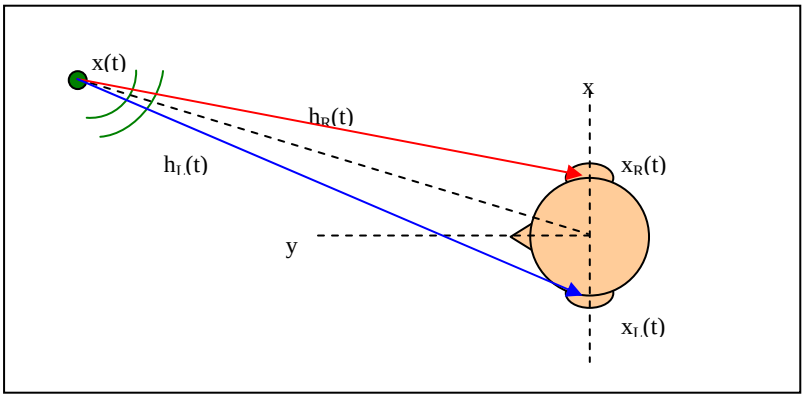


Figura 10. Señales biaurales obtenidas a partir de la señal de la fuente sonora y las HRTF.

$$x_L(t) = \int h_L(\tau)x(t-\tau)d\tau$$
$$x_R(t) = \int h_R(\tau)x(t-\tau)d\tau \qquad \text{Ec. 3}$$

Según la Figura 10, para hallar la presión sonora que una fuente arbitraria x(t) produce en cada oído (x_L y x_R), sólo se necesita conocer la respuesta impulsiva h(t) desde la fuente hasta el oído respectivo (Ec. 3). Esto se llama respuesta impulsiva referida a la cabeza (hrir, head related impulse response), y su transformada de Fourier, H(f), es la HRTF. Estas funciones contienen todos los parámetros físicos de la localización del sonido. Una vez se conoce la HRTF del oído izquierdo y del oído derecho (y por lo tanto la hrir de cada oído), se pueden sintetizar con precisión señales biaurales a partir de una fuente sonora.

Se asume que la HRTF se mide en un entorno anecoico, y por lo tanto no incluye los efectos de las reflexiones del sonido en el entorno, las cuales también proporcionan información de localización.

La HRTF es una función sorprendentemente complicada de cuatro variables: tres coordenadas de espacio y una de frecuencia. En coordenadas esféricas, para distancias más grandes que un metro, se suele decir que la fuente está en campo lejano, y la amplitud de la HRTF cae inversamente con la distancia, según la Ley de Divergencia Esférica. La mayoría de las mediciones de las HRTF se hacen en campo lejano, lo cual reduce esencialmente las HRTF a una función de acimut, elevación y frecuencia.

Para tener una idea de cómo las respuestas varían con el acimut y la elevación, se van a comentar las siguientes gráficas de las hrir y de las HRTF del maniquí KEMAR.

2.3.1 LA HRIR EN EL PLANO HORIZONTAL

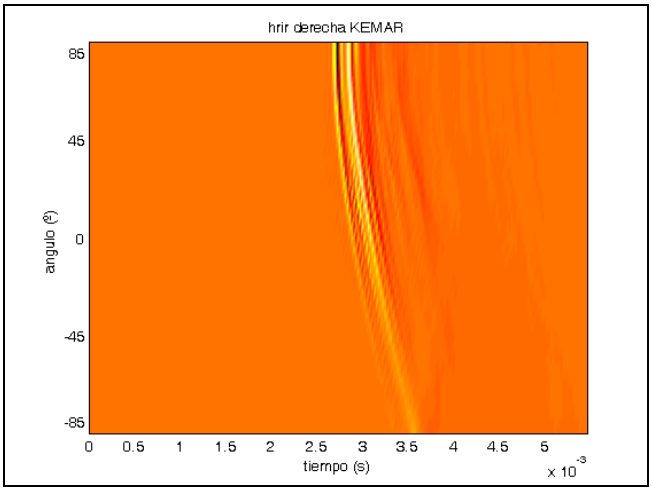


Figura 11. Hrir en el plano horizontal de KEMAR para el oído derecho.

La Figura 11 muestra la respuesta del oído derecho a una fuente impulsiva en el plano horizontal. La intensidad de la respuesta se representa en variaciones del brillo. Por lo tanto, se puede observar que el sonido es más fuerte y llega más pronto cuando procede del lado derecho (acimut $\theta = 90º$). Análogamente, es más débil y llega más tarde cuando procede del lado izquierdo (acimut $\theta = -90º$). Se aprecia que el tiempo de llegada varía con el acimut más o menos de forma sinusoidal. De hecho, el tiempo de llegada coincide bastante bien con la ecuación de la ITD (Ec. 1). En concreto, se aprecia que la diferencia entre el tiempo de llegada más grande y el más pequeño es de unos 0.7 ms, justo lo que se supone teóricamente. Este comportamiento se ha corroborado en medidas subjetivas como las mostradas en la Figura 12.

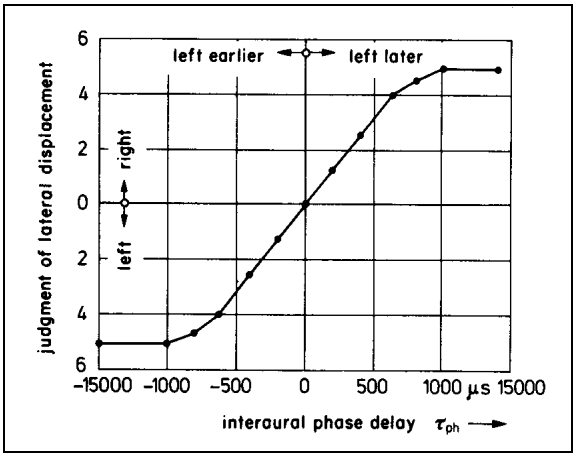


Figura 12. 630 µs es el máximo retardo detectado para impulsos y señales impulsivas (Toole 1965). [Blauert 97].

También es posible explicar algunas de las formas vistas en la imagen pensado en el comportamiento físico. Por ejemplo, la secuencia inicial de cambios rápidos (bandas claras y oscuras) se debe a las reflexiones en la oreja. El pico de llegada a 0.4 ms después del pico inicial es debido a la reflexión del hombro.

2.3.2 LA HRTF EN EL PLANO HORIZONTAL

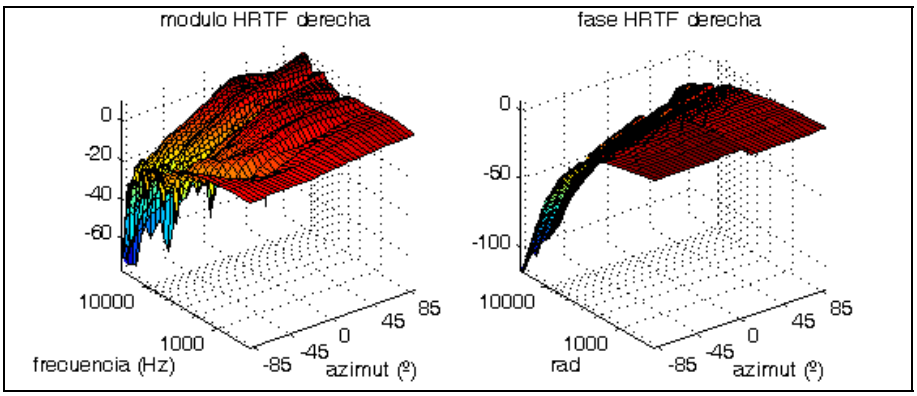


Figura 13. HRTF en el plano horizontal de KEMAR para el oído derecho.

Este gráfico de malla (Figura 13) muestra la respuesta en frecuencia para el oído derecho de KEMAR según se mueve la fuente en el plano horizontal. Aunque la superficie es más bien desigual, si se mira una frecuencia se puede ver un cambio aproximadamente sinusoidal con el acimut θ. Como se esperaba, la respuesta es normalmente mayor cuando la fuente está a θ = 90º (dirigida al oído derecho) y menor cuando la fuente está a θ = -90º (en el lado opuesto de la cabeza) para frecuencias altas (> 1.6 kHz).

El pico alrededor de 4 kHz se debe a la resonancia del canal auditivo. El mínimo a 10 kHz es el famoso "mínimo de oreja", cuya frecuencia cambia con la elevación φ [Blauert 97].

En las HRTF están implícitos los parámetros primarios de información de la localización que es lo que se va a estudiar y analizar para poder estimar el ángulo del evento sonoro.

2.3.3 LA HRTF EN EL PLANO MEDIO

Como se ha explicado anteriormente, en resumen se puede decir que la localización en el plano medio se basa en la información espectral del sonido recibido. Los estudios han demostrado que las distorsiones espectrales causadas por la oreja a las frecuencias altas, por encima de 5 kHz, se comportan como parámetros de localización en el plano medio. El espectro varía sistemáticamente a frecuencias mayores que 5 kHz según varía la elevación de la fuente. Existe un nodo que varía de 6 a 10 kHz cuando la elevación del sonido cambia de -45º a 45º como se puede apreciar en la representación del módulo de la HRTF para diferente ángulo de elevación del maniquí KEMAR (Figura 14).

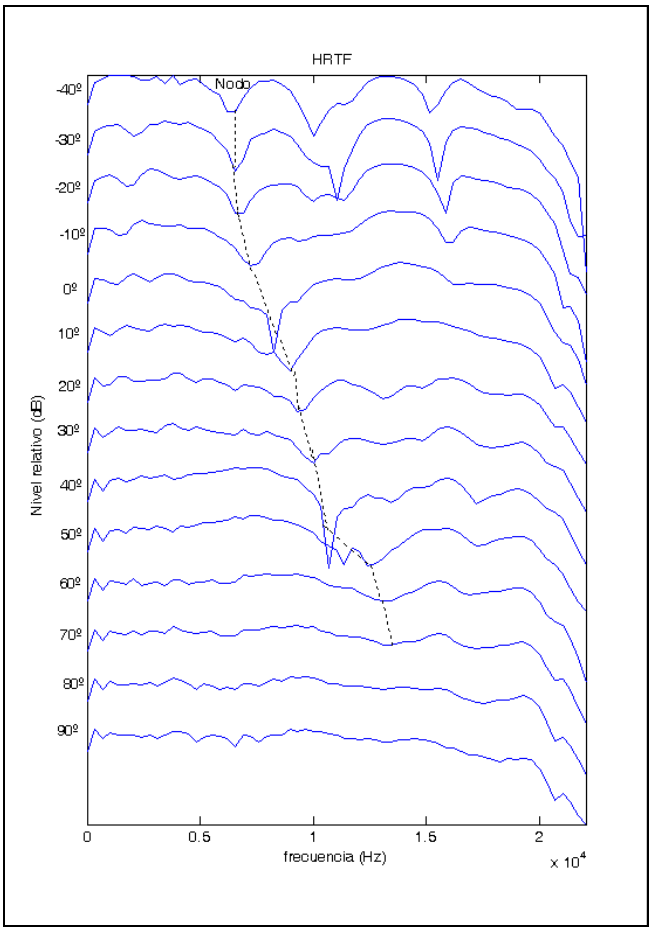


Figura 14. HRTF para ángulos de elevación desde – 45º a 90º, se destaca la posición del primer nodo.

Los parámetros espectrales según los estudios [Blauert 97] existen entre 4 y 16 kHz y se pueden dividir entre parámetros frontales, traseros y superiores. Los parámetros frontales son: un nodo de una octava con frecuencia de corte inferior entre 4 y 8 kHz y un realce de energía para frecuencias superiores a 13 kHz. Los parámetros superiores son: un pico de ¼ de octava entre 7 y 9 kHz. Y por último los parámetros traseros son: un pequeño pico entre 10 y 12 kHz y un decrecimiento de energía por encima y debajo de este pico.

2.3.4 INTERPOLACIÓN DE LA HRTF

Ya se ha comentado la necesidad de interpolar para conseguir las HRTF de los ángulos intermedios, no proporcionadas en las bases de datos medidas. Existe una amplia bibliografía sobre métodos de interpolación [Nishino 96] [Freeland 04] e incluso existe la posibilidad de sintetizar las HRTF en función de características antropométricas de las personas [Rodríguez 05] [Sobreira 01]. El estudio realizado por [Hartung 99] es muy interesante porque compara el método de interpolación de la distancia inversa ponderada y el método de interpolación de los “splines” esféricos; aunque existen otros métodos de interpolación basados en una representación de coeficientes PCA (Principle-Component-Analysis) o transformaciones Karhunen-Loëve o modelos de polos-ceros.

Por lo tanto, la señal biaural será la suma de las contribuciones de cada señal de altavoz o fuente sonora, filtrada por las HRTF interpolada que corresponda al ángulo relativo que forman el altavoz y la cabeza para cada uno de los oídos.

2.4 LOCALIZACIÓN ESPACIAL

En el libro "Spatial Hearing" [Blauert 97] se realiza un compendio muy amplio de los trabajos existentes sobre localización espacial. La mayoría de los estudios evalúan los diferentes parámetros de localización de forma subjetiva para diferentes condiciones de contorno. De dichos estudios se pueden sacar las siguientes conclusiones:

- La precisión en la localización depende del tipo de sonido y del tipo de sala: p.e. un tono sostenido en cámara reverberante no se puede localizar y un clic en cámara anecoica es localizado con precisión. También depende de la posición de la fuente, de los sonidos previos y del paso del tiempo que provoca en el oyente una variación en su capacidad de localización (Figura 15).

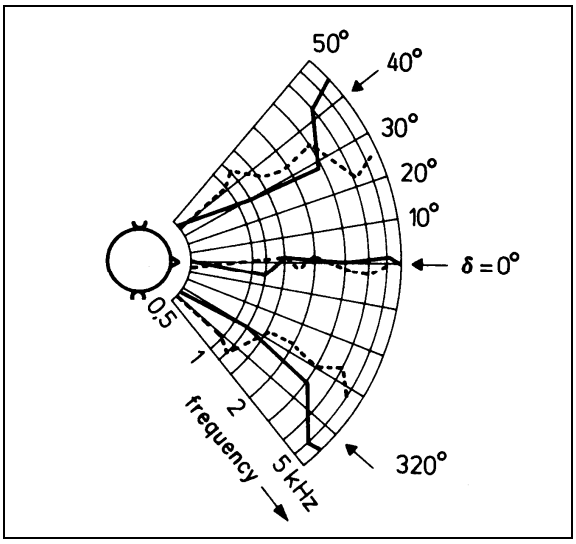


Figura 15. En la figura se muestra la localización de una fuente emitiendo tonos puros (sólido), salvas de tonos gausianas (puntos) en comparación con ruido de banda ancha colocado a 320º, 0º, 40º. Incluso la localización depende de la variación del espectro a lo largo del tiempo (Sandel 1955 y Boerger 1965). [Blauert 97].

- Existen dos fenómenos llamados "adaptación": la pérdida de sensibilidad cuando el sonido es largo, aparece a los pocos segundos y es máximo a los 3-5 minutos; y el periodo de readaptación que lleva 1-2 minutos.
- Existen dos atributos en la localización de un evento sonoro: la posición espacial y su extensión que son evaluados mejor por comparación con otros eventos sonoros.
- **Umbral de percepción de localización**: es el menor cambio de un atributo específico de la fuente sonora suficiente para producir un cambio en la localización del evento sonoro, p.e. dirección o distancia. Por lo tanto, el espacio de los eventos sonoros tiene menos resolución que el de la fuente sonora.
- En resumen, los parámetros que afectan a la localización son:
    - La posición de la fuente.
    - El tipo de señal.
    - Los sonidos previos.
    - Y el sujeto: su umbral de percepción de localización varía con el tiempo.

La mayor resolución se produce para sonidos colocados a azimut 0º, luego se toma el umbral de percepción de 0º como la máxima resolución espacial. Además este umbral mínimo varía con la frecuencia, aunque se mantiene dentro de los 5º de resolución. De la tabla siguiente se puede concluir que la máxima resolución estaría entre 1º y 5º (Figura 16). 5º es la resolución con la que se miden las HRTF, por lo tanto suficiente para representar las señales biaurales desde el punto de vista de resolución de la localización de los eventos sonoros.

Referencia	**Tipo de señal**	**Umbral de percepción de localización**
Klemm (1920)	Impulsos (clicks)	0.75º - 2º
King (1930)	Tren de impulsos (clicks)	1.6º
Stevens (1936)	Sinusoides	4.4º
Schmidt (1953)	Sinusoides	> 1º
Sandel (1955)	Sinusoides	1.1º - 4.0º
Mills (1958)	Sinusoides	1.0º - 3.1º
Stiller (1960)	Ruido de banda estrecha, salvas de tono $\cos^2$	1.4º - 2.8º
Boerger (1965a)	Salvas de tono gaussianas	0.8º - 3.3º
Gardner (1965a)	Voz	0.9º
Perrott (1969)	Salvas de tono con diferentes tiempos de subida y caída y frecuencias	1.8º - 11.8º
Blauert (1970b)	Voz	1.5º
Haustein (1970)	Ruido de banda ancha	3.2º

Tabla 1. Umbral de percepción de localización para θ = 0º. [Blauert 97].

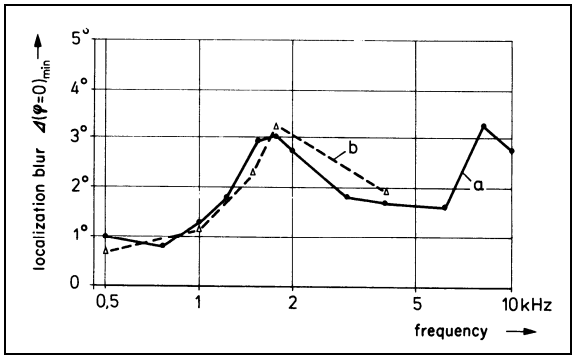


Figura 16. Variación con la frecuencia del umbral de percepción de localización para θ = 0º: a) señal sinusoidal (Mills 1958), b) salvas de tono gaussianas de 1/3 octava (Boerger 1965). [Blauert 97].

- Para θ = 90º el umbral se incrementa de tres a diez veces su valor a 0º (aproximadamente 10º).
- Se han hecho experimentos donde se aprecia que incrementar la duración del sonido hasta 700 ms hace que decrezca el umbral de percepción de localización.

- Con señales de banda estrecha puede suceder que el evento sonoro aparezca simétricamente a la fuente sonora (detrás – delante), debido a que no hay información de tipo espectral para hacer la diferenciación. Esto se solventa con ligeros movimientos de cabeza. Cualquier tipo de sonido se llega a localizar bien si se permite el movimiento de la cabeza.
- La diferencia de nivel interaural depende de la difracción de la cabeza y por lo tanto de la distancia de la fuente sonora (Figura 17).

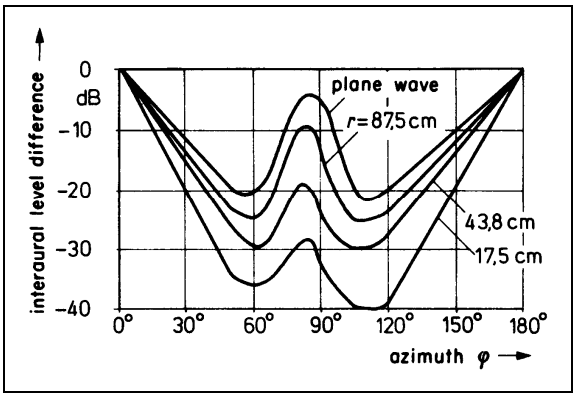


Figura 17. Variación teórica con la distancia de la ILD para 1860 Hz, cabeza esférica y diferentes distancias (Hartley 1921). [Blauert 97].

- El sistema auditivo evalúa además las componentes espectrales de las señales de entrada al oído individualmente con independencia de las diferencias de tiempo interaural.
- Cuando se desplazan dos sinusoides en tiempo aparecen dos formas de medir el retardo debido a la periocidad de la señal (ambigüedad de la ITD) y por lo tanto aparecen dos eventos sonoros, pero predomina aquel que está más cerca del plano medio (Figura 18).

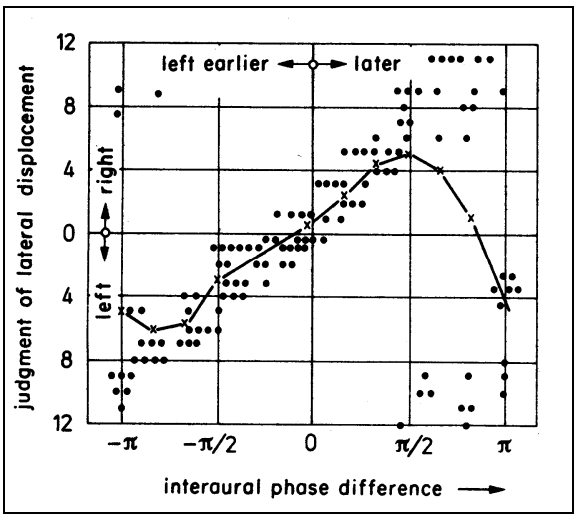


Figura 18. Localización según varía la diferencia de fase interaural de dos sinusoides de 600 Hz (Sayers 1964). [Blauert 97].

- **Periodo refractario**: las neuronas y nervios del sistema auditivo son incapaces de reaccionar de nuevo durante 1-2 ms para poder detectar un umbral y medir el retardo. Esto ocurre para frecuencias mayores a 800-1600 Hz (Stevens 1938). Se ha

llegado a un acuerdo por el cual para frecuencias superiores a 1.5-1.6 kHz no se puede apreciar desplazamientos laterales en señales sinusoidales o modulaciones de tonos puros, sin embargo para salvas o ruidos de una octava si se aprecian debido al retardo interaural de la envolvente. Además la precisión de evaluación de la envolvente para determinar la lateralización depende de si la señal tiene escalones pronunciados o no (Elfner 1968) (Figura 19).

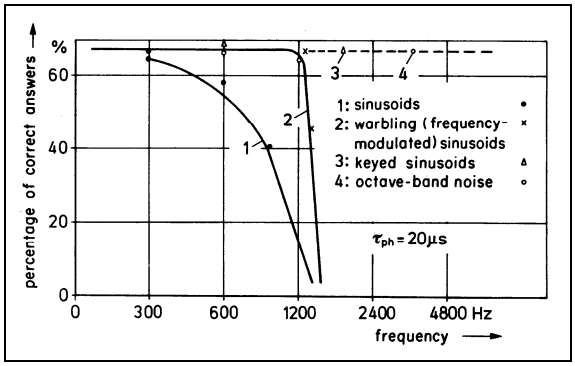
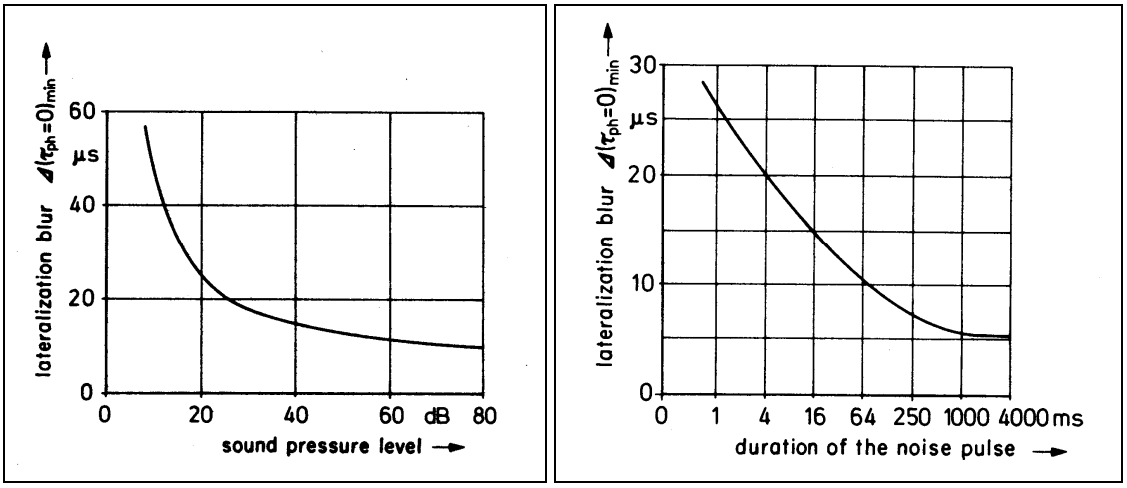


Figura 19. Capacidad de detectar un retardo de fase interaural según las frecuencias y tipos de señales (Scherer 1959). [Blauert 97].

- El umbral de percepción de lateralización es mayor para niveles de señal pequeños y duraciones de la señal pequeñas (Figura 20).



a) b)

Figura 20. Variación del umbral de percepción de lateralización ($\theta = 0^o$) para: a) sinusoide de 500 Hz (Zwislocki 1956, Hershkowitz 1969), b) pulsos de ruido de banda ancha f_{max} = 5 kHz (Tobias 1959). [Blauert 97].

- La lateralización según la diferencia de nivel interaural depende de la frecuencia. Además pueden aparecen múltiples eventos sonoros al variar la ILD o ensanchamiento de la fuente (Figura 21).

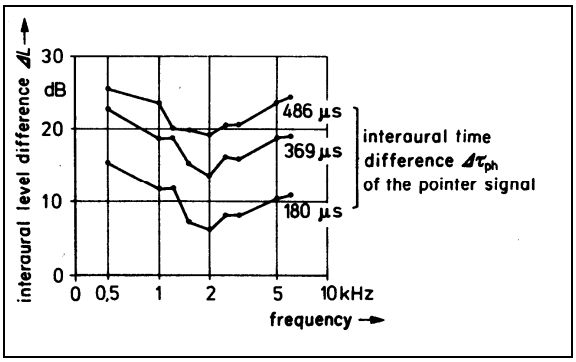


Figura 21. Variación de la ILD en función de la frecuencia (Feddersen 1957). [Blauert 97].

- El umbral de percepción de la ILD puede llegar a ser de unos 0-1 dB para condiciones muy concretas. Además este valor depende del nivel de las señales (Rowland 1967, Hershkowitz 1969), para niveles bajos es pequeño, para niveles altos permanece prácticamente constante según se aumenta el nivel (Upton 1936, May 1964, Babkoff 1969).
- Debido a la fatiga del oído la lateralización va disminuyendo según pasa el tiempo (Urbantschitsch 1881, Thompson 1979, von Bekesy 1930). Aunque existe un proceso de aprendizaje a largo plazo, que ajusta las diferencias de nivel propias de cada individuo a la localización.
- **Trading**: el evento sonoro es desplazado mediante una diferencia de tiempo o nivel y después se determina el valor opuesto de nivel o tiempo que es necesario para volver el evento sonoro al centro. A partir de los resultados de trading se puede concluir que la importancia relativa entre la ILD y la ITD depende del tipo de señal. ILD tiene mayor importancia cuando la señal incluye componentes por encima de 1.6 kHz y el nivel es bajo.
- De acuerdo a lo que se conoce hasta ahora sobre la evaluación subjetiva de la ITD y de la ILD y su influencia en la lateralización, se puede resumir en las siguientes conclusiones:
    - El sistema auditivo emplea al menos dos mecanismos evaluadores que hasta cierto punto trabajan independientemente uno del otro: Teoría Dúplex.
    - El primer mecanismo interpreta los desplazamientos de tiempo interaural de la estructura fina de las señales de entrada del oído. Su influencia en el desplazamiento del evento sonoro está basado solamente en las componentes por debajo de 1.6 kHz de la señal.
    - El segundo mecanismo interpreta las diferencias de nivel de presión sonora y también los desplazamientos en tiempo de las envolventes. Tienen una influencia dominante en el desplazamiento del evento sonoro si la señal tiene componentes importantes por encima de 1.6 kHz.
    - El segundo mecanismo es variante en el tiempo; por ejemplo, puede ser alterado por un proceso de re-aprendizaje (Held 1955).
    - La importancia relativa de los dos mecanismos puede variar según los individuos.
    - El primer mecanismo parece tener mayor persistencia en tiempo que el segundo.

- El desplazamiento interaural de la envolvente de las componentes espectrales de la señal puede ser aproximado por el retardo de grupo promedio en ese rango de frecuencia.
- Cuando se puede mover la cabeza libremente aparecen otros parámetros que ayudan a localizar el evento sonoro.
- Existen otras informaciones que ayudan a localizar el evento sonoro. Entre ellos está el efecto de las teorías visuales. Por ejemplo, está demostrado que si un sujeto está viendo la televisión donde aparece alguien hablando, localiza el evento sonoro en la imagen del locutor. Si cierra los ojos localiza el evento sonoro en la dirección del altavoz.
- Cuando la diferencia de tiempo es mayor a 1 ms, se produce la Ley del Primer Frente de Onda (Cremer 1948) y el evento sonoro se localiza en la dirección de la fuente sonora que llega antes. El conjunto de la localización sumatoria y la Ley del Primer Frente de Onda es lo que se llama efecto precedente (Wallach 1949).
- El oído interno analiza la señal espectralmente en componentes de ancho de banda relativo aproximadamente constante, es lo que se llaman bandas “perceptuales”.
- Se ha propuesto que cuando los parámetros de localización están distorsionados, el mecanismo de percepción usa aquel más consistente. Un parámetro es consistente si sugiere la misma dirección en un ancho de banda ancho (Wightman 1996).

Basándose en el conocimiento de cómo funciona el sistema auditivo desde un punto de vista subjetivo, avalado por los numerosos estudios, se pueden implementar modelos computacionales que simulen la percepción subjetiva. En este sentido, desde hace poco tiempo, existen modelos computacionales del sistema auditivo que se han aplicado en temas de audio y procesado de voz con mucho éxito. La investigación sobre medidas objetivas de calidad subjetiva del sonido basadas en un modelo de percepción ha llevado al resultado final de la norma [ITU Recommendation BS.1387]: “Method for objective measurements of perceived audio quality”, (Task Group 10/4). La idea es comparar la señal original y la señal de salida del dispositivo, que se quiere evaluar, en el dominio espectral de la percepción, es decir, en el dominio tiempo-frecuencia. Basándose en medidas de distancia se expresa un índice que está correlacionado con los resultados de tests subjetivos expresado con los valores MOS (Mean Opinion Store). Sin embargo, en esta norma no se ha tenido en cuenta la calidad del sonido relacionada con aspectos biaurales.

El modelado de la audición biaural por medio computacional se enfrenta a problemas similares a cualquier modelado de la percepción humana, es decir, es difícil conseguir fórmulas explícitas y reglas que describan el comportamiento de estos sistemas no lineales tan complejos [Harma 99]. La mayoría de los modelos tradicionales de la audición biaural analizan sólo la lateralización del sonido basándose en la correlación cruzada interaural.

Se ha avanzado muy poco en el modelado computacional de la percepción de atributos espaciales para evaluar la calidad del sonido reproducido. La principal razón es que el modelado biaural incluye aspectos muy problemáticos como el efecto precedente.

Se podría utilizar un buen modelo biaural para evaluar no sólo los sistemas tradicionales de grabación y reproducción multicanal, sino los nuevos sistemas de audio sintético relacionados con la acústica virtual donde se pueden sintetizar fuentes virtuales de sonido por medio de altavoces reales o de auriculares. Por lo tanto, cada vez son más importantes los temas de audio tridimensional, técnicas biaurales y “auralización”.

La precisión en la localización de las fuentes de sonido depende de un conjunto de parámetros relacionados con la naturaleza del sonido, la antropometría y características del oyente, el movimiento voluntario o involuntario de la fuente o del oyente y la acústica del entorno donde están la fuente y el oyente. Algunos de estos atributos favorecen una correcta localización mientras que otros (como los ecos y la reverberación) son perjudiciales. En nuestro trabajo utilizamos un entorno acústico de referencia sin ecos ni reverberaciones para estudiar los parámetros de localización debidos solamente a las HRTF.

2.5 LOCALIZACIÓN EN EL PLANO HORIZONTAL

Para la localización en el plano horizontal se realiza un procesado de la señal biaural a través de un modelo de localización biaural. El conocimiento que se tiene de cómo el cerebro extrae la información de la dirección del sonido a partir de la sensación sonora es todavía limitado [Blauert 97], aún así se han desarrollado técnicas de análisis de la señales biaurales que simulan el procesado del sistema auditivo humano [Pulkki 99], [Faller 04], [Viste 04], [Blauert 97]…

Este modelo biaural se ha aplicado a diferentes sistemas de reproducción espacial [Pulkki 01], [Pulkki 05], pero nunca a una configuración tipo ventana acústica virtual hasta el comienzo de este trabajo [Gómez-Alfageme 04].

Uno de los modelos que se van a describir, llamado modelo de localización psicoacústico biaural, intenta modelar el sistema de percepción humano [Harma 99], o al menos aquellos atributos del mismo relacionados con la localización de los eventos sonoros. Los modelos que se han desarrollado utilizan básicamente dos fuentes de información espacial:

- Diferencia de nivel que llega a cada oído ILD (Interaural Level Difference).
- Diferencia de tiempo de llegada a cada oído ITD (Interaural Time Difference).

A groso modo, es posible decir que mediante las diferencias de intensidad se distingue la posición a izquierda o derecha de la fuente para frecuencias mayores a 5000 Hz. Esto se debe a que a partir de estas frecuencias la diferencia de intensidad no es debida a la diferente distancia entre oídos y fuente sino al efecto sombra que la cabeza produce sobre el oído más alejado. Para frecuencias bajas, donde la longitud de onda es comparable al tamaño de la cabeza, el sonido sufre difracción y la atenuación es menor. Para frecuencias inferiores a 1600 Hz es el retardo interaural, ITD, quien permite la localización basándose en que el sonido de la fuente llega antes al oído más cercano a la misma.

Rango frecuencias	**Método**
< 1.6 kHz	ITD: de la portadoras
> 1.6 kHz	ITD: envolvente de las señales recibidas
> 1.6 kHz	ITD ambigua: ILD resuelve

Tabla 2. Rango de frecuencias de los parámetros de localización horizontal.

Estos efectos permiten localizar fuentes en el plano horizontal pero no permiten distinguir movimientos verticales de la fuente o saber si el sonido proviene de delante o de un punto simétricamente situado detrás del oyente. Para ello se emplea la información que nos proporciona el pabellón auditivo, cuya respuesta en frecuencia depende de la dirección de

llegada del sonido y que se mide en la HRTF. Estos sistemas trabajan bien para bandas anchas, aunque para bandas estrechas los resultados no son tan buenos.

En el trabajo realizado no ha sido necesario implementar el procesado debido al efecto precedente porque las diferencias de tiempo entre los altavoces del array son mínimas y por lo tanto no se perciben como sonidos retardados. Tampoco hay reverberación en el escenario acústico utilizado.

Los dos modelos que se han implementado y estudiado están basados en la Teoría Dúplex. Según esta teoría, la diferencia de tiempo interaural (ITD) y la diferencia de nivel interaural (ILD) son dos de los parámetros más importantes a la hora de estimar la lateralización de un evento sonoro en el plano horizontal. Estos modelos parten de la señal biaural que llega a cada oído y dan como salida un estimación del ángulo de azimut en el plano horizontal que sería el mismo que percibiría un oyente al que le llegara esa señal biaural (Figura 22).

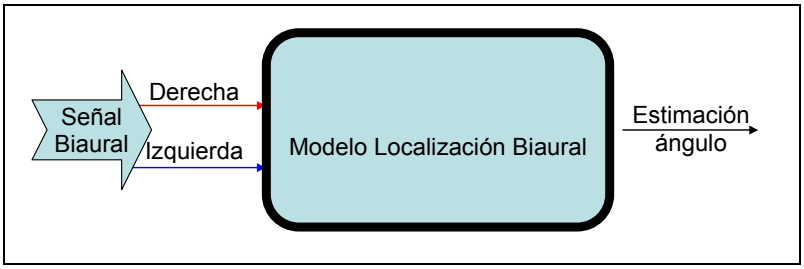


Figura 22. Modelo de localización biaural.

2.6 MODELO PSICOACÚSTICO BIAURAL

Una vez estudiado qué parámetros influyen subjetivamente en el sistema de audición para localizar un evento sonoro, se tiene que modelar el comportamiento del sistema auditivo con algoritmos que trabajen sobre la señal que reciben los oídos. Este apartado pretende mostrar los algoritmos encontrados en la bibliografía. A estos modelos se les llama psicoacústicos porque simulan el procesado que hace nuestro sistema auditivo desde la parte de transducción aérea-neurológica, hasta la parte puramente neurológica en la cual se decide con qué ángulo aparece la imagen acústica creada en nuestro cerebro (evento sonoro). Esta última parte sería la etapa del estimador.

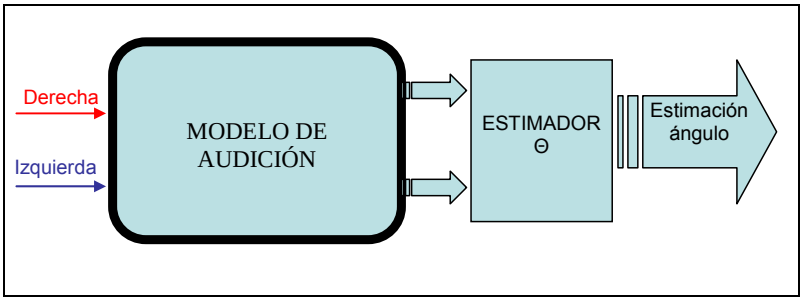


Figura 23.Método de localización biaural basado en un modelo de audición.

El modelo de audición (Figura 23) está dividido en cuatro etapas que modelan cada parte del sistema auditivo humano: oído externo-medio, cóclea, células ciliares y la adaptación neuronal. Cada una de estas partes está modelada con diferentes sistemas que se pueden elegir según los modelos propuestos por diferentes autores (Figura 24).

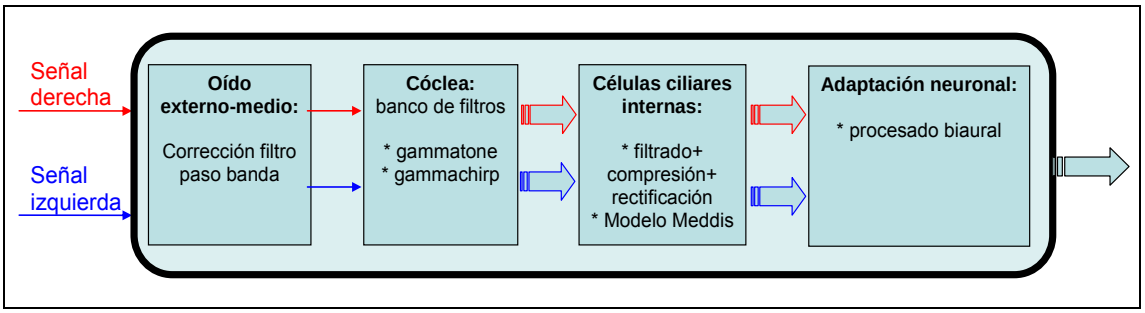


Figura 24. Modelo de audición.

2.6.1 MODELO DEL OÍDO EXTERNO Y DEL OÍDO MEDIO

La función de transferencia del oído externo se modela con filtros y son las HRTF. Como ya se ha visto, estas se pueden medir o sintetizar. La función de transferencia del oído medio generalmente se simula con un filtro lineal fijo.

En el trabajo de [Harma 99] se diseña como un filtro de respuesta similar a las curvas de sonoridad como se muestra en la Figura 25. También es posible diseñar un filtro "frequency-warped" o IIR que modelen la respuesta de esta zona del oído.

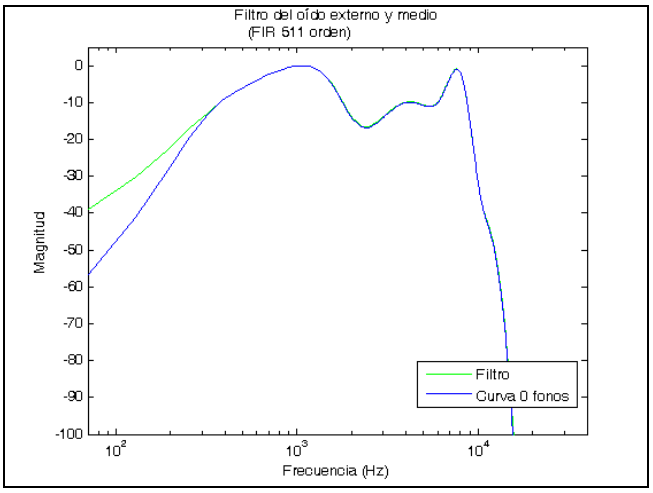


Figura 25. Filtro del oído medio diseñado con las toolbox de HUTear [Harma 00]. Se ha escogido un filtro FIR de 511 coeficientes el cual se aproxima a las curvas de sonoridad de nivel 0 fonos.

Este filtro del oído medio se aplica a las dos señales biaurales, izquierda y derecha, por lo tanto su influencia en la ILD e ITD que miden la diferencia entre las dos señales es casi nula, sobre todo porque después se hace un filtrado en bandas "perceptuales". Este razonamiento nos lleva a no implementarlo.

2.6.2 MODELO DE LA CÓCLEA

El análisis en frecuencia es realizado por la cóclea que está en el oído interno. Hay varias aproximaciones para modelar el comportamiento de este órgano. La forma más general es un banco de filtros con respuesta rectangular o trapezoidal cuyo ancho de banda corresponda a una banda crítica. Existen muchos estudios sobre los ancho de bandas críticos. Un modelado típico es el uso de un banco de filtros "gammatone". Hay varios parámetros a elegir dentro de este banco, el principal es el número de canales de frecuencia

que determinará la resolución en frecuencia de todo el modelo biaural. Dentro de los filtros "gammatone" hay tres alternativas: la primera es usar 42 filtros de ancho de banda equivalente rectangular, ERB [Slaney 98], típicamente se utilizan de 32 a 256 filtros. Las otras dos son filtros de cuarto y sexto orden "complex-valued gammatone". A nivel práctico se puede especificar las frecuencias centrales de todas las bandas o dar la frecuencia central de la primera banda y el número de canales.

Los filtros de ancho de banda rectangular equivalente (ERB) han sido determinados para oyentes jóvenes normales a niveles de presión sonora medios. La escala es similar en concepto pero diferente en números a la de Bark [Blauert 97].

$$\text{ERBnumber} = 21.4 \log(4.37 \cdot f_c(\text{kHz}) + 1) \qquad \text{Ec. 4}$$

Donde ERBnumber es el número del filtro desde 1 hasta 42, por lo tanto la frecuencia central del filtro, f_c, va desde 26 hasta 20767 Hz.

En nuestro caso se ha elegido como frecuencia mínima 200 Hz y 42 canales, calculados según la descripción dada en [Slaney 98] (Figura 26). Para frecuencias muy bajas, la respuesta del oído es muy mala (ver Figura 25 de sonoridad), y por lo tanto se considera que por debajo de 200 Hz los parámetros de localización no proporcionan información. El último filtro calculado tiene una frecuencia central de 20.8 kHz, frecuencia prácticamente no audible. Luego ese margen de frecuencias cubre de sobra el margen útil de trabajo del sistema auditivo.

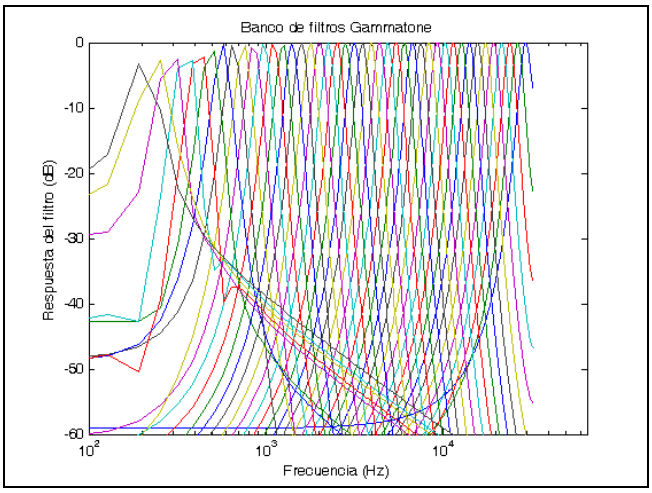


Figura 26. Respuesta del banco de filtros gammatone utilizado con 42 canales.

Otro modelo sería por ejemplo implementar una de línea de transmisión. Además se le puede añadir una compensación asimétrica variable con el tiempo a los filtros [Harma 00]. Otro banco de filtros utilizado podrían ser los filtros gammachirp.

2.6.3 MODELO DE LAS CÉLULAS CILIARES

Existen tres aproximaciones diferentes para implementar la transducción mecánico-neuronal y simular el comportamiento de la respuesta de la excitación de los nervios auditivos, la cual es función de las células ciliares internas de la membrana basilar. En muchos modelos esta fase es implementada como un sistema de rectificación, comprensión y filtrado [Harma 00]. La rectificación se puede elegir entre ninguna, de media onda o de onda completa. Lo más habitual es elegir una rectificación de media onda seguida de un filtro paso bajo de primer orden con τ = 1.25 ms [Blauert 97]. Por encima de 1.6 kHz, esto

significa que la envolvente de las señales es derivada, esto es, las señales se demodulan en amplitud, comportamiento que simula el resultado de los estudios subjetivos, donde a partir de esa frecuencia se analiza la envolvente.

Se puede elegir entre dos compresiones: una función de compresión no lineal basada en el nivel de presión sonora o la clásica regla de compresión $y = x^c$ (típicamente $c = 0.7$).

El filtrado se implementa con un filtro paso bajo para todos los canales de frecuencia de corte 1 kHz. También existe la posibilidad de hacer un filtrado usando un integrador de primer orden con una constante de tiempo de 20 ms o usar el modelo de Meddis [Harma 00] para las células ciliares que simula el modelo de la sinapsis (contacto entre neuronas) o un modelo neuronal refractivo.

2.6.4 MODELO NEURONAL

Este modelo depende de la información que se quiere extraer de la señal. Por ejemplo existen los modelos siguientes: el modelo neuronal refractivo de Meddis; modelo de adaptación neuronal de ventana temporal de Plack; modelo de una red de adaptación no lineal de Dau y el modelo de una red de adaptación no lineal de Karjalainen (referenciados en [Harma 00]). Estos modelos han sido diseñados para predecir algunos efectos específicos que ocurren en la audición que varían con el tiempo.

En el caso de procesado biaural, el modelo a implementar es aquel que extrae los parámetros de localización de ambas señales: ITD e ILD que más adelante se tratan. Algunos modelos utilizan una función de ponderación de la ITD en función del la ILD. Otros modelos trabajan con la ILD e ITD totalmente separados.

2.6.4.1 MODELO DE COINCIDENCIA O CORRELACIÓN CRUZADA

Este modelo se utiliza para calcular la diferencia de tiempo interaural. Las señales de salida del filtrado de las células ciliares son sometidas a una correlación cruzada a corto plazo en cada banda de frecuencia [Blauert 97]:

$$\Phi_{x_R,x_L}(t,\tau) = \int_{-\infty}^{t} x_L(\vartheta)x_R(\vartheta-\tau)G(t-\vartheta)d\vartheta \qquad \text{Ec. 5}$$

donde $G(t-\vartheta)$ es una función de ponderación que da menos importancia a los valores del producto $x_L(\vartheta)x_R(\vartheta-\tau)$ según sean más antiguos. La función es:

$$G(s) = \begin{cases} e^{-s/\tau_{RC}} & s \geq 0 \\ 0 & s < 0 \end{cases} \qquad \text{Ec. 6}$$

donde τ_{RC} es unos pocos milisegundos. La amplitud de la correlación cruzada se hace menor y más ancha según se incrementa la ITD.

Para convertir la IACC (InterAural Cross Correlation) calculada según la Ec. 5 en ángulo, la información de la IACC debe alimentar un procesador de reconocimiento de patrones, el cual decide que patrón (que depende del ángulo) predeterminado concuerda mejor con el valor de la IACC de la entrada. A partir de la información que da el reconocedor de patrones se forma la estimación del evento sonoro.

2.6.4.2 ALGORITMO DE LINDEMANN

Algoritmo diseñado para tener en cuenta la ILD y la ITD. Es un algoritmo simple y que cubre una gran parte de los fenómenos psicofísicos biaurales (Lindemann 1985, 1986, 1982) [Blauert 97]. Este algoritmo (Figura 27) integra elementos inhibidores que el sistema auditivo también usa.

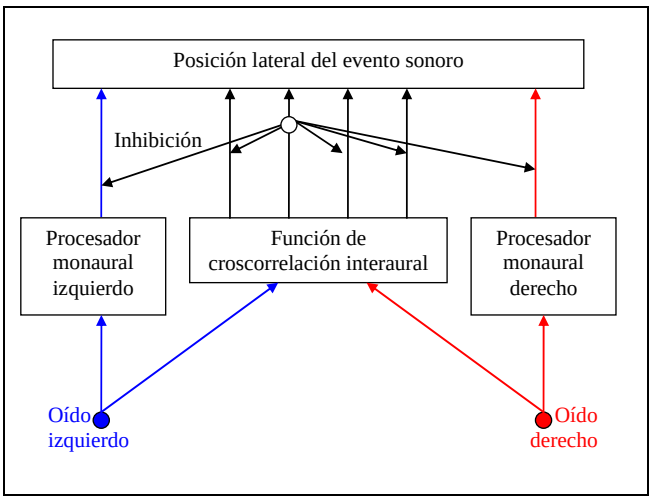


Figura 27. Algoritmo de Lindemann.

Siempre que la línea de croscorrelación genera una salida en cierta posición, la cual corresponde con un cierto retardo, esta salida puede disparar un mecanismo que inhibe la actividad de otras posiciones. Además hay dos caminos monoaurales (izquierdo y derecho), los cuales pueden ser también inhibidos por la actividad biaural (Figura 28).

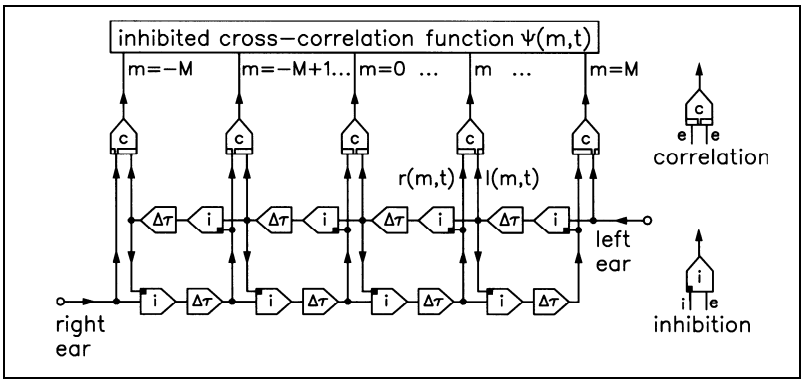


Figura 28. Algoritmo de Lindemann: proceso de inhibición o correlación. [Blauert 97].

El algoritmo de Lidemann hace uso de un mecanismo llamado inhibición contralateral. Cualquier actividad en una de las dos líneas de retardo puede inhibir la actividad en la posición correspondiente de la línea de retardo opuesta. Por ejemplo: los dos oídos son excitados por un impulso que tiene una dirección lateral. Cada línea de retardo recibe el impulso con diferentes retardos. La actividad neuronal viaja en direcciones opuestas por las líneas de retardo. De este modo se causa inhibiciones en los puntos opuestos de la línea de retardo. Finalmente las actividades de ambas líneas se encuentran en algún lugar y estimulan el detector de coincidencia de esa posición. Debido a la inhibición contralateral, la actividad originada en otras posiciones que no sea la que ocurre en el punto de coincidencia se ha eliminado.

Los parámetros a ajustar son los siguientes. La cantidad de inhibición puede ser total o cero. La persistencia se puede ajustar para que sólo haya inhibiciones simultáneas o que una inhibición disparada persista un tiempo, de acuerdo a una función de memoria específica.

Este algoritmo suprime los picos secundarios de la correlación cruzada haciendo que la función sea más picuda. La ILD y la ITD se tienen en cuenta para crear un desplazamiento del pulso de inhibición.

Este modelo puede causar la inhibición de todo el sistema cuando la ILD es grande, por ejemplo a altas frecuencias. Este modelo completo es un sistema de red neuronal artificial adaptable a unos oídos por un procedimiento de aprendizaje.

2.6.4.3 MODELO DE COHERENCIA INTERAURAL

En este modelo [Faller 04] para tener una buena localización se va a utilizar un criterio de selección de la ITD y la ILD que es el siguiente: sólo se van a tener en cuenta las ITD e ILD que claramente indiquen una posición de la fuente, para ello se va a usar la coherencia interaural (IC).

La ITD y la IC se calculan a partir de la correlación cruzada normalizada de cada canal del banco de filtros después de haberle aplicado la transducción neuronal.

Si las señales de salida del procesado anterior son x_L y x_R para cada oído y cada filtro, y n es el índice de la muestra, la correlación cruzada (IACC) se puede calcular como:

$$\mathrm{IACC}(n,m) = \frac{a_{12}(n,m)}{\sqrt{a_{11}(n,m)a_{22}(n,m)}} \qquad \text{Ec. 7}$$

Donde

$$\begin{aligned} a_{12}(n,m) &= \alpha\, x_R\left(n-\max\{m,0\}\right)x_L\left(n-\max\{-m,0\}\right)+(1-\alpha)\,a_{12}(n-1,m)\\ a_{11}(n,m) &= \alpha\, x_R\left(n-\max\{m,0\}\right)x_R\left(n-\max\{-m,0\}\right)+(1-\alpha)\,a_{11}(n-1,m)\\ a_{22}(n,m) &= \alpha\, x_L\left(n-\max\{m,0\}\right)x_L\left(n-\max\{-m,0\}\right)+(1-\alpha)\,a_{22}(n-1,m) \end{aligned} \qquad \text{Ec. 8}$$

y $\alpha \in [0,1]$ determina la constante de tiempo de la caída exponencial de la ventana de estimación:

$$T = \frac{1}{\alpha f_s} \qquad \text{Ec. 9}$$

Donde f_s es la frecuencia de muestreo. IACC (n,m) se evalúa en intervalos de tiempo dentro del rango [-1,1] ms, por lo tanto $m/f_s \in$[-1,1] ms. La ITD (en muestras) se calcula como el intervalo que hay hasta el máximo de la IACC:

$$\mathrm{ITD}(n) = \arg\max_m \mathrm{IACC}(n,m) \qquad \text{Ec. 10}$$

La resolución temporal de la ITD está limitada por el periodo de muestreo.

La normalización de la correlación cruzada se hace para calcular la IC, definida como el valor máximo de la IACC:

$$IC(n) = \max_{m} IACC(n,m) \qquad \text{Ec. 11}$$

Este valor describe la coherencia de las señales del canal izquierdo y derecho. En principio tiene un rango de [0,1] donde 1 ocurre cuando x_L y x_R son perfectamente coherentes.

La ILD se calcula como:

$$\Delta L(n) = 10\log\left(\frac{L_R(n,\tau(n))}{L_L(n,\tau(n))}\right) \qquad \text{Ec. 12}$$

Donde:

$$\begin{aligned} L_R(n,m) &= \alpha\, x_R^2(n-\max\{m,0\}) + (1-\alpha)\, L_R(n-1,m) \\ L_L(n,m) &= \alpha\, x_L^2(n-\max\{m,0\}) + (1-\alpha)\, L_L(n-1,m) \end{aligned} \qquad \text{Ec. 13}$$

Las diferencias de nivel entre canales son más grandes que la ILD calculada porque normalmente se ha utilizado una compresión de la envolvente en el oído medio.

2.6.4.4 MODELO DE SONORIDAD

La ILD se calcula como la diferencia de sonoridad interaural [Pulkki 01]. La sonoridad expresada en sonos viene dada por la fórmula de Zwicker e:

$$L = \sqrt[4]{\langle x^2 \rangle} \qquad \text{Ec. 14}$$

Donde $\langle x^2 \rangle$ es el promedio temporal de la potencia de la señal. La raíz cuarta aproxima el exponente de 0.23 usado en las medidas de la sonoridad según [ISO 532] [Zwicker 99]. Los niveles de sonoridad, LL en fonos, se calculan usando la expresión:

$$LL = 40 + 10 \cdot \log_2 L \qquad \text{Ec. 15}$$

Se calcula la resta de los niveles de sonoridad en cada banda del banco de filtros de la señal biaural, dando como resultado la ILD.

$$ILD = LL_R - LL_L \qquad \text{Ec. 16}$$

2.6.5 TRABAJOS DE MODELOS

2.6.5.1 HUTear

[Harma 99] ha desarrollado una "toolbox" para Matlab llamado HUTear, en la cual se modela el sistema de audición humano. Con esta herramienta se puede simular la percepción del oído, pero no calcular parámetros de localización y mucho menos ángulos de estimación del evento sonoro.

El filtrado del oído medio se diseña como un filtro lineal fijo cuyos coeficientes FIR se generan con la toolbox. También es posible diseñar un filtro "frequency-warped" o IIR que modelen la respuesta de esta zona del oído.

El banco de filtros que simula la respuesta de la cóclea se puede elegir entre los filtros "gammatone" o filtros de cuarto y sexto orden "complex-valued gammatone".

Las células ciliares se modelan con un sistema de rectificación, comprensión y filtrado: la rectificación se puede elegir entre ninguna, de media onda o de onda completa. Se puede elegir entre dos compresiones: una función de compresión no lineal basada en el nivel de presión sonora o la clásica regla de compresión $y = x^c$ (típicamente $c = 0.7$).

El filtrado se implementa con un filtro paso bajo para todos los canales de frecuencia de corte 1 kHz. También existe la posibilidad de hacer un filtrado usando un integrador de primer orden con una constante de tiempo de 20 ms o usar el modelo de Meddis para las células ciliares que simula el modelo de la sinapsis (contacto entre neuronas) o un modelo neuronal refractivo. Además se puede implementar un umbral para los canales de forma que se limite el nivel mínimo a un valor.

Se pueden usar cuatro modelos neuronales diferentes implementados en la toolbox: el modelo neuronal refractivo de Meddis; modelo de adaptación neuronal de ventana temporal de Plack; modelo de una red de adaptación no lineal de Dau y el modelo de una red de adaptación no lineal de Karjalainen.

2.6.5.2 BINAURAL CUE SELECTION TOOLBOX

[Faller 04] [Faller 04a] ha desarrollado una toolbox para Matlab llamado Binaural Cue Selection Toolbox. Se basa en el modelo psicoacústico ya descrito en el modelo de coherencia interaural pero introduce algunos cambios. Esta herramienta determina la ITD, ILD y la IC (Interaural Correlation), pero no estima el ángulo asociado a dichos parámetros. Además este modelo tiene una diferencia significativa con el resto, y es que la potencia de la señal en cada instante de tiempo afecta a la ITD, la IC y la ILD, por eso en escenarios acústicos complejos (con reverberación, varias fuentes, etc…) a veces desestima parte de la señal que también influye en la localización.

Este modelo está diseñado para localizar varias fuentes diferentes en entornos ruidosos, por lo tanto implementa un modelo de efecto precedente. El cálculo de los parámetros de localización se realiza en función del tiempo.

El modelo del oído interno es configurable, se puede elegir una rectificación de media onda más compresiones de diferentes tipos. Se ha escogido una ventana exponencial con T = 10 ms (Ec. 9) que simula la constante de tiempo de inhibición del modelo biaural [Faller 04]. Sólo se considerarán los valores de $|ITD| < 1$ ms y los de $|ILD| > 7$ dB.

Se calculará la ITD y la ILD como el promedio de todos los valores calculados en el tiempo (n).

2.7 MODELOS NO PSICOACÚSTICOS

2.7.1 MODELO BASADO EN REDES NEURONALES

[Venegas 06] presenta un modelo computacional de localización sonora espacial para sonidos de banda ancha en ambiente anecoico inspirado en el sistema auditivo humano e implementado mediante redes neuronales artificiales. Este modelo entrega una estimación del ángulo de azimut y del ángulo de elevación.

Aunque en este modelo se calcula la ILD como:

$$ILD = 10\log\left(\sum_n (x_R[n])^2\right) - 10\log\left(\sum_n (x_L[n])^2\right) \qquad \text{Ec. 17}$$

Se ha comprobado a la hora de implementar el LSE que los resultados de la ILD calculada según la Ec. 17 y según la Ec. 16 por sonoridad no varían.

2.7.2 MODELO BASADO EN FILTRADO INVERSO

[Keyrouz 06] propone un algoritmo de detección del ángulo de llegada de un sonido en campo libre para uso en un robot con dos oídos. Lo que hace es simplificar la base de datos de las HRTF de su maniquí acústico a través de procesamiento con tres técnicas llamadas: ecualización del campo difuso, truncamiento del modelo balanceado y análisis de componentes principales. Después calcula una base de datos con los filtros inversos para cada posición de la base de datos. El algoritmo consiste en filtrar la señal biaural con cada pareja de filtros inversos de cada posición. Se supone que al filtrar con el filtro inverso se recuperará el sonido de la fuente "puro", en ambos oídos, para aquel filtrado que compense la respuesta de las HRTF según el ángulo de llegada. Para saber cuando se ha recuperado el sonido de la fuente, se correlaciona las dos salidas (izquierda y derecha), y cuando la correlación sea máxima (la misma señal) ese es el ángulo de llegada.

Este algoritmo a pesar de no basarse en un modelo psicoacústico da muy buenos resultados excepto para el plano medio, donde las señales biaurales son idénticas y por lo tanto la correlación es máxima para todos los ángulos de elevación.

2.7.3 MODELO BIAURAL STFT

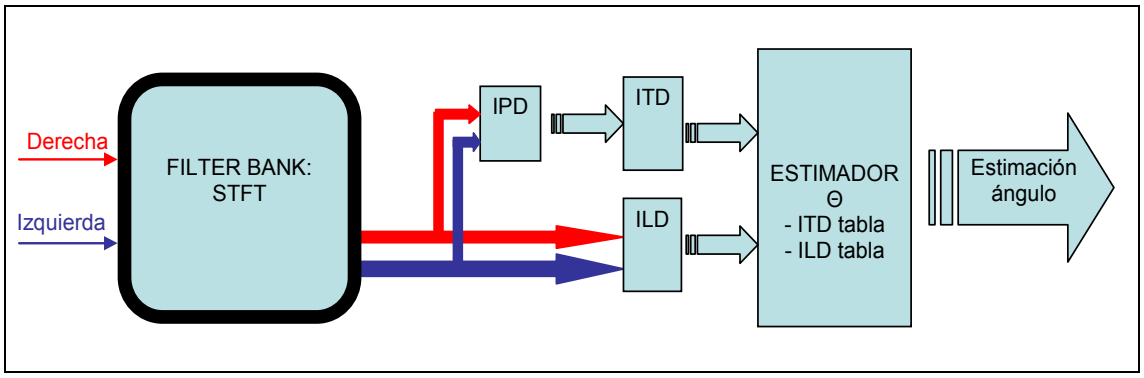


Figura 29. Método de localización biaural basado en la STFT.

Este método biaural utiliza la STFT (Short Time Fourier Transform) para procesar la señal biaural en el dominio de la frecuencia [Viste 04]. La ITD se calcula a partir de la diferencia de fase interaural (IPD), por lo tanto aparece un problema de ambigüedad en el cálculo de la fase debido a su periocidad. El significado práctico de esto es que para una frecuencia, la diferencia de fase que existe entre los dos oídos puede ser debida a la posición de la fuente sonora en varios sitios.

$$IPD(t,f,p) = \arg\frac{X_R(t,f)}{X_L(t,f)} + 2\pi p \qquad \text{Ec. 18}$$

$$ITD(t,f,p) = -\frac{IPD(t,f,p)}{f}\frac{1}{2\pi} \qquad \text{Ec. 19}$$

Donde t es la ventana de tiempo, f es el coeficiente espectral, X_R y X_L son las señales biaurales derecha e izquierda y p es el índice de periocidad.

LA ILD en dB viene dada por:

$$ILD(t,f) = 20\log\left|\frac{X_R(t,f)}{X_L(t,f)}\right| \qquad \text{Ec. 20}$$

La estimación del azimut basada en la ILD, $\theta_{ILD,}$ se usa para seleccionar de entre las posibles estimaciones de azimut basadas en ITD, $\theta_{ITD,}$ la más cercana a ella. La estimación basada en la ITD es más precisa pero ambigua debido a la periocidad, sin embargo la estimación basada en la ILD es menos precisa.

$$\theta(t,f) = \theta_{ITD}(t,f,p)\Big|_{p=\arg\min|\theta_{ILD}(t,f)-\theta_{ITD}(t,f,p)|} \qquad \text{Ec. 21}$$

Se utilizan dos tablas de búsqueda para estimar el azimut a partir del cálculo de la ITD y de la ILD (θ_{ILD}, θ_{ITD}). Ambas tablas tienen los valores de ITD e ILD calculadas a partir de las medidas de las HRTF del maniquí acústico. Dichos cálculos se realizan para cada ángulo de azimut y para cada frecuencia según las siguientes ecuaciones:

$$IPD(\theta,f) = \arg\frac{HRTF_R(\theta,f)}{HRTF_L(\theta,f)} \qquad \text{Ec. 22}$$

$$ITD(\theta,f) = -\frac{IPD(\theta,p)}{f}\frac{1}{2\pi} \qquad \text{Ec. 23}$$

$$ILD(\theta,f) = 20\log\left|\frac{HRTF_R(\theta,f)}{HRTF_L(\theta,f)}\right| \qquad \text{Ec. 24}$$

Las funciones ILD e ITD pueden no ser monotónicas para alguna frecuencia y por lo tanto en la búsqueda del $\theta_{ITD}(t,f)$ y del $\theta_{ILD}(t,f)$ se pueden encontrar varios valores posibles. La solución por la que se ha optado es la elección de la estimación más próxima a 0º (aunque puede que no sea la correcta). Según se ha estudiado subjetivamente [Blauert 97] el sistema auditivo también opta por la fuente más cercana a 0º cuando hay conflicto de interpretaciones.

Este modelo utiliza los dos parámetros de localización de la Teoría Dúplex y por lo tanto combina las precisiones de la ITD y de la ILD que dependen del margen de frecuencias.

2.8 ESTIMADORES DE ÁNGULO

De los modelos de audición se obtiene dos informaciones la ITD y la ILD, que son los parámetros de localización en el plano horizontal, esta información servirá para estimar el ángulo de azimut en el que el oyente localiza el evento sonoro. Se pueden utilizar diversos estimadores que se enumeran a continuación.

2.8.1 ESTIMADOR 1

Propuesto por Woodworth-Schlosberg [Blauert 97] para un modelo de cabeza esférica y ondas planas.

$$ITD \cdot c = \frac{D}{2}(\theta + \text{sen}\theta) \qquad \text{Ec. 25}$$

donde D = 0.152 m. Diámetro de la cabeza.

2.8.2 ESTIMADOR 2

Propuesto por Kuhn [Kuhn 77] para un modelo de cabeza esférica con difracción que depende de la frecuencia y el diámetro de la cabeza.

$$ITD \cdot c = \frac{3}{2} D\,\text{sen}\theta \quad f \le 500\ \text{Hz}$$
Valores intermedios para el otro rango. Ec. 26
$$ITD \cdot c = D\,\text{sen}\theta \quad f > 3\ \text{kHz}$$

2.8.3 ESTIMADOR 3

La base de datos de las HRTF que vienen dadas en función de cada ángulo, se procesan con el modelo biaural. A su salida se obtienen los valores de ITD e ILD para cada filtro del banco de filtros en cada una de las direcciones. Se crean dos tablas de búsqueda donde se colocan dichos valores respectivamente.

Cuando en el procesado de una señal biaural se calcula la ILD y la ITD en función de la frecuencia, se busca el ángulo en las tablas que tenga la ITD o ILD más cercana para cada una de las frecuencias.

A veces estas tablas se suavizan para reducir los errores [Pulkki 01] [Viste 04]. Este suavizado se ha realizado según el eje de frecuencia. A pesar de esto, a veces aparecen dentro de la misma frecuencia dos ángulos con el mismo valor de ITD o ILD, provocando una ambigüedad.

2.8.4 OTROS ESTIMADORES

Existen otros estimadores no evaluados porque según la bibliografía dan peores resultados que los basados en la búsqueda por tabla. Según [Viste 04] se podrían utilizar las expresiones:

$$ITD(\theta, f) = \alpha_q \frac{a(\text{sen}\theta + \theta)}{c} \quad ILD(\theta, f) = \beta_q \frac{\text{sen}\theta}{c} \qquad \text{Ec. 27}$$

Los parámetros α_q y β_q son factores de escala que se calculan a partir de las HRTF, haciendo la media para todas las funciones.

2.9 EVALUACIÓN DE LAS ESTIMACIONES

En la búsqueda de un buen algoritmo para localizar eventos sonoros se han implementado diferentes modelos. Para comparar la fiabilidad de la estimación de unos frente a otros, se

han calculado dos valores como indicadores. Por una parte el ángulo estimado, sabiendo en qué ángulo está la fuente colocada, y por otra parte la frecuencia de dicha estimación. El ángulo estimado θ se calcula como la moda estadística de todos los valores de azimut estimados en todas las bandas "perceptuales". Si existen dos modas se da como estimación la media.

La frecuencia estadística en % se calcula como la cantidad de veces que se repite la moda. No se ha elegido la media porque si en alguna banda el valor del ángulo estimado falla dando un valor muy lateral, este valor desviará mucho la media.

Otras formas de evaluar la estimación obtenida sería la propuesta por [Sontacchi 02]. Propone evaluar varios parámetros en los que se juntan las estimaciones mediante ILD e ITD.

Uno de ellos es la función de localización como la media de las estimaciones de las diferentes bandas ponderadas:

$$L = \frac{1}{2}\left(\frac{1}{\sum w_{ITD}(f)}\sum_{z=1}^{n} w_{ITD}(f)\theta_{ITD}(f)\right) + \frac{1}{2}\left(\frac{1}{\sum w_{ILD}(f)}\sum_{z=1}^{n} w_{ILD}(f)\theta_{ILD}(f)\right) \quad \text{Ec. 28}$$

Siendo w los valores de las ponderaciones (para ILD la ponderación es la unidad para todas las bandas y para la ITD la ponderación cae desde 800 Hz 0.2 unidades por banda). De esta forma se da más valor a las ITD de baja frecuencia y sólo se da valor a las ILD en altas frecuencias. f es cada banda "perceptual".

El otro parámetro es la desviación estándar:

$$BL = \sqrt{\frac{1}{2}\sum_{i=ILD}^{ITD}\frac{1}{\sum w_i(f)}\sum_{z=1}^{n} w_i(f)\left[\theta_i(f) - L\right]^2} \quad \text{Ec. 29}$$

En el LSE, los valores de los ángulos estimados a partir de la ILD son bastantes irregulares, por lo que no se va a usar este tipo de evaluación que fusiona las estimaciones según la ITD y la ILD. Existe mucha diferencia en el error cometido de la estimación basada en la ITD frente a la basada en la ILD.

2.10MODELO LOCALIZACIÓN EN EL PLANO MEDIO

A lo largo del análisis de toda la bibliografía, se ha visto que existe muchos estudios analizando los diferentes aspectos de la localización en el plano horizontal pero pocos en el plano medio o cualquier otro punto del espacio.

Buscando referencias que localicen en el plano medio se han encontrado sólo dos algoritmos, uno es el propuesto por [Blauert 07] (Figura 30), que no realiza una estimación del ángulo de elevación sino que da como resultado solamente una diferenciación entre arriba, delante y detrás. El otro algoritmo encontrado es el de [Keyrouz 06] que ya se ha explicado antes y no estima el ángulo en el plano medio, una solución que da dicho autor en referencias posteriores es utilizar un tercer micrófono alejado un poco del robot para localizar en este plano ("localización triaural").

Por lo tanto, en esta Tesis se ha desarrollado un nuevo algoritmo basado en los parámetros de localización monoaurales, que se explicará en el capítulo siguiente, dando respuesta a este vacío en este campo de la localización espacial.

2.10.1 ALGORITMO DE BLAUERT

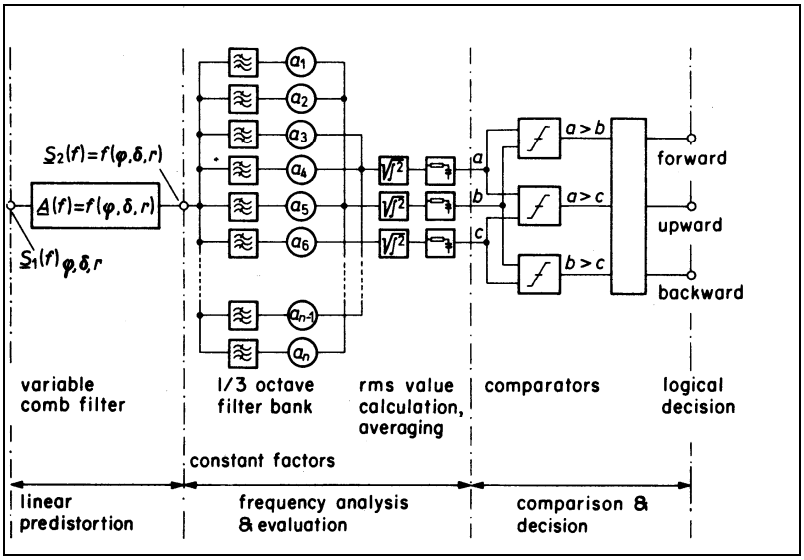


Figura 30. Modelo hipotético de evaluación de la señal de entrada al oído en el caso de escucha en el plano medio [Blauert 97].

La señal recibida se filtra con A(f) que es un filtro variable tipo peine en función del ángulo de llegada. Este filtrado es el que hace la HRTF. Después se filtra en tercios de octava y se pondera en función de la banda direccional a la que pertenece y la influencia que tiene en la estimación de la dirección. Se suman las tres bandas direccionales (delante, detrás y arriba). Se halla la potencia y se promedia en el tiempo cada una de estas tres bandas direccionales para luego compararlas con un valor umbral y decidir si la localización del evento sonoro está en una de estas tres direcciones.

Este modelo no tiene en cuenta la influencia de la familiaridad del sonido en su localización, esta variación a nivel cognitivo es muy difícil de modelar. Otro problema de este modelo es que analiza la potencia en bandas por lo tanto necesita sonido estacionarios o cuasi-estacionarios.

2.11 RESUMEN

El sistema auditivo humano hace uso del conocimiento interno de sus HRTF para identificar la dirección de procedencia del sonido. Por lo tanto es imprescindible contar con unas buenas medidas de HRTF para poder simular señales biaurales (utilizando un buen método de interpolación); y para poder simular un sistema de percepción humana, donde se analizan diferencias de nivel, de tiempo, máximos y mínimos de las señales que llegan a los oídos. Además, el conocimiento profundo de los parámetros de localización espacial, por un lado la Teoría Dúplex en el plano horizontal y por otro, los parámetros espectrales en el plano medio, permiten implementar modelos aproximados que simulan el procesado que realiza nuestro sistema auditivo para estimar la dirección del evento sonoro con una precisión bastante buena.

El modelo dentro del plano horizontal consiste en comparar las dos señales biaurales y determinar diferencias de fase o tiempo de llegada y diferencias de nivel que dependerán del ángulo de azimut y de la frecuencia. Sin embargo, en el plano medio sólo hay una señal a analizar porque la señal biaural es igual para los dos oídos, y el análisis, mucho más complejo, se realiza determinando donde están posicionados los nodos y picos de la señal

monoaural, así como la distribución de energía a lo largo de la frecuencia. Este análisis se realiza en el dominio de la frecuencia según las bandas “perceptuales” y teniendo en cuenta la consistencia de la percepción, que se debe de repetir en un ancho de banda grande.

La plasmación práctica de toda la teoría anterior, la aplicación y el análisis de los distintos modelos anteriormente mencionados, que como consecuencia han llevado a la construcción del Localizador de Eventos Sonoros, LES, están desarrollados en los capítulos siguientes.

Capítulo 3: LOCALIZACIÓN EN EL PLANO HORIZONTAL

3.1 TRABAJO DE INVESTIGACIÓN

En este capítulo se expone el trabajo de investigación desarrollado, así como los estudios comparativos y resultados obtenidos, en la búsqueda de un buen método de localización subjetiva dentro del plano horizontal. Como los mecanismos de localización en este plano y en el plano medio son claramente distintos, se ha dejado para el siguiente capítulo el estudio de la localización en el plano medio. Los métodos de estimación del ángulo en ambos casos se basan en las HRTF, por lo tanto, la primera parte del presente capítulo consistirá en medir la base de datos de nuestro maniquí acústico HATS.

Una vez implementado el Localizador de Eventos Sonoros se va a estudiar la percepción en campo libre del sonido de una fuente sonora, tanto simulada como real para comprobar la eficacia de la estimación. Así se determinará el nivel de precisión de dicha herramienta.

Posteriormente se estudia el escenario acústico objeto de esta Tesis, el sistema de teleconferencia descrito con la ventana acústica virtual, de esta forma se podrán analizar y valorar si los problemas de estimación dependen de la herramienta (problemas acotados en el experimento anterior) o del sistema multicanal. El estudio del escenario acústico planteado en el sistema de teleconferencia se hace a través de un programa de simulación implementado con Matlab donde se puede elegir la estructura y configuración del sistema multicanal pudiéndose, por lo tanto, seleccionar diferentes escenarios con diferentes configuraciones en los arrays que forman la ventana acústica virtual (ver Anexo 3-1).

En una última fase se implementa físicamente la parte de reproducción del sistema de teleconferencia (array de altavoces) captando la señal biaural con un maniquí acústico en una sala anecoica; esta fase permitirá situarse en una reproducción real, descartando los problemas de simulación que pudieran influir en la estimación. Es necesario comprobar la robustez de la herramienta en todos estos escenarios acústicos para analizar su eficacia y su funcionamiento real y simulado. Se pretende como objetivo final que el localizador LES siempre trabaje a partir de la señal biaural, y esta pueda ser simulada dentro de la aplicación de Matlab o medida con un maniquí acústico en una configuración de sistema multicanal real.

Por lo tanto, los siguientes apartados consisten en el desarrollo de la herramienta capaz de estimar el ángulo percibido de un evento sonoro LES (Localizador de Eventos Sonoros). El método de estimación se basa en la señal biaural para el plano horizontal y en la señal monoaural para el plano medio. El procesamiento de dichas señales pretende reproducir el modelo de percepción humana clásico [Blauert 97]. Algunos de los modelos del sistema auditivo a nivel fisiológico y cognitivo que aparecen en la bibliografía han sido estudiados,

implementados y revisados para analizar su eficiencia a la hora de estimar el ángulo de azimut en el plano horizontal, incorporando parte de ellos al modelo implementado en el LES.

Todo el equipamiento utilizado en los apartados siguientes se describe técnicamente en el apartado de "Equipamiento" del Capítulo 4.

3.2 BASES DE DATOS

Tanto para el proceso de simulación de los entornos acústicos y de simulación de la señal biaural captada, como para los métodos de estimación del azimut y la elevación, es necesario el conocimiento de las HRTF (Figura 1). Se ha realizado una base de HRTF propia, basada en el maniquí acústico HATS (fabricado por Head Acoustics), ya que es este maniquí el que se usa para las pruebas en entornos reales.

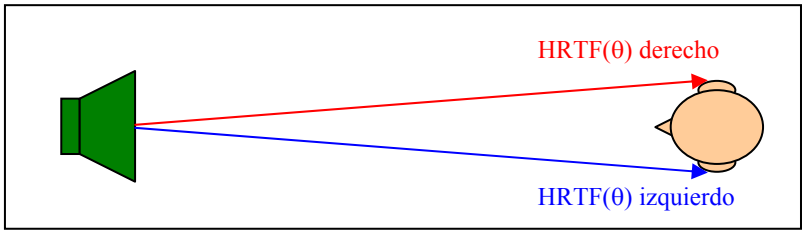


Figura 1. Funciones de transferencia de la cabeza para cada oído.

Además, en las pruebas iniciales realizadas para esta Tesis [Gómez-Alfageme 04] [Blanco-Martín 05] y en las pruebas que se van a exponer para comprobar la robustez del LES frente a diferentes señales biaurales, generadas con diferentes maniquís o bases de datos, se ha usado la base de datos de KEMAR desarrollada en el MIT [Gardner 00], ampliamente conocida e utilizada en este campo.

3.2.1 BASE DE DATOS DE KEMAR

En concreto, las medidas realizadas en el MIT Media Lab [Gardner 00a] para el torso acústico KEMAR en el plano horizontal tienen una precisión de 5º y de 10º en el plano medio. Las hrir de KEMAR se midieron colocando el altavoz siempre a la misma distancia, 1.4 m, y utilizando señales MLS (Maximum Length Sequences). Después se cogió una ventana de 128 muestras donde se encontraba la parte representativa de la respuesta, truncando la respuesta al impulso. La respuesta de la sala y del altavoz fue compensada midiendo la respuesta del altavoz, realizando un filtro inverso de la respuesta y filtrando cada hrir.

Por lo tanto, el catálogo de KEMAR no tiene información relativa al retardo entre el altavoz y el oyente (información de distancia), sólo tiene información del retardo según el ángulo. Para poder tener información de la distancia de la fuente hay que retardar la hrir según la distancia r entre fuente y maniquí. Para ello hay que determinar donde comienza la respuesta de la hrir según el razonamiento siguiente.

La hrir que antes llega al oído es la medida para 90º de azimut en el oído derecho. Se ha supuesto que la respuesta comienza en la primera muestra que alcanza el 10% del valor máximo de la energía (muestra 5). Además como la distancia se calcula al centro de la cabeza se ha adelantado esta posición el número de muestras que llega antes el sonido a la oreja, suponiendo que el diámetro de la cabeza es 0.152 m. Se van a hacer los cálculos para

que la base de datos esté normalizada a 1 m, es decir, crear una base de datos de hrir como si hubieran sido medidas con la fuente a 1 m de distancia. Luego el número de ceros que hay que añadir a todas las hrir al principio es:

$$\text{nº ceros} = \text{round}\left(\frac{1-0.152/2}{343} f_{\text{muestreo}}\right) - 5 = 114 \qquad \text{Ec. 1}$$

Donde f_{muestreo} es 44100 kHz.

La base de datos consistirá en la respuesta en tiempo de 242 (128+114) muestras para 37 ángulos (-90º a 90º cada 5º) para los dos oídos (izquierdo y derecho) en el plano horizontal.

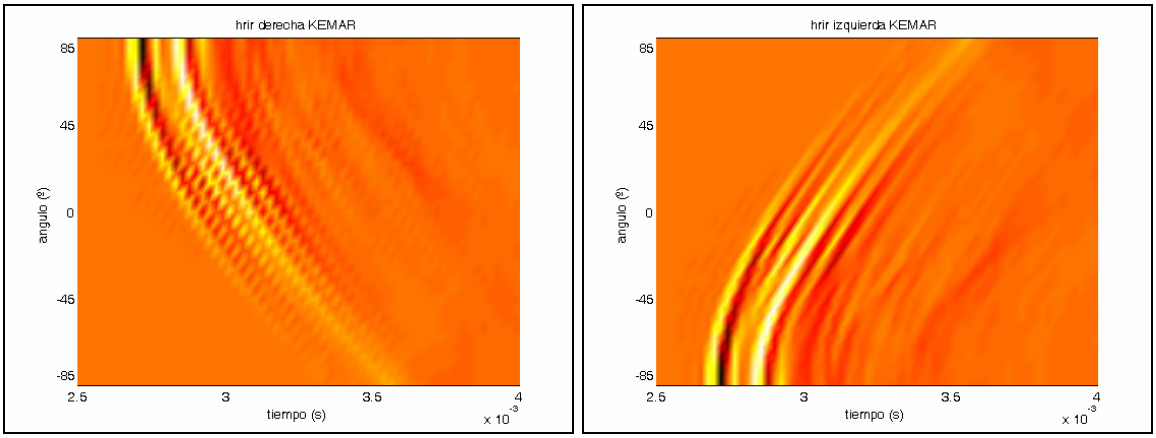


Figura 2. hrir del oído derecho e izquierdo, a 1 m, del maniquí KEMAR en el plano horizontal.

Las HRTF se calculan con la transformada de Fourier, fft, aplicada a las hrir.

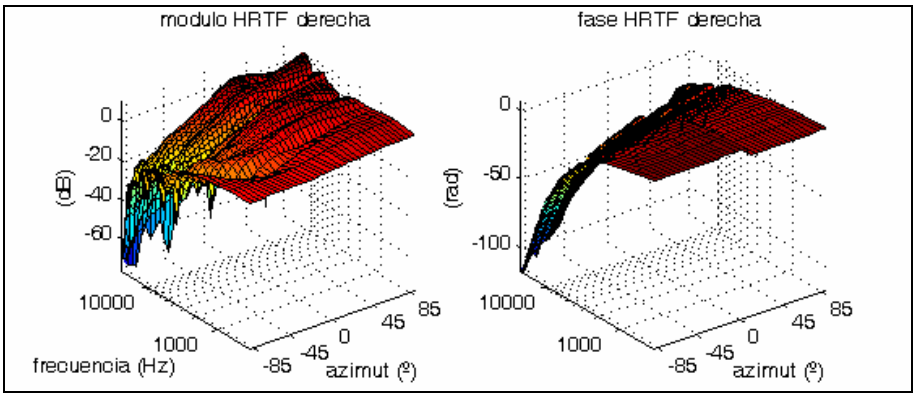


Figura 3. HRTF del oído derecho del maniquí KEMAR en el plano horizontal.

3.2.2 BASE DE DATOS DE HATS

El maniquí acústico con el que se realizan las medidas en esta Tesis es el HATS (de Head Acoustics). Por lo tanto, se ha creído conveniente medir la función de transferencia para dicho maniquí, creando la base de datos necesaria para el LES. En nuestro caso se ha utilizado un método de medida de la HRTF que independiza totalmente el sistema y entorno de medida de la HRTF. Es un método más complejo que el utilizado en la medida de KEMAR, porque surge la necesidad de procesar la señal que proporciona el sistema de medida según se explica a continuación.

Se han realizado las medidas con el sistema de medida PULSE de Brüel & Kjaer en la cámara anecoica del Departamento de Ingeniería Audiovisual y Comunicaciones de la Escuela Universitaria de Ingenieros Técnicos de Telecomunicación, que se ha cualificado y caracterizado convenientemente [Gomez-Alfageme 07]. La resolución de medida en el plano horizontal es de 5º de azimut, desde -90º a 90º; y en el plano medio de 10º, desde -40º hasta 90º. El altavoz se ha colocado a 1.5 m. La señal de prueba utilizada es un ruido blanco y el tiempo promediado 3 s. Según la definición de función de transferencia de la cabeza, la HRTF es la respuesta de cada oído referida a la respuesta cuando no está la cabeza y se coloca un micrófono en la posición central de la cabeza. La medida se realiza en dos fases, primero se mide la función de transferencia de cada oído para cada ángulo (Figura 4).

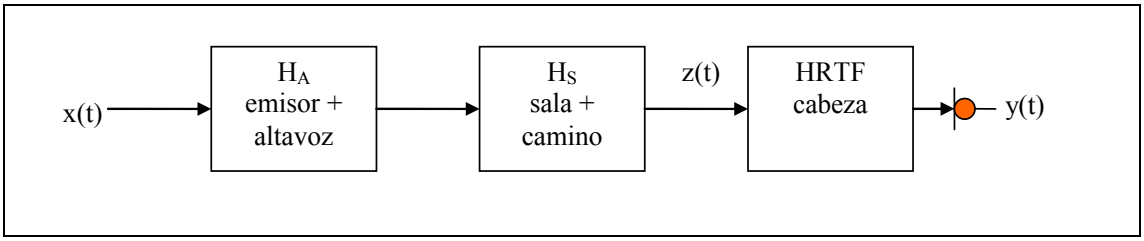


Figura 4. Funciones de transferencia involucradas en la medida con un micrófono dentro del oído del maniquí acústico.

Después se mide la función de transferencia de un micrófono cuando se coloca en la posición del punto medio de la cabeza (sin que esta esté). Hay que recordar que se quiere medir la influencia de la cabeza y del torso en el sonido recibido en cada oído. El diagrama de bloques de medida se muestra en la Figura 5.

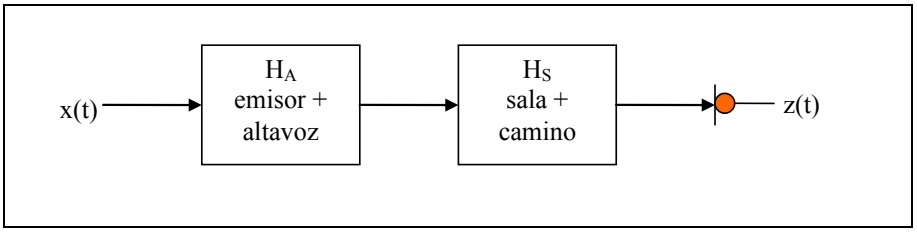


Figura 5. Funciones de transferencia involucradas en la medida con un micrófono en la posición del centro de la cabeza.

El sistema de medida PULSE puede medir los autoespectros de las señales de entrada y salida: x(t), y(t) en el primer caso y x(t), z(t) en el segundo como:

$$C_{XX} = X^{*}X \quad C_{YY} = Y^{*}Y \quad C_{ZZ} = Z^{*}Z \qquad \text{Ec. 2}$$

Además mide los cross-espectros entre la salida y la entrada:

$$\begin{aligned} C_{ZX} &= Z^{*}X \\ C_{YX} &= Y^{*}X \end{aligned} \qquad \text{Ec. 3}$$

De la primera situación se obtiene:

$$\frac{Y(f)}{X(f)} = \left[\frac{Y^{*}}{X^{*}}\right]^{*} = \left[\frac{Y^{*}X}{X^{*}X}\right]^{*} = \left[\frac{C_{YX}}{C_{XX}}\right]^{*} \qquad \text{Ec. 4}$$

De la segunda situación se obtiene:

$$\frac{Z(f)}{X(f)} = \left[\frac{Z*}{X*}\right]^* = \left[\frac{Z*X}{X*X}\right]^* = \left[\frac{C_{ZX}}{C_{XX}}\right]^* \qquad \text{Ec. 5}$$

Luego la HRTF se calculará como:

$$\frac{Y(f)}{Z(f)} = HRTF = \frac{Y(f)}{X(f)}\frac{X(f)}{Z(f)} = \left[\frac{Y*X}{X*X}\right]^*\left[\frac{X*X}{Z*X}\right]^* = \left[\frac{C_{YX}}{C_{XX}}\right]^*\left[\frac{C_{XX}}{C_{ZX}}\right]^* \qquad \text{Ec. 6}$$

En el sistema PULSE se puede programar una función donde a partir de las señales medidas proporcione la HRTF como:

$$HRTF = \left[\frac{C_{YX}}{C_{XX}}\right]^*\left[\frac{C_{XX}}{C_{ZX}}\right]^* = \left[\frac{C_{YX}/C_{XX}}{C_{ZX}/C_{XX}}\right]^* \qquad \text{Ec. 7}$$

El sistema de medida proporciona la HRTF, a continuación se exportan los datos a Matlab y se realiza la ifft (inverse fast Fourier transform) para obtener la respuesta impulsiva de cada ángulo, hrir. Un aspecto muy importante es que hay que tener en cuenta que, para aquellos ángulos donde el oído esté adelantado con respecto a la posición del micrófono en el centro de la cabeza, la hrir comenzará antes del origen del tiempo y por lo tanto, al calcular la transformada inversa, el comienzo de la respuesta aparece en la zona final de la ifft. Luego es necesario introducir un desplazamiento a la hrir que tenga en cuenta el retardo producido por la distancia entre el altavoz y el oído, colocando toda la respuesta hrir a continuación. Se va a normalizar la base de hrir para distancias de 1 m, análogamente a como se hizo anteriormente con la base de KEMAR.

Para normalizar la base de datos a la distancia de 1 m y sabiendo que la muestra donde comienza la respuesta (primera muestra que alcanza el 10% del valor máximo de la energía) es la muestra 2043, habrá que desplazar este comienzo de la hrir al nº de muestra que debería estar para una distancia de 1 m:

$$n^o\,\text{muestra} = \text{round}\left(\frac{1-0.152/2}{343} f_{\text{muestreo}}\right) = 176 \qquad \text{Ec. 8}$$

donde $f_{muestreo}$ es 65536 kHz y 0.152 el radio de la cabeza.

A continuación se muestra en la Figura 6 las hrir en el plano horizontal para los dos oídos, una vez se ha normalizado a 1 m.

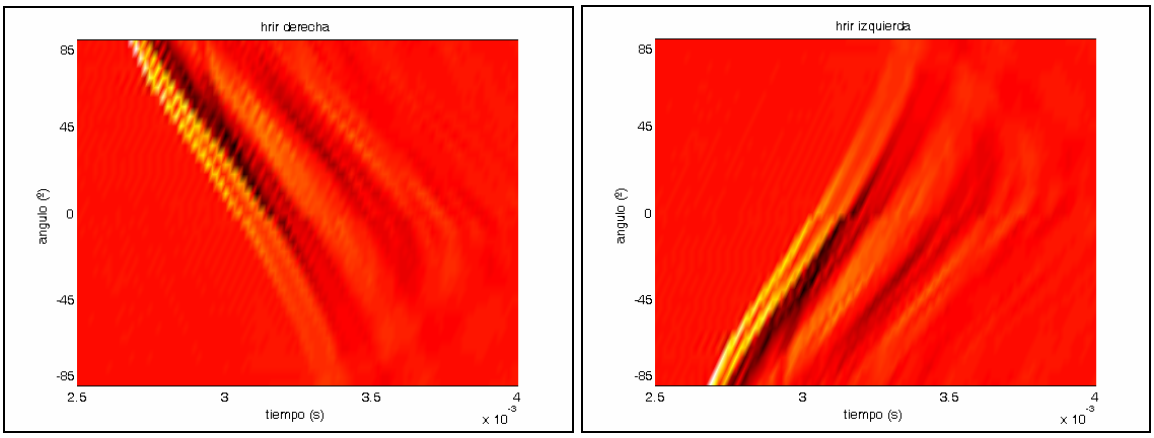


Figura 6. hrir del oído derecho e izquierdo, a 1 m, del maniquí HATS en el plano horizontal.

Además se muestra en la Figura 7 las HRTF del oído derecho, módulo y fase, que se ve que es muy parecida a la de KEMAR mostrada en el apartado anterior.

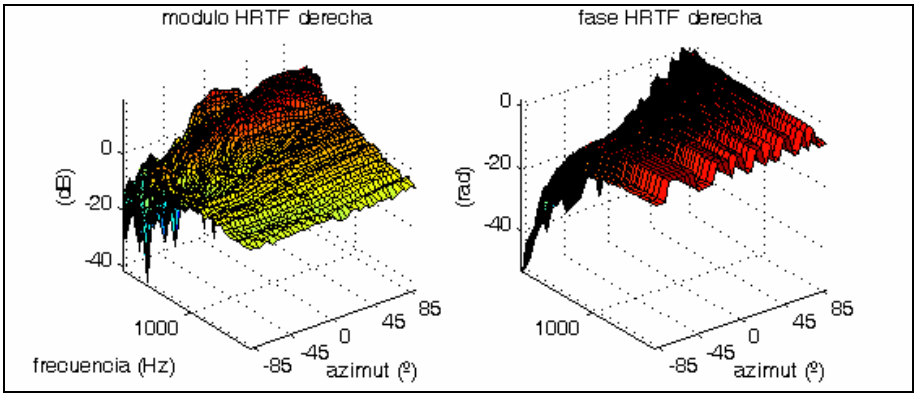


Figura 7. HRTF del oído derecho del maniquí HATS en el plano horizontal.

3.2.3 COMPARACIÓN CON OTRAS BASES

Cada laboratorio que ha generado una base de datos de HRTF lo hace de forma diferente, dependiendo de los medios y sistemas de medida de los que dispone. En cualquier caso es imprescindible conseguir aislar la respuesta en frecuencia de la cabeza del resto de funciones de transferencias involucradas en la medida: sala, sistema de excitación y medida, altavoces, micrófonos, etc. Por ejemplo, en la base de datos de KEMAR [Gardner 95] se realiza una ecualización con la respuesta inversa del altavoz para cada respuesta impulsiva medida para cada ángulo. Este es el método descrito en [Begault 04] para la medida de la hrir. La respuesta de la sala se supone no influye porque las medidas han sido realizadas en una cámara anecoica y además truncan las hrir para evitar las posibles reflexiones. La señal utilizada es una MLS (Maximum Length Sequences).

En la base de datos desarrollada en el CIPIC [Algazi 01] para el KEMAR, se utilizan señales "Golay-code" para medir las hrir. Se aplica una ventana Hanning modificada para quitar las reflexiones de la sala (que no es anecoica). Los resultados son ecualizados con una señal llamada campo libre que es medida en la posición del centro de la cabeza, para corregir las características espectrales de los transductores. Esta señal es limitada a 15 dB para evitar las colas de la respuesta.

En ambas bases de datos se utilizan medidas de respuesta al impulso que luego son ecualizadas evitando en mayor o menor medida la influencia de la sala, en ambas se hace

una limitación temporal para evitar las reflexiones de la sala. En el método propuesto en esta Tesis lo que se mide es la HRTF directamente, independizando la medida de la respuesta de la sala, camino aéreo, transductores, sistemas de medida, etc. Además se utiliza un sistema de medida de doble canal que permite calcular el espectro cruzado instantáneamente, independizando la medida de la señal de excitación. En los métodos de medida donde se mide la hrir se utilizan sistemas de un solo canal y por lo tanto la señal de medida deben ser MLS o Golay-code, necesitando un procesado posterior para obtener la respuesta impulsiva.

3.2.4 INTERPOLACIÓN DE LAS HRIR

La necesidad de tener las HRTF para cualquier ángulo posible lleva al estudio de las interpolaciones de las HRTF. Este tema sigue muy candente produciéndose continuas aportaciones científicas como se puede ver en las últimas ediciones del JAES Marzo 2008 [Ajdler 08] y JAES Enero/Febrero 2008 [Keyrouz 08], ya que una HRTF mejor interpolada lleva a la síntesis de una señal biaural más real, y por lo tanto a un sonido 3D más natural.

Hemos realizado un estudio de dos de los métodos de interpolación y se ha llegado a la conclusión que el mejor, sin incrementar en exceso el cálculo numérico, es el método de la distancia inversa [Hartung 99].

Para obtener la señal biaural en la posición del oyente es necesario utilizar la función de transferencia de la cabeza, HRTF, para cada uno de los oídos. Estas HRTF dependen de la posición relativa entre el altavoz y el oído. Las posiciones posibles son infinitas, pero las medidas realizadas en las bases de datos se han llevado a cabo para un conjunto discreto de posiciones.

En nuestro trabajo la posición relativa oídos-altavoz puede ser cualquiera, por lo que las HRTF utilizadas deben interpolarse entre las medidas de las bases de datos para el plano horizontal o el plano medio. Así que es necesario establecer alguna herramienta de interpolación. Hemos estudiado la interpolación por el método de ponderación de la distancia inversa frente a una interpolación por media [Gómez-Alfageme 04]. Teniendo en cuenta que el error en la estimación del azimut en pruebas subjetivas [Blauert 97] es mayor o igual a 5º, creemos innecesario utilizar otros métodos de interpolación más complejos que no mejoran sustancialmente los resultados de la interpolación, ya que la propia resolución humana en la localización es peor.

Método de la ponderación de la distancia inversa

El método de la ponderación de la distancia inversa [Hartung 99] se basa en un algoritmo que estima un valor x_j como la suma ponderada de los valores medidos en las dos posiciones más próximas x_1 y x_2.

$$x_j = \frac{w_{1j}x_1 + w_{2j}x_2}{w_{1j} + w_{2j}} \qquad \text{Ec. 9}$$

Los valores están ponderados con la distancia linear inversa:

$$w_{ij} = \frac{1}{d_{ij}} \qquad \text{Ec. 10}$$

En una esfera de radio uno, la d_{ij} se calcula como la distancia sobre la circunferencia:

$$d_{ij} = \cos^{-1}\left[\cos\left(\theta_i - \theta_j\right)\right] \qquad \text{Ec. 11}$$

Donde θ es el azimut, o en su caso la elevación, en el sistema de coordenadas referido a la cabeza.

Método de la media

Este algoritmo estima un valor x_j como la media de los valores medidos en las dos posiciones más próximas x_1 y x_2:

$$x_j = \frac{x_1 + x_2}{2} \qquad \text{Ec. 12}$$

3.2.4.1 ESTUDIO DE LAS INTERPOLACIONES

La base de datos de KEMAR consiste en las respuestas al impulso relativas a la cabeza, hrir, con 128 muestras por respuesta y cuya frecuencia de muestreo es 44.1 kHz. Por lo tanto, se ha realizado el estudio de las interpolaciones tanto en el dominio de la frecuencia como en el dominio del tiempo. En el dominio de la frecuencia, se ha calculado la transformada de Fourier discreta de la respuesta al impulso y se ha interpolando el nivel del módulo y la fase.

$$\text{Nivel} = 20\log\left|\text{HRTF}\right| \qquad \text{Ángulo} = \angle\text{HRTF} \qquad \text{Ec. 13}$$

Es decir, se han interpolado las hrir y las HRTF para comparar los resultados y ver en qué dominio es mejor interpolar. Se han realizado dos series de interpolaciones. En la primera serie se han interpolado a partir de las hrir separadas 10º las hrir intermedias (separadas 5º), para compararlas con las correspondientes que han sido medidas y están en la base de datos. En la segunda, se han interpolado a partir de las hrir separadas 15º las hrir intermedias separadas 5º.

	Ángulos
Medidos KEMAR	-90º -85º -80º -75º -70º -65º -60º -55º -50º -45º -40º -35º -30º -25º -20º -15º -10º -5º 0º 5º 10º 15º 20º 25º 30º 35º 40º 45º 50º 55º 60º 65º 70º 75º 80º 85º 90º
Interpolados 1ª serie	-90º -85º -80º -75º -70º -65º -60º -55º -50º -45º -40º -35º -30º -25º -20º -15º -10º -5º 0º 5º 10º 15º 20º 25º 30º 35º 40º 45º 50º 55º 60º 65º 70º 75º 80º 85º 90º
Interpolados 2ª serie	-90º -85º -80º -75º -70º -65º -60º -55º -50º -45º -40º -35º -30º -25º -20º -15º -10º -5º 0º 5º 10º 15º 20º 25º 30º 35º 40º 45º 50º 55º 60º 65º 70º 75º 80º 85º 90º

Tabla 1. Ángulos medidos e interpolados.

La comparación se basa en el error cuadrático medio entre la función hrir medida, de la base de datos, de KEMAR y la interpolada para un mismo ángulo de azimut.

$$ECM = \sqrt{\frac{1}{j}\sum_{j}\left(x_{medido} - x_{interpolado}\right)^2} \qquad \text{Ec. 14}$$

En la tabla siguiente se muestran los resultados de las interpolaciones. En la primera fila aparecen los errores cuadráticos medios (ECM) de las interpolaciones de las hrir hechas con el método de "media" y con el método de la distancia inversa, "distancia"; para la primera serie "10" y la segunda "15".

En la segunda fila aparece el ECM calculado sobre las hrir cuando se interpola en el dominio de la frecuencia. Para ello las hrir originales se han transformado en HRTF, se ha calculado el módulo y la fase, se ha interpolado el módulo y la fase para las dos series y se han vuelto a convertir las HRTF interpoladas en hrir calculando el ECM.

En la tercera fila aparece el ECM calculado para el módulo y la fase de la HRTF.

	Media 10	**Distancia 10**	**Media 15**	**Distancia 15**
ECM hrir	0.0544	0.0544	0.0775	0.0733
ECM hrir (interpolando HRTF)	0.0822	0.0822	0.0948	0.0862
ECM módulo (HRTF)	1.6277	1.6277	1.9029	1.7851
ECM fase (HRTF)	1.9819	1.9819	2.0410	2.0602

Tabla 2. Error cuadrático medio de las interpolaciones para la base de datos de KEMAR.

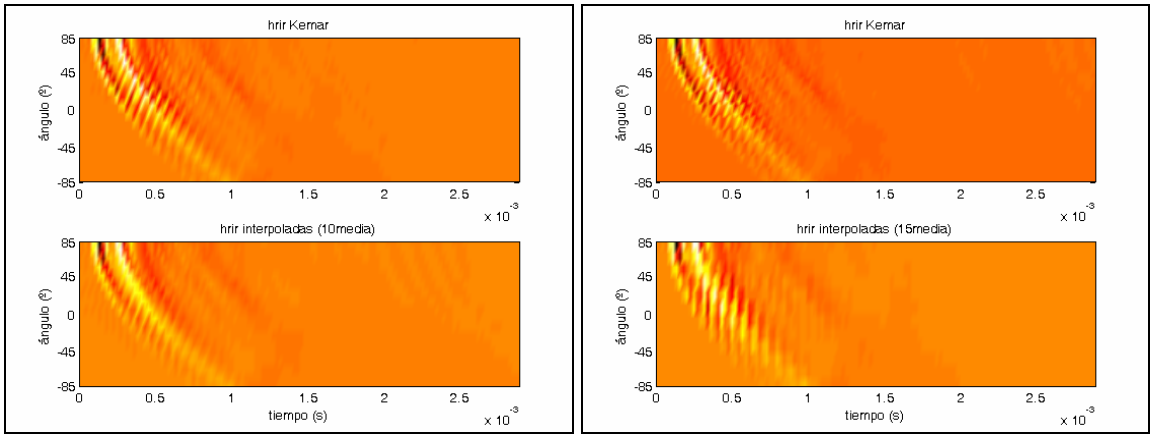


Figura 8. hrir medidas e interpoladas con el método de la media, serie 10 y 15.

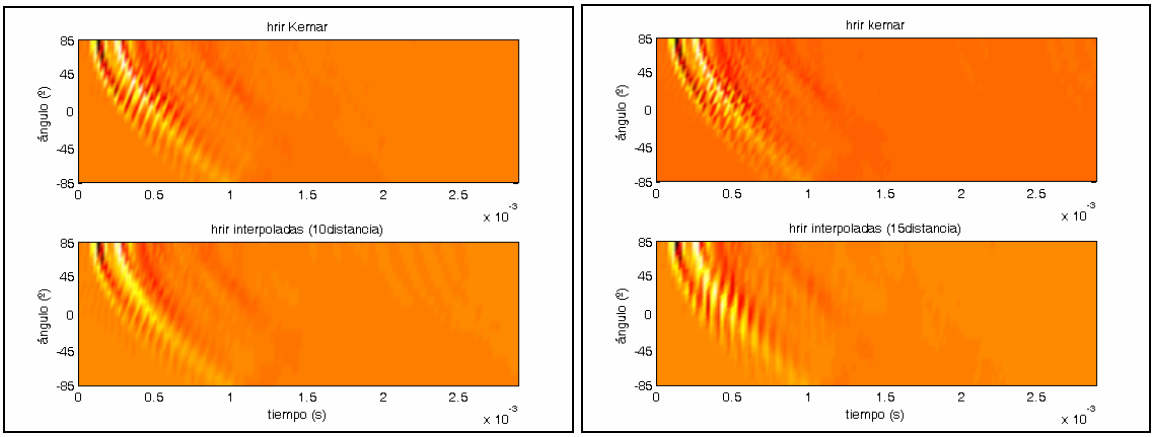


Figura 9. hrir medidas e interpoladas con el método de la distancia inversa, serie 10 y 15.

Como se demuestra matemáticamente, el método de la distancia inversa es idéntico al método de la media para la serie 10, donde las interpolaciones están equi-espaciadas de las hrir medidas. El método de la distancia inversa proporciona un ECM menor en el caso de la serie 15.

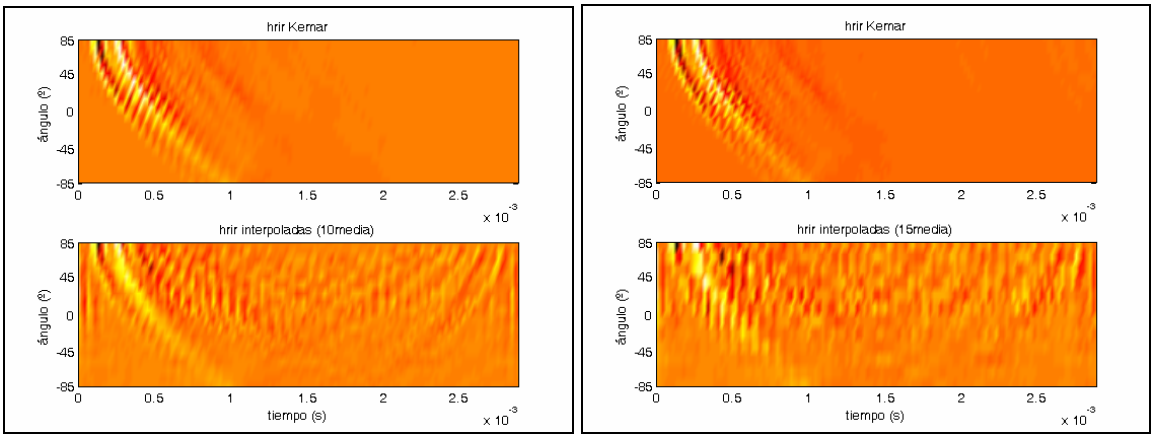


Figura 10. hrir calculada a partir de la interpolación de la HRTF con el método de la media, serie 10 y 15.

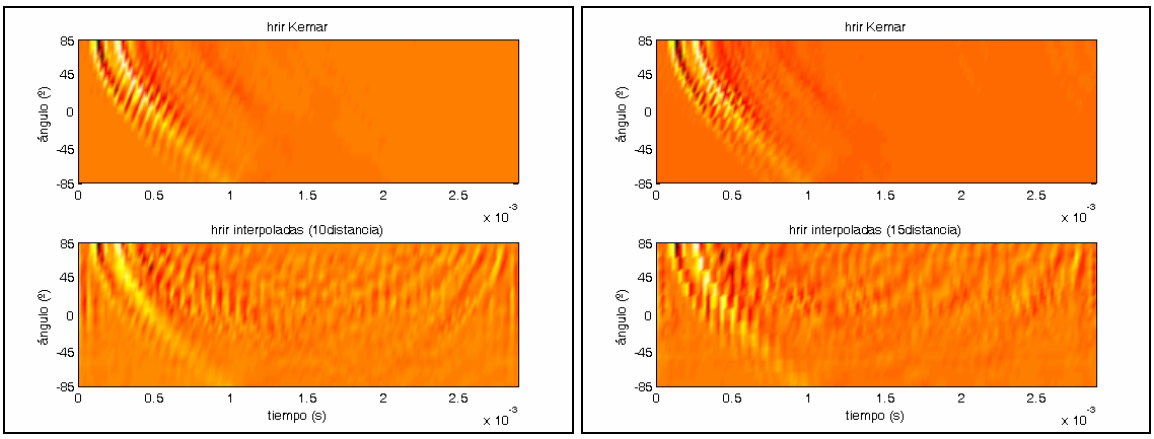


Figura 11. hrir calculada a partir de la interpolación de la HRTF con el método de la distancia, serie 10 y 15.

Se aprecia que el interpolar las HRTF es peor ya que al interpolar la fase de las HRTF que tienen un comportamiento no muy regular, el resultado no es bueno.

A continuación se presentan los resultados en el dominio de la frecuencia de las interpolaciones tanto para el módulo como para la fase.

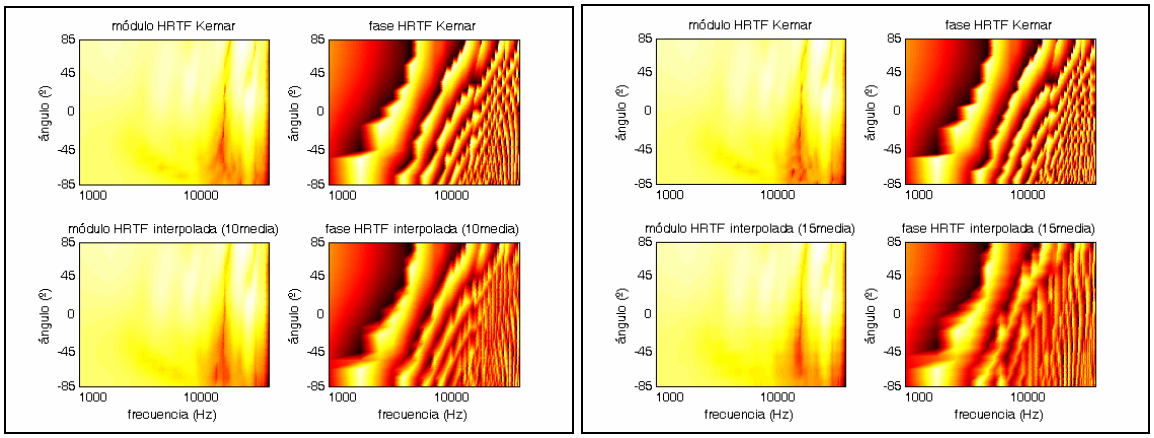


Figura 12. HRTF interpoladas con el método de la media, serie 10 y 15.

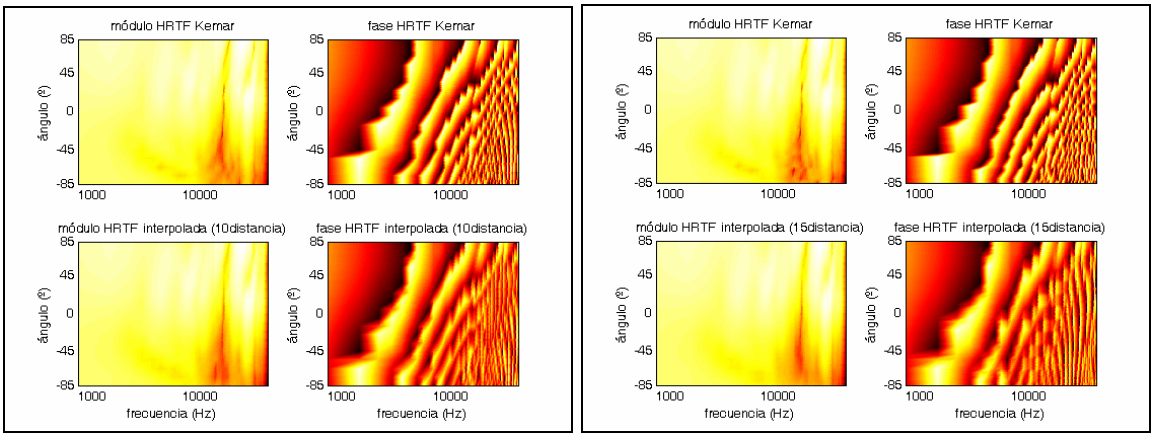


Figura 13. HRTF interpoladas con el método de la distancia, serie 10 y 15.

Como antes, para la serie 10 los dos métodos dan resultados idénticos, y para la serie 15 en el caso del módulo es mejor la interpolación por distancia y en el caso de la fase es mejor la interpolación de la media, esto debe obedecer al comportamiento no regular de la fase.

Conclusión

Los resultados nos llevan a inferir que las mejores interpolaciones se obtienen en el dominio del tiempo, realizadas directamente sobre las hrir medidas. Por otra parte, como se muestra en la representación de la fase, esta toma valores dispares que se van desplazando según la frecuencia. Al intentar interpolar estos máximos y mínimos se van cancelando y se produce un "suavizado" en los valores. Por lo tanto es mejor interpolar en tiempo. Estos resultados han sido publicados en [Gómez-Alfageme 04].

También se ha realizado el mismo estudio para la base de datos de nuestro maniquí. La tabla de los ECM para la base de datos de HATS es la siguiente:

	Media 10	Distancia 10	Media 15	Distancia 15
ECM hrir	0.0271	0.0271	0.0289	0.0286
ECM hrir (interpolando HRTF)	0.0349	0.0349	0.0372	0.0353
ECM módulo (HRTF)	3.3355	3.3355	3.7021	3.6145
ECM fase (HRTF)	2.2137	2.2137	2.1860	2.2346

Tabla 3. Error cuadrático medio de las interpolaciones para la base de datos de HATS.

A la vista de los resultados se pueden sacar las mismas conclusiones para las dos bases de datos. Por lo tanto, el LES utilizará el método de interpolación de la distancia inversa aplicado a las hrir, cuando necesite calcular las hrir en posiciones intermedias de los ángulos proporcionados en las bases de datos.

3.3 ESCENARIOS ACÚSTICOS

Como el objetivo es implementar una herramienta de medida del parámetro subjetivo de la localización espacial, y comprobar su funcionamiento, se han recreado de forma simulada y de forma real, diferentes escenarios acústicos para valorar la eficacia de la herramienta y hacer comparaciones.

En primer lugar se han comparado los resultados de localización para el caso en el que sólo hay una fuente dentro de un entorno anecoico. Este escenario nos sirve para evaluar la eficacia del LES a la hora de estimar el ángulo, ya que la localización del evento sonoro y de la fuente sonora deben de coincidir. Este simplificado escenario, que se va a llamar radiación directa, se ha simulado y también se ha recreado dentro de la cámara anecoica para comparar los resultados entre la señal biaural simulada y la señal biaural medida con el HATS. Una vez verificado que el LES tiene una buena precisión de estimación, tanto para entornos simulados como reales, se aplica a los sistemas de reproducción de ventana acústica virtual.

Para poder realizar pruebas en las que se simulen diferentes configuraciones de sistemas de audio multicanal para teleconferencias, e incluso diferentes codificaciones de las señales que se transmiten, se simula en Matlab un entorno acústico en campo libre (anecoico). En el se puede elegir la posición de la fuente, el número de transductores, su posición y la posición del oyente, dentro de la configuración de ventana acústica virtual. Posteriormente se ha construido la etapa de reproducción del sistema de ventana acústica virtual con un array de altavoces y se ha medido la señal recibida por el HATS dentro de una cámara anecoica real. Estos dos escenarios acústicos servirán para comprobar la capacidad del sistema de ventana acústica virtual de reproducir eventos sonoros y en concreto los parámetros que determinan la localización de los eventos sonoros.

La simulación consiste en calcular la señal biaural que recibiría el maniquí acústico para diferentes posiciones de una o varias fuentes en el plano horizontal o en el plano medio. En este proceso de validación se utilizan las mismas hrir para crear la señal biaural y para crear las tablas de búsqueda que utilizan los modelos psicoacústicos de audición para la estimación del ángulo.

En el plano horizontal, obviamente las estimaciones son muy buenas en el caso de radiación directa, ya que la información de ILD e ITD proporcionada por las hrir es la misma para formar las tablas y para formar la señal biaural, por lo tanto los errores pueden

ser relacionados con la limitación del modelo, y consecuentemente sería la mejor estimación posible.

A continuación se describe el proceso que se hace para calcular la señal biaural para el caso en el que se colocan las fuentes en el plano horizontal. El método de cálculo en el plano medio es análogo.

3.3.1 SIMULACIÓN RADIACIÓN DIRECTA

Para ensayar y estudiar los diferentes modelos psicoacústicos de localización, que han dado como fruto el LES, se ha simulado la percepción en campo libre del sonido (llamada radiación directa) como método de calibración de la herramienta de estimación (LES). En este caso la información de localización proviene sólo de las hrir. Una vez que se filtra la señal de la fuente con esta información se obtiene la señal biaural. A continuación se estima el ángulo de llegada a partir de la señal biaural. Este proceso de obtención de la señal biaural es el que se utiliza para generar audio 3D o "auralización" [Begault 04].

El proceso es el que sigue: se coloca una fuente ideal frente a un oyente (maniquí) a una cierta distancia r y en el plano horizontal con un cierto azimut θ.

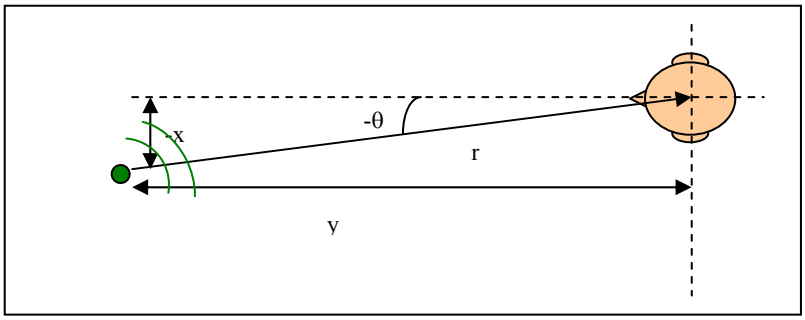


Figura 14. Disposición del altavoz y del oyente para la simulación directa.

$$\theta = \text{arctg}\left(\frac{x}{y}\right) \qquad r = \sqrt{x^2 + y^2} \qquad \text{Ec. 15}$$

Se calcula la $hrir_R(\theta)$ y la $hrir_L(\theta)$ de cada oreja para el ángulo de azimut θ, a partir de las hrir de la base de datos más próximas en azimut, mediante la interpolación anteriormente descrita. La atenuación producida por la distancia se simula dividiendo cada muestra por la distancia r, ya que suponiendo divergencia esférica (al considerarse campo lejano - ondas esféricas) el valor de la señal disminuye a razón de 1/r.

Para el cálculo de la señal biaural se utiliza el siguiente esquema (Figura 15).

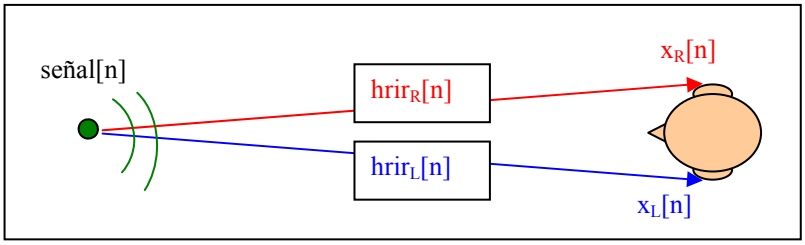


Figura 15. Cálculo de la señal biaural izquierda y derecha a partir de la señal emitida.

La señal biaural ($x_R[n]$ y $x_L[n]$) es la señal que llega a cada uno de los oídos. La señal emitida por la fuente se convoluciona con la hrir interpolada para el azimut concreto, retardada y atenuada en función de la distancia, información que como se ha explicado anteriormente no está incluida en las hrir.

3.3.1.1 CÁLCULO DE LA SEÑAL BIAURAL

El retardo de la señal se realiza añadiendo un número de ceros, proporcional a la distancia r, en la parte anterior de la hrir normalizada a 1 m. El número de ceros a añadir es:

$$n^o ceros(r) = round\left(\frac{r-1}{343} f_{muestreo}\right) \qquad \text{Ec. 16}$$

La atenuación se calcula como:

$$hrir_{R,L}[n,r] = \left[n^o ceros(r) \quad \frac{hrir_{R,L}}{r} \right] \qquad \text{Ec. 17}$$

Y la señal biaural obtenida a partir de la señal de la fuente altavoz[n]:

$$\begin{aligned} x_R[n] &= hrir_R[n,r] * altavoz[n] \\ x_L[n] &= hrir_L[n,r] * altavoz[n] \end{aligned} \qquad \text{Ec. 18}$$

ENSAYO 1

Figura 16. Grabación de la señal biaural real para diferentes ángulos en el plano horizontal dentro de una cámara anecoica.

Para comprobar la fiabilidad de la simulación de la señal biaural, se va a aplicar el LES a una señal biaural real captada con el maniquí y a la señal simulada para comparar los resultados. El sonido emitido por la fuente es un ruido blanco de 4 s de duración en el caso de señal real y 0.5 s en el caso de señal simulada. La fuente sonora se va a colocar a 2.5 m del oyente y a diferentes azimut: 0º, 10º, 20º, 30º, 45º, 60º y 90º para la señal medida y de 0º a 90º cada 10º para la señal simulada. Como se explicará en apartados siguientes, el LES estima el ángulo del evento sonoro basándose en los dos parámetros de localización ILD e ITD, por eso se dan dos estimaciones: roja para ITD y azul para ILD (Figura 17). Además

para valorar la eficacia de la estimación se mide la frecuencia estadística de dicha estimación a lo largo de todo el margen de frecuencia del banco de filtros (ver Capítulo 2).

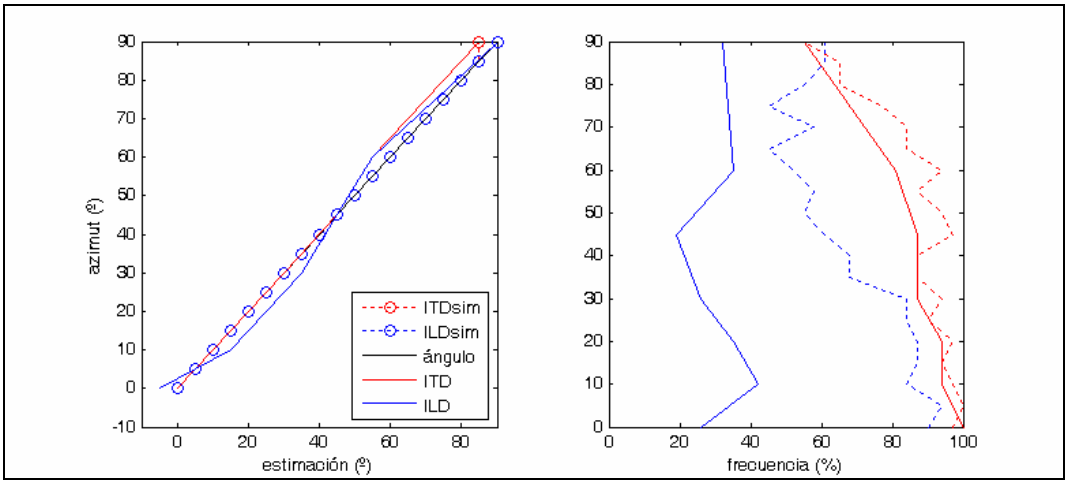


Figura 17. Representaciones de las estimaciones y su frecuencia obtenidas para la señal biaural simulada (ITDsim e ILDsim) y la señal biaural real (ITD e ILD) de una fuente colocada a diferentes ángulos de azimut.

	Señal simulada				Señal real			
	Estimación ITD		Estimación ILD		Estimación ITD		Estimación ILD	
Ángulo señal biaural (º)	**Moda (º)**	**Frec. (%)**	**Moda (º)**	**Frec. (%)**	**Moda (º)**	**Frec. (%)**	**Moda (º)**	**Frec. (%)**
0	0	97	0	90	0	100	-5	26
10	10	97	10	84	10	94	15	42
20	20	97	20	87	20	94	25	35
30	30	94	30	84	30	87	35	26
45	45	94	45	65	45	87	45	19
60	60	94	60	52	55	81	55	35
90	85	55	90	58	85	55	90	32

Tabla 4. Estimaciones del LES para señal biaural real y señal biaural simulada en el caso de radiación directa.

Analizando los resultados (Figura 17) para la estimación que se basa en la ITD (curvas rojas): la estimación tanto para señal real como simulada es correcta para todos los ángulos, excepto dos casos: para 60º, donde la estimación de la señal simulada es correcta pero la de la señal real es 55º; y para 90º donde se estima 85º en ambos casos. Este error está por debajo del umbral de percepción de localización (resolución auditiva), por lo tanto se considera que la estimación basada en la ITD es muy buena, tanto para señales simuladas como reales. Las frecuencias estadísticas de estimación son mejores para el caso de señal simulada (curva punteada) en 4 ángulos aunque la diferencia con la señal real medida (curva continua) no es muy grande.

Analizando los resultados para la estimación basada en la ILD (curvas azules): la estimación para la señal biaural simulada es correcta para todos los ángulos, sin embargo para la señal real no es correcta en 5 ángulos y si en los otros 2. Los errores son mínimos

(de 5º). En cuanto a la frecuencia estadística de la estimación se aprecia que es bastante mayor para el caso de señal biaural simulada que para el caso de señal real.

También se puede hacer un análisis de la distribución estadística de las estimaciones.

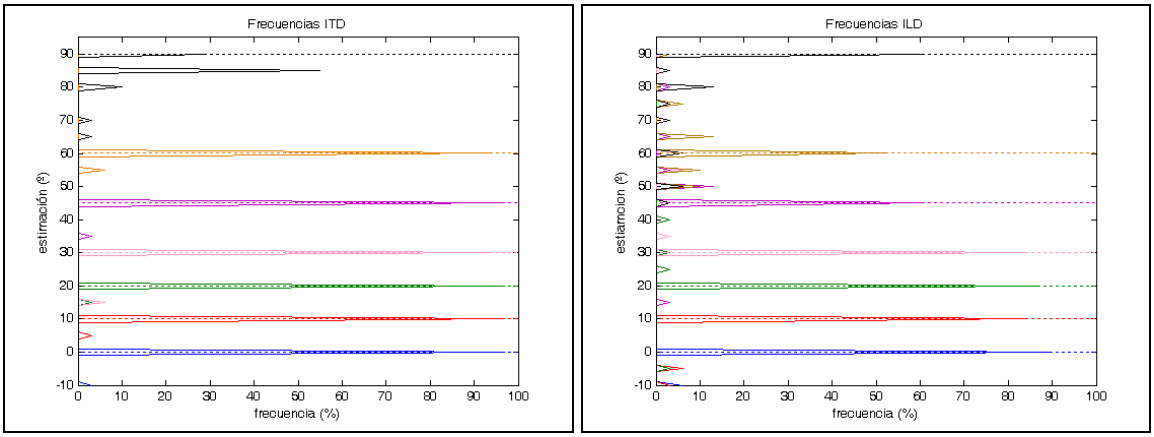


Figura 18. Distribución de las frecuencias de estimación para el caso de señal biaural simulada y los ángulos 0º, 10º, 20º, 30º, 45º, 60º, 90º. (1ª figura estimación ITD, 2ª figura estimación ILD).

Como se puede observar en la Figura 18 las frecuencias de estimación son mayores para la ITD que para la ILD. Vamos a analizar los datos del ángulo que da una estimación ITD errónea (90º): en el 55% de los casos da 85º y este valor es la moda y por lo tanto la estimación proporcionada, pero la estimación de 90º aparece en el 29% de los casos; luego entre 85º y 90º suponen el 84% de los casos. Se puede concluir que la desviación de los resultados alrededor del correcto es muy pequeña (aproximadamente 5º).

En la Figura 19 se muestran la distribución de las frecuencias de estimación para cada posición angular de la fuente para la señal biaural real. Cómo en la señal simulada, las frecuencias son mayores para el caso de ITD que para el caso de ILD.

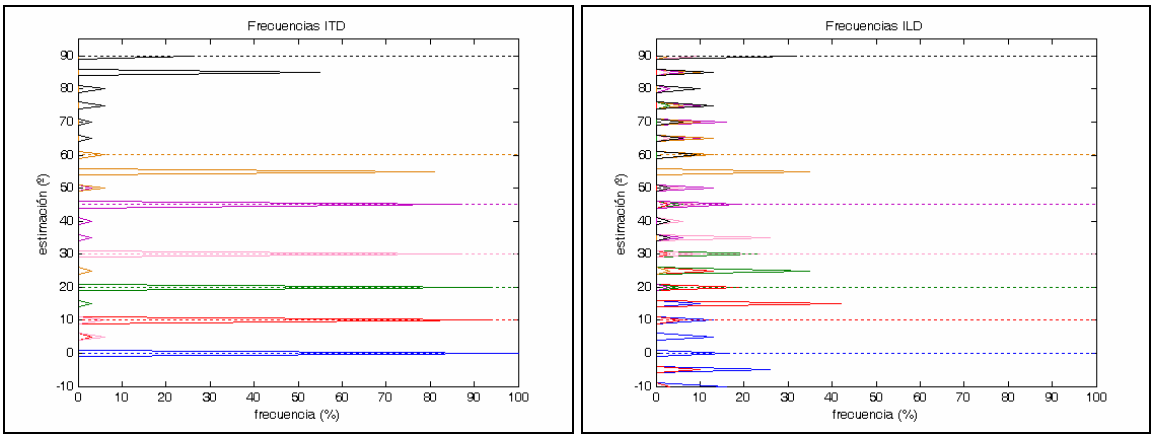


Figura 19. Distribución de las frecuencias de estimación para el caso de señal biaural real y los ángulos 0º, 10º, 20º, 30º, 45º, 60º, 90º. (1ª figura estimación ITD, 2ª figura estimación ILD).

La simulación de la señal biaural descrita anteriormente es el caso más óptimo para aplicar el LES porque la propia base de datos de hrir que se utiliza para los algoritmos de estimación es la que se utiliza para simular la señal biaural.

El análisis en función de la frecuencia del banco de filtros de los resultados anteriores se deja para el apartado del desarrollo del LES.

CONCLUSIÓN ENSAYO 1

Se puede decir que la simulación de señales biaurales se hace con bastante precisión en lo relativo a los parámetros de localización espacial de la fuente que simulan, ya que los resultados son muy parecidos con señal real y con señal simulada. Para los dos casos, las estimaciones del ángulo de llegada son muy buenos y dentro de la resolución del sistema auditivo. Por lo tanto se puede concluir que esta herramienta proporciona una buena medida del ángulo de azimut de una fuente sonora percibido por un oyente.

3.3.2 SEÑALES DE PRUEBA UTILIZADAS

Aunque en un principio se han utilizado muchos tipos de señales para evaluar la herramienta en los diferentes escenarios acústicos (tonos puros, ruidos filtrados de banda estrecha, ruido rosa y blanco y voz masculina y femenina), se ha llegado a la conclusión que, al igual que hace el sistema auditivo, los mejores resultados en la estimación se presentan con señales de banda ancha [Blauert 97] [Gómez-Alfageme 04]. Esto es debido a que el análisis de la ITD se realiza de forma diferente para bajas frecuencias y para altas, pero es la conjunción de todos los valores lo que se evalúa a nivel cognitivo a la hora de efectuar una estimación del ángulo. Las señales utilizadas como fuentes sonoras tanto para la simulación como para la reproducción real serán por tanto:

- Ruido blanco de banda ancha.
- Señal de voz: los archivos de audio utilizado (formato wav) son los la que vienen en la norma [ITU Recommendation BS.1387-1]: "Method for objective measurements of perceived audio quality" dentro de sus ficheros de prueba, llamados nrefsfe.wav y krefsme.wav.

3.4 MODELO PSICOACÚSTICO BIAURAL DEL LES PARA EL PLANO HORIZONTAL

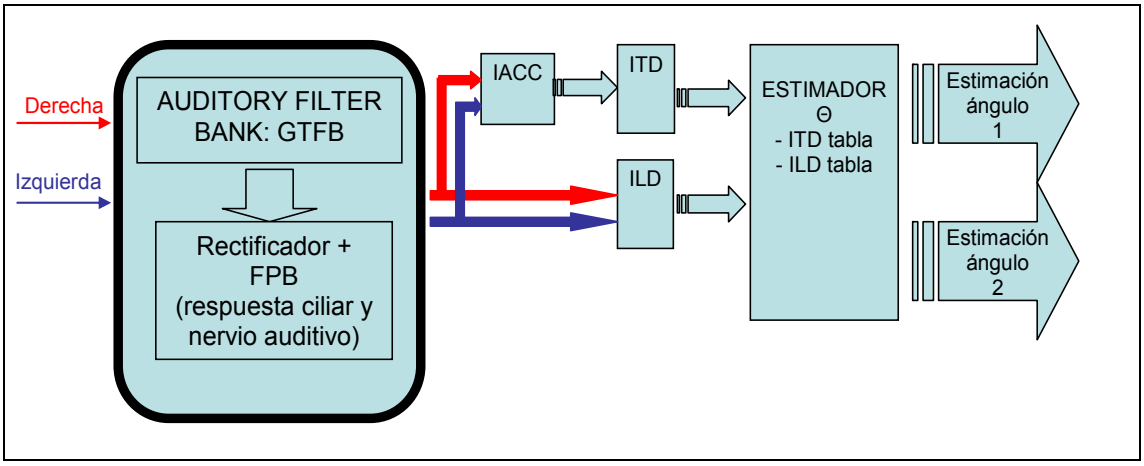


Figura 20.Método de localización biaural basado en un modelo de audición simplificado.

El modelo de audición implementado en el LES es un modelo psicoacústico biaural de audición espacial simplificado, aplicado a la estimación de la localización del evento sonoro. Este modelo biaural utiliza un banco de filtros gammatone para modelar la percepción en frecuencia del oído interno para cada una de las señales biaurales. Por lo tanto, cada señal de entrada al oído es espectralmente separada en rangos de ancho de banda relativos aproximadamente iguales que simulan el filtrado de la coclea. La función de transferencia de estos filtros paso banda representa la sensibilidad del oído. El banco de filtros utilizado es un gammatone (GTFB), formado por 42 filtros paso banda ERB (Equivalent Rectangular Bandwidth) [Slaney 98].

A partir de aquí, el procesado se hace para cada canal de frecuencia, por lo tanto, la estimación del azimut se hace para las 42 frecuencias centrales de los filtros. Dicho procesado consiste en una rectificación de media onda y un filtro paso bajo que modela el comportamiento del nervio auditivo, en concreto, la respuesta de las células ciliares de la membrana basilar [Harma 00] cuando se realiza la transducción mecánico-neuronal. Por encima de 1.6 kHz esto significa que la envolvente de las señales es derivada, esto es, las señales se demodulan en amplitud y por lo tanto, el siguiente paso que es la correlación cruzada se realiza a la envolvente. Este filtrado simula el comportamiento detectado en las pruebas subjetivas [Blauert 97].

La ITD se calcula como el máximo de la correlación cruzada interaural (InterAural Cross Correlation, IACC):

$$\Phi(\tau) = \int_0^{t_0} x_L(t) x_R(t-\tau) dt \qquad \text{Ec. 19}$$

donde τ es la diferencia de tiempo entre las dos señales y t_0 es la longitud de una ventana de tiempo rectangular. En el sistema de audición esta ventana no es rectangular, pero como vamos a usar señales estacionarias, la forma de la ventana no influye en el resultado [Pulkki 99]. La función de la IACC modela el funcionamiento neuronal del método de coincidencia [Blauert 97].

Este modelo psicoacústico biaural ha sido verificado mediante test subjetivos en [Pulkki 01] y [Pulkki 05]. Aparecen unas desviaciones sistemáticas entre las estimaciones del modelo y los resultados de los tests de escucha, sobre todo para lateralizaciones alejadas del plano medio. Estas desviaciones están dentro del llamado umbral de percepción de localización observado por muchos estudios [Blauert 97] cuando se lateraliza la fuente sonora. El sistema auditivo humano es más preciso para ángulos de azimut pequeños (posiciones frontales) que para ángulos de azimut laterales (90º).

El modelo de audición se ha simplificado tanto como ha sido posible manteniendo los parámetros más importantes de la percepción sonora relativos a la localización espacial. El efecto del oído medio no se tiene en cuenta ya que su influencia es a muy baja frecuencia, y la frecuencia menor utilizada en el procesado es 200 Hz (frecuencia central del primer filtro del banco de filtros). Además, este efecto (filtro paso banda) es simétrico en los dos oídos y por lo tanto no cambiará ni la ITD ni la ILD (parámetros de localización biaurales calculados como diferencias) [Faller 04]. El eje de frecuencia según esta escala, es la escala de audición humana dada por las frecuencias centrales del banco de filtros.

En un modelado más profundo se podría incluir un modelo adaptativo de la respuesta del cerebro, según pasa el tiempo, según el nivel, o según sea familiar o no la fuente, etc. En el LES las señales utilizadas de prueba son estacionarias (ruido blanco y voz), por lo tanto no es necesario un sistema adaptativo temporal.

3.4.1 CÁLCULO DE LA IACC Y DE LA ITD

La correlación cruzada interaural (IACC), es un parámetro importante para determinar la dirección percibida de una fuente sonora. Esta función modela el proceso neuronal de coincidencia que sigue el sistema auditivo para determinar la ITD [Blauert 97]. Cuando la correlación cruzada interaural normalizada tiene un máximo estrecho (≈ 1) la dirección de la fuente se define muy bien. Para sistemas multicanal, la IACC no sólo depende de la dirección de los altavoces, sino también de la correlación entre las señales emitidas [Damaske 72]. Cuando se usan señales incoherentes, la IACC puede no tener máximos aparentes, dependiendo de la posición de los altavoces. Cuando se usan señales coherentes, la fuente se localizará en función del retardo - atenuación entre las señales y sus posiciones, proceso llamado de localización sumatoria.

La IACC normalizada se calcula como:

$$\varphi_{LR}(\tau) = \frac{\Phi_{LR}(\tau)}{\sqrt{\Phi_{LL}(0)\Phi_{RR}(0)}} \qquad \text{Ec. 20}$$

La correlación cruzada se calcula para cada banda del banco de filtros gammatone entre las dos señales, izquierda y derecha. La posición del máximo de cada curva IACC indica la ITD para la frecuencia correspondiente a cada banda. En concreto, la ITD se calcula como la posición del primer máximo más cercano a 0 de la IACC. La ITD es aproximadamente ≤ 0.75 ms para el tamaño de cabeza del maniquí, por lo tanto recortamos la respuesta de la IACC a ±0.8 ms (Figura 21). Un ejemplo del conjunto de IACC calculadas se muestra en la figura siguiente como función del retardo τ y del canal ERB. La posición del máximo en una de la curvas de la IACC es la ITD a esa frecuencia.

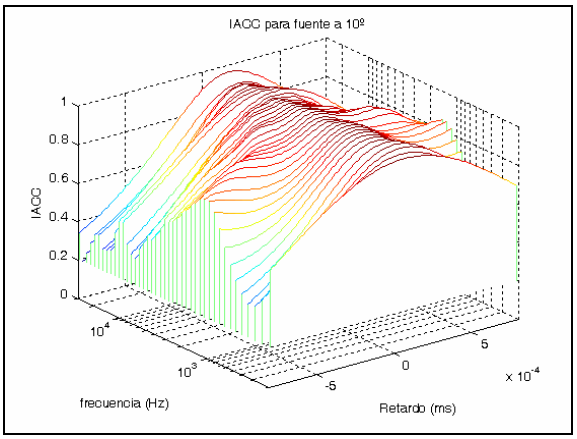


Figura 21.Curvas de IACC para una posición de fuente a 10º para los 42 filtros del banco.

Una vez obtenida la señal biaural (medida o simulada) se calcula su correlación cruzada y el máximo de esta función dentro del margen 0.8 ms es la ITD. Hay que tener en cuenta que la primera reflexión que llegaría en una sala no anecoica sería la del suelo y el retardo debido a la diferencia de caminos se podría estimar en 5 ms y 1 ms para posiciones de fuentes alejadas 1 m o 2.5 m respectivamente; luego este modelo también valdría para condiciones no muy reverberantes sin tener que tener en cuenta el efecto de precedencia.

3.4.2 CÁLCULO DE LA ILD

El otro parámetro evaluado en el modelo es la sonoridad L, la cual se calcula para cada canal ERB en cada oído con la fórmula de Zwicker:

$$L = \sqrt[4]{\langle x^2 \rangle} \qquad \text{Ec. 21}$$

Donde $\langle x^2 \rangle$ es el promedio temporal de la potencia de la señal. Los niveles de sonoridad, LL en fonos, se calculan usando la expresión:

$$LL = 40 + 10 \cdot \log_2 L \qquad \text{Ec. 22}$$

Se calcula la resta de los niveles de sonoridad en cada banda ERB de la señal biaural, dando como resultado la ILD.

$$ILD = LL_R - LL_L \qquad \text{Ec. 23}$$

3.4.3 TABLAS DE BÚSQUEDA

Una vez establecido cómo se mide la ITD y la ILD a partir de las señales biaurales, se crean unas tablas de búsqueda para el algoritmo de estimación del ángulo. Son dos tablas, una para la ITD y otra para la ILD, en las que se almacenan los valores de la ITD y la ILD calculadas sobre las hrir, según se ha descrito anteriormente. Para cada ángulo de azimut las hrir deberán tener la información de diferencia de nivel y de retardo introducida por la cabeza. Por lo tanto, para cada ángulo habrá 42 valores en función del banco de filtros de ITD o ILD. Las tablas creadas para la base de datos HATS se representan en la Figura 22.

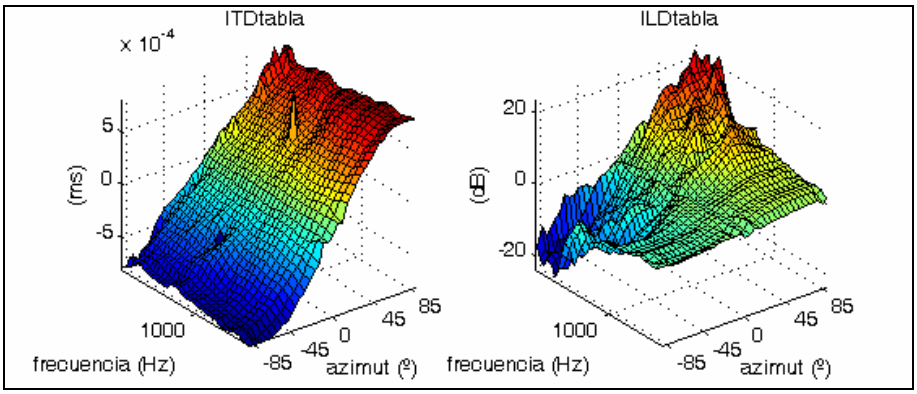


Figura 22. Representación de la ITD e ILD en función del azimut y de la frecuencia para la base de datos de HATS.

Como se aprecia en las curvas, la ITD es positiva para ángulos positivos (el sonido llega antes al oído derecho) y negativa para ángulos negativos (llega antes al oído izquierdo). Además la diferencia de llegada entre oídos depende muy poco de la frecuencia. Se puede apreciar que el comportamiento de estas curvas es bastante monótono.

En cambio, para la ILD casi no hay diferencia de nivel a frecuencias bajas según varía el ángulo; pero, para frecuencias altas la diferencia de nivel es bastante grande (hasta 20 dB) para ángulos de azimut cercanos a 90º. Esto es debido al efecto sombra de la cabeza para frecuencias altas. El comportamiento de estas curvas siempre es monótono, por lo tanto, para una misma frecuencia una determinada diferencia de nivel puede encontrase en

ángulos azimut diferentes a la vez, esto provocará errores de estimación en algunas frecuencias.

3.4.4 COMPARATIVA ENTRE ESTIMADORES

Una de las decisiones tomadas a la hora de implementar el modelo psicoacústico biaural del LES fue decidir qué estimador de azimut se utilizaba. En la bibliografía aparecen diferentes fórmulas para calcular el azimut en función de la ITD, pero ninguna para calcular el azimut en función de la ILD. En la mayoría se los modelos psicoacústicos biaurales estudiados no se estima un ángulo de azimut, sino que el modelo se queda en el cálculo de las ITD y las ILD. A continuación se describen los estimadores estudiados, se ha determinado cuál da mejores resultados de estimación y ese ha sido el implementado en el LES.

Para comparar los estimadores se ha realizado el siguiente análisis. A partir de los parámetros de localización (ITD, ILD) se estiman los ángulos de azimut con los diferentes estimadores. Esta estimación se ha realizado directamente sobre las hrir (ya que en ellas está toda la información de localización, el resto de funciones de transferencia involucradas en una sonorización real no proporcionan información de localización). La base de datos sobre la que se ha hecho el estudio comparativo de los diferentes estimadores es la de HATS.

Hay que tener en cuenta que, aunque la influencia de la ITD está claramente establecida y estudiada en relación a la localización en azimut, todavía no se ha conseguido una estimación buena (que coincida el azimut de la fuente y del evento sonoro percibido) a partir del cálculo de la ITD sobre todo a acimuts cercanos a 90º [Busson 05].

ITD_estimación1

Propuesto por Woodworth-Schlosberg [Blauert 97] para un modelo de cabeza esférica y ondas planas.

$$ITD \cdot c = \frac{D}{2}(\theta + \text{sen}\theta) \qquad \text{Ec. 24}$$

D = 0.152 m, diámetro de la cabeza de HATS.

En las gráficas se llamará estimador Wood.

ITD_estimación2

Propuesto por Kuhn [Kuhn 77] para un modelo de cabeza esférica con difracción que depende de la frecuencia y del diámetro de la cabeza.

$$\begin{aligned} ITD \cdot c &= \frac{3}{2} D\,\text{sen}\theta \quad f \le 500\ \text{Hz} \\ &\text{Valores intermedios para el otro rango.} \\ ITD \cdot c &= D\,\text{sen}\theta \quad f > 3\ \text{kHz} \end{aligned} \qquad \text{Ec. 25}$$

donde D es el diámetro de la cabeza esférica (utilizamos el diámetro de HATS).

En las gráficas se llamará estimador Kuhn.

ITD_estimación3 e ILD_estimación4

A partir de la tabla de ITD-ILD creada para la base de datos del HATS, se hace una búsqueda dentro de la tabla del valor de ITD o ILD más cercano para una determinada frecuencia; el ángulo de ese valor es la estimación de ITD o la estimación de ILD.

En las gráficas se llamarán estimadores ITDtabla e ILDtabla.

La curva llamada ángulo indica el valor ideal de la estimación que sería el ángulo de azimut de la fuente sonora.

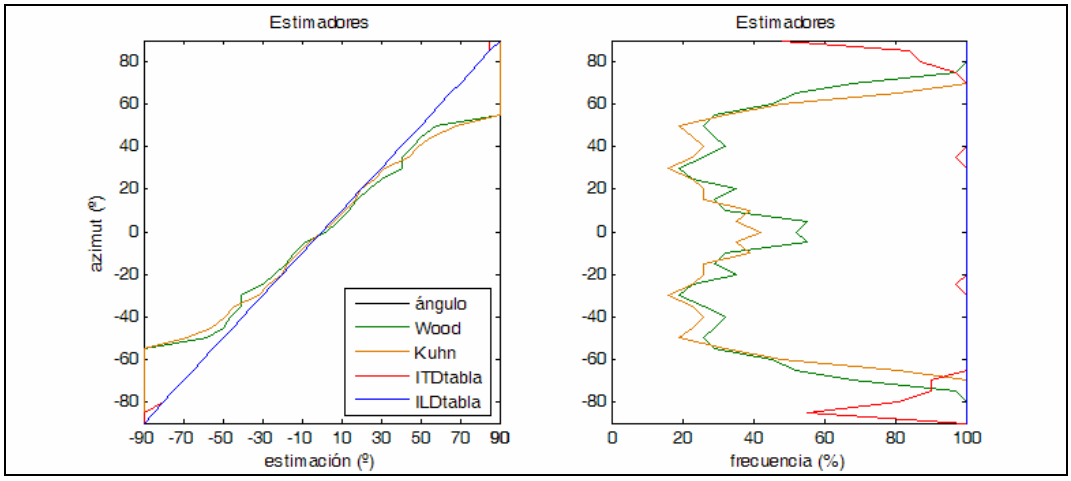


Figura 23. Representación de las estimaciones realizadas por los cuatro estimadores y su frecuencia estadística en función del ángulo sobre la base de datos de HATS.

También se ha realizado el análisis sobre la señal biaural en vez de las hrir. Las señales biaurales han sido simuladas para el caso de radiación directa. En la tabla siguiente se representa el error cuadrático medio entre la estimación y el ángulo.

	Estimación1 (Wood)	Estimación2 (Kuhn)	Estimación3 (ITDtabla)	Estimación4 (ILDtabla)
ECM	2.3744	2.4546	0.1911	0.0000

Tabla 5. Error cuadrático medio para las diferentes estimaciones (señal biaural simulada).

Las disparidades de las estimaciones de cabeza esférica (estimación1 y estimación2) pueden ser debidas a la falta de similitud al aproximar la cabeza humana a una esfera.

Los errores en las estimaciones basadas en la tabla son debidos a que la ITD y la ILD no son funciones monótonas y por lo tanto pueden repetir valores para diferentes ángulos en la misma frecuencia, provocando el error.

Según [Blauert 97], cuando se desplazan dos sinusoides en tiempo existen dos formas de medir el retardo y por lo tanto aparecen dos eventos sonoros, pero predomina aquel que está más cerca del plano medio. Por eso en nuestro algoritmo de búsqueda dentro de las tablas de ITD-ILD si aparecen dos ángulos que tienen el mismo valor de ITD o ILD se da como estimación aquel más cercano a 0º de azimut (plano medio).

A continuación se muestra una comparativa para diferentes ángulos de la estimación en función de la frecuencia.

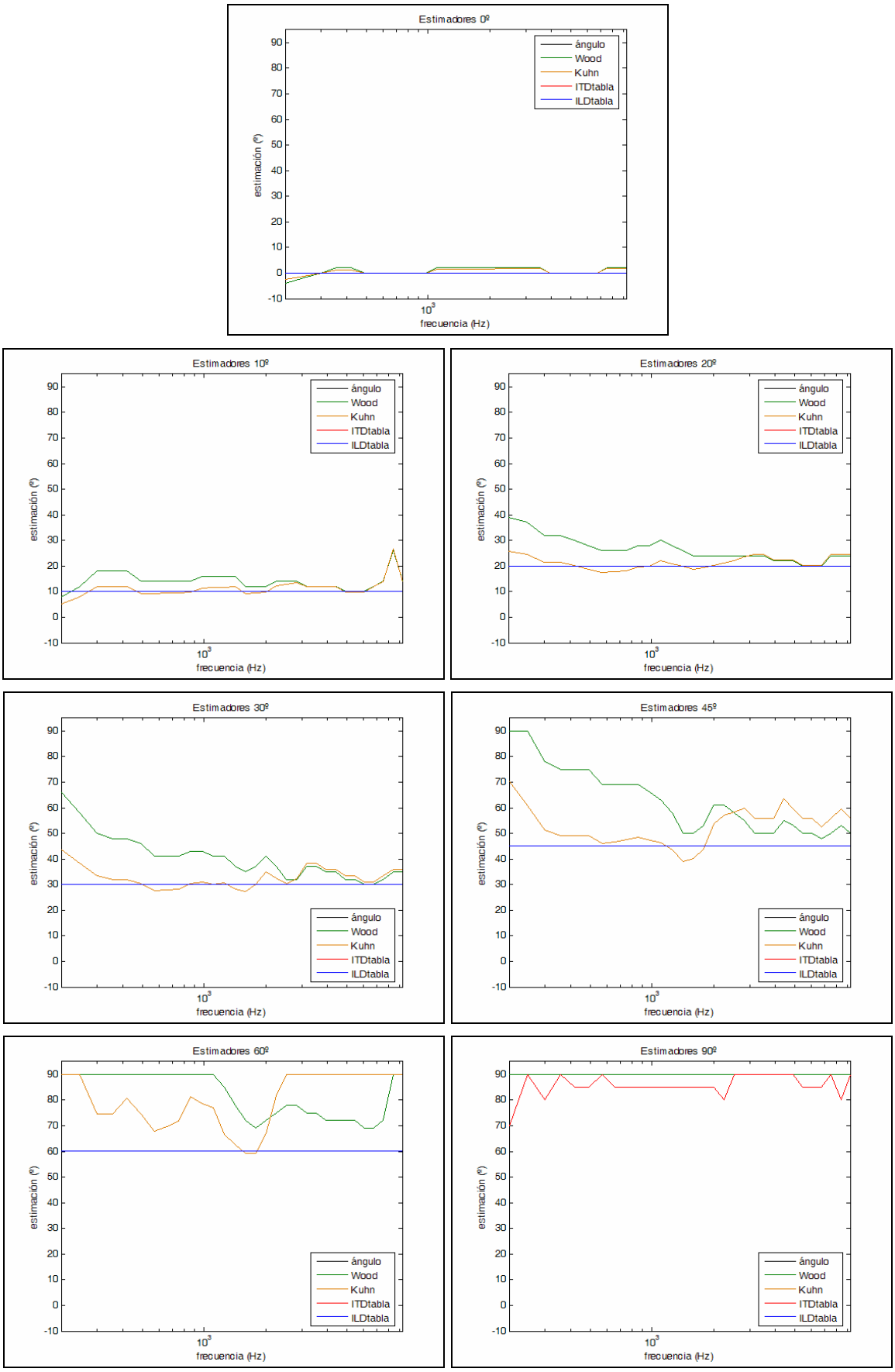


Figura 24. Representación de las estimaciones para los diferentes estimadores en función de la frecuencia para varios ángulos de la fuente, aplicados sobre las hrir de la base de datos.

Como ya se había dicho, los mejores estimadores son los basados en las tablas. En todos los ángulos, excepto en 90º, la estimación basada en la tabla es correcta tanto para ITD cómo para ILD. Para 90º, como es el extremo de la resolución, cuando hay un error se da esta estimación y por lo tanto no se puede valorar este ángulo. Hay que tener en cuenta que el umbral de percepción de localización aumenta con la frecuencia y el azimut. Los modelos de cabeza esférica son válidos para acimuts pequeños, donde el error cometido al medir la diferencia de caminos a los oídos para una cabeza esférica o para una cabeza real son pequeños. Según aumenta el ángulo este error aumenta (ver Anexo 3-2).

Conclusión

Implícita en la señal biaural, viene la información de las HRTF de la cabeza, por lo tanto el mejor estimador es aquel que se basa en las tablas de ILD e ITD de la propia cabeza. Se sabe que las personas realizan un aprendizaje desde que son niños, aprendiendo a localizar las fuentes sonoras según sus propias HRTF. Por eso, existe un gran esfuerzo investigador en encontrar unas HRTF adaptables a cada persona según sus medidas antropométricas.

3.4.4.1 COMPARATIVA ENTRE ESTIMACIONES CON SEÑAL BIAURAL REAL

Para comprobar la eficacia de los diferentes estimadores sobre la señal biaural, se va a utilizar la señal biaural real grabada según se ha explicado anteriormente (recordando, 0.5 s de señales biaurales a diferentes ángulos emitiendo ruido blanco en un entorno anecoico). Se han calculado las estimaciones cuando se usan los diferentes estimadores:

	Estimación1 (Wood)		Estimación2 (Kuhn)		Estimación3 (ITDtabla)		Estimación4 (ILDtabla)	
Ángulo señal biaural (º)	**Moda (º)**	**Frec. (%)**	**Moda (º)**	**Frec. (%)**	**Moda (º)**	**Frec. (%)**	**Moda (º)**	**Frec. (%)**
0	0	77	0	77	0	100	-5	26
10	12	42	12	39	10	94	15	42
20	24	23	22	29	20	94	25	35
30	35	23	33	16	30	87	35	26
45	50	23	56	19	45	87	45	19
60	90	32	90	35	55	81	55	35
90	90	100	90	100	85	55	90	32

Tabla 6. Estimaciones y frecuencias de estimación para los cuatro tipos de estimadores analizados (señal biaural real).

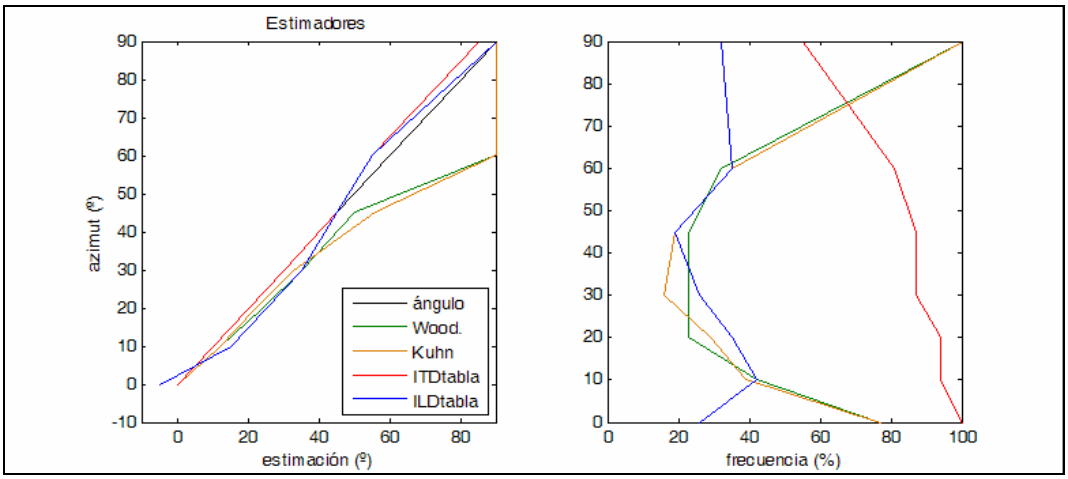


Figura 25. Representación de las estimaciones realizadas por los cuatro estimadores y su frecuencia estadística, en función del ángulo para el caso de señal biaural real.

Conclusión

Las conclusiones son las mismas que cuando se utilizó directamente las hrir en el cálculo de las estimaciones. Por lo tanto, en la elección del estimador que se ha puesto en el LES, se descartan las estimaciones 1 y 2 por cometer mucho error; excepto para 90º, que como el límite del cálculo de la estimación es 90º, sale ese valor para todas las frecuencias. Las mejores estimaciones son para el estimador basado en la tabla de la ITD. Este hecho se repite también subjetivamente. Según los estudios subjetivos [Blauert 97], el sistema auditivo humano realiza mejores localizaciones basándose en diferencias de tiempo interaural.

3.4.4.2 RESPUESTA EN FRECUENCIA DE LOS ESTIMADORES BASADOS EN TABLAS

A continuación, se hace un análisis en frecuencia de las estimaciones obtenidas para la señal biaural cuando se usan los estimadores ITDtabla e ILDtabla. La posición de la fuente se ha variado entre 0º de azimut y 90º de azimut en pasos de 10º en el caso de señal biaural simulada (Figura 26) y 0º, 10º, 20º, 30º, 45º, 60º y 90º en el caso de señal biaural medida (Figura 27).

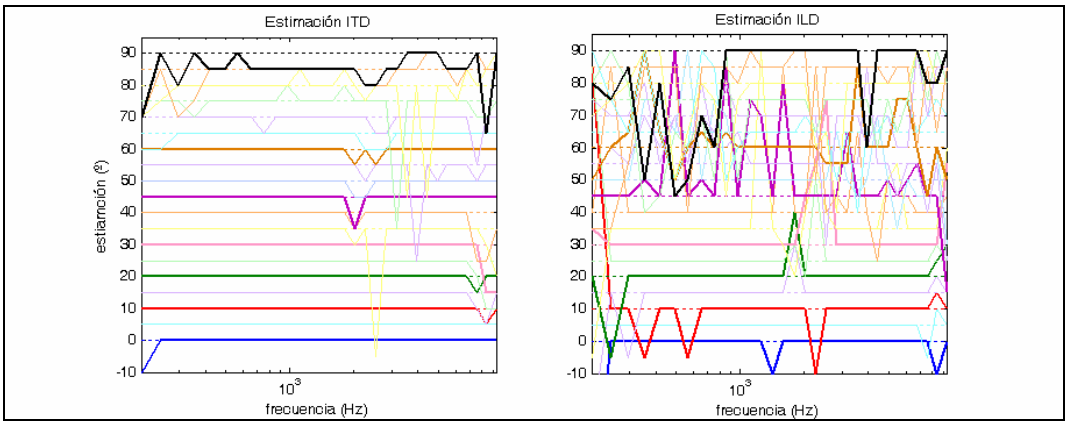


Figura 26. Las líneas punteadas son la posición de la fuente, las líneas continuas son el azimut estimado a partir de la señal simulada basándose en la ITD y en la ILD en función de la frecuencia (señal biaural simulada).

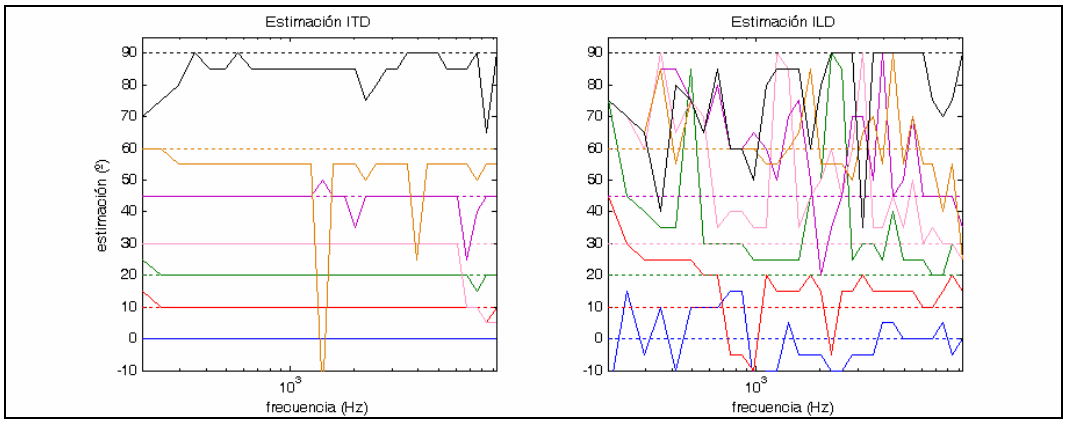


Figura 27. Las líneas punteadas son la posición de la fuente, las líneas continuas son el azimut estimado a partir de la señal simulada basándose en la ITD y en la ILD en función de la frecuencia (señal biaural real).

Conclusión

Como puede observarse en las figuras anteriores, en el caso de estimación por ITD se comete error a partir de 2000 Hz, donde la ITD deja de proporcionar información; este efecto se nota más para ángulos de azimut elevados, donde la resolución auditiva es menor. Para el caso de estimación por ILD existen muchos errores para ángulos de acimut elevados, esto es debido a la respuesta irregular de la ILD en función del azimut y de la frecuencia, que ya se ha comentado antes. Aún así, la moda de estas estimaciones es bastante buena y el azimut estimado corresponde con el real.

3.4.4.3 ESTIMADORES BASADOS EN TABLAS SUAVIZADAS

Se ha probado a suavizar la tabla de ILD e ITD para ver si disminuían los errores de la estimación, según se ha hecho en las referencias [Pulkki 01] [Viste 04]. Como se vio en el apartado anterior, una fuente de error en la estimación está originada por el comportamiento irregular en frecuencia de las tablas, luego se han suavizado para evitar en lo posible el mismo valor de ITD e ILD para dos ángulos, dentro de la misma frecuencia. Este suavizado se ha realizado en frecuencia.

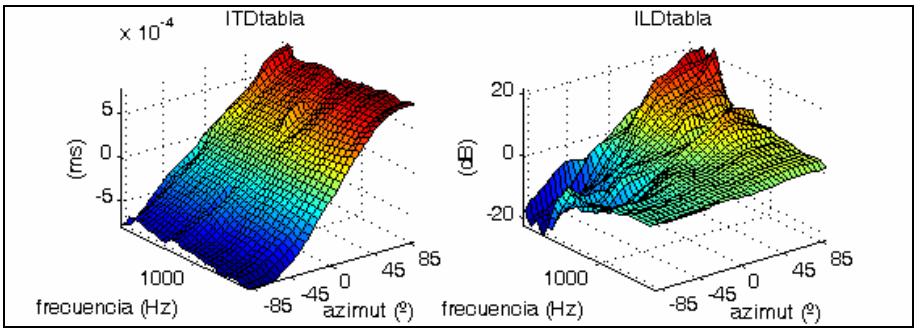


Figura 28. Representación de las tablas suavizadas de la ITD e ILD en función del azimut y de la frecuencia para la base de datos de HATS.

A continuación se presentan las estimaciones utilizando las tablas suavizadas y sin suavizar. Se han hecho pruebas para el caso de utilizar las hrir directamente o las señales biaurales medidas.

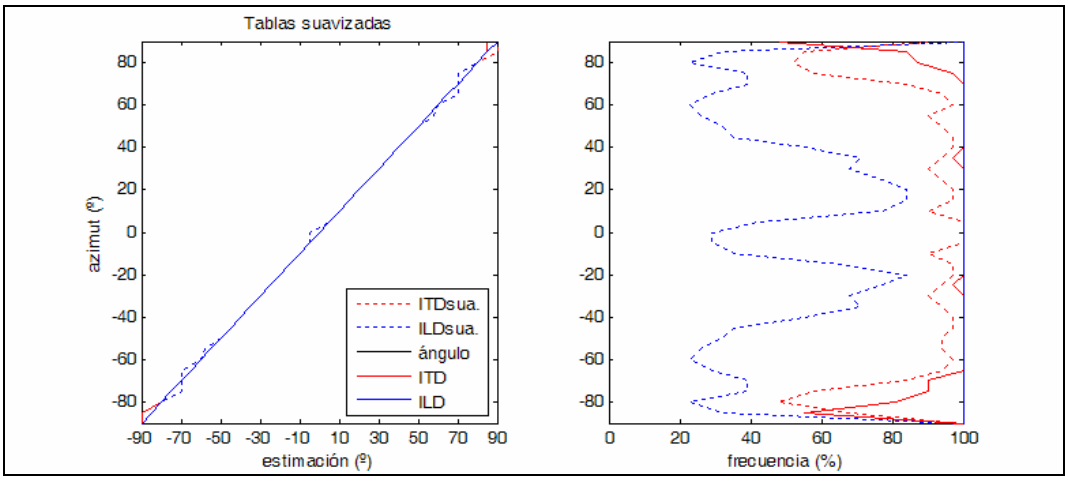


Figura 29. Representación de las estimaciones ITD e ILD realizadas con las tablas suavizadas (curvas punteadas) y sin suavizar (curvas continuas) y su frecuencia estadística respectiva, sobre las hrir de la base de datos de HATS.

	ITDsuavizada	ILDsuavizada	ITDtabla	ILDtabla
ECM	0.1911	0.3232	0.1911	0.0000

Tabla 7. Error cuadrático medio para las estimaciones basadas en tabla suavizada y sin suavizar, estimaciones hechas sobre las hrir.

Cuando se usan las tablas suavizadas, el resultado es igual a cuando no se usan para la estimación por ITD y peor para la estimación por ILD, ya que crece el ECM (error cuadrático medio) de la estimación. Sin embargo, las frecuencias estadísticas de las estimaciones son claramente peores para cualquier ángulo de estimación.

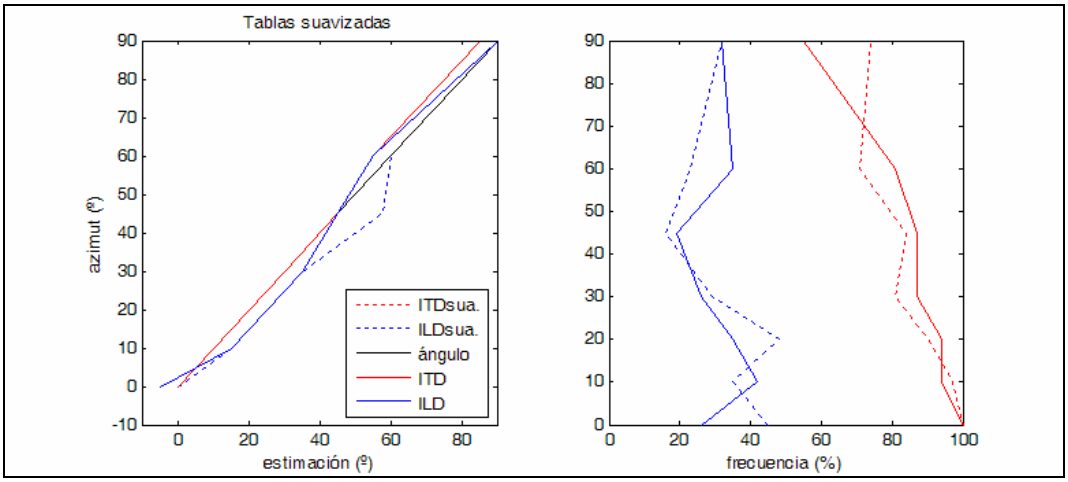


Figura 30. Representación de las estimaciones ITD e ILD realizadas con las tablas suavizadas (curvas punteadas) y sin suavizar (curvas continuas) y su frecuencia estadística respectiva (señal biaural real).

En el caso de la señal biaural medida, en términos generales las estimaciones son iguales en el caso de la ITD y ligeramente mejores en el caso de la ILD cunado no se usa el suavizado de la tabla de búsqueda.

Conclusión

Si las hrir provocan tienen irregularidades en las curvas de la ITD y la ILD, estas irregularidades también se verán reflejadas en las tablas y por lo tanto, la búsqueda será más precisa si no se suavizan los datos. Esta es una de las diferencias entre los métodos de Pulkki y Viste y el implementado en el LES.

3.4.4.4 ITD_ESTIMACIÓN3 CON BÚSQUEDA A PARTIR DE LA ITD GLOBAL

Una de las posibles razones de incertidumbre a la hora de estimar el ángulo basándose en la ITD es la valoración de la posición del máximo en las curvas de la IACC. Para mejorar esta valoración se propone usar el dato de la ITD global de la señal biaural sin filtrar por el banco de filtros. Como en la IACC pueden existir varios máximos, se busca aquel que esté más cerca al valor de la ITD global, para cada una de las frecuencias del banco.

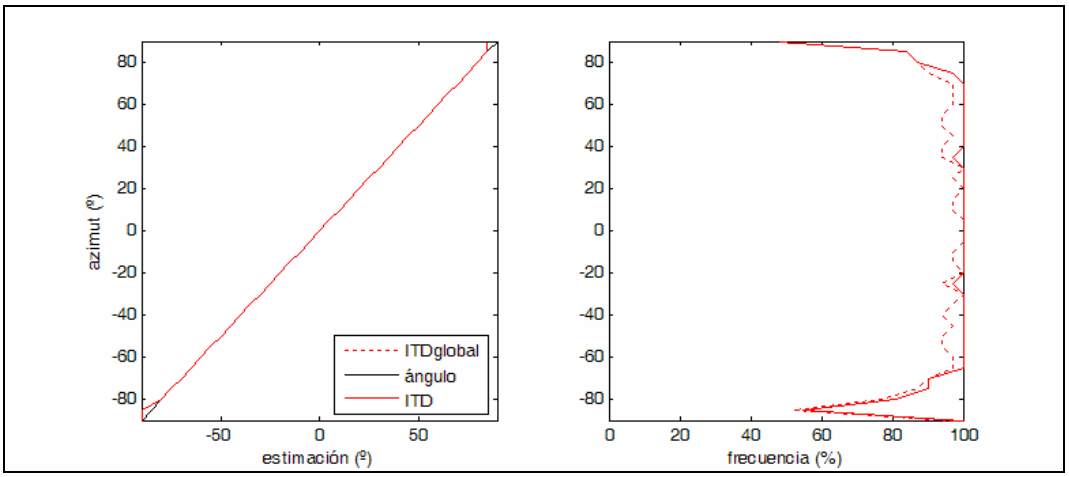


Figura 31. Representación de las estimaciones y su frecuencia estadística basadas en la ITD (curva continua) y en la ITD a partir de la ITDglobal (curva punteada), sobre las hrir de la base de datos de HATS.

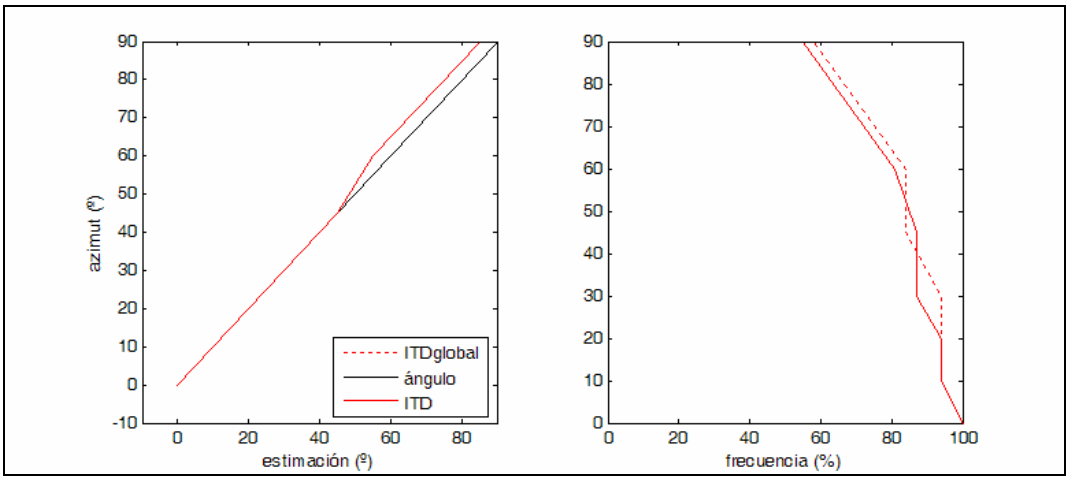


Figura 32. Representación de las estimaciones y su frecuencia estadística basadas en la ITD (curva continua) y en la ITD a partir de la ITDglobal (curva punteada) (señal biaural real).

	ITDglobal	**ITDtabla**
ECM	0.1911	0.1911

Tabla 8. Error cuadrático medio para las diferentes estimaciones (señal biaural simulada).

Conclusión

Aunque el ECM es igual, la frecuencia de la estimación es menor en general cuando se usa la ITDglobal, luego se considera una estimación peor y no se va a usar este procedimiento.

3.4.5 LATENCIA DE LA IACC

A veces, hay varios picos prominentes en la IACC localizados a diferentes ITD, según la teoría biaural aquellos valores consistentes o que se repiten a lo largo de la frecuencia son más relevantes en la localización. Para implementar esto, se ha utilizado un segundo orden de coincidencia que consiste en multiplicar la función IACC de una frecuencia con la inmediatamente superior e inferior, así los picos consistentes se mantienen y refuerzan como se ve en la figura siguiente.

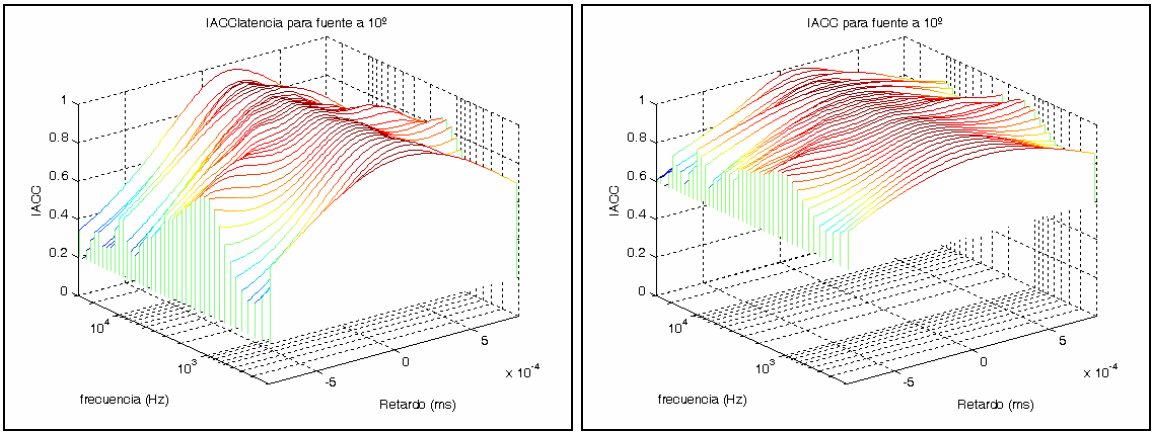


Figura 33.Curvas de IACC para una posición de fuente a 10º, para los 42 filtros del banco cuando se utiliza latencia y cuando no se utiliza.

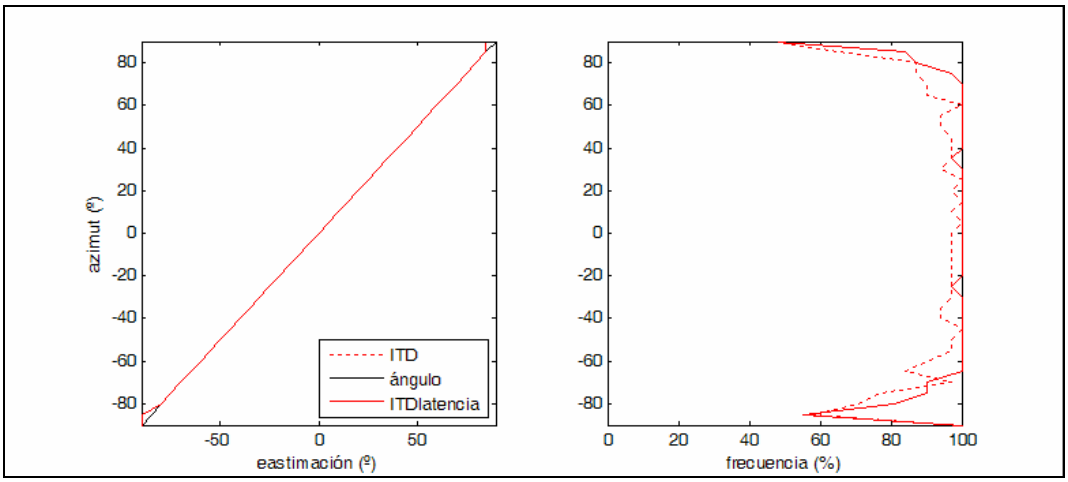


Figura 34. Representación de las estimaciones y su frecuencia estadística basadas en la ITD con latencia (curva continua) y en la ITD sin latencia (curva punteada), sobre las hrir de la base de datos de HATS.

	ITDlatencia	**ITD**
ECM	0.1911	0.0000

Tabla 9. Error cuadrático medio para las diferentes estimaciones (señal biaural simulada).

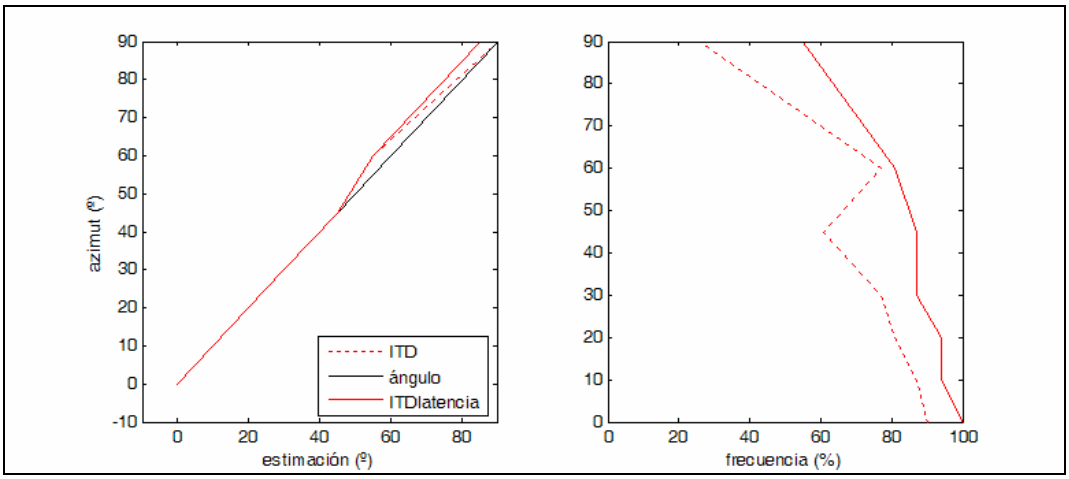


Figura 35. Representación de las estimaciones y su frecuencia estadística basadas en la ITD con latencia (curva continua) y en la ITD sin latencia (curva punteada), (señal biaural real).

	ITDlatencia		ITD	
Ángulo señal biaural (º)	**Moda (º)**	**Frec. (%)**	**Moda (º)**	**Frec. (%)**
0	0	100	0	90
10	10	94	10	87
20	20	94	20	81
30	30	87	30	77
45	45	87	45	61
60	55	81	55	77
90	85	55	90	26

Tabla 10. Estimaciones y frecuencias estadísticas de estimación cuando se usa latencia y cuando no se usa (señal biaural real).

Conclusión

Como se ve en la Figura 35 la frecuencia estadística de la estimación basada en la ITD con latencia es ligeramente mayor para todos los ángulos, cuando se hace la prueba sobre las hrir de la base de datos.

Para la señal biaural real esto también se confirma, el valor de la estimación es el mismo pero se mejora en la frecuencia estadística, luego se va a utilizar este estimador con latencia en el cálculo de la ITD, en la implementación del LES.

3.4.6 ITD BASADA EN EL RETARDO DE GRUPO

[Sontacchi 02] propone calcular la ITD a partir del retardo de grupo. Si se evalúa el retardo de grupo a partir de la derivada negativa de la fase de la respuesta impulsiva aparecen errores. Psicoacústicamente, se ha visto por experimentos que los oyentes no son sensibles a los retardos entre señales si estas aparecen en diferentes bandas. Por lo tanto, se propone calcular el retardo de grupo T_g de cada una de las bandas "perceptuales" de la siguiente manera:

$$T_g(f) = \frac{\sum_n n \cdot h_F^2(t,f)}{\sum_n h_F^2(t,f)} \qquad ITD(f) = T_{g,L}(f) - T_{g,R}(f) \qquad \text{Ec. 26}$$

Siendo h_F la señal biaural filtrada de cada banda f .

Después se estima el ángulo de llegada haciendo una búsqueda en la ITD de la base de datos que más se parezca. Esta búsqueda debe estar limitada alrededor del ángulo que se va a estimar.

Este método ha sido implementado y da unos pésimos resultados cuando la señal es real. Además no se puede saber a priori cual es el ángulo de la fuente, ya que el estimador no debe conocer su posición.

3.4.7 INFLUENCIA DEL FILTRO DEL OÍDO EXTERNO Y MEDIO EN LOS RESULTADOS

Este filtro del oído externo y medio (OEM) se aplica a las dos señales biaurales, izquierda y derecha, por lo tanto su influencia en la ILD e ITD, que miden la diferencia entre las dos señales es casi nula, como se ve en la figura siguiente.

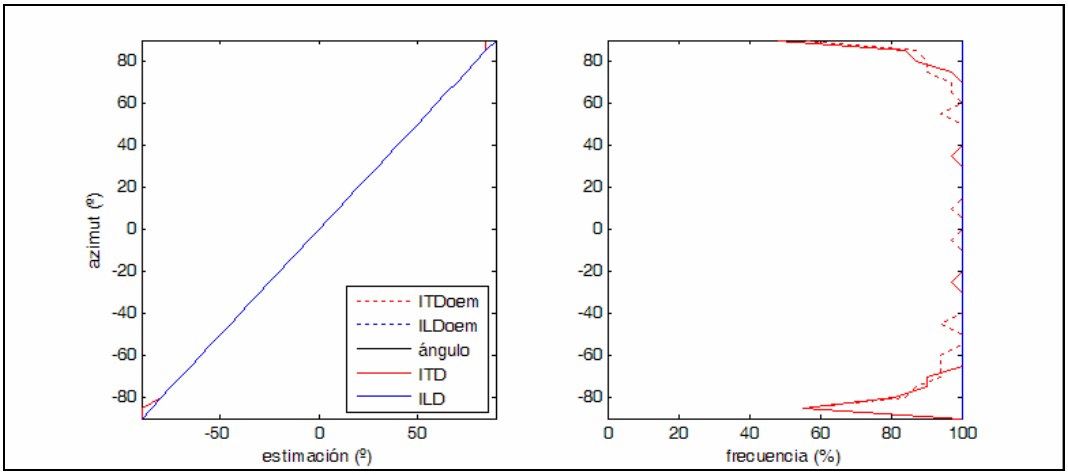


Figura 36. Representación de las estimaciones y su frecuencia estadística cuando se utiliza el filtro OEM (ITDoem - ILDoem) y cuando no se usa (ITD - ILD) para las hrir de la base de datos de HATS.

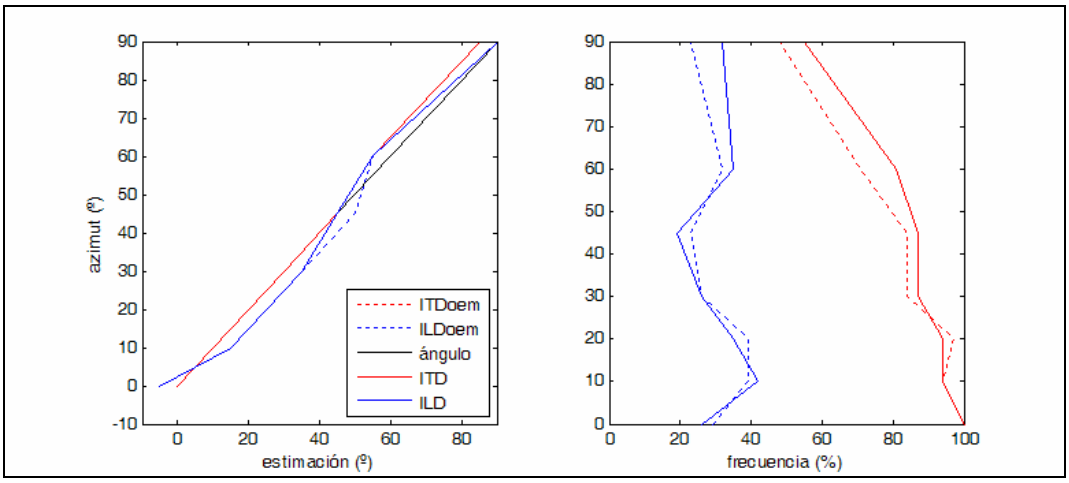


Figura 37. Representación de las estimaciones y su frecuencia estadística cuando se utiliza el filtro OEM (ITDoem - ILDoem) y cuando no se usa (ITD - ILD) para la señal biaural real.

Conclusión

En la primera gráfica (Figura 36) se representa la estimación del ángulo que se obtiene a partir de las hrir directamente; con filtro o sin filtro los resultados son correctos en ambos casos. En la segunda gráfica (Figura 36) se representa la frecuencia estadística de la estimación con filtro y sin filtro, como se ve son prácticamente iguales. Es con señales biaurales reales donde se nota un empeoramiento al utilizar el filtro OEM (Figura 37). Esto puede ser debido a que en la captación de la señal se utiliza un maniquí acústico que ya tiene simulado un oído externo y medio, luego introducir está etapa es redundante. Este filtrado viene implícito en las hrir según han sido medidas con el maniquí.

3.4.8 COMPARACIÓN ENTRE BANCOS DE FILTROS DEL MODELO DE CÓCLEA

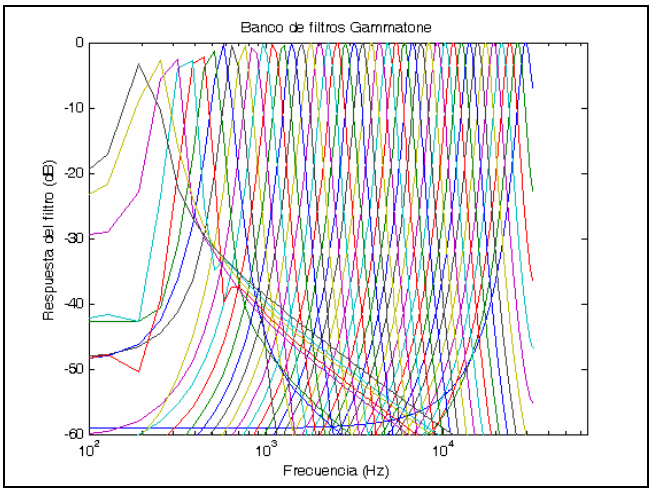


Figura 38. Respuesta del banco de filtros gammatone utilizado en el LES de 42 canales.

Se va a utilizar el nuevo banco de filtros propuesto por [Slaney 98] en lugar del que se usa en la toolbox de HUTear [Harma 00] porque da mejor respuesta de los filtros en baja frecuencia como se puede apreciar en las figuras anterior y posterior.

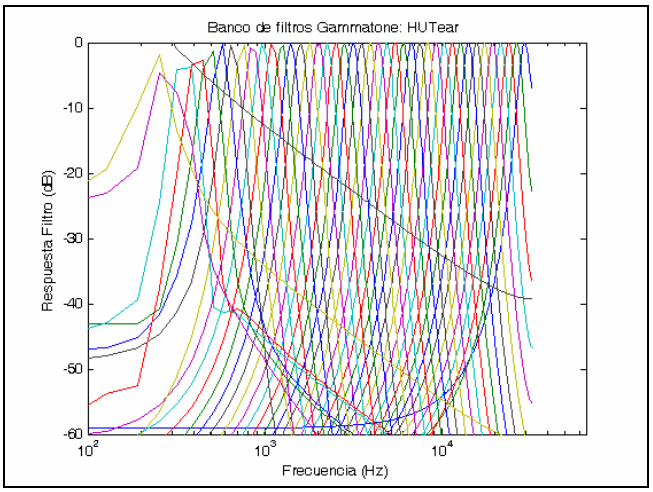


Figura 39. Respuesta del banco de filtros gammatone utilizado en HUTear de 42 canales.

3.4.9 COMPARATIVA ENTRE OÍDOS INTERNOS DEL MODELO DE CÉLULAS CILIARES

En el modelado de la respuesta del oído interno se utilizan diferentes técnicas. Una de las posibilidades es realizar una rectificación de media onda como la utilizada en el LES o una rectificación de onda completa como la utilizada en la toolbox de HUTear [Harma 00]. Se ha realizado una comparativa entre ambas opciones y el resultado es el siguiente:

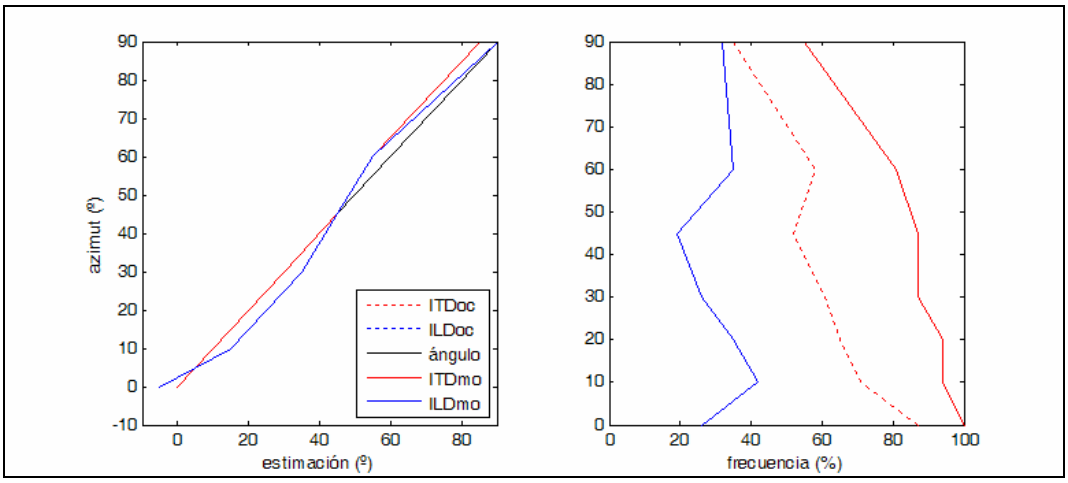


Figura 40. Representación de las estimaciones y su frecuencia estadística cuando se utiliza rectificación de onda completa (ITDoc - ILDoc) y media onda (ITDmo - ILDmo) para la señal biaural real.

La frecuencia estadística de la estimación es mayor para la rectificación de media onda que se ha implementado en el LES.

Se puede elegir también entre dos compresiones: una función de compresión no lineal basada en el nivel de presión sonora o la clásica regla de compresión $y = x^c$ (típicamente c = 0.7) [Blauert 97].

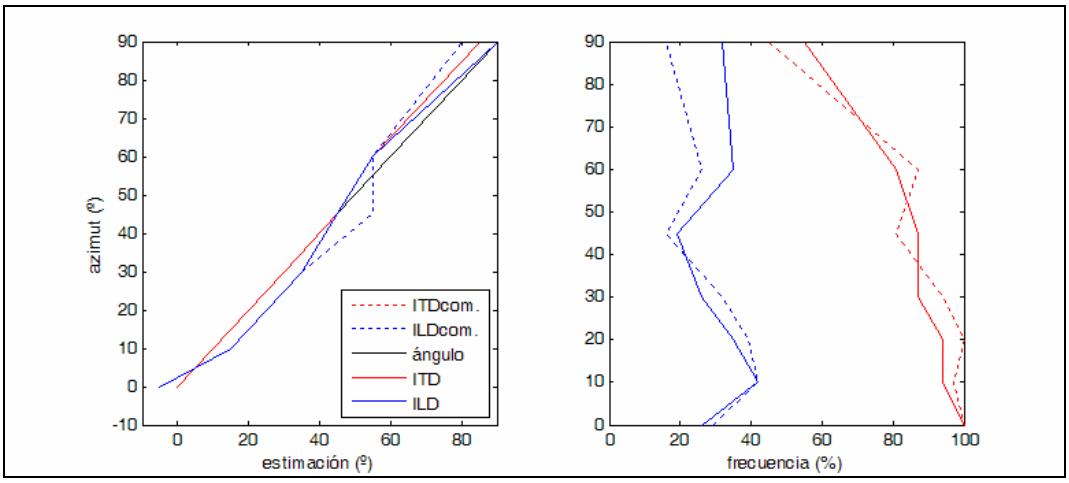


Figura 41. Representación de las estimaciones y su frecuencia estadística cuando se utiliza comprensión (ITDcom. - ILDcom.) y cuando no se usa (ITD - ILD) para la señal biaural real.

Conclusión

La compresión no supone ninguna mejora en los resultados por lo que en el modelo final simplificado implementado en el LES no se utiliza.

Otra de las opciones a la hora de simular el comportamiento de las células filiares es el filtrado, que en el LES se ha implementado como un filtro paso bajo para todos los canales de frecuencia de corte 1 kHz. También existe la posibilidad de hacer un filtrado usando un integrador de primer orden con una constante de tiempo de 20 ms o usar el modelo de Meddis [Harma 00] para las células ciliares que simula el modelo de la sinapsis (contacto entre neuronas) o un modelo neuronal refractivo. En las pruebas realizadas se ha visto que ninguno de estos tres modelos funciona cuando se hace un procesado biaural.

3.4.9.1 MODELO DEL OÍDO INTERNO SEGÚN EL BINAURAL CUE SELECTION

Esta parte del modelo del oído puede tener muchas implementaciones. En las toolbox desarrolladas en el Bianural Cue Selection [Faller 04a] se puede configurar y elegir una rectificación de media onda más compresiones de diferentes tipos. Se han evaluado las posibilidades y en concreto se presenta a continuación los resultados de la configuración elegida por [Faller 04]. En este modelo la envolvente de la señal de salida de cada banda del banco de filtros es comprimida elevando la señal a la potencia de 0.23. Después se hace una rectificación de media onda, seguido de una elevación al cuadrado y un filtrado paso bajo de cuarto orden con una frecuencia de corte de 425 Hz.

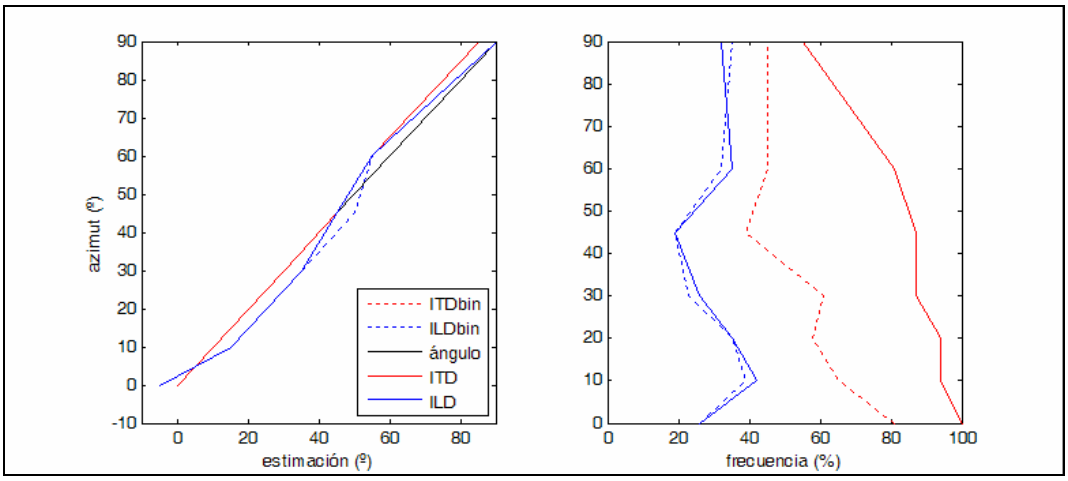


Figura 42. Representación de las estimaciones y su frecuencia estadística con el modelo de oído interno descrito por Faller (ITDbin - ILDbin) y el modelo utilizado en el LES (ITD - ILD) para la señal biaural real.

Conclusión

Como se ve en las figuras, aunque la estimación a partir de la ITD es igual de buena, la frecuencia estadística de la estimación disminuye mucho. La estimación a partir de la ILD es peor excepto en dos ángulos y la frecuencia estadística de la estimación muy parecidas. Por lo tanto no se va utilizar este modelo de oído interno.

3.4.10 SENSIBILIDAD DEL LES A LA BASE DE DATOS

El Localizador de Eventos Sonoros se ha optimizado utilizando siempre la base de datos de hrir del maniquí HATS. Para comprobar como influye la base de datos utilizada a la hora de localizar los eventos sonoros se ha utilizado en el LES la base de datos de KEMAR.

Esta base de datos está muestreada a 44100 Hz, por lo tanto ha sido necesario un remuestreo a 65536 Hz que es la frecuencia de muestreo con la que se trabaja en el LES.

Se han analizado dos situaciones: la capacidad de estimación del LES sobre la propia base de datos, ya que los parámetros de localización ILD e ITD están implícitos en las hrir. Y la capacidad de estimación sobre una señal biaural real medida con el maniquí HATS. Los resultados se muestran comparándolos siempre con los obtenidos si se usa la base de datos de HATS.

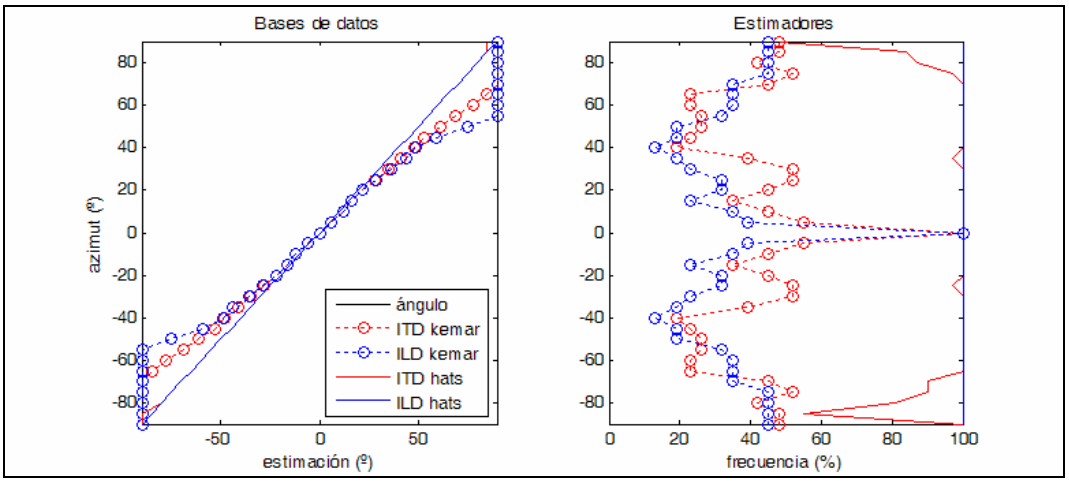


Figura 43. Representación de las estimaciones y su frecuencia estadística para la base de datos de KEMAR (ITD kemar – ILD kemar) y la de HATS (ITD hats – ILD hats) cuando se utiliza una u otra en el LES. Análisis de la estimación sobre las hrir de las bases de datos.

Como puede verse la estimación es más acertada para la base de datos del HATS. Para la base de KEMAR existen desviaciones a ángulos por encima de 40º y la frecuencia de repetición de la estimación es mucho menor.

Directamente, visualizando las tablas de búsqueda de las ITD e ILD, se ve cómo para KEMAR, estas tienen un comportamiento no monótono tanto en frecuencia como en azimut, para ángulos de azimut grandes provocando un error en la estimación.

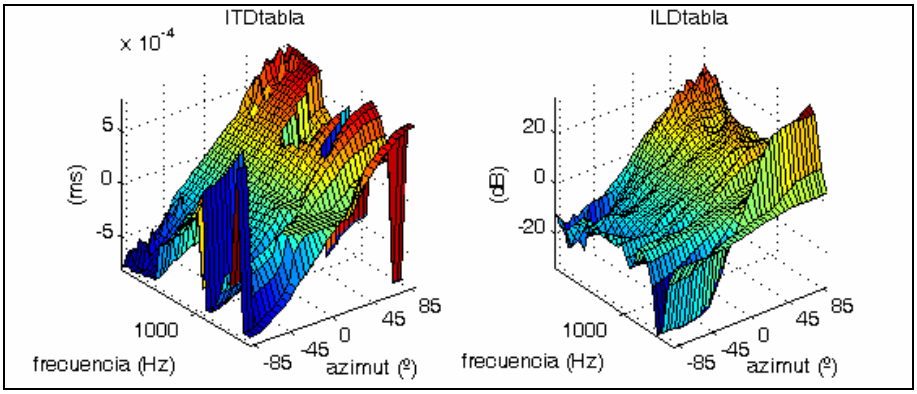


Figura 44. Representación de las tablas de búsqueda de las ITD e ILD para la base de datos de KEMAR.

En segundo lugar vamos a comprobar la capacidad de estimación del LES basándose en una base de datos de un maniquí diferente al maniquí utilizado para captar la señal biaural. Se ha grabado una señal biaural con el maniquí de HATS y se utiliza la base de datos de KEMAR.

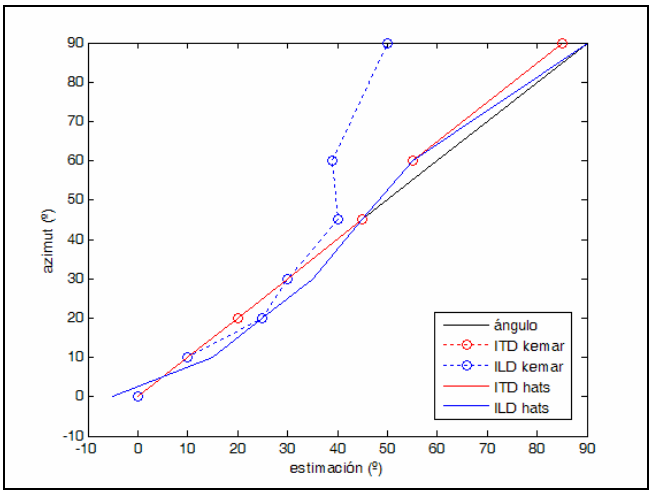


Figura 45. Representación de las estimaciones cuando se usa la base de datos de KEMAR (ITD kemar – ILD kemar) y la de HATS (ITD hats – ILD hats) para una señal biaural grabada con un maniquí HATS.

Los resultados son bastantes parecidos para la estimación de la ITD independientemente de la base de datos utilizada, sin embargo, la estimación basada en la ILD cuando se utiliza la base de datos de KEMAR es muy grande. Esto demuestra que la variación del nivel de llegada a cada oído depende mucho de la difracción y absorción provocada por la cabeza y hombros. El material y la forma de la cabeza por lo tanto influirán mucho en estas variaciones del nivel. En la siguiente foto se puede apreciar la diferencia en la forma de la cabeza y de los hombros para los dos maniquís.

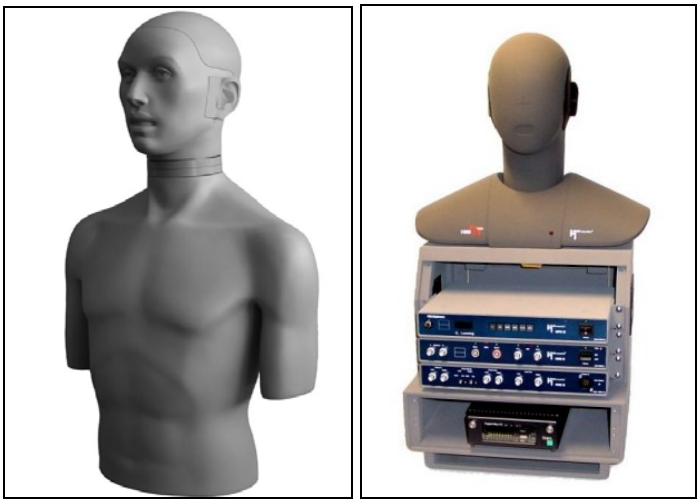


Figura 46. Fotografía del maniquí KEMAR (derecha) y del maniquí HATS (izquierda).

3.4.11 SENSIBILIDAD DEL LES AL TIPO DE SONIDO UTILIZADO

En los ensayos se ha utilizado ruido blanco porque al ser una señal estacionaria, la localización es independiente de la forma de onda de la señal. Se sabe que con señales de voz o señales musicales los ataques ayudan a localizar las fuentes. Se va a comprobar como funciona el LES cuando la señal emitida por la fuente es de voz; comparándolo con las mismas posiciones de la fuente cuando se usa una señal de ruido blanco.

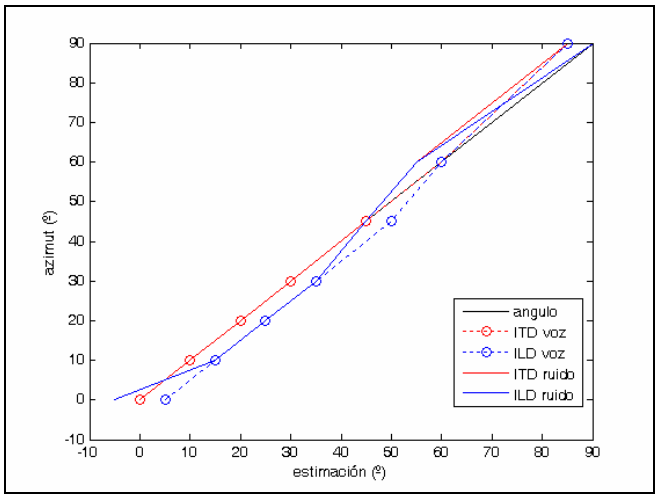


Figura 47. Representación de las estimaciones con dos tipos de sonido, señal de voz (ITD voz – ILD voz) y ruido blanco de banda ancha (ITD ruido – ILD ruido) con señal biaural real.

La sensibilidad del LES al tipo de señal utilizada es baja, las diferencias con respecto al ángulo correcto son de 5º para algunos ángulos y con respecto a la estimación de ruido blanco también son de 5º. En el caso de la estimación basada en ITD, la respuesta es correcta en todos los ángulos excepto en 90º y la estimación basada en la ILD está desviada 5º excepto en un ángulo.

Conclusión

Se puede concluir que no existen grandes diferencias a la hora de utilizar estos dos tipos de señales cuando la estimación se basa en diferencias de nivel y diferencias de tiempo de llegada de la señal biaural. Hay que tener en cuenta que se analiza un 0.5 s de señal, suficiente en el caso de voz para apreciar la diferencia de nivel y tiempo.

3.5 APLICACIÓN DEL LES A LA VENTANA ACÚSTICA VIRTUAL

Teniendo en cuenta que el objetivo de la Tesis es valorar la capacidad de reproducción de los sistemas de teleconferencia, se va a simular una ventana acústica virtual colocada entre la fuente y el oyente. La idea de ventana acústica virtual consiste en reproducir el campo sonoro que se capta en una sala en otra sala, de forma que la sensación acústica sea que hay una ventana o hueco físico real separando las dos salas [Harma 02].

En esta simulación, las señales aplicadas a los altavoces son calculadas simulando un array de micrófonos. Los micrófonos simulados son ideales. El efecto de la no idealidad de los transductores electroacústicos en la percepción direccional se deja para futuros estudios; ya que habría que valorar la influencia de su directividad.

En el escenario acústico simulado se coloca una fuente en una posición relativa a la cabeza (x,y). Se muestrea el campo sonoro que se propaga con un array lineal de 10 micrófonos (el número de micrófonos elegido es el número de altavoces del array construido para realizar las medidas). Se reproduce dicho campo con un array lineal de 10 altavoces (conexión “hard-wired”) (Figura 48). La distancia entre la fuente y el array de micrófonos será la mitad de y. Y la distancia entre el array de altavoces y el oyente también la mitad. Si se supone que el muestreo del frente de onda y la reproducción del frente de onda son

perfectos, según el Principio de Huygens, el camino acústico será el mismo que si se estuviera haciendo una radiación directa con una fuente colocada a una distancia r.

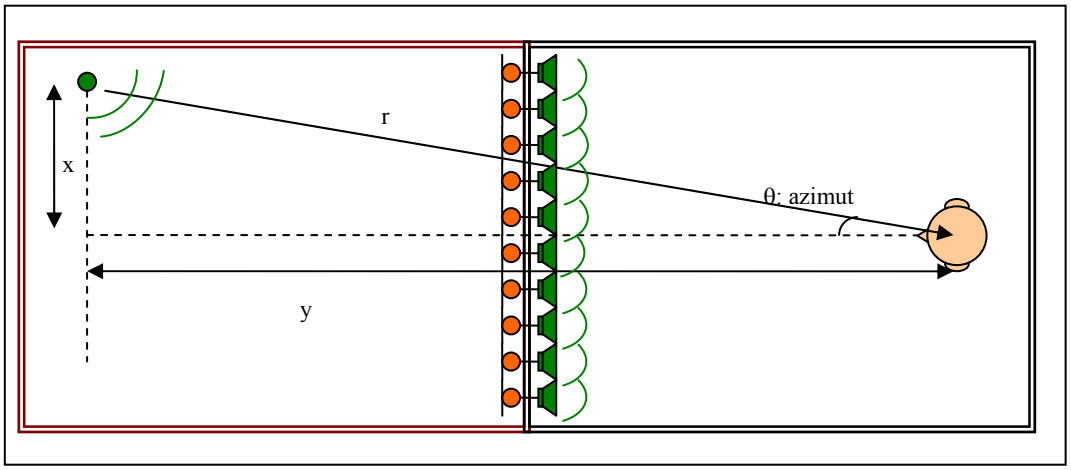


Figura 48. Simulación de ventana acústica virtual a través de un array de micrófonos y de altavoces. Posiciones relativas oyente-fuente.

En este escenario, la simulación se lleva a cabo en dos partes. La primera consiste en simular las señales captadas por los diez micrófonos teniendo en cuenta la posición relativa de cada micrófono a la fuente según su posición en el array (cada señal sufre un retardo y una atenuación diferente). La segunda consiste en simular la emisión de diez altavoces y calcular la señal biaural recibida. Para ello se aplica la señal captada por cada micrófono directamente a cada altavoz.

3.5.1 SEÑAL DEL ARRAY DE MICRÓFONOS SIMULADA

En la figura anterior se puede ver una representación del array lineal de N micrófonos a los que llega un frente de ondas incidiendo con un ángulo θ respecto al oyente. Se supone que las ondas son esféricas (distancias superiores a 1 m) y la propagación es en campo libre. Así pues a cada micrófono i le llega una versión retardada de la señal que se emite (sonido[n]) según la distancia, siendo el retardo:

$$\tau_i = \frac{\mathrm{dis}(\mathrm{fuente} - \mathrm{micro}_i)}{c} \qquad \text{Ec. 27}$$

El número de ceros a añadir será:

$$\mathrm{n^o ceros}(r,i) = \mathrm{round}\left(\tau_i \cdot f_{\mathrm{muestreo}}\right) \qquad \text{Ec. 28}$$

Por lo tanto, la señal en cada micrófono se calcula teniendo en cuenta la atenuación por distancia:

$$\mathrm{altavoz}_i[m] = \left[\mathrm{n^o ceros}(r,i) \quad \frac{\mathrm{sonido}[n]}{\mathrm{dis}(\mathrm{fuente} - \mathrm{micro}_i)} \right] \qquad \text{Ec. 29}$$

Esta señal, $\mathrm{altavoz}_i$, es la que se utiliza para ser emitida por el array de altavoces, tanto en el caso de simulación de la sala receptora como en el caso de implementación real del array en la cámara anecoica.

3.5.2 SEÑAL BIAURAL SIMULADA EMITIENDO CON ARRAY

La señal biaural se calculará como la suma del filtrado digital de las señales emitidas por los altavoces con las hrir interpoladas, correspondientes a cada posición del altavoz-oyente, según la teoría de la localización sumatoria [Blauert 97].

A continuación se muestran los resultados de las estimaciones basadas en la ITD y en la ILD, cuando se aplica el LES a la señal biaural simulada, para el array descrito anteriormente. El ángulo de la fuente irá desde 0º hasta 90º en pasos de 5º.

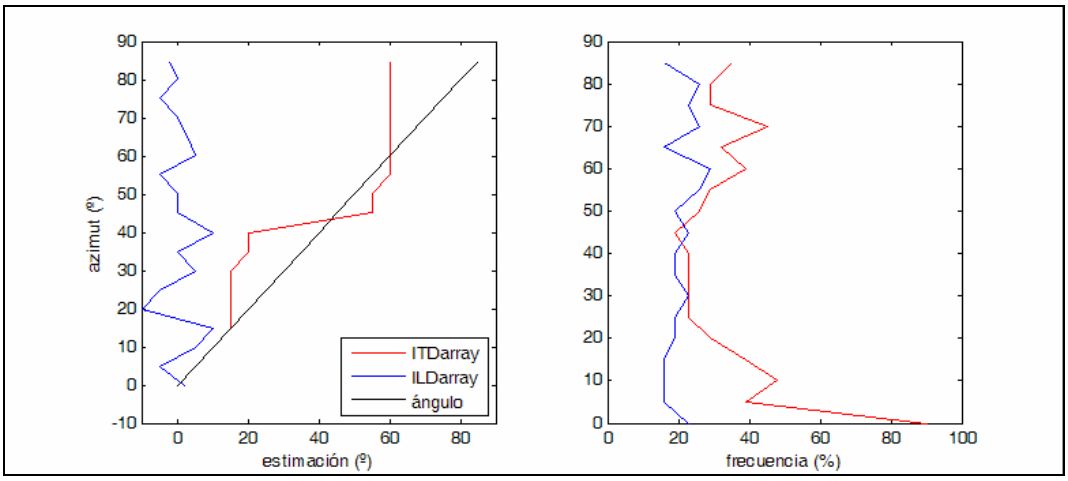


Figura 49. Representaciones de las estimaciones y su frecuencia estadística obtenidas para la señal biaural simulada emitida por un array para los ángulos de 0º a 90º en pasos de 5º.

Ángulo evento sonoro (º)	**0**	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**
Estimación ITD (º)	0	5	10	15	15	15	15	20	20	55	55	60	60	60	60	60	60	60
Estimación ILD (º)	2	-5	5	10	-10	-5	5	0	10	0	0	-5	5	3	0	-5	0	-2

Tabla 11. Estimaciones del LES para señal biaural simulada radiada por un array.

También se muestran a continuación las estimaciones del array simulado comparándolas con las estimaciones para el caso de radiación directa en una cámara anecoica (0º, 10º, 20º 30º 45º y 60º). Con un array lineal como el de la configuración, no se pueden reproducir fuentes colocadas a 90º, por eso se representan los resultados hasta 60º.

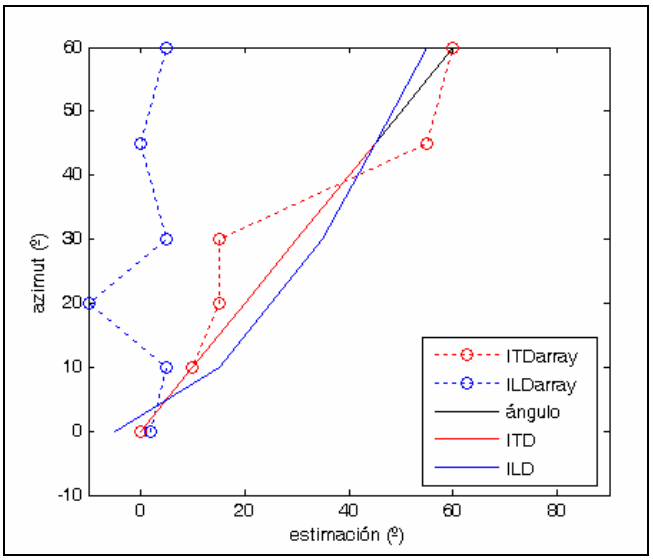


Figura 50. Comparación entre las estimaciones para una señal biaural simulada cuando se emite por un array (ITDarray - ILDarray) y la señal biaural real cuando la emisión es directa y real (ITD - ILD).

Para esta configuración del array de altavoces, la apertura acústica del array es:

$$\theta_{max} = \text{arctg}\left(\frac{((n^{o}_{altavoces}-1)*d_{alt})/2}{y/2}\right) = \text{arctg}\left(\frac{((10-1)*0.175)/2}{5/2}\right) = 17.5^{o} \quad \text{Ec. 30}$$

El ángulo que forma el oyente con el altavoz más alejado es 17.5º, luego por la teoría acústica, el array no podrá representar eventos sonoros fuera de este ángulo.

Analizando la estimación basada en la ITD, del evento sonoro representado por el array se ve que coincide con la posición de la fuente fielmente hasta 15º. A partir de una posición de la fuente de 20º hasta 40º la estimación se queda en la apertura acústica del array. Por encima de esta posición la estimación tiene un valor constante de 60º.

Analizando la estimación basada en la ILD se ve que para ángulos de la fuente menores a la apertura acústica del array, la estimación es cercana a la posición real, pero por encima de la apertura acústica la estimación de la ILD no es buena.

Para analizar los resultados en función de la frecuencia y analizar la influencia de la frecuencia de aliasing del array se va a representar la estimación de la ITD y de la ILD para los distintos ángulos.

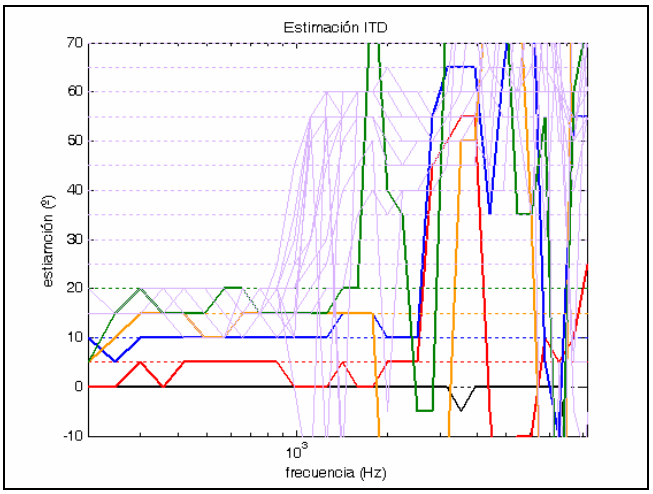


Figura 51. Representación de las estimaciones ITD en función de la frecuencia para la señal simulada emitida por un array (0º, 5º, 10º, 15º, 20º, de 25º a 70º en pasos de 5º).

Como se puede observar la estimación dentro de la apertura acústica del array (< 20º) es buena para frecuencias menores que la de aliasing (2000 - 6000 kHz). Para los demás ángulos las estimaciones en baja frecuencia no pasan de los 20º.

Para alta frecuencia la capacidad del array de simular la ventana acústica virtual debido a la teoría del WFS es mala, y por lo tanto los parámetros de localización no se conservan. Hay que destacar que existe una tendencia en este margen de frecuencias a estimar 60º, también observado en la Figura 49.

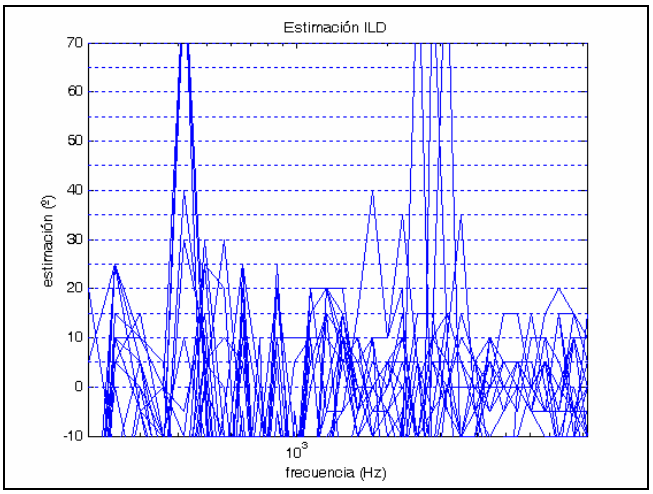


Figura 52. Representación de las estimaciones ILD en función de la frecuencia para la señal simulada emitida por un array.

Como se aprecia en la figura anterior el parámetro de localización basado en la ILD no se conserva, quedando prácticamente limitados todos los resultados de la estimación a ángulos menores que la apertura acústica.

3.5.3 SEÑAL BIAURAL MEDIDA EMITIENDO CON ARRAY

ENSAYO 2

Figura 53. Fotografías del montaje del array en la cámara anecoica.

Con este ensayo se pretende comprobar que la simulación del escenario de ventana acústica virtual y la realidad se parecen bastante a nivel de parámetros de localización sonora. Para ello, se va a montar dentro de la cámara anecoica la parte de reproducción de la ventana acústica virtual. Se construye un array de 10 altavoces y se emite las señales captadas por el array de micrófonos que han sido simuladas. Por lo tanto la parte de recepción será simulada y la de reproducción será real. Se utiliza el maniquí acústico para captar la señal biaural y comprobar la capacidad del array de simular fuentes para distintos ángulos de azimut.

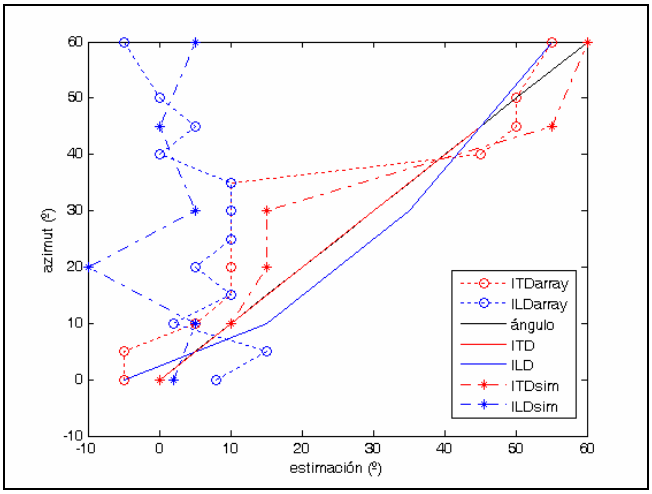


Figura 54. Representación de las estimaciones para la señal biaural real para emisión con el array (ITDarray - ILDarray), la señal biaural simulada emitida con el array (ITDsim - ILDsim) y la señal biaural real para radiación directa (ILD - ITD).

CONCLUSIÓN ENSAYO 2

El comportamiento de la estimación en el caso de radiación con array real es similar al del caso de radiación con array simulado; tanto para la estimación basada en ITD como la

estimación basada en ILD. Se puede llegar a las mismas conclusiones con array simulado y array real: la incapacidad de estimar el azimut si se hace la estimación basada en la ILD, la doble limitación de la estimación basada en la ITD a la apertura acústica del array y a la frecuencia de aliasing. Por lo tanto podemos concluir que la simulación es bastante fiel a la realidad, y se puede utilizar este escenario acústico para estudiar la influencia de la configuración del array en la localización del evento sonoro.

La percepción sonora que se tiene cuando se utiliza el array real es de desplazamiento del sonido hacia el último altavoz, según va variando la posición de la fuente simulada.

Quedaría como trabajo futuro realizar tests subjetivos de la capacidad de reproducir con el array de altavoces diferentes localizaciones del evento sonoro.

3.5.4 VARIACIÓN DE LA ESTIMACIÓN EN FUNCIÓN DE LA CONFIGURACIÓN DE LA VENTANA ACÚSTICA VIRTUAL

Se han visto dos claras limitaciones que presenta la ventana acústica virtual a la hora de reproducir los parámetros de localización del sonido. Para ver como influyen estas dos limitaciones sobre los parámetros de localización se van a simular diferentes tipos de array con configuración distinta. Para comenzar, se va a variar la apertura acústica del array (es decir, el número de altavoces para la misma distancia entre altavoces), para comprobar que la capacidad de simular eventos sonoros sólo se produce dentro de la apertura. Se han simulado 5 arrays con distancia entre altavoces constante de 0.175 m y número de altavoces variable. La apertura acústica se calcula según la Ecuación 30.

Nº altavoces	Apertura acústica
5	8º
10	17º
15	26º
20	34º
25	40º

Tabla 12. Apertura acústica del array en función del nº de altavoces.

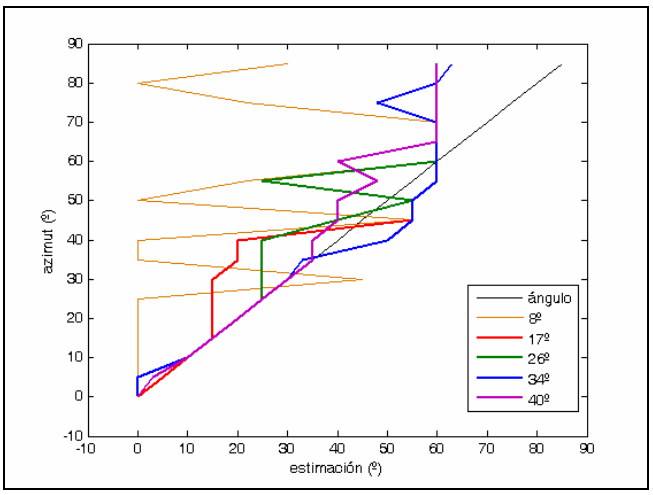


Figura 55. Representación de las estimaciones basadas en la ITD para la señal biaural simulada cuando el array varía el nº de altavoces.

Para 5 altavoces (curva naranja), 8º de apertura acústica, la estimación es totalmente errónea. Para 10 altavoces (curva roja) la estimación se mantiene bien hasta 15º, la apertura es de 17º. Para 15 altavoces (curva verde) la estimación se mantiene bien hasta 25º, la apertura es de 26º. Para 20 altavoces (curva azul) la estimación se mantiene bien hasta 35º, la apertura es de 34º. Y por último, para 25 altavoces (curva morada) la estimación se mantiene bien hasta 35º siendo la apertura de 40º.

En cuanto a la estimación según la frecuencia, los resultados son los siguientes. Se ha representado en curvas de color malva los ángulos que están fuera de la apertura acústica en cada caso.

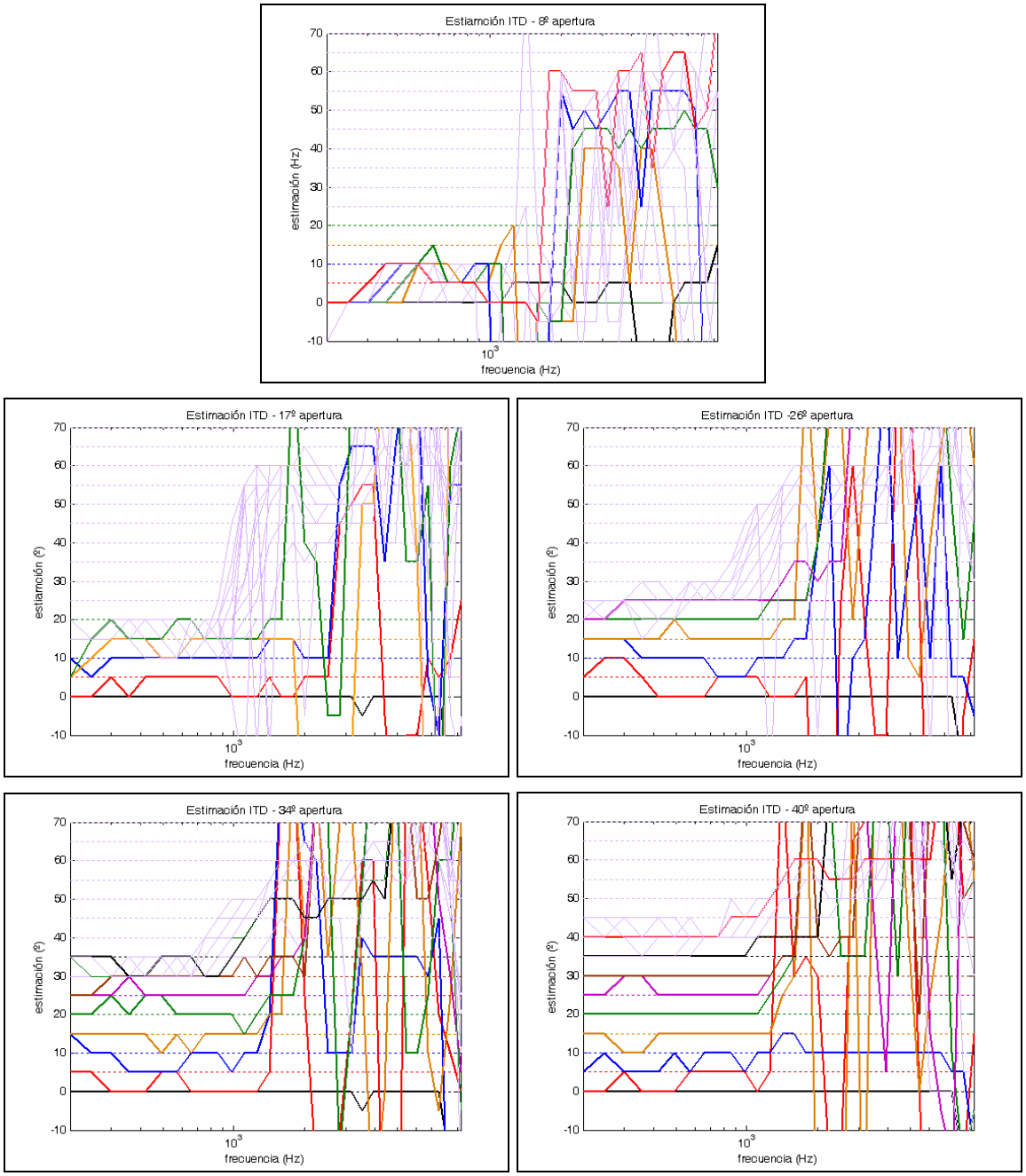


Figura 56. Representación en frecuencia de las estimaciones basadas en la ITD para la señal biaural simulada cuando el array varía el nº de altavoces.

Como se aprecia en la Figura 56, según va aumentando la apertura la capacidad de representar eventos sonoros en diferentes acimuts se amplía. En todos los casos la limitación por frecuencia de aliasing se mantiene.

Conclusión

Queda comprobada la limitación de representar eventos sonoros por un array a la apertura acústica de dicho array. En cuanto la estimación basada en la ILD da estimaciones erróneas como ya se había visto antes para todas las configuraciones. No se mejora esta estimación cuando se aumenta el número de altavoces.

La segunda limitación es la frecuencia de aliasing del WFS [Boone 95]: para frecuencias mayores a la de aliasing el campo sonoro no se reproduce correctamente. El parámetro que influye en la frecuencia de aliasing es la distancia entre los altavoces y micrófonos del array, el ángulo entre la fuente y el plano del array de micrófonos y el ángulo entre el oyente y el plano del array de altavoces. Se va a variar la distancia entre altavoces - micrófonos, manteniendo la apertura acústica a 17º, para ello hay que variar el número de micrófonos - altavoces para cubrir la misma longitud del array. Por lo tanto, se va a analizar cómo influye la distancia entre los transductores sobre la frecuencia aliasing y por consiguiente sobre la capacidad de estimar el ángulo de azimut según aumenta la frecuencia. La expresión de la frecuencia de aliasing es la dada por [De Brujin 03]:

$$f_{max} = \frac{c}{d_{alt}\left(\sin\theta_{max,M} + \sin\theta_{max,L}\right)} \qquad \text{Ec. 31}$$

Se han simulado 4 arrays con distancia entre altavoces variable y apertura acústica constante:

Distancia entre altavoces (m)	**Frecuencias aliasing (Hz) 85º - 0º de azimut**	**Nº altavoces**
0.050 m	23168 - 6866	32
0.100 m	11937 - 3433	16
0.175 m	6524 - 1962	10
0.200 m	6361 - 1373	8

Tabla 13. Rango de las frecuencias de aliasing en función de la distancia entre altavoces.

El cálculo de la frecuencia de aliasing se muestra en el Anexo 1-1.

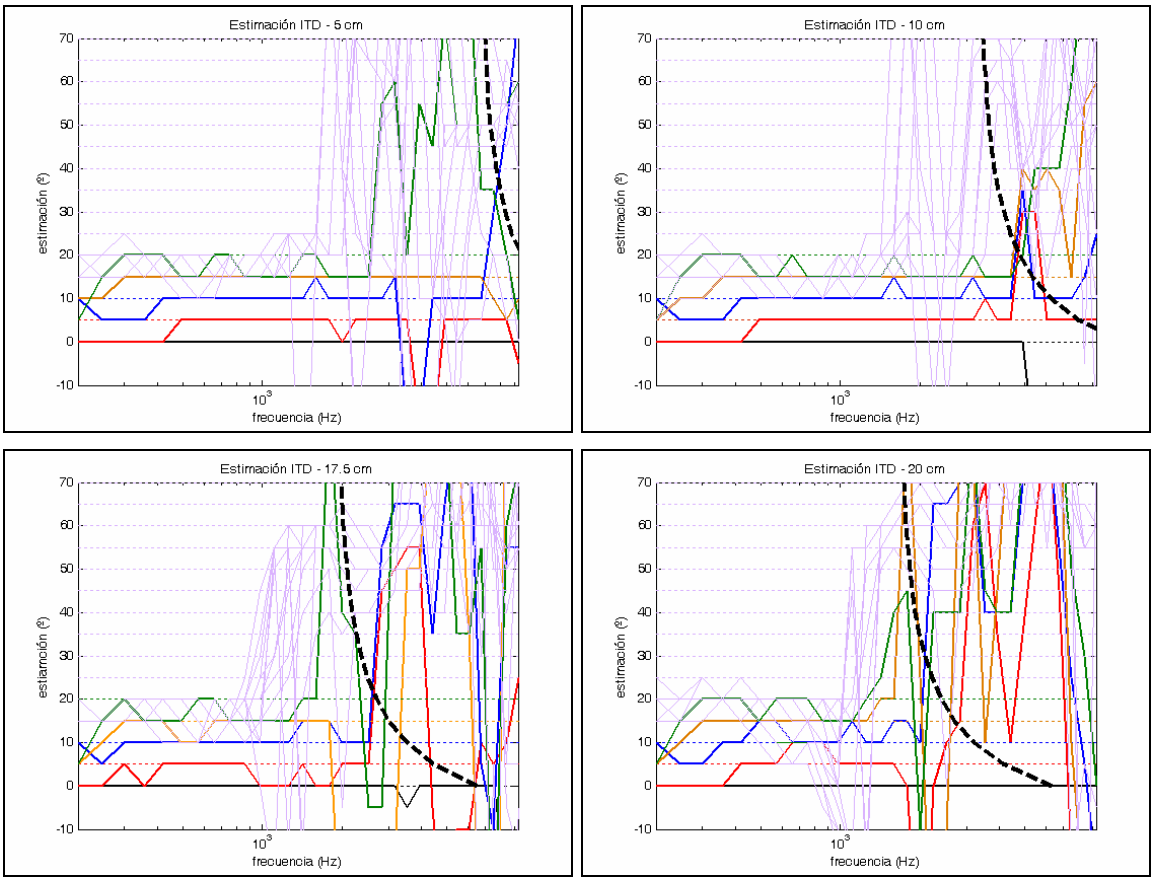


Figura 57. Representación en frecuencia de las estimaciones basadas en la ITD para la señal biaural simulada cuando el array varía la distancia entre altavoces.

En la figura anterior se muestra la frecuencia de aliasing calculada para cada ángulo de posición del evento sonoro, azimut de 0º a 85º, en una línea negra punteada. Se ve como esta curva se desplaza hacia las bajas frecuencias cuando la distancia entre altavoces va creciendo. Además, se aprecia como el mal comportamiento en la estimación del azimut va siguiendo el desplazamiento de las frecuencias de aliasing, siendo su límite en frecuencia para una correcta estimación. Esto se cumple para los ángulos dentro de la apertura acústica (líneas gruesas y no malvas).

Conclusión

Por tanto, queda comprobada la limitación de representar eventos sonoros por un array a la apertura acústica de dicho array y a la frecuencia de aliasing, basándonos en los parámetros de localización. En cuanto a la estimación basada en la ILD sigue dando estimaciones erróneas como ya se había visto antes para todas las configuraciones. No se mejora esta estimación cuando se varía la frecuencia de aliasing.

3.6 EJEMPLO DE APLICACIÓN DEL LES

El origen de este trabajo de investigación fue elaborar una herramienta que evaluara la calidad del sonido reproducido por un WFS con diferentes técnicas de procesado, en cuanto a la capacidad de recrear el evento sonoro en la posición equivalente. Para ello se implementó el LES, que evalúa los parámetros de localización sonora.

En la siguiente prueba se aplica el localizador de eventos sonoros a diferentes escenarios acústicos de WFS y se comparan los resultados de la estimación, para determinar cual de ellos es mejor con respecto al criterio de localización. El escenario acústico simulado se basa en la ventana acústica virtual. Se tiene una sala emisora con dos fuentes, una deseada y otra interferente. El sonido se capta con un array de 10 micrófonos con una distancia entre ellos de 0.175 m. La fuente deseada está situada a 3.938 m del array y en el eje central perpendicular a este (x = 0), emite una señal de voz. La fuente interferente está situada a 2.938 m del array y desplazada x = -3 m del eje central, emitiendo otra señal de voz.

En la sala emisora se tiene un array de 10 altavoces separados 0.175 m con un maniquí acústico situado a 2.5 m del array en el eje central.

Se van analizar cuatro casos diferentes de procesado de la señal que se transmite entre el array de micrófonos y el de altavoces:

- Caso 1: Se capta sonido por el micrófono nº 5 y sólo se emite sonido por el altavoz nº 5 del array que está colocado -2º del oyente. (En las gráficas estas curvas se llamarán "mono").
- Caso 2: Se capta sonido por los 10 micrófonos del array y directamente se emiten por el array de altavoces. (Curvas "hw").
- Caso 3: Se utiliza un procesado de la señal captada por el array microfónico basado en "Fixed Beamforming + Wave Field Synthesis" que mejora la percepción de la señal deseada. (Curvas "fb").
- Caso 4: Se utiliza un procesado de la señal captada por el array microfónico basado en "Robust GSC + Wave Field Synthesis" que mejora la percepción de la señal deseada. (Curvas "rgsc").

El procesado de señal ha sido desarrollado por J. Beracochea dentro de su Tesis Doctoral "Codificación de Audio Multicanal para entornos de tipo ventana acústica virtual" [Beracoechea 07]. El LES va analizar la señal biaural simulada captada por el maniquí acústico HATS.

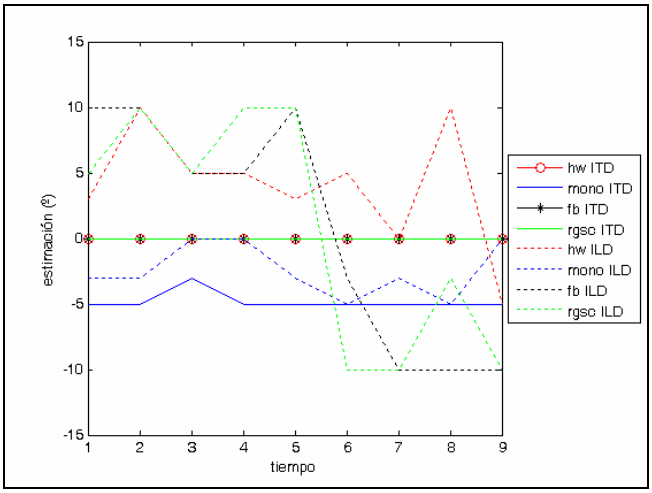


Figura 58. Estimación del azimut del evento sonoro en función del tiempo.

En la Figura 58 se muestra la estimación del azimut basado en la ITD y en la ILD (eje vertical). Como la señal emitida es de larga duración, se va calculando la estimación a lo largo del tiempo para intervalos temporales de 0.5 s, representados en el eje horizontal.

Se puede ver como la estimación basada en la ITD es correcta: la fuente deseada está a 0º de azimut cuando se reproduce con un array y a 5º de azimut cuando la reproducción es mono. Sin embargo la basada en la ILD muestra resultados variables, como ya se ha visto anteriormente; excepto en el caso mono donde sólo emite un altavoz y por lo tanto la estimación de la ILD también es correcta (está entre 0º y 5º, margen de estimación muy bueno). Por otra parte, se aprecia como cuando la emisión es de un solo altavoz (curva mono) la localización del evento se sitúa entorno a -5º,-2.5º cerca del la posición real del altavoz que es 2º.

A continuación se presentan los resultados en función de la frecuencia. En las gráficas siguientes se muestra el resultado de la estimación basada en la ITD (curvas rojas) y la ILD (curvas azules) para cada intervalo temporal de 0.5 s según el eje horizontal de frecuencia (banco de filtros "preceptúales" ERB).

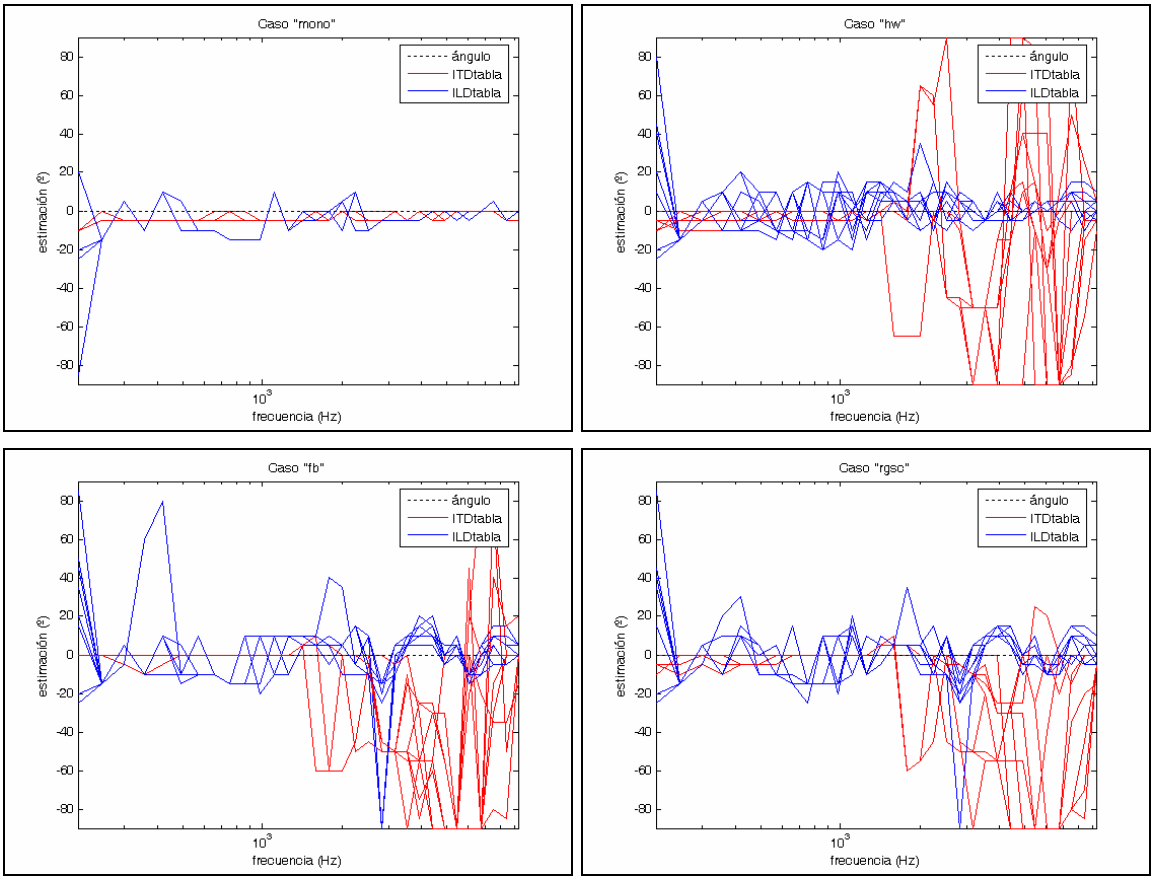


Figura 59. Estimación del azimut del evento sonoro en función de la frecuencia para cada una de las ventanas de tiempo y cada uno de los casos.

Como se puede ver los resultados de la ILD son bastante erráticos a lo largo de todo el margen de frecuencia (el array no conserva la información de la ILD). Sin embargo los resultados de la ITD muestran dos claros comportamientos:

- Primero, cuando no se usa array (curva mono) la estimación es correcta a lo largo de todo el ancho de banda estando centrada a -5º, como ya se ha dicho anteriormente, tanto para la estimación basada en ILD como la basada en ITD.

- Segundo, cuando se usa array debido al aliasing que aparece en el WFS a partir de 1500 Hz la estimación es errónea (comportamiento ya analizado en el estudios anteriores).

A pesar de estos resultados, la sensación sonora real sigue siendo correcta porque el cerebro aplica la condición de consistencia: "un resultado de estimación es consistente cuando ese mismo resultado aparece en un ancho de banda extenso [Wightman-Kistler]".

Si se quiere cualificar la localización según el procesado de la señal, parece ser que el basado en Fixed Beamforming + Wave Field Synthesis presenta menor desviación del valor de azimut verdadero para el margen de bajas frecuencias.

3.7 MODELO BIAURAL STFT

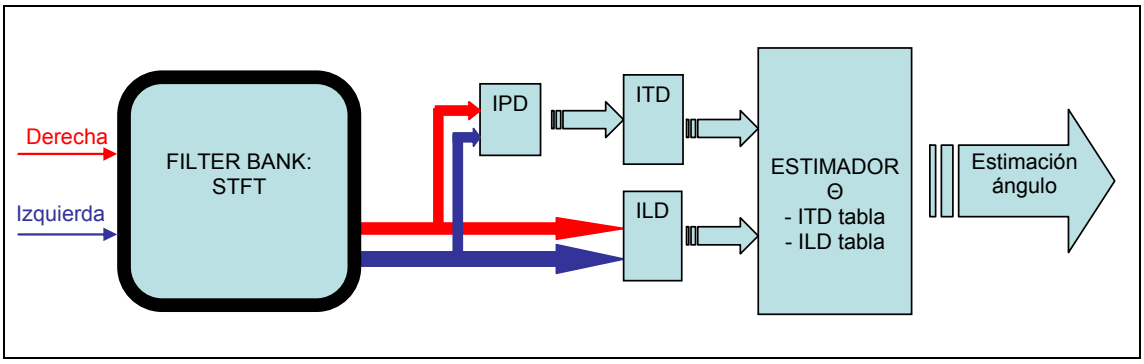


Figura 60. Método de localización biaural basado en la STFT.

En el proceso de búsqueda de un modelo biaural se implementó el siguiente método basado en una propuesta de [Viste 04]. Este modelo a priori pareció muy interesante ya que combinaba la información de la ILD y la ITD. El proceso biaural se basa en la STFT (Short Time Fourier Transform) para procesar la señal biaural en el dominio de la frecuencia. La ITD se calcula a partir de la diferencia de fase interaural (IPD), por lo tanto aparece un problema de ambigüedad en el cálculo de la fase debido a su periocidad. El significado práctico de esto es que para una frecuencia, la diferencia de fase que existe entre los dos oídos puede ser debida a la posición de la fuente sonora en varios sitios.

$$\mathrm{IPD}(t,f,p) = \arg\frac{X_R(t,f)}{X_L(t,f)} + 2\pi p \qquad \text{Ec. 32}$$

$$\mathrm{ITD}(t,f,p) = -\frac{\mathrm{IPD}(t,f,p)}{f}\frac{1}{2\pi} \qquad \text{Ec. 33}$$

donde t es la ventana de tiempo, f es el coeficiente espectral, X_R y X_L son el espectro de las señales biaurales derecha e izquierda y p es el índice de periocidad. Esta forma de calcular la ITD también es usada en [Sontacchi 02].

LA ILD en dB viene dada por:

$$\mathrm{ILD}(t,f) = 20\log\left|\frac{X_R(t,f)}{X_L(t,f)}\right| \qquad \text{Ec. 34}$$

La estimación del azimut basada en la ILD se usa para seleccionar de entre las posibles estimaciones de azimut basadas en ITD la más cercana a ella. La estimación basada en la ITD es más precisa pero ambigua debido a la periocidad, sin embargo la estimación basada en la ILD es menos precisa.

$$\theta(t,f) = \theta_T(t,f,p)\Big|_{p=\arg\min|\theta_L(t,f)-\theta_P(t,f,p)|} \qquad \text{Ec. 35}$$

Se utilizan dos tablas de búsqueda para estimar el azimut a partir del cálculo de la ITD y de la ILD (θ_L, θ_T). Ambas tablas tienen los valores de ITD e ILD calculadas a partir de las medidas de las HRTF del maniquí acústico. Dichos cálculos se realizan para cada ángulo de azimut y para cada frecuencia según las siguientes ecuaciones:

$$IPD(\theta,f) = \arg\frac{HRTF_R(\theta,f)}{HRTF_L(\theta,f)} \qquad \text{Ec. 36}$$

$$ITD(\theta,f) = -\frac{IPD(\theta,p)}{f}\frac{1}{2\pi} \qquad \text{Ec. 37}$$

$$ILD(\theta,f) = 20\log\left|\frac{HRTF_R(\theta,f)}{HRTF_L(\theta,f)}\right| \qquad \text{Ec. 38}$$

Después de procesar la base de datos del HATS con este método por STFT, las tablas de búsqueda que se usan son las siguientes:

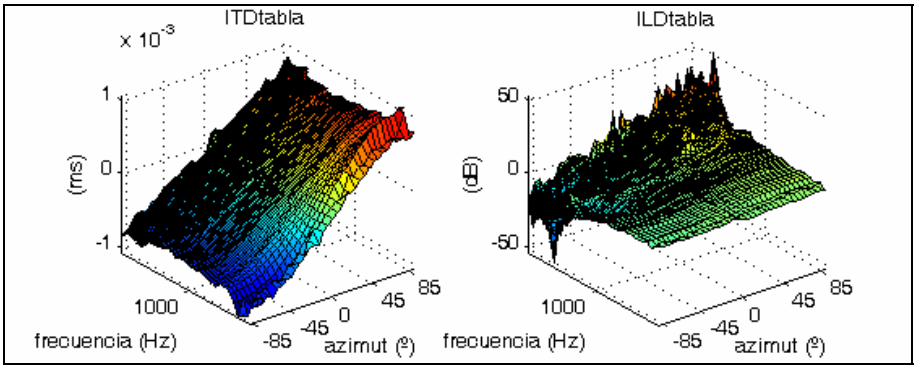


Figura 61. Representación de la ITD e ILD en función del azimut y de la frecuencia para la base de datos del HATS.

Las funciones ILD e ITD pueden no ser monotónicas para alguna frecuencia y por lo tanto en la búsqueda del $\theta_T(t,f)$ y del $\theta_L(t,f)$ se pueden encontrar varios valores posibles de azimut para una misma frecuencia. La solución por la que se ha optado es la elección de la estimación más próxima a 0º (aunque puede que no sea la correcta). Esta solución también es utilizada por el sistema auditivo según se muestra en algunos estudios subjetivos [Blauert 97].

Este modelo utiliza los dos parámetros de localización de la Teoría Dúplex y por lo tanto combina las precisiones de la ITD y de la ILD que dependen del margen de frecuencias.

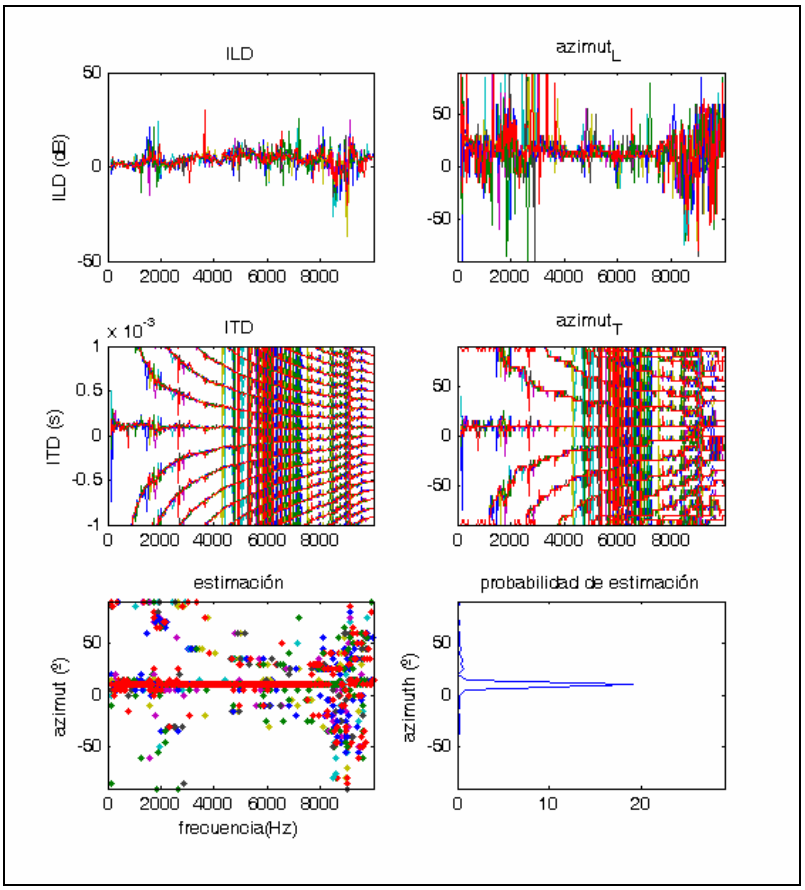


Figura 62. Representación del procesado seguido en el modelo de la STFT para una fuente real colocada a 10º.

La figura anterior muestra los resultados del modelo e ilustra el procesado seguido. Se han analizado 10 ventanas de 2048 muestras mediante la STFT (cada color es una ventana). La señal biaural ha sido grabada por el maniquí acústico dentro de una cámara anecoica colocando una fuente de ruido blanco a 10º. En la primera gráfica se representa la ILD calculada para las ventanas y en la segunda la estimación del azimut a partir de la ILD. La ILD para baja frecuencias es muy errática y por lo tanto en este margen de frecuencias no es útil para determinar el azimut.

En la gráfica tercera se representa la ITD para cada ventana y cada valor de p y en la cuarta la estimación del azimut a partir de la ITD. Como se ve en estos resultados la estimación basada en la ILD tiene una desviación mayor que la basada en la ITD. Además, la estimación por encima de los 2000 Hz es ambigua, hay diferentes estimaciones del azimut para una misma frecuencia y ventana (según p). Según crece la frecuencia esta ambigüedad también crece y por lo tanto la precisión del estimador decrece haciendo que sea poco precisa la estimación. Gráficamente la estimación basada en la ILD es usada para determinar sobre la estimación de la ITD cual p es la solución correcta. La última gráfica representa el histograma de las estimaciones en tanto por ciento, como se ve hay un pico elevado a 10º que es el azimut real de la fuente sonora.

Se ha aplicado este procesado a las señales biaurales grabadas y se compara los resultados con los obtenidos por el modelo psicoacústico biaural utilizado en el LES.

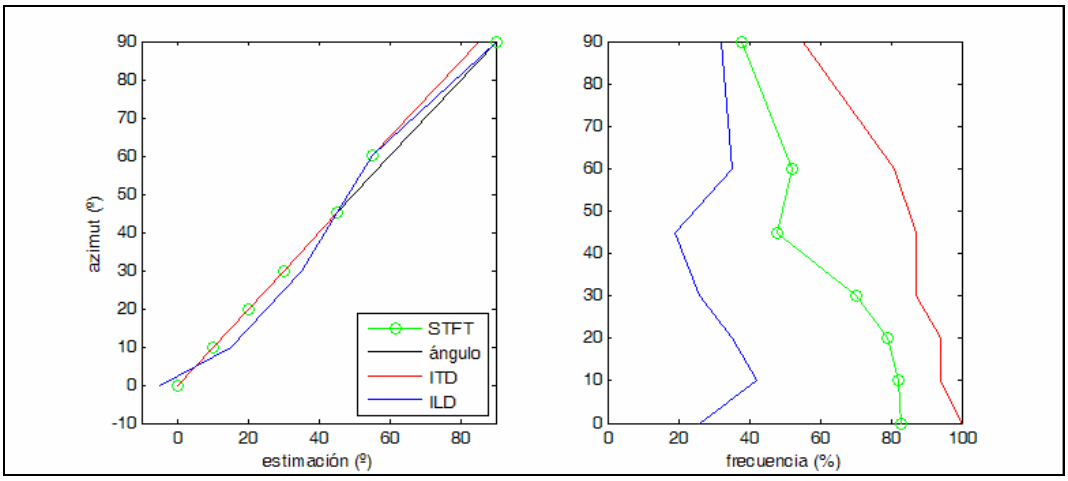


Figura 63. Representación de los estimadores y frecuencia para los dos modelos biaurales para diferentes ángulos de la fuente de una señal biaural real.

	Estimación ITD		Estimación ILD		STFT	
Ángulo fuente (º)	**Moda (º)**	**Frec (%).**	**Moda (º)**	**Frec. (%)**	**Moda (º)**	**Frec. (%)**
0	0	100	-5	26	0	83
10	10	94	15	42	10	82
20	20	94	25	35	20	79
30	30	87	35	26	30	70
45	45	87	45	19	45	48
60	55	81	55	35	55	52
90	85	55	90	32	90	38

Tabla 14. Estimaciones y su frecuencia estadística de estimación cuando se usa STFT o el LES en el caso de radiación directa real.

A diferencia de lo propuesto por [Viste 04] se ha comprobado que un suavizado en las tablas de búsqueda empeora la estimación del ángulo, por lo tanto no se va usar el suavizado descrito en este modelo.

	STFT tabla suavizada		STFT	
Ángulo fuente (º)	**Moda (º)**	**Frec. (%)**	**Moda (º)**	**Frec. (%)**
0	0	80	0	83
10	10	79	10	82
20	20	73	20	79
30	30	66	30	70
45	45	48	45	48
60	55	43	55	52
90	90	32	90	38

Tabla 15. Estimaciones y frecuencias de estimación cuando se usa suavizado y cuando no en el caso de radiación directa real.

Conclusión

En general la precisión decrece según crece el azimut para los dos modelos como sucede con oyentes reales. El modelo psicoacústico biaural utilizado por el LES da mayores frecuencias de estimación, pero en cualquier caso ambos métodos son buenos para el caso de radiación directa.

Este método no se puede aplicar al caso de array, porque cómo ya se ha demostrado el parámetro de ILD no se conserva en el caso de array, luego este método de estimación no funciona al utilizar tanto la información de ILD como la de ITD.

3.8 RESUMEN

En esta capítulo se ha presentado el modelo psicoacústico biaural implementado en el Localizador de Eventos Sonoros, LES. Este modelo hace una estimación de la localización en el plano horizontal del evento sonoro percibido por el sistema auditivo. Este modelo simula el comportamiento de dicho sistema a nivel físico y psíquico, teniendo en cuenta aspectos como la latencia. Los parámetros en los que se basa el LES para determinar el azimut son, según la Teoría Dúplex, la diferencia de tiempo interaural y la diferencia de sonoridad interaural. Para comprobar la eficacia de la herramienta implementada se han hecho ensayos con señales biaurales simuladas y señales biaurales reales medidas con un maniquí acústico, dando unos resultados muy buenos para ambos casos. Además se ha aplicado el LES al caso de ventana acústica virtual, quedando de manifiesto las limitaciones de este sistema a la hora de reproducir los parámetros de localización horizontal.

El siguiente paso sería implementar una herramienta análoga para el plano medio. Como los mecanismos de localización que utiliza el sistema auditivo en este plano son muy diferentes, se va a abordar en el capítulo siguiente.

Capítulo 4: LOCALIZACIÓN EN EL PLANO MEDIO

4.1 TRABAJO DE INVESTIGACIÓN

En este capítulo se expone el trabajo de investigación desarrollado así como los estudios comparativos y resultados obtenidos, en la búsqueda de un buen método de localización subjetiva dentro del plano medio. Este capítulo surge de la necesidad de un modelo psicoacústico para este plano que modele el sistema auditivo análogamente al desarrollado en el plano horizontal.

Como ya se dijo, los métodos de estimación del ángulo se basan en las HRTF, por lo tanto, la primera parte del capítulo consistirá en medir la base de datos de nuestro maniquí acústico HATS.

Una vez implementado el Localizador de Evento Sonoro se va a estudiar la percepción en campo libre del sonido de una fuente sonora, tanto simulada como real para comprobar la eficacia de la estimación. Así se determinará el nivel de precisión de dicha herramienta. Posteriormente se comprueba como el sistema de teleconferencia descrito en la ventana acústica virtual no se puede aplicar para reproducir fuentes en elevación.

Por último se hace una descripción del equipamiento utilizado en la realización de todos los ensayos.

4.2 ESTIMACIÓN EN ELEVACIÓN

La localización de fuentes sonoras se puede analizar desde dos puntos de vista, la localización de las fuentes que emiten la potencia sonora y la localización del evento sonoro que es la posición imaginaria de una fuente desde donde es percibido que se emite la potencia sonora.

Cómo ampliación al trabajo desarrollado hasta este momento, se cree conveniente desarrollar un modelo de estimación del ángulo de elevación. Este debe simular el proceso psicoacústico que desarrolla el sistema auditivo a la hora de determinar la elevación a partir de una señal biaural. Hay que tener en cuenta que la señal biaural es idéntica para los dos oídos cuando se está en el plano medio. Los parámetros de localización ya no son la ITD y la ILD, sino otros llamados monoaurales porque son iguales para los dos oídos.

El sistema auditivo se va entrenando a lo largo de toda la vida para aprender a localizar el sonido. Existe una componente psicológica muy importante en la localización del sonido en el plano medio: p.e. el sonido de un helicóptero se suele localizar arriba aunque la fuente sonora se haya colocado en una elevación inferior. Además, los sonidos de banda

ancha se localizan en diferentes elevaciones dependiendo de la frecuencia de la banda: cuanto más componentes de alta frecuencia tenga el sonido este tiende a desplazarse hacia arriba [Ferguson 05]. Este efecto está relacionado con la forma del módulo de la HRTF para los ángulos superiores, donde aparece un incremento del nivel en las frecuencias entre 7 y 9 kHz.

Por lo tanto, un paso más sería localizar eventos sonoros en el plano medio a partir de la señal monoaural que percibimos. En este campo hay pocos avances porque se parte de dos señales que en el plano medio son iguales (oído derecho e izquierdo) y a partir de ellas hay que sacar los parámetros espectrales de la oreja que son los que determinan la elevación.

En una propuesta de [Keyrouz 06] se propone utilizar el filtro inverso de las hrir de la base de datos para filtrar la señal biaural intentando recuperar la señal emitida por la fuente. Luego, se hace la correlación cruzada entre ambas recuperaciones (derecha e izquierda) y la que dé mayor nivel es la pareja de hrir de la posición de la fuente. Este método que parece sencillo tiene dos problemas, uno es que la inversión de las hrir no es estable y hay que procesar la respuesta en frecuencia para convertir las hrir a filtros de fase mínima que sean invertibles. El segundo problema por el cual no da buenos resultados en el plano medio es que en este caso la señal izquierda y derecha son iguales en todos los ángulos de elevación y por lo tanto su correlación cruzada es máxima en todos los ángulos. La solución que presentó el autor es usar un tercer micrófono para la localización, modelo que se aleja del sistema auditivo real.

En consecuencia, en esta tesis se propone un método para estimar la elevación a partir de una señal monoaural. Los estudios, [Blauert 97] [Iida 07] [Keyrouz 06] [Toledo 08], llevados a cabo para determinar como es el mecanismo de percepción de la elevación muestran que las distorsiones espectrales que causa la oreja, en el sonido que llega al oído en el rango de alta frecuencia (f > 5 kHz), contribuyen a la localización. En general, existen parámetros espectrales de localización para el plano medio entre 4 y 16 kHz. Los parámetros se pueden dividir, cómo ya se ha visto en el capítulo segundo, entre parámetros que determinan las posiciones frontales, las superiores o las posteriores:

- Parámetros frontales: son un nodo de aproximadamente una octava de ancho, cuya frecuencia de corte inferior está entre 4 y 8 kHz (le vamos a llamar N1) y un incremento de energía por encima de los 13 kHz.
- Parámetros superiores: es un pico de ¼ de octava entre 7 y 9 kHz (le llamaremos P1).
- Además, se aprecia claramente que existe un nodo espectral que varía entre 6 y 10 kHz cuando el ángulo de elevación varía entre -45º y 45º.

Estos parámetros espectrales son detectados por los oyentes como parámetros de localización en el plano medio. La forma y el tamaño de la oreja están relacionados con las frecuencias de alguno de estos nodos.

En el plano medio, la capacidad de determinar la elevación depende de los parámetros espectrales de la HRTF. El sonido recibido sufre una coloración debido a la cabeza-torso-orejas que depende del ángulo de elevación. En el trabajo presentado por [Iida 07] se intenta caracterizar las curvas del módulo de las HRTF, en función de los picos y nodos (parámetros espectrales). Además, se evalúa la importancia en la localización de estos picos y nodos con estudios subjetivos. Lo que hace es sintetizar unas HRTF en función de los nodos (llamados N) y los picos (llamados P), para cada ángulo de elevación. Los resultados muestran que las HRTF sintetizadas usando el primer y segundo nodo (N1 y N2) y el primer pico (P1) proporcionan casi la misma precisión de localización subjetiva

que si se usan las HRTF medidas. Analizando las curvas se aprecia que N1 y N2 cambian mucho según varía la elevación de la fuente. En conclusión, N1 y N2 pueden ser calificados como parámetros espectrales, y el sistema de audición puede utilizar P1 como referencia para analizar la información de N1 y N2. La figura siguiente muestra cómo cambia el espectro sistemáticamente en las frecuencias mayores de 5 kHz según varía el ángulo de elevación de la fuente.

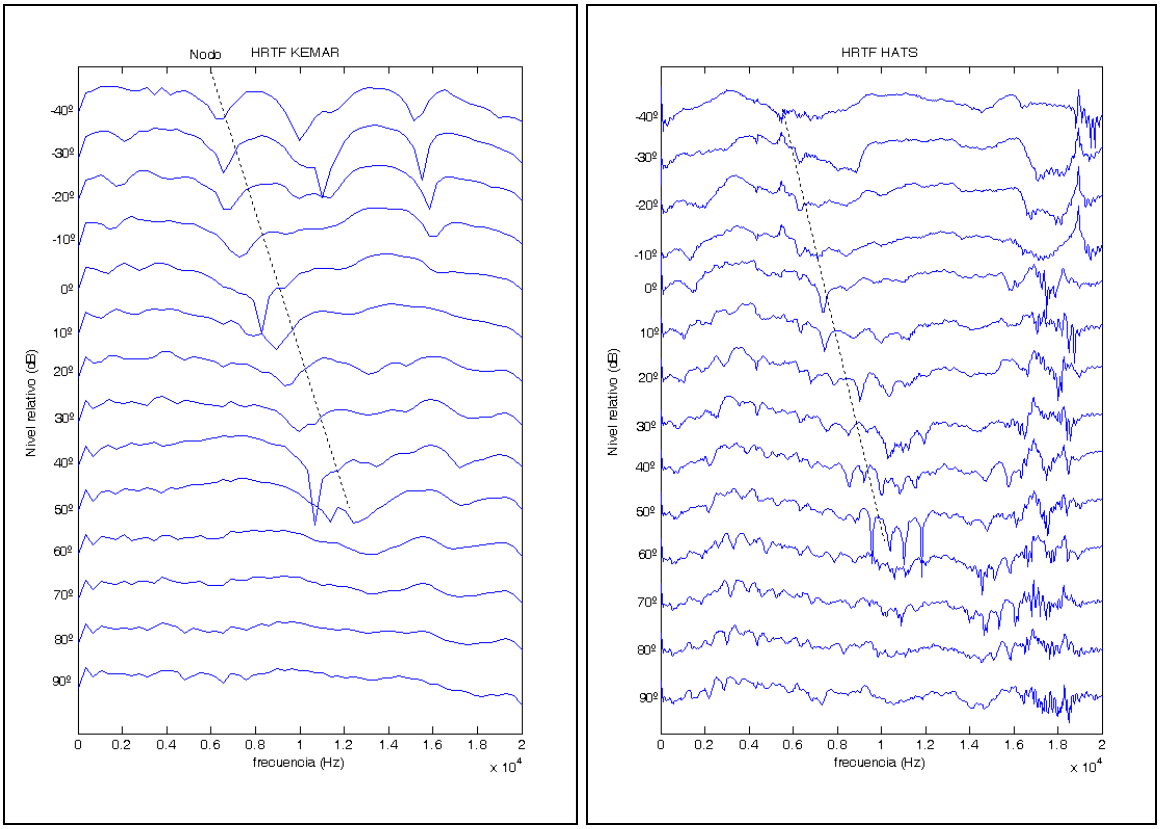


Figura 1. Representación del módulo de la HRTF de la base de datos de KEMAR y de HATS para los ángulos de elevación -40º a 90º.

4.2.1 BASE DE DATOS DE KEMAR EN ELEVACIÓN

En el plano medio la hrir de los dos oídos son iguales, y por lo tanto, la base de datos que se ha utilizado consiste en la respuesta en tiempo de un oído para 14 ángulos, desde -40º a 90º en pasos de 10º [Gardner 00]. Las hrir son de 242 muestras, normalizadas a 1 m de distancia. La base de datos proporciona la hrir de esos ángulos de elevación, luego es necesario calcular el módulo de la HRTF para caracterizar las curvas espectralmente. La frecuencia de muestreo es de 44100 Hz y la duración de las HRTF una vez realizada la transformada de Fourier (fft) es de 65 muestras.

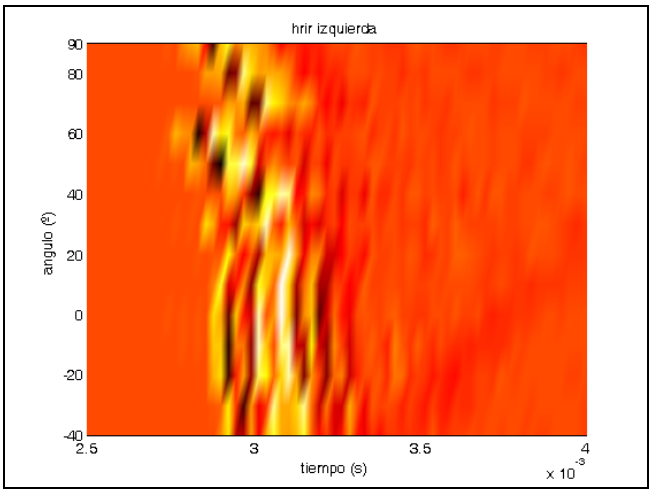


Figura 2. Hrir del oído derecho a 1 m del maniquí KEMAR en el plano medio.

4.2.2 BASE DE DATOS DE HATS EN ELEVACIÓN

Las HRTF en el plano medio fueron medidas en la cámara anecoica de la EUITT en pasos de 10º desde -40º hasta 90º. La distancia a la que se colocó la fuente fue de 1.5 m. El proceso de medida ya ha sido explicado anteriormente para el caso de la medida en el plano horizontal. Se muestran unas fotos del montaje de medida, el cual no resulta nada sencillo, ya que hay que medir el ángulo del altavoz hacia el centro de la cabeza, la distancia y colocar su radiación en el plano medio del maniquí. Esto se realizó con un puntero láser y dos medidores de ángulos en cuadratura.

Figura 3. Grabación de la señal biaural real para diferentes ángulos en el plano medio dentro de una cámara anecoica. Detalle del medidor de ángulos en cuadratura.

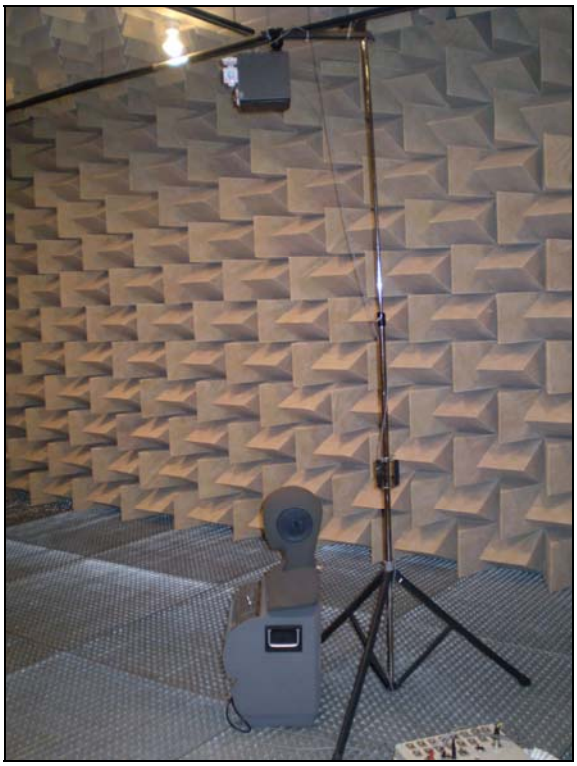


Figura 4. Grabación de la señal biaural real para una elevación de 90º.

Las medidas se realizaron con el sistema PULSE de Brüel & Kjaer, cuya frecuencia de muestreo es de 65536 Hz y el número de muestras de cada HRTF es de 1024.

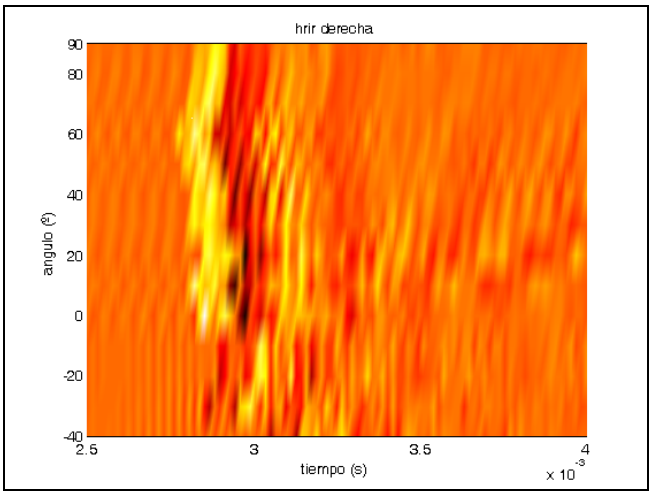


Figura 5. Hrir del oído derecho a 1 m del maniquí HATS en el plano medio.

Después de medir la HRTF (Figura 1) es necesario procesar estas respuestas para calcular la hrir y crear una base de datos para 1 m de las hrir. Este procesamiento es el mismo que se utilizó en el caso del plano horizontal, y que ha sido explicado en el capítulo tercero.

4.2.3 MÉTODO DE ESTIMACIÓN DE LA ELEVACIÓN

El trabajo expuesto a continuación tiene dos propósitos. Uno es desarrollar un método que estime la elevación de una fuente sonora a partir de la señal monoaural en el plano medio. Y el otro es analizar cómo son reproducidos los parámetros espectrales cuando se usa un

sistema de reproducción multicanal, como es el caso de la ventana acústica virtual, en el plano medio.

Se han simulado los mismos escenarios acústicos que en el plano horizontal, de forma análoga a lo descrito anteriormente. Por lo tanto, se parte del análisis de una señal monoaural captada por un maniquí acústico (simulada o medida). De las dos señales de la señal biaural proporcionadas por el maniquí, sólo se procesa una de ellas ya que la señal derecha e izquierda son iguales.

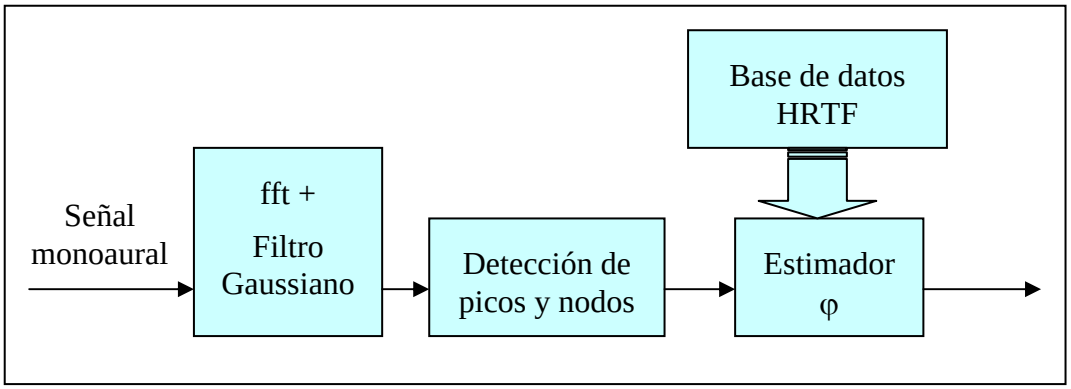


Figura 6. Método de localización en el plano medio.

El conocimiento que tiene el sistema auditivo, que ha sido aprendido a lo largo de la vida, se simula con la base de datos de los parámetros espectrales. El ser humano compara la percepción que recibe con la almacenada en su experiencia para determinar el ángulo de llegada de los sonidos.

El proceso de localización es el siguiente. Se realiza la fft de la señal monoaural recibida. Sobre el módulo de la respuesta en frecuencia se detectan los nodos y picos (Figura 6). Se ha procesado con anterioridad el módulo de las HRTF, de cada ángulo de elevación de la base de datos, para determinar los nodos y los picos y caracterizarlos por varios parámetros. El proceso de obtención de los parámetros de los nodos y los picos es el mismo para la señal monoaural que para la base de datos. Los parámetros espectrales que se obtienen de la señal que llega a los oídos son comparados con los parámetros espectrales de la base de datos. El ángulo de elevación de la base de datos que tenga los parámetros más parecidos a los de la señal monoaural se establecerá como el ángulo estimado de elevación.

4.2.3.1 EXTRACCIÓN DE LOS PICOS Y NODOS ESPECTRALES

Como ya se ha mencionado anteriormente, los picos y nodos por encima de 5 kHz son los que contribuyen a la localización en elevación. Por lo tanto, se ha limitado la búsqueda de picos y nodos a frecuencias mayores de 4 kHz. A partir de esta frecuencia se van a etiquetar los picos y nodos que aparezcan en orden creciente (p.e. P1, N1, P2, N2,…). Cada pico o nodo se caracteriza con su frecuencia central, su nivel relativo y su ancho de banda a -3 dB. Las curvas son normalizadas de forma que el valor máximo sea 0 dB.

Para conseguir que sólo los picos y nodos más profundos, y por lo tanto característicos, no se confundan con las pequeñas fluctuaciones del módulo del espectro, se realiza un filtrado Gaussiano al módulo de la fft de la señal.

$$H_W(f) = \sum_{n=-n_1}^{n_1} H(f+n)W(n)$$
$$W(n) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-n^2}{2\sigma^2}} \qquad \text{Ec. 1}$$

Donde W(f) es el filtro Gaussiano, $H_W(k)$ es la HRTF filtrada. f y n son el índice de frecuencia. H(f) es el módulo de la fft de la señal monoaural o de la HRTF, en su caso. El número de muestras de la HRTF es de 65 para la base de datos de KEMAR a 44100 Hz de frecuencia de muestreo; y de 1024 para la base de datos de HATS cuya frecuencia de muestreo es de 65536 Hz. Los parámetros n y σ se han ajustado para obtener unas curvas suavizadas que permitieran una correcta caracterización de los nodos y picos principales. Se han elegido unos valores para n = 4 y σ = 1.3 en el caso de KEMAR; y de n = 43 y σ = 5 en el caso de HATS. (4 y 43 muestras corresponden respectivamente a un margen de frecuencias de 1378 Hz).

Una vez filtrada la señal, los picos y nodos se han establecido como los mínimos y máximos del módulo del espectro de las HRTF. Se ha normalizado el nivel máximo a 0 dB para independizar el nivel de los picos y nodos del volumen del sonido. Así, sólo se caracteriza el nivel de cada pico o nodo con respecto al nivel máximo de la curva (0 dB). Los picos y nodos extraídos se muestran como círculos en la figura siguiente.

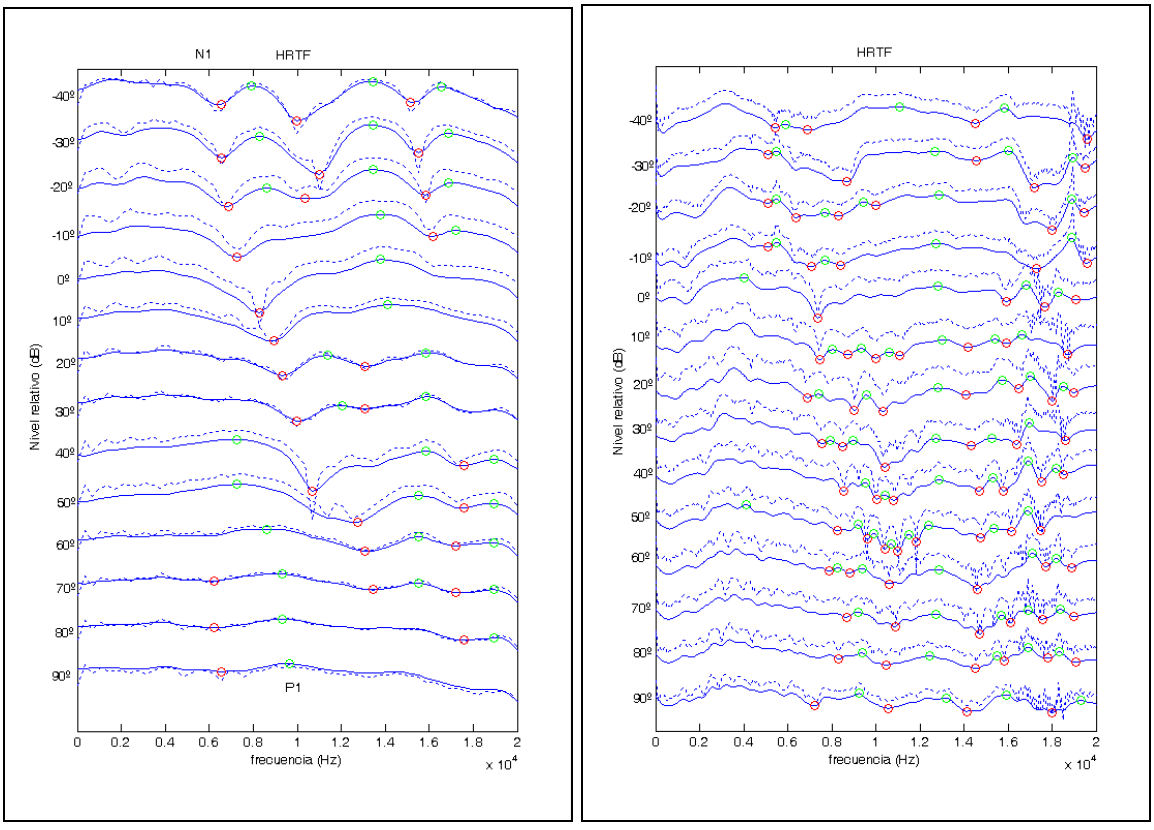


Figura 7. Representación del módulo de la HRTF de la base de datos de KEMAR (derecha) y de HATS (izquierda) para los ángulos de elevación -40º a 90º. Línea continua: señal filtrada, línea discontinua: señal no filtrada. Nodos: círculos rojos, picos: círculos verdes.

4.2.4 LOCALIZACIÓN EN ENTORNOS SIMULADOS

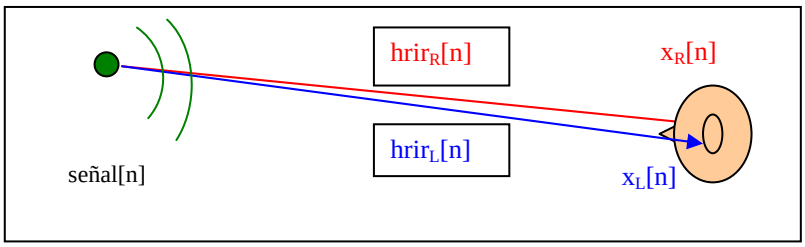


Figura 8. Representación del escenario acústico simulado para la obtención de la señal monoaural.

Se simula una fuente sonora colocada en diferentes posiciones en el plano medio, a 2.5 m de distancia desde -40º hasta 90º en pasos de 10º. La señal monoaural se calcula filtrando el sonido de la fuente con la hrir apropiada, análogamente a como se explicó en el plano horizontal. Hay que tener en cuenta la atenuación de la señal debida a la distancia.

Las dos señales utilizadas son ruido blanco que tiene un espectro plano en frecuencia y voz que tiene un espectro tipo ruido rosa (descendiente con la frecuencia).

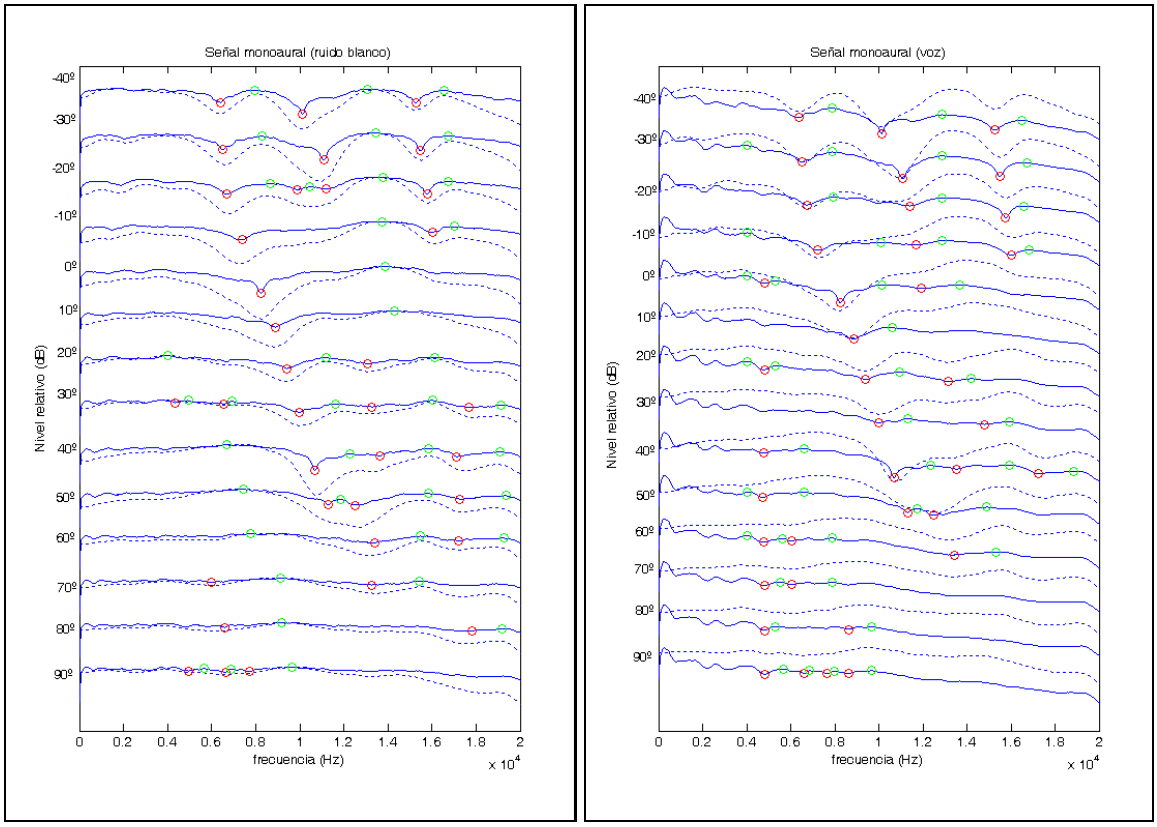


Figura 9. Representación del módulo de la señal monoaural con una fuente sonora de ruido blanco y voz para los ángulos de elevación de -40º a 90º (línea continua: señal monoaural, línea discontinua: HRTF de la base de datos con la que comparar). Círculos rojos: nodos, círculos verdes: picos. (Base de datos de KEMAR).

Como se aprecia en la Figura 9, algunos de los nodos y picos que aparecen en la señal monoaural coinciden con los de la base de datos.

4.2.4.1 ESTIMADORES

Basado en la distancia

Para establecer cual curva de las HRTF de la base de datos se parece más a la del espectro de la señal monoaural, se mide la distancia entre los picos y nodos de las dos curvas. Este estimador está basado en minimizar el error cuadrático medio.

$$ECM = \sqrt{\frac{\sum_f \left(X(f) - X_{HRTF}(f)\right)^2}{\sum_f X(f)^2}} \qquad \text{Ec. 2}$$

Donde X(f) es el valor del espectro de la señal monoaural para la frecuencia f, X_{HRTF} es el valor del espectro, para esa frecuencia, de la HRTF de la base de datos. El ángulo estimado será aquel cuya curva de HRTF proporcione un ECM menor.

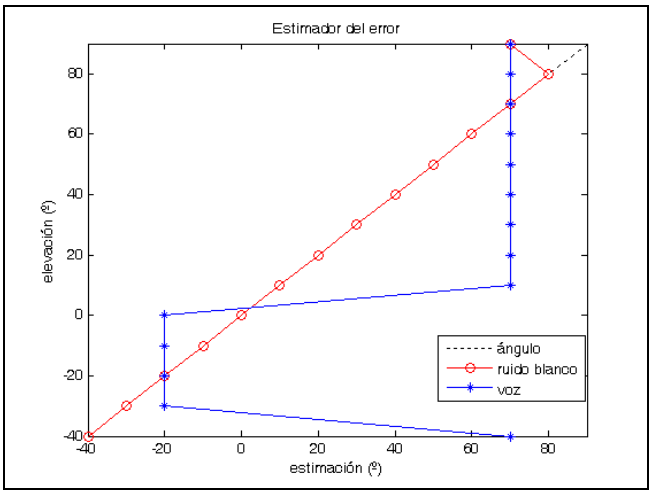


Figura 10. Estimación del ángulo de elevación para una fuente colocada a diferentes elevaciones, para los dos tipos de señales: ruido blanco (o) y voz (*).

Los resultados obtenidos para ruido blanco son muy buenos ya que su espectro es plano y al filtrar con la hrir, los nodos y picos que aparecen son debidos exclusivamente a la propia hrir. Sin embargo, los resultados mostrados para voz no son buenos porque su espectro tiene una caída en frecuencia que hace que la distancia entre la curva de la señal monoaural y la curva de la HRTF sea cada vez mayor, y por lo tanto el error cometido sea mayor.

Basado en la caracterización espectral de los nodos y picos

Uno de los mayores problemas es establecer cuáles de los nodos y picos de la señal monoaural pertenecen a las HRTF y cuáles pertenecen al propio sonido. En la Figura 9 se muestra la caracterización de picos y nodos de la señal monoaural, para dos tipos de sonido: ruido blanco y señal de voz. Algo que visualmente es muy sencillo de analizar y decidir, si dos curvas son parecidas o no, en cuanto a sus nodos y picos, algorítmicamente resulta muy complicado.

Como se puede apreciar la señal monoaural reproduce todos o muchos de los nodos y picos de las HRTF. El conjunto de picos y nodos de la señal monoaural se compara con el conjunto de cada una de las curvas de la base de datos.

El algoritmo que evalúa el grado de parecido entre las dos curvas lo hace en base a lo siguiente. Tanto las HRTF de las bases de datos como la señal monoaural tienen una serie de picos P y nodos N, caracterizados por su índice f que representa la frecuencia a la que aparecen, en función de la frecuencia de muestreo utilizada:

$$\text{HRTF}(\varphi) = \{f_{Ni}, f_{Pi}\} \qquad \varphi = -40^o, -30^o, ..., 90^o$$
$$\text{Señal} = \{f_{Nj}, f_{Pj}\}$$

Sólo se van a analizar los picos y nodos que aparecen entre 4000 Hz y 15000 Hz, que es donde se encuentran los picos y nodos que contribuyen a la localización. Cada curva HRTF(φ) obtendrá una valoración V(φ), en función de la distancia de sus picos y nodos a los de la señal monoaural con la que queremos comparar.

$$V(\varphi) = \sum_{Ni} w_{Ni} + \sum_{Pi} w_{Pi} \qquad \begin{array}{l} w_i = 1 \text{ si } f_i = f_j \\ w_i = 0.5 \text{ si } |f_i - f_j| = 1 \quad \forall f_j \\ w_i = 0.25 \text{ si } |f_i - f_j| = 2 \\ w_i = 0, \text{ resto} \end{array}$$

En esta expresión se pondera el valor w inversamente a la distancia en frecuencia, entre los picos y nodos. Esto significa que si dos picos están en la misma frecuencia el valor es máximo y según se van alejando su valor va disminuyendo.

El ángulo de elevación estimado será aquel cuya HRTF obtenga el mayor valor V(φ).

$$\varphi_{\text{estimado}} = \text{mean}(\varphi|_{\max V(\varphi)})$$

Si hay varias ángulos de elevación, φ, para los cuales el valor V(φ) es máximo, se hace la media de los valores φ estimados.

Este estimador se aplica al módulo de la fft de cada ventana de la señal monoaural, que ha sido "enventanada" con una ventana hanning con solapamiento. Primero se calcula el promedio del módulo de la respuesta en frecuencia de un grupo de ventanas (el número de ventanas promediado se puede elegir); y después se caracterizan sus nodos y picos. Una vez estimado un ángulo de elevación, para cada grupo promediado, se define la moda de dichos grupos como el ángulo estimado de dicha señal.

Los resultados de la aplicación del procesado descrito se presentan en la siguiente figura para el caso de señal monoaural simulada con la base de datos de KEMAR.

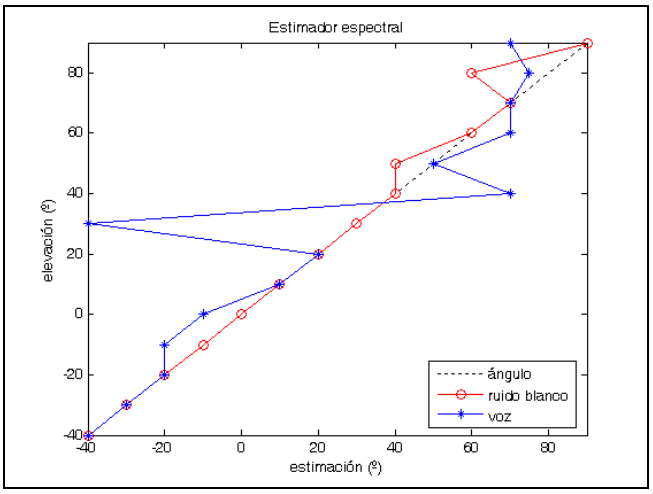


Figura 11. Estimación del ángulo de elevación para una fuente colocada a diferentes elevaciones: ruido rosa (o) y voz (*). (Señal monoaural simulada).

Este método de localización simula el comportamiento del sistema auditivo, ya que busca los picos y nodos de las HRTF que caracterizan la elevación. Los resultados son bastantes buenos. Los resultados son mejores para ruido blanco ya que el espectro plano del ruido no afecta a la posición (en amplitud) de los picos y nodos. Para la señal de voz los resultados también son buenos porque al promediar ventanas de tiempo de la señal monoaural, el espectro resultante de la señal de voz es plano con caída (similar al ruido rosa), y no afecta demasiado a la posición en frecuencia de los nodos y picos de las HRTF. En este tipo de señal aparece como parámetro muy importante, a la hora de establecer la estimación, el tamaño de los grupos de promedio de las ventanas temporales.

El comportamiento de este estimador tiene su equivalencia con el sistema auditivo, ya que como la señal monoaural de voz tienen poca energía a frecuencias altas, la localización de este tipo de sonidos (sin energía a frecuencias altas) tiende a producirse a ángulos de elevación menores [Ferguson 05]. Se localiza el evento sonoro alrededor de 60º para posiciones superiores de la fuente.

Conclusión

Se ha propuesto un método de localización monoaural para el plano medio basado en los parámetros espectrales. Se ha analizado la capacidad de estimar para diferentes tipos de señal y diferentes tipos de estimadores. El estimador mejor para los dos tipos de señales es aquel que analiza los nodos y los picos, independientemente del nivel relativo, y los identifica como pertenecientes a las HRTF para un cierto ángulo.

4.2.4.2 LOCALIZACIÓN CON ARRAY

En este experimento se ha simulado una ventana acústica virtual pero en lugar de en el plano horizontal en el plano vertical con 10 transductores por array y distancia entre ellos de 0.175 m. Se busca analizar la capacidad de la ventana acústica vertical de reproducir los parámetros espectrales de un evento sonoro, necesarios para su localización en el plano medio. Los resultados se muestran en la figura siguiente:

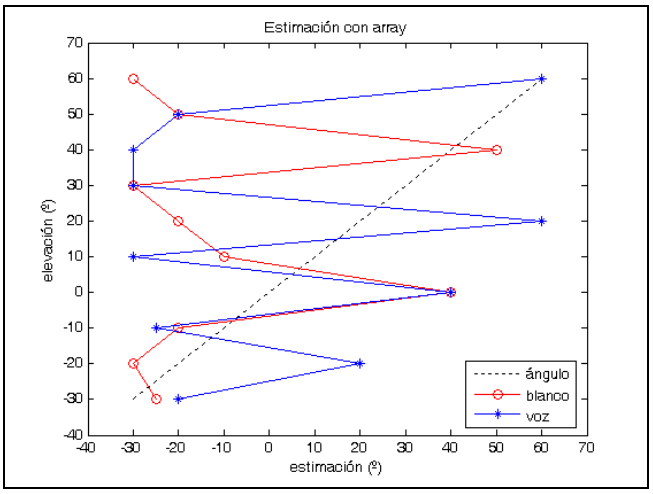


Figura 12. Estimación del ángulo de elevación para una evento sonoro de una ventana acústica virtual colocado a diferentes elevaciones: ruido rosa (o) y voz (*).

La figura anterior muestra el mal comportamiento del array a la hora de reproducir los parámetros espectrales. Este comportamiento era esperado debido a que los parámetros espectrales se encuentran a frecuencias por encima de los 4 - 5 kHz y el campo sonoro reproducido por el array a esas frecuencias es muy malo ya que está por encima de la frecuencia de aliasing (1 – 1.5 kHz).

4.2.5 LOCALIZACIÓN EN ENTORNOS REALES

Para estudiar la precisión del modelo de localización propuesto para el plano medio, se van a realizar grabaciones de señales biaurales reales con el maniquí acústico HATS, en una cámara anecoica colocando la fuente a diferentes ángulos de elevación. Las señales utilizadas son ruido blanco (tipo de sonido idóneo debido a que tiene el módulo plano) y la señal de voz (tipo de sonido más complicado, para el cual será necesario utilizar ventanas temporales y promediar la potencia de la señal, para más tarde calcular el nivel de cada grupo de ventanas).

En la figura siguiente representamos el módulo de las señales monoaurales en comparación con las curvas HRTF. Se han marcado los nodos y picos que el algoritmo detecta para esas señales monoaurales.

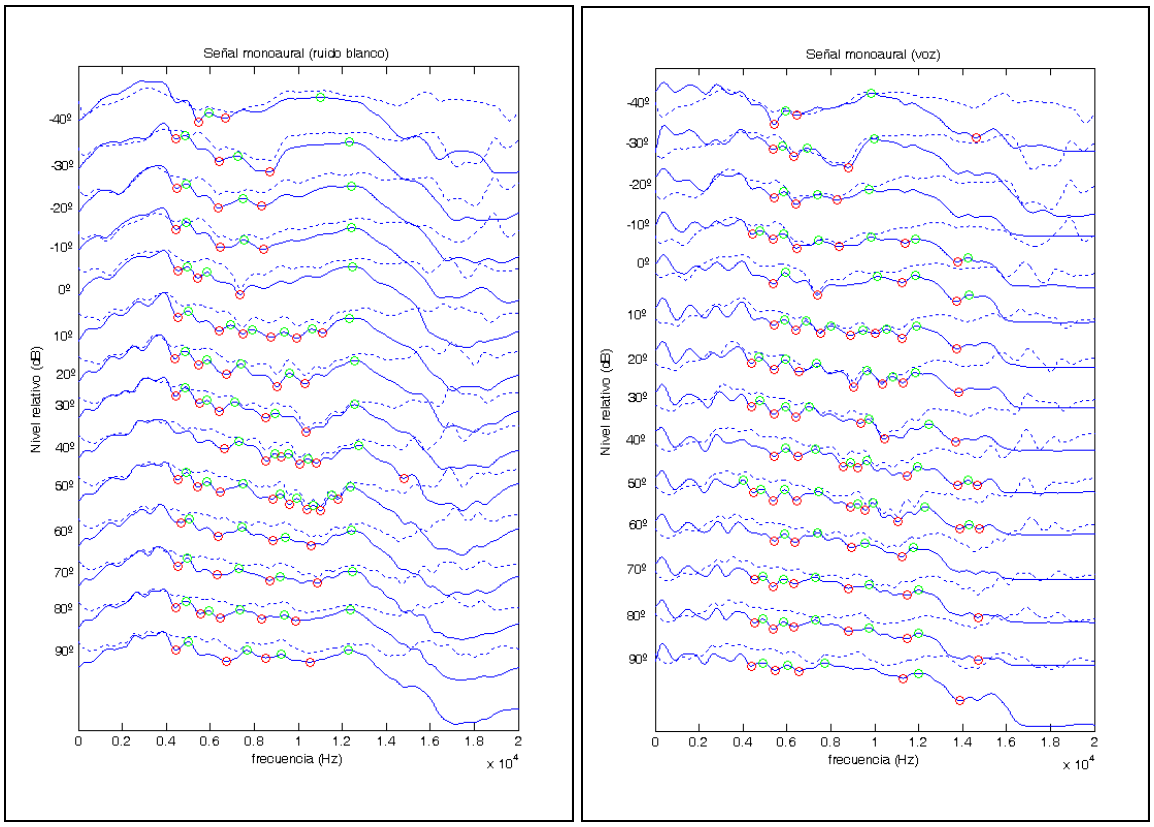


Figura 13. Representación del módulo de la señal monaural con una fuente sonora de ruido blanco y voz, para los ángulos de elevación -40º a 90º (línea continua: señal monoaural, línea discontinua: HRTF con la que comparar). Círculos rojos: nodos, círculos verdes: picos.

Las señales monoaurales reales muestran una clara caída a partir de 1600 kHz, por lo tanto, al algoritmo de estimación se le ha indicado que sólo valore los picos y nodos que hay entre 4 kHz y 16 kHz.

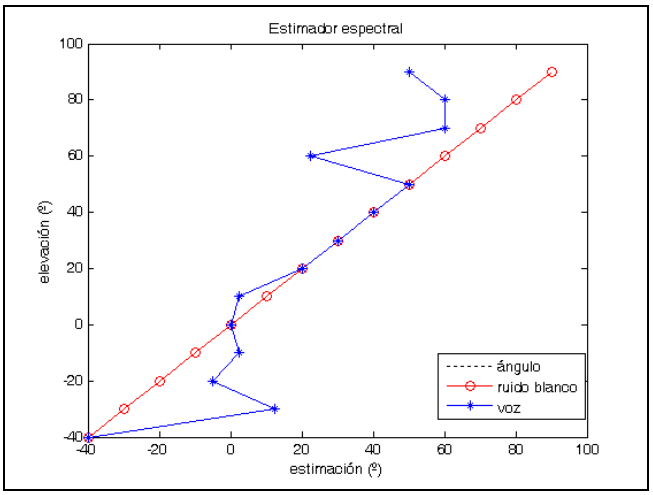


Figura 14. Estimación del ángulo de elevación para una fuente colocada a diferentes elevaciones: ruido rosa (o) y voz (*). (Señal monoaural real).

Cómo puede comprobarse, la estimación para el caso de ruido blanco es muy buena. Para el caso de voz, la estimación depende mucho del nº de ventanas que se promedien para calcular el módulo de la señal monoaural, esto es debido a que la señal de voz es más irregular, y dependerá mucho del tramo de fichero analizado.

Conclusión

El método propuesto para la localización monoaural en el plano medio basado en los parámetros espectrales funciona para cualquiera de las bases de datos utilizadas (KEMAR y HATS). Además, no sólo funciona para señales monoaurales simuladas, donde sólo actúan las HRTF correspondientes y la señal de la fuente; sino que también funciona para señales monoaurales reales donde además afecta la respuesta del dispositivo de reproducción-grabación y la de la sala, introduciendo variaciones en la respuesta en frecuencia.

4.3 EQUIPAMIENTO

Para la realización de la Tesis Doctoral se ha contado con los siguientes medios:

- Medios bibliográficos, cuya recopilación se detallan en el capítulo de bibliografía.
- Medios informáticos de simulación. Toda la simulación informática previa a la implementación del sistema se realiza utilizando el programa MATLAB. La herramienta de estimación del azimut y elevación se desarrolla en MATLAB.
- Equipamiento electroacústico y de procesado de medida e infraestructura disponible en los Laboratorios de Sonido del DIAC en la E.U.I.T. Telecomunicación de la U.P.M.:
    - Maniquí acústico HATS (Head And Torso Simulator): sirve para grabar las señales biaurales. El maniquí simula la cabeza y el torso de una persona, sus orejas y el conducto auditivo donde se encuentran dos pequeños micrófonos de precisión. El fabricante es Head Acoustics, y el modelo es HMS II.3 que sigue los requerimientos de la [Recomendación ITU P.58] “Head and torso simulator for telephonometry”. El HATS HMS II.3 contiene dos simuladores de oído HIS L e HIS R para el lado izquierdo y derecho respectivamente. El oído incorpora un micrófono propiedad de Head Acoustics comparable al modelo GRAS 40AG. Además incorpora dos simuladores de oreja o pabellón auditivo con las mismas propiedades acústicas y rigidez de una oreja humana, HEL IV y HER IV, sigue los requerimientos de la [Recomendación ITU P.57] “Artificial ears”.
    - Micrófono 4188 de B&K: micrófono de precisión de condensador de ½ pulgada.
    - Preamplificador 2669 de B&K: preamplificador de ½ pulgada.
    - Sistema de grabación y reproducción multicanal PARIS de ENSONIQ: sistema de reproducción de 10 canales.
    - Sistema de reproducción multicanal MOTU 896HD: sistema de reproducción de 10 canales, controlado por Matlab.
    - 5 Amplificadores de potencia de doble canal Yamaha A100a.
    - Mesa giratoria B&K modelo 3922 que permite girar la cabeza en el plano horizontal.

Figura 15. Derecha: sistema de reproducción multicanal MOTU con los respectivos amplificadores de potencia. Izquierda: sistema de medida PULSE de Brüel and Kjaer.

- Altavoz de Yamaha MSP5: altavoz tipo bass-reflex, de dos vías, auto amplificado.
- Array de altavoces: las cajas de los altavoces fueron hechas a medida con altavoces de 6 pulgadas.
- Cámara anecoica: sala de 5x7x3 m, rodeada por cuyas de lana mineral recubiertas que absorben la energía y evitan las reflexiones acústicas.
- Analizador PULSE de B&K: sistema de medida modelo 3560C. Consta de un panel de hardware para la adquisición de datos denominado Front-End que se conecta a un PC mediante interfaz LAN tipo 7533 y del software Pulse LabShop para el tratamiento de datos.

A continuación se muestran algunas fotos del array de altavoces y del maniquí acústico dentro de la cámara anecoica.

Figura 16. Array de altavoces dentro de la cámara anecoica de la EUITT.

Capítulo 5: CONCLUSIONES

5.1 CONCLUSIONES

El objetivo inicial de esta Tesis era la implementación de una herramienta que realizara medidas objetivas de localización de un evento sonoro (cualidad subjetiva) para sistemas de teleconferencia. Este objetivo, que estaba asociado a la investigación desarrollada en el Grupo GAPS, fue cumplidamente superado y se fue ampliando para dar alcance a otros sistemas de reproducción multicanal. Puesto que es de esperar que los resultados de la Tesis también sean de aplicación para otros sistemas multicanal, actuales o que puedan ser desarrollados en el futuro, podemos concluir que los objetivos planteados en el inicio han sido ampliamente superados.

Los sistemas de teleconferencia estudiados se basan en el concepto de ventana acústica virtual, donde los sistemas de captación y reproducción están constituidos por un array de micrófonos y un array de altavoces, respectivamente. Este escenario acústico determinó que, en un principio, centráramos nuestro estudio en la localización en el plano horizontal, pero con posterioridad el objetivo se hizo más ambicioso, llegando a englobar también la localización en el plano medio.

Tradicionalmente los sistemas de teleconferencia con más de una fuente reproductora, se han mantenido limitados al plano horizontal, ya que incluso los sistemas de reproducción multicanal actuales se suelen mantener anclados en este plano. Aún así, se ha creído necesario ampliar el desarrollo de la herramienta de localización LES al plano medio, de forma que permita evaluar futuros sistemas, donde se podrían colocar hasta tres o cuatro fuentes en elevación.

Como fruto de los estudios realizados, se ha demostrado que la utilización de arrays de altavoces en el plano medio no es útil dada la incapacidad de estos sistemas de reproducir los parámetros de localización en elevación, debido a la frecuencia de aliasing de los sistemas WFS. Sin embargo, no se descarta la posibilidad de que la colocación de tres o cuatro fuentes pueda permitir trasladar el evento sonoro fuera del plano horizontal, por la ley de primer frente de onda o efecto Haas.

Los nuevos sistemas de reproducción de audio espacial o 3D que se puedan desarrollar en el futuro, necesitarán herramientas que evalúen parámetros de calidad objetivos pero además, es sobradamente conocida la necesidad de obtener también parámetros de calidad subjetivos. Una de estas cualidades subjetivas a evaluar es la capacidad de reproducir los parámetros de localización que ayudan al sistema auditivo a situar los eventos sonoros. En este marco, disponer de una herramienta como el LES, que permite analizar la dirección del evento sonoro reproducido por un sistema multicanal, se estima de una gran utilidad gracias a la ayuda que puede prestar en el desarrollo de estos sistemas comprobando a

priori si los parámetros de localización son correctamente reproducidos con gran inmediatez. De no contar con este tipo de herramientas sería necesario realizar costosos y lentos ensayos subjetivos con sujetos, que no siempre arrojan resultados fiables por la gran cantidad de variables que pueden polarizar los resultados finales.

No menos importante resulta la capacidad que presenta la herramienta desarrollada para ayudar en la búsqueda de la mejor configuración posible del sistema multicanal. Es muy interesante poder variar el número y disposición de los elementos electroacústicos de un sistema de reproducción multicanal y evaluar, sin necesidad de hacer el montaje en una sala real, la capacidad de dicha configuración de reproducir los eventos sonoros en la localización buscada.

La calidad del audio reproducido se puede medir con diferentes métodos, en función de diversos parámetros que describen la señal de audio (todavía es señal eléctrica) o el sonido (una vez convertida en señal acústica). Sin embargo, el receptor final del sistema de telecomunicación es el oyente, y por lo tanto será muy importante analizar que es lo que él percibe (calidad subjetiva). El poder estimar la localización del evento sonoro que percibiría un oyente cuando se usa un sistema de reproducción multicanal, así como, el poder analizar y estudiar los diversos parámetros que utiliza al sistema auditivo en el proceso de localización, se convierte en una herramienta muy útil a la hora de diseñar, mejorar e implementar dichos sistemas multicanal. Primero se harían estudios simulados de estimación y solamente en una etapa final de los diseños se necesitaría realizar estudios subjetivos.

Una vez concluida su utilidad en los sistemas tradicionales de grabación y reproducción multicanal, es importante resaltar que cada vez son más importantes los asuntos relacionados con audio tridimensional, técnicas biaurales y “auralización”. Estos sistemas de audio sintético relacionados con la acústica virtual, pueden sintetizar fuentes virtuales de sonido por medio de altavoces reales o de auriculares, por lo que poder contar para su desarrollo con herramientas como la implementada en esta Tesis, redunda en una simplificación del proceso de desarrollo al permitir aplicar las mismas medidas en diferentes sistemas o configuraciones.

Todo esto nos lleva a situar a esta Tesis en un marco de conocimientos bastante diverso, donde se conjugan conocimientos de acústica y de psicoacústica con tecnologías de procesado de señal, así como metodologías de simulación de entornos acústicos, reconocimientos de patrones y modelado del sistema auditivo humano. Como fruto del enfoque eminentemente práctico que se ha mantenido a lo largo del desarrollo de la Tesis, se ha buscado la aplicación práctica de los temas anteriormente mencionados. Algunos de los cuales están todavía en etapas iniciales de estudio e investigación, sobre los que realmente existen pocos trabajos publicados como por ejemplo la localización en elevación.

En puntos posteriores, se desarrolla una síntesis de las conclusiones más relevantes que se han ido obteniendo a lo largo de la Tesis. Dichas conclusiones han sido expuestas con mayor profundidad en los capítulos anteriores como conclusión de cada ensayo. Se empezará por la adecuación de la herramienta en la determinación de la localización en el plano horizontal, y su aplicación a los sistemas de teleconferencia y se terminará por su aplicación en la localización en el plano medio. Además se expondrán las conclusiones relacionadas con la simulación de señales biaurales y la medida de funciones de transferencias de la cabeza, HRTF.

5.1.1 LOCALIZACIÓN EN EL PLANO HORIZONTAL

La Tesis tiene como punto de partida el trabajo desarrollado por Blauert, que recoge y analiza los estudios de las pruebas subjetivas de localización realizadas durante décadas por numerosos investigadores. Este conocimiento ha llevado a establecer relaciones bien conocidas y admitidas como la Teoría Dúplex, la cuál establece la relación entre la localización subjetiva de una fuente sonora y los parámetros de localización biaural (ITD e ILD). Las pruebas subjetivas recogidas en los estudios mencionados analizan cómo se comporta la percepción subjetiva en función de la variación de estos parámetros. Basándonos en la evaluación recogida en más de un centenar de estudios, admitiremos que estas relaciones existen y son correctas lo que nos servirá como punto de partida para el desarrollo de una herramienta que simula el proceso realizado por el sistema auditivo para determinar la localización del evento sonoro. Todo el sistema toma como punto de partida la señal biaural de un maniquí acústico que simulará la función de oyente.

A consecuencia de las numerosas pruebas realizadas en la Tesis se han podido extraer conclusiones que se corroboran con los abundantes estudios de pruebas subjetivas referidas en la bibliografía. Entre ellas podemos mencionar que la estimación basada en ITD es más fiable que la basada en la ILD, o también que la estimación obtiene mejores resultados para señales cuyo espectro sea ancho y estable.

Existen diversos modelos que simulan el comportamiento del sistema auditivo humano desde el punto de vista de localización espacial. En la presente Tesis se ha hecho un profundo estudio de los mas destacables, de forma que el análisis de sus diversas configuraciones nos ha permitido sintetizar un modelo biaural basado en el análisis en bandas “perceptuales” de la señal biaural y en la teoría Dúplex. La base del modelo desarrollado es la información de localización que está implícita tanto en la diferencia de nivel interaural como en la diferencia de tiempo interaural.

La herramienta desarrollada en la Tesis, a la que hemos denominado LES (Localizador de Eventos Sonoros), implementa el modelo biaural referido anteriormente. Al LES se le presenta como entrada la señal biaural, proporcionando como salida la estimación del ángulo en el que percibiría un oyente el evento sonoro, si tuviera dicha señal biaural en sus oídos. La precisión de la herramienta obtenida ha podido ser verificada tanto para señales reales (medidas con un maniquí acústico) como para señales simuladas.

Para profundizar en la validación de la herramienta LES, también se ha comprobado su robustez frente a diferentes bases de datos de HRTF, puesto que el modelo biaural de localización se fundamenta en este tipo de información. Las bases de datos empleadas han sido la creada para esta Tesis sobre un maniquí denominado HATS, y la ampliamente conocida del maniquí KEMAR.

A la vista de los resultados obtenidos, podemos afirmar que el LES es capaz de localizar eventos sonoros dentro del plano horizontal en el margen de -90º a 90º de azimut con una elevada precisión, ya que conseguimos operar con una resolución de 2.5º. La precisión es mayor cuando la señal de test empleada es ruido blanco en lugar de señal de voz (al igual que ocurre en los estudios subjetivos). Esta precisión está en el rango de 5º-10º, tolerancia aceptable a la hora de localizar los eventos sonoros.

Una de las ventajas de la herramienta desarrollada es que la señal biaural que se introduce en el LES puede ser medida bien mediante un maniquí acústico en un entorno real, o bien ser simulada mediante la base de datos de HRTF de ese mismo maniquí. Como la señal biaural puede tener una duración elevada, la estimación del ángulo se realiza para ventanas temporales pequeñas, obteniendo como resultado la moda estadística de todas las

estimaciones. Para poder analizar la fiabilidad de la estimación obtenida, el LES permite obtener la frecuencia estadística con la que se repite dicha estimación.

Otra de las conclusiones obtenidas es que la configuración de ventana acústica virtual resulta de gran interés para sistemas de teleconferencia, ya que es capaz de reproducir la localización de las fuentes originales. Este hecho ayudará a localizar el evento sonoro, por ejemplo un interlocutor, en el mismo sitio que lo reproduciría la imagen proyectada en la pantalla. Para llegar a esta conclusión nos hemos basado tanto en los ensayos simulados como en los ensayos reales que se han realizado para un sistema de teleconferencia que emplea la configuración de ventana acústica virtual, mediante arrays de micrófonos y de altavoces. Aunque, debido a la frecuencia de aliasing característica de los arrays, la reproducción del campo sonoro con este tipo de sistemas no es muy buena para frecuencias altas-medias, si está indicada para frecuencias bajas, donde se encuentran ubicadas las principales componentes de la voz. Por lo tanto, esta limitación en frecuencia no afectaría de forma notable a los sistemas de teleconferencia, cuyo principal uso es la transmisión de señales de voz.

Gracias a las pruebas realizadas con la herramienta desarrollada, se ha podido determinar que la representación del sonido mejora mediante el uso de arrays donde la distancia entre altavoces disminuya y donde la apertura acústica del array aumente. De forma práctica esto significaría emplear muchos altavoces con poca separación entre ellos, lo que acarrea el inconveniente del incremento en la cantidad de canales a alimentar, y el consiguiente problema de la transmisión de dichos canales entre origen y destino del sistema de comunicación. Profundizando en esta idea, ya existen en la actualidad sistemas de reproducción multicanal que, mediante técnicas de procesado de señal y un array de altavoces, permiten “proyectar” el sonido en la dirección deseada, sin necesidad de incrementar el número de altavoces.

5.1.2 LOCALIZACIÓN EN EL PLANO MEDIO

Como ampliación del objetivo principal de la Tesis, nos pareció muy interesante estudiar la localización en el plano medio. Este campo está en sus primeras etapas de investigación. En la búsqueda bibliográfica se encontraron pocos trabajos. Hay diversos artículos sobre estudios subjetivos de localización en elevación cuando se varían ciertos parámetros de la fuente, pero sólo encontramos uno sobre un modelo de simulación de localización del sonido en elevación. Dicho modelo no funciona justo para el plano medio. Hay que tener en cuenta que el audio espacial, está pensado principalmente para el plano horizontal.

El sistema auditivo humano tiene para el plano medio unos principios de funcionamiento muy distintos de los que emplea para el plano horizontal. Como resulta obvio, una fuente localizada en el plano medio producirá la misma señal en ambos oídos, por lo que la señal biaural o “biauralidad” no puede aportar ninguna información de localización. El proceso que sigue el sistema auditivo para obtener la localización en el plano medio es analizar la forma del espectro de la señal recibida para determinar el ángulo desde el que le llega el sonido.

Para poder obtener el ángulo de elevación de la fuente sonora, el cerebro ha ido aprendiendo a lo largo de la vida la asociación existente entre el ángulo de elevación y una serie de picos y nodos que aparecen en el módulo del espectro de la señal recibida en los oídos. Estos parámetros espectrales que dependen del ángulo de elevación son debidos a la influencia de la forma de la oreja, los hombros y el torso. Para reproducir este mecanismo, en el LES se ha desarrollado una técnica que simula este conocimiento a través del uso de una base de datos de las HRTF en elevación. La técnica empleada analiza el espectro de la

señal monoaural de entrada y como resultado determina la distribución de nodos y picos característicos asociada a cada ángulo. Aunque en la comparación visual de los espectros resulta bastante sencilla la identificación de nodos y picos, su implementación en un algoritmo matemático que compare estas características morfológicas ha requerido el estudio y análisis de diferentes técnicas hasta la obtención de un método cuyos resultados son bastantes buenos.

El análisis espectral que determina la localización en el plano medio está en gran medida influenciado por el tipo de señal que se pretende localizar, ya que el propio espectro de la señal de la fuente puede enmascarar o modificar los parámetros de localización monoaural (eminentemente espectrales). Como ejemplo podemos observar como la señal de la voz tiende a localizarse siempre por debajo de 90º debido a que posee menor energía en las altas frecuencias.

Este método de localización funcionará mejor con señales de banda ancha, en la cuales aparecen los nodos y picos característicos del ángulo. Si esta señal fuera de banda estrecha no se podrían identificar todos los nodos y picos pertenecientes a un ángulo, y daría resultados de localización erróneos. En los estudios subjetivos referenciados aparece este comportamiento de mala localización con señales de banda estrecha.

Casi siempre que se habla de sonido envolvente, multicanal, se entiende que es en el plano horizontal; donde por otra parte el sistema auditivo es más preciso debido a que ambos oídos se encuentran en este plano con una cierta separación. Una vez desarrollada e implementada la herramienta de análisis de parámetros de localización horizontal, se encontró de gran interés estudiar su aplicación y uso para reproducir un campo acústico mediante la técnica de ventana acústica virtual en el plano medio. Se trabajó y analizó la evolución de los parámetros de localización en elevación, campo en el que no se ha encontrado ningún estudio sobre WFS verticales entre la bibliografía consultada. Como consecuencia de dicho estudio se puede concluir que, puesto que dichos parámetros se encuentran en frecuencias medias-altas, la localización en el plano medio resulta prácticamente imposible puesto que el array de altavoces es incapaz de reproducir bien el campo sonoro para estas frecuencias.

Para futuros trabajos e investigaciones se podría abordar la aplicación de la herramienta en el estudio y desarrollo de sistemas que puedan desplazar los eventos sonoros en elevación. Con este objetivo se puede apuntar la utilización de una configuración tipo estereofónica o con tres altavoces balanceados,… Esta ampliación al plano medio permitiría colocar, por ejemplo, el sonido de un helicóptero encima de un oyente y no sólo delante o detrás del mismo, lo que permitiría recrear un campo sonoro virtual más realista. Un ejemplo de este tipo de sistemas es el nuevo 22.2 que se ha propuesto para el sistema de televisión “Ultra High Definition”.

En resumen, en esta Tesis se presenta un modelo de estimación de la elevación en el plano medio de un evento sonoro. Que sepamos, nunca antes se había realizado un modelo de localización en elevación basado en los parámetros espectrales que utiliza el sistema auditivo. Se ha demostrado que el sistema propuesto funciona bien tanto para señales biaurales reales como simuladas, y los resultados de estimación están dentro de los márgenes de tolerancia que proporcionan los estudios subjetivos referenciados. Por tanto, este trabajo supone una contribución importante al estado del arte actual, en temas de localización objetiva de fuentes sonoras en este plano.

5.1.3 SÍNTESIS DE LA SEÑAL BIAURAL

Una de las líneas de trabajo desarrolladas en la Tesis es la simulación de un sistema de ventana acústica virtual. Para ello se ha procedido a sintetizar la señal biaural recibida por un maniquí acústico ante la presencia de uno o varios altavoces radiando colocados en el plano horizontal o en el plano medio. Esta señal biaural se calcula mediante el uso de la función de transferencia de la cabeza del maniquí y la propiedad sumatoria del sonido. En un principio la investigación se basó en el uso de la base de datos de hrir de KEMAR, pero más tarde se detectó la necesidad de crear una base de datos propia basada en el maniquí acústico de nuestros ensayos, que pudiera permitir la obtención de unos resultados más precisos. Esta base de datos propia fue también empleada posteriormente para es desarrollo del modelo de localización.

Tomando como punto de partida los sistemas de medida de los que se disponía, se ha buscado el procedimiento más apropiado para lograr medir HRTF. Como consecuencia de esta búsqueda se ha desarrollado y propuesto un nuevo método de medición donde el sistema proporciona la función de transferencia de la cabeza de forma completamente independizada del sistema de medida que la rodea, incluida la sala, el sistema de reproducción y el sistema de captación. Resultan evidentes las ventajas de este método respecto al procedimiento de medida seguido en otras bases de datos, en las que esta separación de funciones de transferencia se realiza mediante un postprocesado basado en la ecualización o compensación, por ejemplo, de la respuesta del altavoz.

El proceso de medida de las HRTF presenta numerosas complicaciones. Mientras que desde el punto de vista del sistema de medida no se encuentran grandes inconvenientes, desde el punto de vista de configuración física de la cabeza y del altavoz resulta bastante complejo asegurar la correcta colocación del altavoz con respecto a la cabeza para un determinado ángulo. El proceso de ajuste para cada posición medida conlleva un importante consumo de tiempo. Sin embargo, el procedimiento de medida simplemente consiste en repetir la medición para cada ángulo; una vez configurado e implementados los algoritmos que calculan a partir de las HRTF proporcionadas por el sistema medida, las hrir adecuadas para el uso dentro del LES.

Debido a lo costoso de medir las hrir, las bases de datos se suelen limitar a un número determinado de ángulos, realizando un muestreo espacial. La resolución suele ser de 5º para el plano horizontal y de 10º grados para el plano medio. Por lo tanto, a la hora de calcular la señal biaural de fuentes colocadas en ángulos que no aparecen en la base de datos, es necesario interpolar las hrir más cercanas a dicha posición.

En la búsqueda de un buen sistema de interpolación se podrían utilizar las nuevas propuestas presentadas en los últimos JAES. Sin embargo, según el estudio realizado en esta Tesis, para poder conseguir la precisión de localización que se necesita, hemos comprobado que en el resultado no influye tanto la interpolación, ya que el umbral de percepción de localización (o precisión humana) es peor que la resolución de la base de datos. La precisión del sistema auditivo es muy variable, tal y como se analiza profundamente en el libro de [Blauert 97] “Spatial Hearing”, (capítulo 2, apartado 2.1. “Localization and localization blur”, pp. 37-50). En dichos estudios se ha comprobado como los resultados dependen de muchos factores, como tipo de sonido de la fuente (ancho de banda, tiempos de ataque, niveles, componentes espectrales,…), presentación al oyente, entrenamiento del oyente, etc.…; concluyendo que una precisión mínima de 5º sólo se alcanza en circunstancias muy especiales.

Por todo ello, y aunque se ha estudiado la interpolación para la implementación de un buen método, el error cometido siempre será menor a 5º, ya que ésta es la resolución de la herramienta LES, y en consecuencia, la interpolación no será un aspecto crítico porque está por debajo de la precisión del sistema auditivo. El estudio de los diferentes métodos de interpolación nos lleva a inferir que los mejores resultados se obtienen a partir de las interpolaciones en el dominio del tiempo, realizadas directamente sobre las hrir.

5.2 APORTACIONES

A lo largo de esta Tesis se ha desarrollado una búsqueda continua de la herramienta más apropiada para la localización de eventos sonoros. A continuación enumeraremos los conocimientos más destacables que hemos obtenido en dicho proceso:

- Nuevo método de medida de las HRTF. En este método sustituimos la medida de las hrir por la medición de las HRTF directamente. Con esta técnica evitamos la influencia del resto de las funciones de transferencia que están involucradas cuando se mide con un maniquí acústico dentro de una cámara anecoica. Esto supone un gran avance frente a otros métodos donde se graba la respuesta impulsiva de la cabeza en tiempo y se procesa más tarde una "ecualización" que compense la respuesta de lo que no es cabeza. El procedimiento ha sido creado para el sistema de medida multicanal PULSE de Brüel & Kjaer, mediante el desarrollo y programación de funciones específicas. Hay que resaltar la importancia de efectuar la medida directa de las HRTF, sin necesidad de realizar postprocesado por el notable ahorro de trabajo que acarrea en el proceso global. Además el procedimiento de medida seguido es el descrito en la teoría y definición de las HRTF, donde se debe relacionar la función de transferencia cuando está la cabeza y cuando no está colocando un micrófono en su lugar.
- LES (Localizador de Eventos Sonoros). Se ha diseñado e implementado una herramienta, basada en el paquete informático Matlab, de forma que a partir de la señal biaural, ya sea simulada o medida con un maniquí acústico, proporcione una estimación de la posición del evento sonoro, bien en el plano horizontal o bien en el plano medio. Esta herramienta proporciona la capacidad de evaluar un parámetro subjetivo de forma objetiva, por lo tanto, las ventajas asociadas son:
    - Repetibilidad de ensayos.
    - Abaratamiento de ensayos.
    - Disminución de tiempos de ensayos.
    - No resulta necesario utilizar un grupo de personas a las que hay que entrenar.
    - Utilización de un maniquí acústico en lugar de una persona, con la consiguiente total disponibilidad espacio-temporal de un "oyente".

  La herramienta LES es fácil de usar y puede proporcionar a su salida además de un ángulo estimado de localización del evento sonoro reproducido, la evolución de dicha estimación en el tiempo, según se van analizando las ventanas temporales de la señal de entrada, o la representación en frecuencia, según se analiza en el banco de filtros "perceptuales" del oído interno, así como la distribución estadística de estas estimaciones.

Debemos destacar la capacidad de la herramienta LES para simular configuraciones de escenarios acústicos con diferentes tipos de array o diferentes tipos de distribución de las posiciones de las fuentes. En el estudio se han tomado los resultados de alguna de estas configuraciones y han sido validados con medidas reales, comprobando que en ambos casos las estimaciones son similares. Esta simulación previa a la implementación de sistemas multicanal podría permitir un ahorro de tiempo-dinero, evitando sucesivas pruebas en las primeras etapas de desarrollo. Este procedimiento permitiría, sólo al final del desarrollo de los sistemas, recurrir puntualmente a su montaje real o a experimentos subjetivos más refinados.

- Tanto para el plano horizontal como para el plano medio, se ha hecho un estudio profundo acerca de la capacidad de los arrays de altavoces de reproducir los parámetros de localización, pensando sobre todo en su utilización para sistemas de teleconferencia.
- El hecho de poder disponer para los ensayos de una cámara anecoica, ha permitido realizar los mismos en un entorno ideal, donde se evita la influencia de la sala a la hora de medir la señal biaural.

5.3 LÍNEAS FUTURAS

La ampliación de este trabajo de Tesis, al ser multidisciplinar, presenta numerosas líneas que podrían seguirse. Entre ellas podríamos citar la aplicación de la herramienta LES a entornos acústicos reales, donde existen reflexiones en superficies y reverberación. En este sentido sería necesario aplicar un procesado temporal adicional que incluyera el efecto de precedencia. Este fenómeno todavía es poco conocido y no existen modelos fiables, obteniendose resultados satisfactorios sólo para ciertos tipos de sonidos. También podría resultar interesante la aplicación de los conocimientos existentes del enmascaramiento auditivo de unas bandas con otras en función del nivel relativo entre ellas.

En otra línea de trabajo se podría ampliar el LES a las tres dimensiones, conjugando la localización en el plano horizontal para 360º con la localización vertical para 180º. Desde el punto de vista subjetivo, el sistema auditivo tiene problemas de localización para ciertos rangos de ángulos, siendo más preciso en el rango estudiado en esta Tesis. De igual forma la precisión del sistema auditivo es muy dispar, pudiendo variar de forma bastante significativa en función del tipo de sonido utilizado y de su localización. Teniendo en cuenta estas limitaciones, resultaría bastante interesante una ampliación del LES a la semiesfera superior, que podría servir para analizar sistemas multicanal de sonido envolvente en 2D o en 3D. Mediante la conjugación de los parámetros ILD e ITD de localización horizontal y los parámetros espectrales de localización en elevación, se podría extender la localización más allá del plano horizontal y medio. Esto ayudaría de forma significativa al interesante desarrollo de la simulación de un entorno acústico donde se pudieran colocar fuentes sonoras en cualquier lugar respecto de un oyente, alimentar dichas fuentes con sonidos procedentes del sistema multicanal correspondiente y además, poder registrar el sonido que percibiría un oyente, para analizarlo.

La herramienta puede ser empleada también en el desarrollo de posibles mejoras del modelo de audición biaural. Teniendo en cuenta que el ser humano sufre un aprendizaje en su percepción espacial, este mecanismo se podría implementar con redes neuronales, donde la capacidad de localización variaría en función de la memoria.

Otro de los numerosos frentes de trabajo abiertos es la búsqueda de las HRTF apropiadas para un oyente determinado. En este campo de investigación siguen apareciendo estudios sobre diferentes métodos de interpolación en función del tipo de señal y del dominio tiempo-frecuencia; o diferentes métodos de síntesis de HRTF en función de los parámetros antropométricos.

Si nos fijamos ahora en la teoría del WFS habría que estudiar y analizar porque estos sistemas no son capaces de reproducir el parámetro de localización biaural basado en la ILD. Se sabe que este parámetro influye para frecuencias medias-altas, justo donde la reproducción del campo sonoro de la WFS no funciona. La continuación de los presentes trabajos permitiría profundizar en la búsqueda de un método que permitiera mejorar el WFS en este aspecto.

Por último, podemos resaltar que puesto que la calidad del sonido percibido tiene diversos aspectos que la cuantifican como pueden ser la extensión de la fuente, profundidad, presencia, claridad,… sería muy interesante desarrollar herramientas objetivas, basadas en los mismos principios de funcionamiento subjetivos que se han implementado en el LES, que fueran capaces de evaluar parámetros subjetivos de calidad a partir de una señal biaural. Aunque están apareciendo estudios subjetivos sobre qué parámetros objetivos parecen influir sobre estas características subjetivas de la audición, todavía existe mucho desconocimiento a la hora de simular el comportamiento cognitivo-neurológico del sistema auditivo humano, por lo que éste es un campo por explorar.

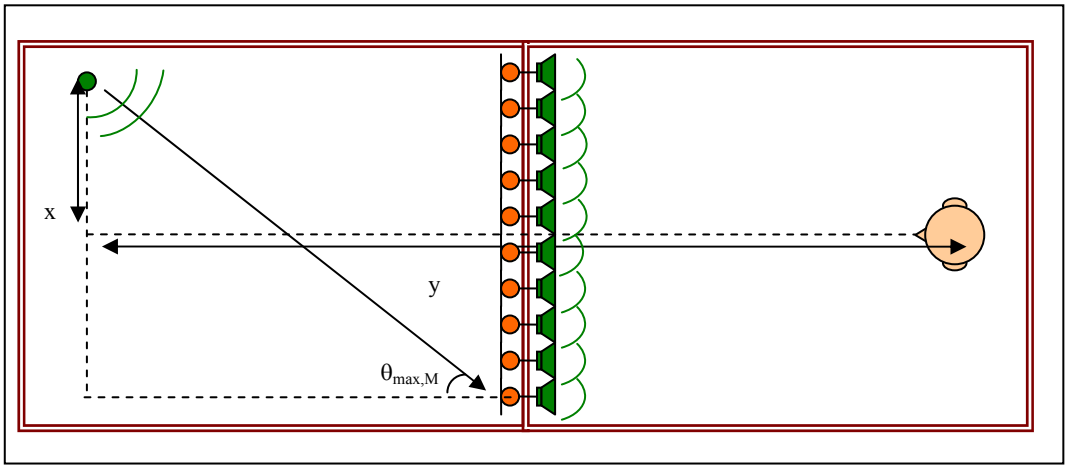
ANEXO 1-1

CÁLCULO DE LA FRECUENCIA DE ALIASING PARA EL ARRAY IMPLEMENTADO EN LA TESIS.

Basándose en la configuración del array y sus medidas se ha calculado la frecuencia de aliasing [De Brujin 03] a partir de la cual se considera que el WFS no produce una reconstrucción del campo sonoro buena:

$$f_{max} = \frac{c}{d_{alt}\left(\sin\theta_{max,M} + \sin\theta_{max,L}\right)}$$

Esta frecuencia depende de la posición concreta de la fuente y el oyente, para conocer el orden de magnitud de dicha frecuencia se va a calcular para las diferentes posiciones que se utilizan como muestra durante este trabajo. En la figura siguiente se representa un esquema de los arrays.



c: velocidad del sonido, 343 m/s.

d_{alt}: distancia entre altavoces, 0.175 m.

$\theta_{max,L}$: para la posición del oyente en el centro es 0º.

$\theta_{max,M}$: se calcula a partir de la distancia x e y:

$$\theta_{max,M} = \text{arctg}\left(\frac{4.5 \cdot d_{alt} + x}{y/2}\right)$$

La distancia entre el array de altavoces y el oyente es 2.5 m y la misma distancia hay entre el array de micrófonos y la fuente:

θ (º)	x (m)	$\theta_{max,M}$ (º)	f aliasing (Hz)
0	0.000	9	6524
10	0.882	17	3530
20	1.820	23	2715
30	2.887	28	2371
45	5.000	33	2135
60	8.660	38	2027

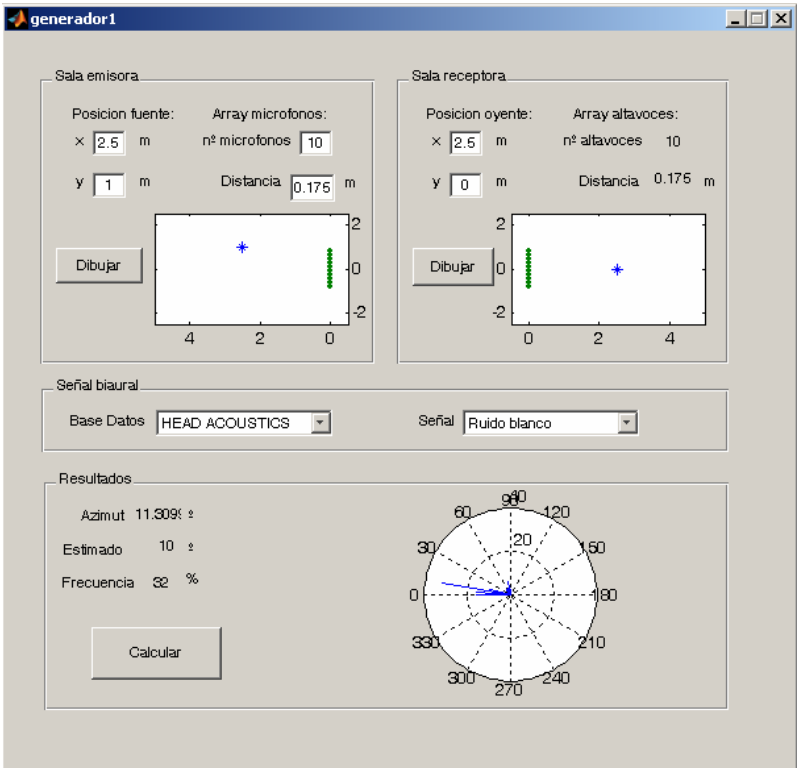
Frecuencias de aliasing a diferentes distancias entre altavoces y diferente nº de altavoces:

θ (º)	x (m)	$\theta_{max,M}$ (º)	f aliasing (Hz)			
			d_{alt} = 5 cm	d_{alt} = 10 cm	d_{alt} = 17.5 cm	d_{alt} = 20 cm
0	0.000	9	23168	11937	6524	6361
5	0.437	13	15721	7995	4455	4141
10	0.882	17	12419	6276	3530	3208
15	1.340	20	10622	5348	3024	2713
20	1.820	23	9526	4785	2715	2416
25	2.332	26	8805	4417	2512	2223
30	2.887	28	8306	4162	2371	2090
35	3.501	30	7946	3979	2269	1996
40	4.195	32	7679	3843	2193	1926
45	5.000	33	7475	3740	2135	1873
50	5.959	35	7318	3660	2090	1832
55	7.141	36	7194	3598	2055	1800
60	8.660	38	7097	3549	2027	1775
65	10.723	39	7020	3510	2006	1756
70	13.737	40	6961	3481	1989	1741
75	18.660	41	6917	3458	1976	1729
80	28.356	43	6885	3443	1967	1721
85	57.150	44	6866	3433	1962	1717

ANEXO 3-1

INTERFAZ DEL LES.

Se ha creado en Matlab un interfaz de usuario gráfico (GUI) para el diseño de diferentes configuraciones de ventana acústica virtual. Este interfaz está dividido en cuatro zonas: configuración de la sala emisora, configuración de la sala receptora, selección del tipo de sonido y base de datos utilizada, y por último, representación de los resultados.



En la sala emisora se puede elegir la posición de la fuente, el nº de micrófonos del array y la distancia entre ellos. El punto medio del array lineal siempre es el origen de coordenadas. Pulsando dibujar aparece representada la sala con las posiciones de la fuente y el array.

En la sala receptora se puede elegir la posición del oyente respecto del array. La configuración del array de altavoces es igual a la del array de micrófonos.

La base de datos utilizada para el cálculo de la señal biaural, y para las tablas de búsqueda de la etapa de estimación, se puede elegir entre la de KEMAR o la de HATS (de Head Acoustics). La señal es seleccionable entre ruido blanco, voz masculina y voz femenina.

En los resultados se presentan el azimut relativo oyente- fuente, el azimut estimado (moda de la estimación), la frecuencia estadística de la estimación, y una representación de la frecuencia estadística de los diferentes valores estimados en función del ángulo.

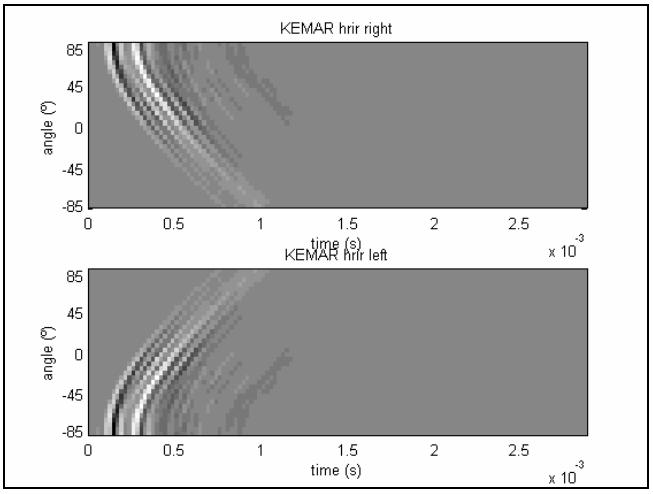
Este interfaz, aunque práctico, se queda corto a la hora de analizar el comportamiento de los parámetros de localización horizontal ILD e ITD en función de la frecuencia.

ANEXO 3-2

ESTUDIO COMPARATIVO CABEZA ESFÉRICA – CABEZA ELIPSOIDE.

Al principio del desarrollo de la Tesis se buscó una fórmula que relacionara mejor la ITD con el ángulo de azimut, que las basadas en el modelo de cabeza esférica. Para ello se modeló la cabeza como una elipse, y se calculó la diferencia de caminos entre las dos orejas.

La ITD es la diferencia de llegada a cada oído, esta se calcula a partir de la hrir de cada oído. Por ejemplo para 45º se ve como llega mucho antes el frente de onda al oído derecho que al oído izquierdo.



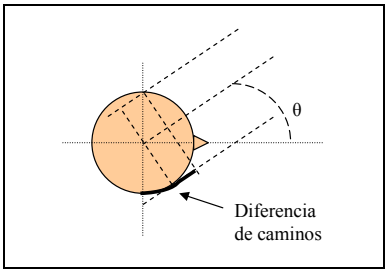
Cabeza esférica

Si se aproxima la cabeza a una esfera de radio b, la ITD para una fuente sonora tan distante que el frente de onda se considera plano se calcula como [Blauert 97]:

$$\mathrm{ITD} = \frac{b}{c}\left(\theta + \sin\theta\right)$$

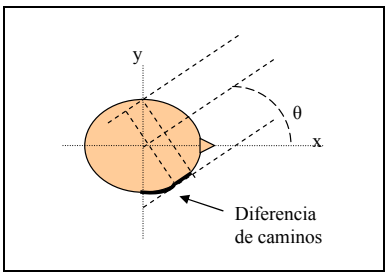
donde θ es el ángulo de azimut y c la velocidad del sonido. Esta fórmula no tiene solución analítica por lo tanto se estima el azimut a partir de la ITD basándose en una tabla con dos entradas, por una parte el valor de θ desde -90º hasta 90º con pasos de 1º y por otra las ITD calculadas con la fórmula. Luego para estimar el azimut comparamos la ITD medida con las de la tabla, dando como resultado aquel azimut cuya ITD se parezca más a la medida.

Esta fórmula sale de medir la diferencia de caminos recorrida por la onda para llegar a cada oído.

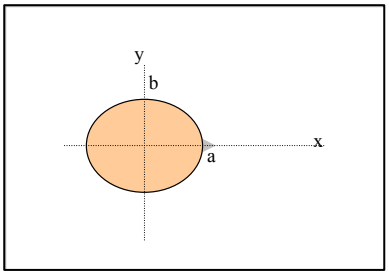


Cabeza elipsoidal

El modelo se basa en una elipse donde la diferencia de caminos tiene una parte de camino recto y otra de difracción, homólogo al caso de cabeza esférica.



En este caso no existe solución analítica para calcular el camino debido a la difracción. Se ha resuelto por métodos numéricos. A continuación se expone el proceso matemático.



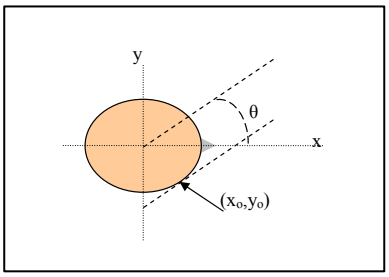
Una elipse queda definida por su ancho, 2a y su alto 2b, según la fórmula de una elipse en coordenadas cartesianas:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$

Poniéndolo en función de y:

$$y=\pm\sqrt{1-\frac{x^2}{a^2}}\cdot b=f(x)$$

Se calcula el valor de y_o para $x_o = 0 : 0.0001 : a$.



La pendiente de llegada de la onda m es:

$$m=tg\theta$$

La recta tangente a una curva es:

$$y=f'(x_o)(x-x_o)+f(x_o)$$

siendo el punto de tangencia (x_o,y_o). La pendiente de dicha tangente es:

$$f'(x)=\mp\frac{bx_o}{a^2}\left(1-\frac{x_o^{\,2}}{a^2}\right)^{-1/2}$$

para una elipse. Luego, se calcula el ángulo θ de la recta tangente a la elipse igualando las pendientes:

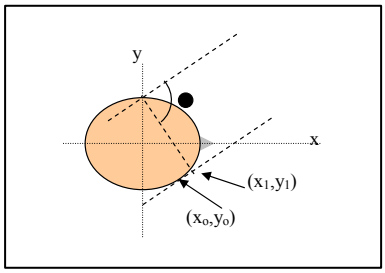
$$tg\theta=\mp\frac{bx_o}{a^2}\left(1-\frac{x_o^{\,2}}{a^2}\right)^{-1/2}$$

La recta tangente también se puede definir como:

$$y=mx+g$$

Se calcula el valor de g para la recta tangente:

$$g=y_o-tg\theta\cdot x_o$$



La recta perpendicular a la llegada del oído izquierdo será:

$$y = nx + b$$

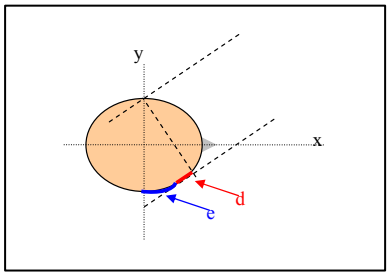
siendo n la pendiente igual a:

$$n = \left(\theta + \frac{\pi}{2}\right)$$

Esta recta cruza la recta tangente en el punto (x_1,y_1). Se calcula este punto con el sistema de ecuaciones formado por las dos rectas (la perpendicular y la tangente):

$$y_1 = tg\left(\theta + \frac{\pi}{2}\right)x_1 + b$$

$$y_1 = tg(\theta)x_1 + g$$



La diferencia de caminos recorrida está compuesta por d, la distancia de propagación recta entre el punto (x_1,y_1) y el punto de tangencia (x_o,y_o) y e, la distancia de propagación de difracción alrededor de la cabeza hasta el oído derecho.

La distancia de propagación recta es:

$$d = \sqrt{(x_1 - x_o)^2 + (y_1 - y_o)^2}$$

El cálculo del perímetro de la elipse entre (x_o,y_o) y (0,-b) será la distancia difractada. Este cálculo se realiza sobre la ecuación de la elipse parametrizada:

$$x = a\cos(\varphi)$$

$$y = b\sin(\varphi)$$

siendo φ el ángulo de cada punto de la elipse (x,y). La longitud de una elipse es la integral de la norma de las derivadas de las ecuaciones parametrizadas:

$$e = \int_{\frac{\pi}{2}}^{\varphi(x_o,y_o)} \sqrt{a^2\sin^2(\varphi) + b^2\cos^2(\varphi)}\, d\varphi$$

Luego se puede calcular la diferencia de tiempo de llegada del sonido a cada oído ITD para un modelo de cabeza elíptica como:

$$ITD = \frac{dis\tan cia}{c} = \frac{e+d}{c}$$

b es el radio de la cabeza 0.093 m y a se ha ajustado de forma iterativa para dar la mejor estimación según las medidas.

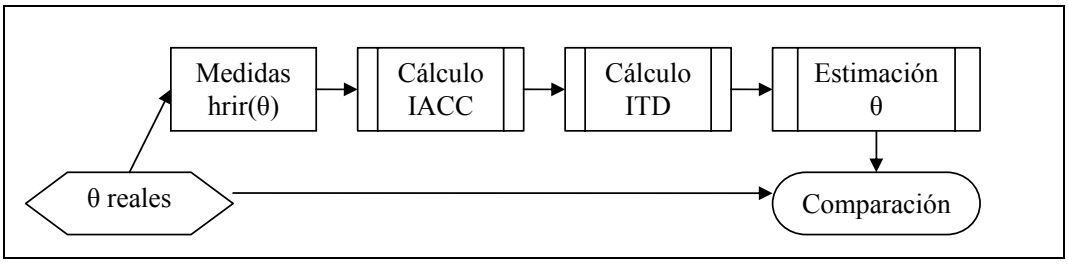
Búsqueda de la mejor forma de cabeza y sus dimensión: estimación del valor a

Para calcular la ITD se usan diferentes fórmulas que intentan modelar la diferencia de tiempo de llegada de la onda a cada oído dependiendo de la diferencia de camino recorrido. Estas fórmulas están basadas en modelos esféricos de la cabeza, realizando diferentes aproximaciones, según sea alta frecuencia o baja frecuencia.

Según va creciendo el ángulo de azimut los errores en la estimación de este son mayores, por esto se ha buscado una fórmula que tenga en cuenta la forma de la cabeza elipsoidal.

En un estudio anterior se utilizó un modelo de ovoide para calcular la diferencia de tiempos para diferentes ángulos de azimut y elevación [Duda 99]. En este trabajo, como sólo se trabaja en el plano horizontal se ha utilizado un modelo elipsoidal.

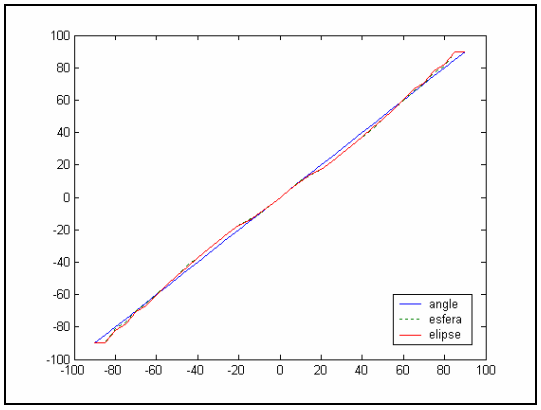
Se ha medido sobre varias cabezas la hrir y se ha calculado la ITD a partir de la IACC. Aplicando la fórmula de cabeza elipsoidal y de cabeza esférica se calcula a partir de la ITD medida la estimación del ángulo y se compara con el ángulo real.



Se ha comparado el modelo esférico y elipsoidal sobre las hrir de KEMAR. b es el radio de la cabeza y se ha ajustado de forma iterativa para dar el menor ECM del modelo esférico.

a se ha ajustado de forma iterativa para dar la mejor estimación de θ para el modelo elíptico con el radio b estimado para cabeza esférica.

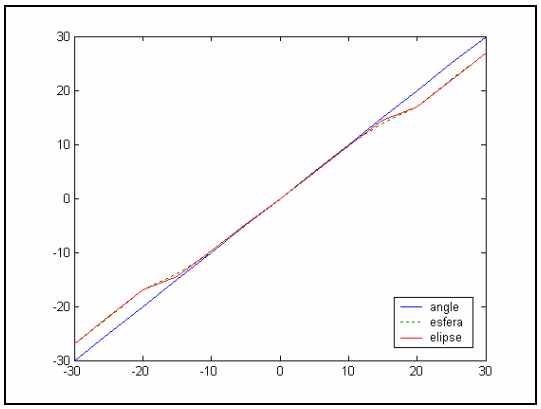
Sobre la base de datos del maniquí acústico KEMAR se realiza la comparación. Resultados:



Parámetros	ECM
Esfera b = 0.094 m	4.9189
Elipse a = 0.091 m	5.1804

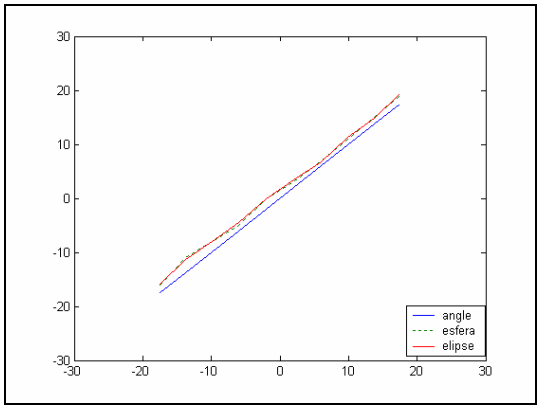
Para KEMAR es mejor el modelo esférico. El mejor modelo elíptico es prácticamente una esfera.

Se reducen los datos al margen de frecuencia utilizado para el array para comparar.



Parámetros	ECM
Esfera b = 0.094 m	4.3077
Elipse a = 0.091 m	4.3645

Sobre las medidas realizadas en la cámara anecoica del maniquí acústico HATS se realiza la comparación. Resultados:

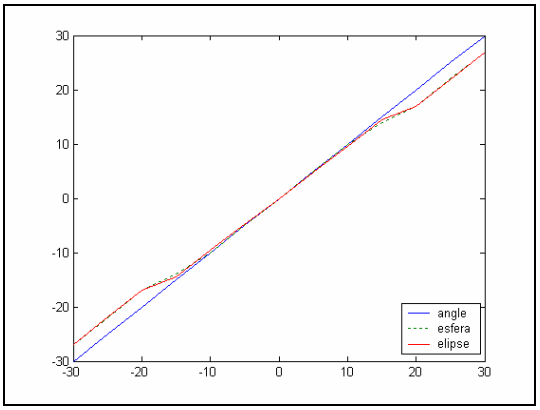


Parámetros	ECM
Esfera b = 0.169 m	2.5547
Elipse a = 0.171 m	2.7646

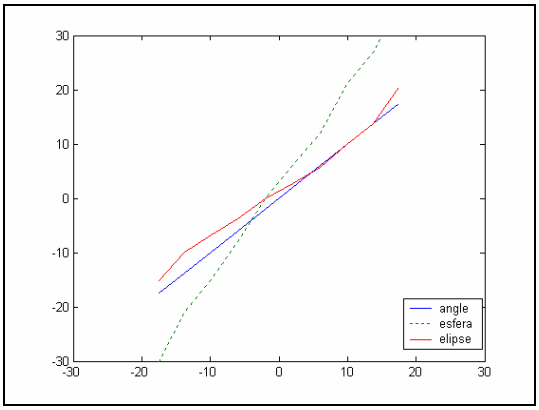
Conclusiones

- No se aprecian mejorías evidentes en el uso del modelo elíptico frente al modelo esférico.
- Los modelos para HATS arrojan menores ECM en la estimación del ángulo de azimut que los modelos de KEMAR.
- Para HATS las dimensiones de las cabezas son mucho más grandes que el tamaño real. Por lo tanto no sirve este modelo. El parámetro b del radio de la cabeza lo vamos a fijar al tamaño medio, b = 9.3 cm.

Búsqueda de la mejor forma de cabeza para un radio de 9.3 cm



KEMAR



HATS

Medidas	Parámetros	ECM (esfera)	ECM (elipse)
KEMAR	a = 0.093 m	4.3077	4.6422
HATS	a = 0.147 m	93.5371	4.8270

Conclusiones

- Para KEMAR, la elipse que mejor resultado da es aquella cuyo dos ejes son iguales (esfera).
- Para HATS, el valor de a varía entre 0.147 y 0.160, valores cercanos al tamaño natural del maniquí. El modelo esférico presenta más error.

Comprobación del error de muestreo

Para $f_{muestreo}$ = 44100 Hz, un error de una muestra es de 22.7 μs (frecuencia muestreo de la base de datos de KEMAR). Se va a comprobar que error induce esta desviación.

Parámetros	ECM (esfera)	ECM (elipse)
a = 0.093 cm b = 0.093 cm	5.1475	6.0224

Para la $f_{muestreo}$ = 51200 Hz un error de muestra es de 19.5 μs (frecuencia muestreo con la que se midieron estas hrir de HATS). Se va a comprobar que error induce esta desviación.

Parámetros	ECM (esfera)	ECM (elipse)
a = 0.147 cm b = 0.093 cm	4.0000	81.9027

Conclusiones

Para las medidas de HATS, la desviación de una muestra en el cálculo de la ITD proporcionar errores en la estimación del ángulo de azimut mayores al error del modelo elíptico. Para el caso esférico los errores en las medidas son mucho mayores que el error debido a la desviación de una muestra.

Para la cabeza HATS se consigue mejores estimaciones con el modelo elíptico. Se van a utilizar los parámetros a = 0.147 cm y b = 0.093 cm.

Capítulo 6: BIBLIOGRAFÍA

Publicaciones propias

[Gómez-Alfageme 04] Gómez-Alfageme J.J., Blanco-Martín E., Torres-Guijarro S., Casajús-Quirós F.J., "Objective measurements of sound source localization in a multichannel transmission system for videoconferencing", AES 116th Convention, Berlin, Germany, 2004.

[Blanco-Martin 05] Blanco-Martín E., Gómez-Alfageme J.J., Torres-Guijarro S., Casajús-Quirós F.J., "Spatial sound localization measures from a dummy head with a loudspeaker array in anechoic chamber", AES 118th Convention, Barcelona, Spain, 2005.

[Gómez-Alfageme 07] Gómez-Alfageme J.J., Sánchez-Bote J.L., Blanco-Martín E., "Design, construction and qualification of the new anechoic chamber at Laboratorio de Sonido, Universidad Politécnica de Madrid", AES 122nd Convention, Viena, Austria, May 2007.

[Blanco-Martin 08] Blanco-Martín E., Casajús-Quirós F.J., Gómez-Alfageme J.J., Ortiz-Berenguer L.I., "Loss of subjective localization cues in virtual acoustic opening", AES 124th Convention, paper nº 7443, Amsterdam, The Netherlands, May 2008.

[Blanco_Martín 08] Blanco-Martín E., Casajús-Quirós F.J., Gómez-Alfageme J.J., Ortiz-Berenguer L.I., "Sound Event localization in horizontal and median planes for virtual acoustic opening", pendiente de publicación en Journal of Audio Engineering Society. Presentado en Marzo 2008, revisado en Septiembre de 2008, presentado para segunda revisión en Noviembre 2008, revisado en Enero de 2008, presentado en Marzo 2009.

[Blanco_Martín 09] Blanco-Martín E., Casajús-Quirós F.J., Gómez-Alfageme J.J., Ortiz-Berenguer L.I., "Simulation of subjective localization of a sound event in the median plane", pendiente de publicación en Applied Acoustics. Presentado en Diciembre 2008, revisado en Abril de 2009.

Artículos:

[Ajdler 08] Ajdler T., Faller C., Sbaiz L., Vetterli M., "Sound field analysis along a circle and its applications to HRTF interpolation", Journal Audio Engineering Society, vol. 56, no. 3, pp. 156-174, March 2008.

[Algazi 01] Algazi V. R., Duda R. O., Thompson D. M., and C. Avendano, "The CIPIC HRTF Database," Proc. 2001 IEEE Workshop on Applications of Signal Processing

to Audio and Electroacoustics, pp. 99-102, Mohonk Mountain House, New Paltz, NY, October, 2001.

[Becker 01] Becker J., Sapp M., "Synthetic soundfields for the rating of spatial perceptions" Applied Acoustics 62, pp 217-228, 2001.

[Beracoechea 07] Beracoechea, J.A., et al., "Overview of coding alternatives for virtual acoustic opening based applications" AES 28th International Conference: paper nº. 6-3, Pitea, Sweden, June 2006.

[Beracoechea 07] Beracoechea, J.A., "Codificación de audio multicanal para entornos de ventana acústica" Tesis Doctoral, SSR, ETSIT, UPM, Madrid 2007.

[Berg 03] Berg J., Rumsey F., "Systematic Evaluation of Perceived Spatial Quality", 24th International Conference: Multichannel Audio, paper nº. 43, May 2003.

[Boone 95] Boone M.M., Verheijen E.N.G., Van Tol P.F., "Spatial sound-field reproduction by wave-field synthesis", Journal Audio Engineering Society, vol. 43, no. 12, pp. 1003-1012, December 1995.

[Busson 05] Busson S., Nicol R., Katz B.F.G., "Subjective investigations of the interaural time difference in the horizontal plane", AES 118th Convention, paper nº 6324, Barcelona, Spain, May 2005.

[Damaske 72] Damaske P., Ando Y., "Interaural cross correlation for multichannel loudspeaker reproduction", Acustica, vol. 27. pp. 232-238, 1972.

[De Bruijn 03] De Bruijn W.P.J., Boone M.M., "Application of Wave Field Synthesis in life-size videoconferencing", AES 114th Convention, paper nº 5801, Amsterdam, Netherlands, March 2003.

[Duda 99] Duda R.O., Avendano C, Algazi V.R., "An adaptable ellipsoidal head model for the Interaural Time Difference", ICASSP'99, Proc. IEEE International Conference on Acoustics Speech and Signal Processing, pp. II-965-968, Mar.1999.

[Faller 04] Faller C., Merimaa J., "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence", Journal of Acoustical. Society of America, vol. 116, no. 5, November 2004.

[Ferguson 05] Ferguson S., Cabrera D., "Vertical localization of sound from multiway loudspeakers", Journal Audio Engineering Society, vol.53, no. 3, pp. 163-173, March 2005.

[Freeland 04] Freeland F.P., Biscainho L.W.P., Diniz P.S.R., "Interpositional transfer function for 3D-sound generation" J. Audio Engineering Society, vol.52, no. 9, pp. 915-930, September 2004.

[Gardner 95] Gardner B., Martin K.D., "HRTF Measurements of a KEMAR", Journal of the Acoustical Society of America, vol. 97, issue 6, pp. 3907-3908, June, 1995.

[Gardner 00] Gardner B., Martin K.D., "HRTF Measurements of a KEMAR Dummy-Head Microphone", Tech. Rep. 280, MIT Media Lab. Perceptual Computing, Cambridge, MA, 1994.

[Harma 99] Härmä A., Palomäki K., "HUTear - a free matlab toolbox for modelling of auditory system", in Proceedings. of Matlab DSP, Conference 1999, pp. 96-99, Espoo, Finland, November 1999.

[Harma 02] Harma A., "Coding principles for virtual acoustic openings" Proceedings of the AES 22nd International Conference: Virtual, Synthetic, and Entertainment Audio, Finland, paper nº 240, pp. 159-165, Espoo, Finland, June 2002.

[Hartung 99] Hartung K., Braasch J., Sterbing S.J., "Comparison of different methods for interpolation of head-related transfer functions", AES 16th International Conference, pp. 319-329, Finland, April 1999.

[Iida 07] Iida, K., Itoh, M., Itagaki, A., Morimoto, M., "Median plane localization using a parametric model f the head-related transfer function base don spectral cues", Applied Acoustics 68, pp. 835-850, 2007.

[Itoh 07] Itoh, M., Iida, K., Morimoto, M., "Individual differences in directional bands in median plane localization", Applied Acoustics 68, pp. 909-915, 2007.

[Keyrouz 06] Keyrouz F., Naous Y., Diepold K., "A new method for 3-D binaural localization based on HRTFs", IEEE International Conference on Acoustic, Speech, and Signal Processing, pp. V-341-344, May 2006.

[Keyrouz 08] Keyrouz F., Diepold K., "A new HRTF interpolation approach for fast synthesis of dynamic environmental interaction", Journal Audio Engineering Society, vol. 56, no. 1/2, pp. 28-35, January/February 2008.

[Kuhn 77] Kuhn G. F., "Model for the interaural time difference in the azimuthal plane", Journal of Acoustical Society of America, vol.62, nº 1, pp. 157-167, July 1977.

[Miyasaka 91] Miyasaka E., "Methods of quality assessment of multichannel sound systems" AES 12th International Conference: The perception of reproduced sound, pp. 188-196, May 1993.

[Nishino 96] Nishino T., Mase S., Kajita S., "Interpolating HRTF for Auditory virtual reality" Journal Acoustical Society of America, vol. 100, issue 4, pp 2602, October 1996.

[Pulkki 99] Pulkki V., Karjalainen M., Huopaniemi J., "Analyzing virtual sound source attributes using a binaural auditory model", Journal Audio Engineering Society, vol. 47,no. 4, pp. 203-217, April 1999.

[Pulkki 01] Pulkki V., Karjalainen M., "Localization of amplitude-panned virtual sources I: stereophonic panning", Journal Audio Engineering Society, vol. 49, no. 9, pp. 739-752, September 2001.

[Pulkki 05] Pulkki V., Hirvonen T., "Localization of virtual sources in multichannel audio reproduction", IEEE Transactions on Speech and Audio Processing, vol. 13, no. 1, pp. 105-119, January 2005.

[Rodriguez 05] Rodriguez S.G., Ramirez M.A., "Extracting and modelling approximated pinna-related transfer functions from HRTF data", Proccedings of ICAD 05 – 11th Meeting of International conference on Auditory Display, pp. 269-273, Limerick, Ireland, July 6-9, 2005.

[Rodriguez 05a] Rodriguez S.G., Ramirez M.A., "HRTF Individualization by solving the least squares problem", AES 118th Convention, Barcelona, Spain, 2005.

[Slaney 98] Slaney M., "Auditory toolbox version 2", Technical Report 1998-010, Interval Research Corporation, 1998.

[Sobreira 01] Sobreira-Seoane M.A., Juhl P., Henriquez V.C., "Calculation of transfer functions related to a head and torso simulator", Proceedings of the Eighth

International Congress on Sound and Vibration, Hong Kong, People's Republic of China, pp 135-142, July 2001.

[Sontacchi 02] Sontacchi A., Noisterning M., Majdak P., Höldrich R., "An objective model of localisation in binaural sound reproduction systems", Proceedings of the AES 21st International Conference: Architectural Acoustics and Sound Reinforcement, paper nº 136, St. Petersburg, Russia, June 2002.

[Toledo 08] Toledo, D., Moller, H., "The role of spectral features in sound localization" AES 124th Convention, paper nº 7450, Amsterdam, The Netherlands, May 2008.

[Torres 03] Torres-Guijarro S., Beracoechea-Alava J.A., Casajús-Quirós F.J., Ortiz-Berenguer L.I., "Multichannel Audio Decorrelation for Coding", Proceedings of the 6th International Conference On Digital Audio Effects (DAFX-03), London, UK, September 2003.

[Venegas 06] Venegas R., Correa R., Floody. S., "Modelo computacional de localización sonora espacial", Ingelectra 2006, Universidad Austral de Chile, Valdivia, Chile, 23-25 Agosto 2006.

[Viste 04] Viste H., Evangelista G., "Binaural source localization", Proceedings 7th International Conference on Digital Audio Effects (DAFX-04), pp. 145-150, Naples, Italy, October 2004.

[Yang 01] Yang W., "Overview of the head-related transfer functions (HRTFs)" ACS 498B Audio Engineering, Summer Program 2001, Graduate Program in Acoustics, The Pennsylvania State University, July 2001.

[Zwicker 99] Zwicker, E., Fastl, H., "Psycoacoustics: facts and models", Ed. Springer-Verlag, Berlin, 1999.

Libros:

[Blauert 97] Blauert J.; "Spatial hearing. The psychophysics of human sound localization", MIT Press, Cambridge, Massachusetts, 1997.

[Begault 94] Begault D.R.; "3-D sound for virtual reality and multimedia", Academic Press, Cambridge, Massachusetts, 1994.

Normas:

ITU Recommendation BS.1387-1 "Method for objective measurements of perceived audio quality".

ITU Recommendation P.800: "Methods for subjective determination of transmission quality". Former Rec. P.80.

ITU Recommendation P.800.1: "Mean Opinion Score (MOS) terminology".

ITU Recommendation BS.1116: "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems".

ITU Recommendation BS.775-1: "Multi-channel stereophonic sound system with or without accompanying picture".

ITU Recommendation P.58: "Head and torso simulator for telephonometry".

ITU Recommendation P.57: “Artificial ears”.

ISO 532: “Acoustics - Method for calculating loudness level”.

WEBS:

[DUDA 00] Duda R.O., “3-D Audio for HCI”, Department of Electrical Engineering, San Jose State University, http://interface.cipic.ucdavis.edu/CIL_tutorial/3D_home.htm.

[Faller 04a] Faller C., Merimaa J., “Binaural Cue Selection Toolbox”, [Faller 04], http://www.acoustics.hut.fi/software/cueselection/.

[Gardner 00a] Gardner B., Martin K.D., “HRTF Measurements of a KEMAR Dummy-Head Microphone”, [Gardner 00], http://sound.media.mit.edu/KEMAR.html.

[Harma 00] Harma A., ”HUTear Matlab Toolbox version 2.0”, March 2000, [Harma 99], http://www.acoustics.hut.fi/software/HUTear/

[Slaney 98a] Slaney, M., “Audotory Toolbox”, [Slaney 98] http://rvl4.ecn.purdue.edu/~malcolm/interval/1998-010.

[Pulkki 99a] Pulkki V., “Binaural Auditory”, [Pulkki 99], http://www.acoustics.hut.fi/~ville/software/auditorymodel.