

UNIVERSIDAD POLITECNICA DE VALENCIA

ESCUELA POLITECNICA SUPERIOR DE GANDIA

MASTER EN POSTPRODUCCION DIGITAL



UNIVERSIDAD
POLITECNICA
DE VALENCIA



ESCUELA POLITECNICA
SUPERIOR DE GANDIA

“TÉCNICAS DE SONIDO BINAURAL EN LA POSTPRODUCCIÓN AUDIOVISUAL”

TRABAJO FINAL DE MASTER

Autor: **RODRÍGUEZ MARIÑO, RAMÓN**

Director: **PAYRI, BLAS**

Gandia, Septiembre de 2011

RESUMEN

La escucha binaural es un recurso que puede ser empleado en aplicaciones audiovisuales para obtener una experiencia inmersiva que mejora el estereó clásico y es comparable al sonido estereó envolvente multicanal. Para ello es necesario un procesamiento específico a través de filtrado de tipo HRTF (*función de transferencia derivada de la cabeza*). Este Proceso consiste en aplicar a cualquier sonido monofónico o estereofónico las características que el sistema auditivo necesita para interpretar la localización espacial tridimensional del sonido (indicios binaurales). Estos indicios dependen principalmente de características antropomórficas individuales, lo que introduce ciertos problemas debidos al uso de HRTF no individualizadas, como dificultad para localizar sonidos fuera de nuestra cabeza (*internalizaciones*) o confusiones entre delante y detrás (*inversiones*). Una buena comprensión de las características de la audición humana es necesaria para optimizar este procesamiento.

En este trabajo se realiza un estado de la cuestión y un repaso de estudios realizados sobre la materia. Se constata que experimentos previos (Kistler,1999; Mackensen, 2004; Werséni, 2009) apuntan a una mejora de la localización espacial con el uso de sonidos del mundo real, o a través de la simulación de movimientos relativos fuente sonora-oyente (indicios dinámicos). Se propone un diseño experimental consistente en un test de percepción con el fin de comparar el comportamiento de distintas señales de audio con distintas características temporales y espectrales, la influencia del movimiento y la fiabilidad de distintas herramientas de postproducción y grabación de campo binaurales. Se han probado sobre 11 sujetos, 3 tipos de sonido (espectralmente difuso, impulsivo y voz) y 4 técnicas de recreación binaural: procesamiento mediante software de postproducción de audio (*panorama 5, binaural panner*), filtrado con *matlab* y grabación de campo con microfónica binaural (*Soundman OKM*) sobre muestras estáticas y con movimiento simulado.

Los resultados son muy similares para todas las técnicas de recreación, e indican que los sonidos impulsivos son detectados con más precisión y menos errores. El movimiento parece mejorar los problemas de internalización (hasta un 40%), y las confusiones detrás-delante, pero aumenta las confusiones delante-detrás y la precisión en la localización en general.

Se constata que un mejor conocimiento de las características de la escucha espacial humana es de interés en la mejora de las estrategias de diseño de sonido y aplicable a la realización de productos audiovisuales novedosos de alta calidad.

ABSTRACT

Binaural hearing can be used in audiovisual applications allowing an immersive experience that goes beyond the classical stereo and comparable to surround multichannel stereo sound. It requires a specific processing through HRTF (head-related transfer function). This processing consists on applying the characteristics that hearing system needs to interpret the spatial location of the sound (binaural cues) to any monophonic or stereophonic sound. These cues depend on individual anthropomorphic characteristics, what led to problems due to non-individualized HRTF processing, as *inside-the-head locations* or *front-back confusions*. A good understanding of the characteristics of human hearing is needed to improve the processing.

In this work, a state-of-the-matter and a background review is done. Previous research (Kistler, 1999; Mackensen, 2004; Wersény, 2009) point to an improvement of spatial location using real-world sounds or relative listener-source movement simulation (dynamic cues). An experimental design is proposed consisting on a perceptual test to compare the behavior of different time and space shaped audio signals, the influence of movement and the reliability of different binaural audio postproduction and field recording tools. Eleven subjects have been tested with 3 different sorts of sounds (diffuse spectrum, impulsive and speech) and 4 different binaural simulation technics: audio postproduction software (*panorama 5*, *binaural panner*), *matlab* filtering and binaural field recording (*Soundman OKM*) through static and moving samples.

The results are very similar regarding the simulation technics, and show that impulsive sound is more easily and precisely detected with less confusions. Movement seems to improve the inside-the-head location (up to 40%) and the back to front confusions, but increase the front-to back and the localization blur in overall.

It is verified that a better understanding of the human spatial hearing characteristics is a field of interest on improving the sound design strategies and applicable to produce high quality original audiovisual products.

ÍNDICE

RESUMEN

ABSTRACT

1. INTRODUCCIÓN	4
2. REVISION TEÓRICA	8
2.1. ANTECEDENTES	8
2.1.1. Audición y escucha.....	8
2.1.2. Audición espacial. Las diferencias interaurales.....	9
2.1.3. ITD e ILD.....	9
2.1.4. Indicios espectrales	12
2.1.5. Indicios dinámicos	14
2.1.6. Indicios adicionales	19
2.2. HERRAMIENTAS DE ESPACIALIZACIÓN DE AUDIO	22
2.2.1. Herramientas clásicas de mezcla	22
2.2.2. Procesado HRTF	24
2.2.3. HRTF no individualizadas. Confusiones	25
2.3. TECNOLOGÍAS BASADAS EN FILTRADO HRTF	29
3. DISEÑO EXPERIMENTAL	35
3.1. OBJETIVOS.....	36
3.2. ESTÍMULOS.....	37
3.2.1. Elección del material sonoro.....	37
3.2.2. Obtención del material sonoro.....	41
3.2.3. Elección del tipo de procesado.....	41
3.2.4. Simulación de movimientos de cabeza	50
3.3. PROCEDIMIENTO.....	52
3.3.1. Sujetos.....	52
3.3.2. Condiciones de escucha.....	52
3.3.3. Tarea.....	53
3.4. RESULTADOS	55
3.4.1. Aciertos/posición	55
3.4.2. Influencia del tipo de material sonoro.....	61
3.4.3. Influencia del tipo de procesado	66
3.4.4. Influencia del movimiento.....	68
4. DISCUSIÓN	72
5. CONCLUSIONES	76
6. referencias bibliográficas.....	77
ANEXO I	83
ANEXO I	87

INTRODUCCIÓN

1. INTRODUCCIÓN

Uno de los objetivos del diseño de sonido audiovisual es aprovechar las posibilidades narrativas y técnicas que ofrecen los soportes de producción, postproducción y reproducción existentes, lo que condiciona en gran medida las estrategias utilizadas en la creación de un producto.

Las técnicas de grabación y producción permiten maximizar la calidad del material sonoro del que se parte. Las técnicas de posproducción aportan posibilidades expresivas y permiten optimizar la calidad de sonido mediante estrategias de mezcla, espacialización y masterización.

Los soportes de reproducción determinan el tipo de escucha que utilizará el destinatario del producto y ésta condiciona el tipo de experiencia que obtendrá.

El sistema de escucha por excelencia es el estéreo, presentado en 1933 por Alan Blumlein (Blumlein, 1933) como una mejora del sistema de reproducción-escucha monofónico, mediante la utilización de dos canales de sonido independientes de grabación-reproducción, en vez de solo uno. Esto permitió por primera vez la espacialización de sonidos mediante la grabación en dos canales, en los que el sonido quedaba registrado con ligeras variaciones que contenían información espacial simple.

Curiosamente, Blumlein patentó el sistema bajo el nombre de “sistema binaural” argumentando que los humanos oímos por dos oídos, por lo que la escucha por dos altavoces era lógicamente más natural.

El sonido estéreo permitió el registro de información espacial y además desarrollar técnicas que permitían mezclar señales monofónicas con diferentes características espaciales simples, utilizando herramientas básicas como el potenciómetro panorámico o el desplazamiento de fase, que imitaban de forma elemental propiedades de la escucha humana binaural. Esto permitió una mejora considerable de la calidad del sonido, ya que las señales se separaban en el

espacio virtual de escucha, mejorando en gran medida enmascaramientos mutuos.

Sin embargo, la reproducción en dos canales, impedía imitar fielmente la escucha espacial humana, lo que llevó al desarrollo de sistemas de reproducción que utilizaban más canales, primero alrededor del sujeto horizontalmente, y luego en todo el espacio tridimensional.

La imitación de la escucha estereo envolvente mediante sistemas de múltiples altavoces es popular y exitosa, pero impone una serie de limitaciones que la convierten en excesivamente costosa y rígida.

Paralelamente al desarrollo del sonido multicanal, una serie de investigaciones llevadas a cabo en el campo de la acústica perceptiva, han revelado propiedades de la escucha humana que permiten una simulación bastante aproximada de ésta de forma más simple, utilizando técnicas de procesado de la señal de audio no triviales denominados sistemas de sonido binaural.

Estos sistemas permiten reproducir sonido con características espaciales tridimensionales mediante simples auriculares, consiguiendo una experiencia más inmersiva que el estéreo convencional y que bajo ciertas condiciones incluso supera al sonido multicanal multialtavoz.

Los sistemas de sonido binaural permiten, además de una experiencia inmersiva, otras mejoras como más separación de fuentes de sonido, ya que el *espacio virtual* de reproducción es mucho más amplio, disminuyendo los enmascaramientos y mejorando la calidad de sonido final. Además, tienen ventajas con respecto a los sistemas multialtavoz como menor coste, mayor simplicidad de configuración del sistema de reproducción, mayor movilidad, ya que solo se necesitan unos auriculares, y teóricamente mejor calidad y nitidez ya que se evitan diafonías entre altavoces.

Además introducen nuevas posibilidades narrativas y estéticas, ya que la localización de fuentes no se restringe a la región delantera, trasera o superior, y

posibilita la utilización de la usual localización interior de sonidos como una opción más.

Sin embargo, este sistema también tiene ciertas limitaciones, ya que se restringe a la escucha individual en auriculares, y la simulación no es totalmente fiel.

En todo caso la escucha binaural tiene claras aplicaciones en el audiovisual, y ha sido utilizada exitosamente por un número todavía limitado de productos audiovisuales.

Una simple aproximación a la materia no parece suficiente dada la complejidad de las características de la escucha binaural y la falta de teorías unificadas, por lo que se aprecian carencias en la comprensión y desarrollo posibilidades de este tipo de escucha.

Este trabajo, intenta aportar un enfoque a la materia desde el punto de vista de la posproducción. Para ello se ha realizado una aproximación a los antecedentes teóricos y experimentales y una revisión del estado de la tecnología y las aplicaciones de síntesis binaural. Esto nos ha conducido a la ideación de un diseño experimental que tratará de estudiar el comportamiento de herramientas de posproducción de audio con el fin de explorar las posibilidades que esta tecnología aporta al diseño de sonido audiovisual.

Parte de este trabajo de exploración se ha realizado en "The Hong Kong University of Science and Technology", durante el spring semester de 2011, en el marco del programa de intercambio PROMOE, y bajo la supervisión del Dr. H.R.Y. So, y el Dr. Blas Payri. El resultado es el documento "Design of an experiment to study the influence of dynamic cues on binaural listening localization accuracy. Application to film sound postproduction". Incluido como anexo en el CD de entrega.

REVISIÓN TEÓRICA

2. REVISIÓN TEÓRICA

2.1. ANTECEDENTES.

2.1.1. Audición y escucha.

Los humanos tenemos dos oídos, con una envoltura externa irregular llamada oído externo, pabellón auricular o pina, que son los órganos sensoriales encargados de la *audición* o percepción auditiva. A los oídos puede llegar sonido emitido por una fuente (*sonido monofónico*) o por dos o más fuentes (*sonido estereofónico*). Independientemente del número de fuentes que intervengan en la emisión, el sonido puede ser recibido por un solo oído (*audición monoaural*) o por los dos (*audición binaural*).

En el caso común de *audición binaural*, en que ambos oídos reciben sonido, se da el caso, de que debido a su distinta ubicación en el espacio, se presentan en el sonido percibido algunas leves diferencias denominadas *diferencias interaurales*, que son el principal mecanismo utilizado por el sistema auditivo, junto con algunos otros mecanismos monoaurales y cognitivos para discernir la localización de la o las fuentes emisoras en el espacio tridimensional. Podemos llamar a esta habilidad *audición espacial u holofónica* (del griego "hólos", todo).

El término *escucha*: "acción o acto de prestar atención a lo que se oye", tiene en el ámbito audiovisual otra acepción común, aplicada al método o soporte que limita la propia escucha, y que depende del método para grabar-reproducir el sonido finalmente escuchado, tenemos entonces *escucha monofónica* si el sonido proviene de una fuente o fue grabado por un solo captador de sonido, o *escucha estereofónica* si fue captada y además reproducida por más de uno.

El mismo criterio debe ser aplicado en el caso de la captación humana de sonido y su posterior reproducción. Por tanto, si recreamos el hecho de captar sonido por solo un oído tenemos una escucha *monoaural*; y si recreamos el sonido captado por los dos oídos tenemos una *escucha binaural*, que permite al receptor escuchar con cualidades espaciales u holofónicas.

En este texto nos referiremos a esta condición como **escucha binaural**.

2.1.2. Audición espacial. Las Diferencias Interaurales

El sistema auditivo humano utiliza una compleja combinación de mecanismos o indicios para determinar la localización de una fuente sonora en una determinada posición angular. Los indicios más efectivos tienen relación con las diferencias relativas en las características de la onda acústica que alcanza los dos oídos. La ubicación de los oídos a la misma altura en el plano horizontal (Figura 2.1), maximiza las diferencias para eventos sonoros que ocurren alrededor del oyente, respecto de las que ocurren arriba o abajo. Desde un punto de vista evolutivo, este hecho está combinado con el de la disposición física de los ojos, “diseñada” en la zona frontal para maximizar la percepción de la profundidad y la distancia (visión estereoscópica), pero dejando descuidada la percepción periférica, quedando ésta como responsabilidad del sistema auditivo, a través de la escucha de sonidos a ras de suelo fuera del campo visual.

Esto, junto con un mecanismo de adquisición altamente especializado en la inmediatez, con un periodo de integración de entre 150 y 300 milisegundos sobre bandas críticas (Zwicker, 2007), y la imposibilidad de omitir la percepción a voluntad, hace que podamos describir al oído como *el sentido de la consciencia y la alerta*.

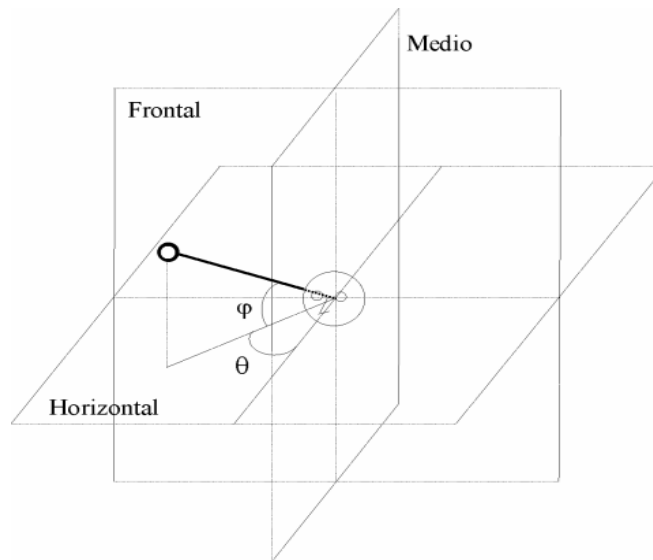


FIGURA 2.1. SISTEMA DE EJES Y PLANOS DE USO COMÚN Y UTILIZADO A LO LARGO DE ESTE TRABAJO. LA VARIACIÓN ANGULAR EN EL EJE HORIZONTAL SE DENOMINA AZIMUT (θ) Y EN EL EJE VERTICAL ELEVACIÓN (ϕ). EL PLANO FRONTAL DIVIDE EL ESPACIO EN DELANTE Y DETRÁS, Y EL PLANO MEDIO EN IZQUIERDA Y DERECHA. LOS PLANOS HORIZONTAL Y FRONTAL ATRAVIESAN LOS OÍDOS, INTERSECIÁNDOSE EN EL DENOMINADO EJE AURAL.

La **teoría dúplex** de Rayleigh fue la primera, hace entorno a un siglo, en sugerir que las diferencias interaurales de tiempo (*ITD*) y las diferencias interaurales de nivel (*ILD*), son los principales indicios que utiliza el sistema auditivo para discriminar direcciones de procedencia.

2.1.3. *ITD* e *ILD*

Las diferencias interaurales de tiempo se basan en que cualquier sonido alcanza cada oído con una diferencia de temporal debida a la diferencia de caminos a un oído y otro (Figura 2.2). Esa diferencia de tiempo es utilizada por el sistema auditivo para calcular la dirección de la que proviene ese sonido.

Esta hipótesis funciona correctamente mientras la diferencia de fase es tal que la distancia entre máximos del frente de onda es mas pequeña que la distancia que separa ambos oídos, es decir, la longitud del perímetro de la cabeza entre éstos (una diferencia de aproximadamente 0.8 a 0.65 milisegundos). Por encima de ese valor, una longitud de onda se vuelve más pequeña que el tamaño de la cabeza, con lo que la información de fase relativa al tiempo de llegada a cada oído deja de ser significativa para transmitir la información sobre cuál es el

frente de onda adelantado (Begault, 2000). De aquí se puede extraer que las ITD son utilizables principalmente para frecuencias por debajo entre 1 y 1.5 KHz aproximadamente.

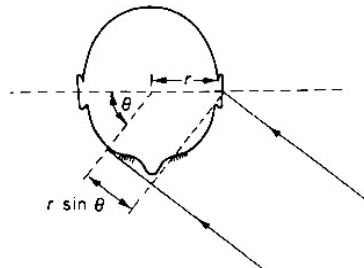


FIGURA 2.2. DIFERENCIA DE CAMINOS ENTRE LOS DOS OÍDOS EN AZIMUT (PLANO HORIZONTAL).
(BLAUERT)

Para frecuencias más altas el mecanismo sincronización es más complejo, dado que dos puntos de frentes de ondas sucesivos pueden llegar a ambos oídos con un mismo valor de fase. La estrategia consiste en detectar diferencias de amplitud en las envolventes de la onda, como durante el transitorio inicial, que se puede considerar como un pulso de frecuencia muy inferior modulando en amplitud la señal original. El sistema auditivo extrae la información de la envolvente global de amplitud de las componentes de alta frecuencia en ambos oídos y mide la diferencia en el tiempo de llegada para cada una de las dos envolventes, de esta manera consigue diferenciar cual es el frente de onda que llega antes (Figura 2.3). Se suele hacer referencia a este tipo de indicios como *diferencias interaurales de envolvente de ITD* o IED (*ITD interaural envelope difeffences*)

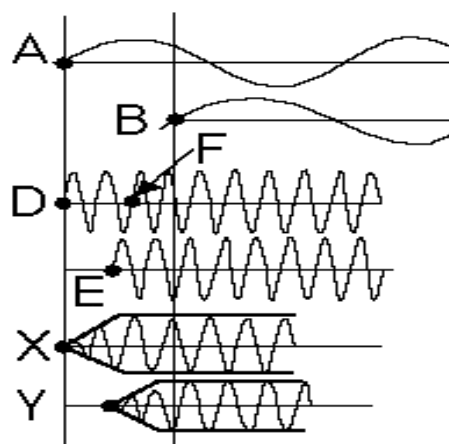


FIGURA 2.3. ITD Y ENVOLVENTE DE ITD (BEGAULT, 2000). RESULTA AMBIGUO SI E PRECEDE A D O A F, SIN EMBARGO X PRECEDE A Y INEQUÍVOCAMENTE.
(BEGAULT)

Las **diferencias interaurales de nivel**, por otro lado, utilizan las diferencias en intensidad de la señal en ambos oídos para discriminar las direcciones de sonidos entrantes. El sonido suena más fuerte en el oído más próximo a la fuente de sonido, no tanto por la distancia entre oídos sino por el denominado **efecto sombra** (*head shadow effect*). Este efecto se incrementa cuanto más alta es la frecuencia (o inversamente proporcional a la longitud de onda) resultando en niveles relativamente altos de atenuación: de 10 dB en 3 KHz para 90 grados de azimut; hasta unos 35 dB en 10 KHz (Begault, 2000). Por debajo de 1-1.5 KHz aproximadamente, las ILDs dejan de ser relevantes como indicio de localización espacial natural porque las ondas comienzan a difractarse alrededor de la cabeza en lugar de reflejarse, evitando el efecto sombra.

Las ITDs e ILDs se consideran los indicios más importantes para la localización espacial, y cada uno de ellos tiene su específico rango de operación en el mecanismo global de localización. Las ITD prevalecen a frecuencias más bajas, hasta la mencionada frecuencia de corte. Además dominan a las ILDs siempre que el contenido en bajas frecuencias sea suficiente, como por ejemplo en señales de banda ancha (Wightman & Kistler, 1991). Las ILDs son significativas en ausencia de bajas frecuencias y son suplidas por las ITD a frecuencias bajas, especialmente para señales de banda estrecha (Middlebrooks & Green, 1991). Esta complementariedad funcional se suele denominar teorías *time-intensity trading* (negociación tiempo-intensidad).

El Cono de confusión.

Si tenemos en cuenta solo las ITDs y las ILDs, que dependen básicamente de la distancia de una fuente a cada oído, hay un cierto número de puntos a la misma distancia de ambos oídos que deriva en una forma cónica con foco en cada oído (Figura 2.4). Es el denominado **cono de confusión**; debido a la imposibilidad de las ITDs y las ILDs para desambiguar el origen de la fuente. Esta indeterminación deriva en las denominadas confusiones delante-detrás o frontal-posterior (también conocidas como inversiones) y confusiones arriba-abajo.

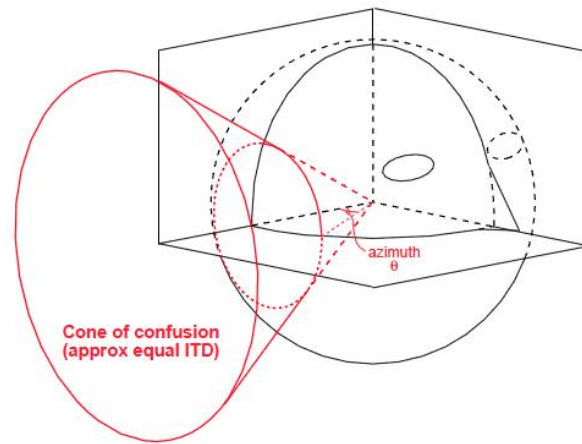


FIGURA 2.4. CONO DE CONFUSIÓN (MANDEL)

2.1.4. Indicios Espectrales

De lo dicho hasta aquí podemos extraer que como en el cono de confusión no existen diferencias interaurales, debe haber otro tipo de características que aporten información a cerca de la diferencia entre fuentes procedentes de la parte frontal, posterior y superior.

La forma del torso, cabeza y el oído externo, produce variaciones espectrales basadas en los fenómenos de reflexión, difracción, sombra, interferencia y resonancia, aportando la información necesaria para resolver en parte las ambigüedades previamente mencionadas. Estas son los denominados **indicios espectrales** o **direccionales**, ya que según la dirección de procedencia, el tipo de variación espectral será distinto.

Se considera que estos indicios funcionan de forma monoaural a modo de filtros que dependen del ángulo de incidencia y la distancia en cada oído por separado, con menos peso que diferencias interaurales de tiempo o de intensidad (Middlebrooks & Green, 1991).

Algunos experimentos han demostrado que el efecto del torso y hombros amplifica los sonidos frontales respecto de los traseros aproximadamente entre 250 y 599 Hz, particularmente en el plano medio y los atenúa entre 800 y 1200

Hz. Resonancias en el canal auditivo estimulan bandas de frecuencia en 2.5, 7 y 14.5 KHz y atenúan en 3.2,10 y 18 (Begault, 2000).

Blauert fue mas lejos mostrando a partir de fuentes sinusoidales y otras de tipo banda estrecha, que las posiciones de los *eventos auditivos* se forman en máximos localizados localmente, llamados **bandas direccionales** (Blauert, 1996, 2005). El espectro de las señales en los dos oídos es atenuado o estimulado de forma característica para estas bandas concretas para cada ángulo de incidencia, debido a los efectos de difracción y dispersión en la cabeza y las orejas. Esto implica que la percepción de la dirección de procedencia podría depender en gran medida contenido espectral del sonido incidente.

Perceived location	Center frequency kHz	Bandwidth kHz
overhead	8	4
forward (band #1)	0.4	0.2
forward (band #2)	4	3
rear (band #1)	1	1
rear (band #2)	12	4

TABLA 2.1. BANDAS DIRECCIONALES SEGÚN BLAUERT

Bateau (1967) elaboró un modelo para las reflexiones producidas en las rugosidades del oído externo, que provocan retardos entre el sonido directo y el reflejado en ellas, produciendo patrones de refuerzo y cancelación en función del ángulo de incidencia.

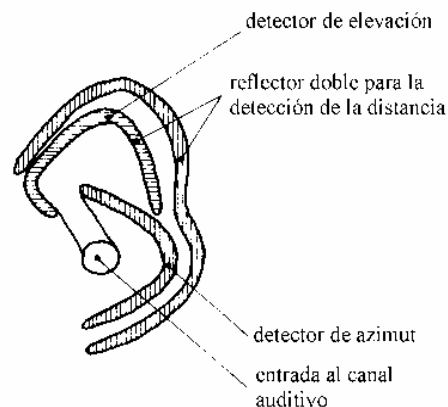


FIGURA 2.5. FUNCIONES DEL PAVELLÓN AUDITIVO SEGÚN BATEAU (EN CETTA).

2.1.5. Indicios Dinámicos.

En la percepción cotidiana, es común encontrarse fuentes en movimiento, o en su defecto, solemos mover la cabeza voluntaria o involuntariamente para apoyar el sentido visual con el fin de adquirir referencias de la fuente sonora. Por lo tanto, en el caso general, si movemos nuestra cabeza hacia la fuente para visualizarla, además conseguimos minimizar las diferencias interaurales (de tiempo y de intensidad), con lo que la indeterminación disminuye ya que para ángulos frontales la precisión en la localización y la localización borrosa (*localization blur* en Blauert, 1996) es mejor. El sistema cognitivo aprende de alguna manera cómo interpretar estas variaciones como pistas en coordinación con las visuales.

Movimiento de la cabeza.

Existen tres movimientos básicos de la cabeza: el movimiento de **rotación** a lo largo del plano medio, para localizar movimientos en azimut; el movimiento de **ladeo** o inclinación de la cabeza a lo largo del plano frontal para localizar movimientos encima y debajo del plano horizontal, y el movimiento de **cabeceo** o rotación del plano auditivo, que no implica ningún desplazamiento de ejes por lo que ninguna característica interaural varía, pero sí la influencia espectral del pabellón auditivo.

De esta manera, los movimientos de rotación de la cabeza (a lo largo del plano horizontal) contribuyen a desambiguar las confusiones delante-detrás y el movimiento de ladeo (a lo largo del plano medio) las confusiones arriba-abajo relacionadas con el cono de confusión.

La importancia de los movimientos de cabeza en la localización del sonido fue planteada por primera vez por Thurlow (1967), quién fotografió sujetos mientras localizaban sonidos de 5 segundos de duración, y descubrió que el movimiento de rotación (acimutalmente) es el más importante, especialmente para bajas frecuencias. También descubrió que incluso movimientos de cabeza inducidos involuntariamente a los sujetos tienen una gran influencia positiva en

la localización. Sin embargo, también observó grandes diferencias entre unos individuos y otros respecto de la máxima longitud de los movimientos.

Consideraciones acerca del Movimiento.

En el caso estático, el cambio más pequeño detectable en la posición angular de un evento auditivo debido a cambios en la posición de la fuente (*localización borrosa* según Blauert; o *mínimo ángulo audible*, según Begault), se ha medido en torno a 1 grado para la mejor situación, en la dirección frontal (Tabla 2.2).

Reference	Type of signal	Localization blur (approximate)
Klemm (1920)	Impulses (clicks)	0.75°–2°
King and Laird (1930)	Impulse (click) train	1.6°
Stevens and Newman (1936)	Sinusoids	4.4°
Schmidt et al. (1953)	Sinusoids	> 1°
Sandel et al. (1955)	Sinusoids	1.1°–4.0°
Mills (1958)	Sinusoids	1.0°–3.1°
Stiller (1960)	Narrow-band noise, cos ² tone bursts	1.4°–2.8°
Boerger (1965a)	Gaussian tone bursts	0.8°–3.3°
Gardner (1968a)	Speech	0.9°
Perrott (1969)	Tone bursts with differing onset and decay times and frequencies	1.8°–11.8°
Blauert (1970b)	Speech	1.5°
Haustein and Schirmer (1970)	Broadband noise	3.2°

TABLA 2.2. RESOLUCION DE LOCALIZACIÓN PARA LAS DIRECCIONES FRONTALES (BLAUERT, 1996)

De la misma manera, se ha medido el *mínimo ángulo de movimiento audible*, entorno a 3 grados (Strybel, Manlias & Perrott, 1992; en Begault), ángulo que se incrementa con la velocidad de rotación desde 2 a 8 grados por segundo hasta un máximo de 1000 grados por segundo (Féron, 2010), hasta el cual somos capaces de seguir un movimiento en trayectoria continua, y a partir del cual el tiempo de integración del sistema auditivo se excede (Begault 2000). Este ángulo mínimo también depende del área en la que la fuente esta ubicada, que normalmente es mas pequeño para sonidos de banda ancha y en la dirección centro-frontal.

La importancia de los indicios dinámicos.

Los indicios dinámicos han sido dignos de consideración dentro del estudio de la escucha espacial desde prácticamente sus principios. Young, en 1931 sugirió que eran mas importantes que los indicios espectrales tras reemplazar el pabellón auricular de varios sujetos por un embudo y comprobar que los movimientos de cabeza podían compensar la ausencia de esa pina individual (en Makensen, 2004).

Wallace planteó en 1939, tras un experimento con fuentes móviles, que la localización de sonidos no solo depende del “factor pina”, y que los movimientos de cabeza son mas efectivos en la localización de sonido, especialmente para sonidos con ángulos en azimut y elevación (Wallace, 1939, 1940).

Esta hipótesis fue posteriormente probada por wightman & Kistler (1999) mostrando que los indicios dinámicos reducen la frecuencia de las confusiones delante-detrás incluso para aquellos oyentes que registran un gran numero de ellas. Es más, sugieren que no se requieren específicamente movimientos de cabeza para producir los indicios dinámicos y lo demuestran en un experimento en que la cabeza de los sujetos está fija, y un altavoz se mueve. Sin embargo, el experimento solo fue totalmente exitoso cuando al oyente se le permitía mover el altavoz voluntariamente. Aparentemente, el oyente debe ser consciente de la dirección de movimiento para extraer la información necesaria.

Podemos decir, por lo tanto, que los movimientos de cabeza producen un cambio en la posición del evento auditivo, que es equivalente al cambio relativo en la posición de la fuente (i.e. un movimiento azimutal dextrógiro u horario de la cabeza equivale a un movimiento relativo levógiro o anti horario en la posición de la fuente en las mismas circunstancias). En general, esta afirmación debe ser matizada, por ejemplo, cuando tenemos ambos movimientos, de cabeza y fuente. Boer & Urk (1941, en Makensen), presentaron a través de auriculares un sonido grabado con un dummy esférico que se hacía rotar en ambos sentidos a horario y anti horario. Los sujetos fueron capaces de localizar la fuente correctamente cuando movían sus cabezas en la misma dirección que el dummy, sin embargo, si se movían en sentido opuesto, incurrían en confusiones delante- detrás.

De hecho, los movimientos de cabeza implican el uso de otro mecanismo que determina la orientación de la cabeza o un cambio en esta: **el órgano vestibular o del balance**, responsable del sentido del equilibrio, que de acuerdo con Blauert (1996) utiliza diferentes mecanismos para detectar inmovilidad (a través de gravedad) o posición de la cabeza y cambios en el movimiento (a través de fuerzas centrífugas). Por lo tanto, es razonable pensar que este mecanismo vestibular también forma, de alguna manera, parte del sistema auditivo y se puede considerar como indicio binaural incluido dentro del grupo de indicios dinámicos.

Los indicios dinámicos pueden solucionar problemas de localización

El uso de indicios dinámicos ha sido planteado y demostrado como solución de los problemas propios de localización dentro de la cabeza y confusiones delante-detrás. Los experimentos más destacables se han llevado a cabo mediante técnicas de *head-tracking* (seguimiento de cabeza).

Se ha planteado, mediante la teoría de los indicios vestibulares que el movimiento relativo entre cabeza y fuente requiere cierto conocimiento de las características del movimiento para que éste sea efectivo. Por un lado, el uso de auriculares conlleva que los movimientos de cabeza no concuerden con los movimientos de la fuente por lo que el mundo visual pierde su correlación con el auditivo. Por otro lado, un movimiento virtual de la cabeza (como en grabaciones con cabeza de maniquí), sin ninguna otra referencia, no pueden ser interpretadas correctamente. Los sistemas de head-tracking sincronizan la cabeza del oyente con la del maniquí para conseguir grabaciones con movimiento correlado entre cabeza-oyente.

Entre numerosos experimentos que muestran la efectividad de una fuente móvil de esta manera son, por ejemplo, Mackensen (2004), que testeó un sistema de tracking con cabeza de dummy, y encontró que las confusiones delante-detrás (que aparecían sobre todo en el rango de -30° a 30°), desaparecían casi completamente cuando el sistema de seguimiento estaba habilitado, lo cual es un resultado comparable con el caso de movimiento libre voluntario. Esto se

cumplía incluso con un dummy de forma esférica sin características de tipo pina-HRTF incluidas.

Hay otro resultado interesante llevado a cabo por Wersényi (2009) usando interpolación para simular una fuente móvil a lo a través de movimientos angulares reducidos (1° - 7°) a lo largo de la posición centro-frontal. Movimientos de 1° - 2° se mostraban efectivos para resolver problemas de internalización en un 21% de los sujetos; y solo en un 14% para inversiones. Para estos pequeños ángulos no se detectó consciencia de movimiento por debajo de movimientos angulares de 4° .

Wenzel (1995), mostro que con movimiento de cabeza, al contrario que en la situación estática, hacemos uso de la variación de ILD para resolver confusiones delante-detrás de una manera más efectiva que las ITDs, probablemente porque son más consistentes en el tiempo, lo que sugiere que los sonidos con contenido en altas frecuencias serían favorecidos.

Wightman & Kistler (1999), como ya se ha mencionado, sugirieron que los indicios dinámicos necesarios para desambiguar una fuente en posición frontal de una trasera no es necesario que sean debidos a un movimiento del sujeto pero es necesario que éste sea consciente del movimiento de alguna manera (indicios vestibulares). En el mismo informe sugieren además que algunos oyentes no cometen confusiones delante-detrás con fuentes reales (sonidos ecológicos) y en general el ratio de éstas es mayor que con fuentes no “reales” incluso sin movimiento.

2.1.6. Indicios Adicionales.

Aparte de los mencionados ITD, ILD e indicios espectrales (también conocidos como direccionales o HRTF), e indicios dinámicos (incluyendo vestibulares-orientacionales), existen algunos otros con cierta repercusión en la percepción, especialmente porque algunos pueden superponerse a otros, haciendo difícil definir un modelo. *Indicios visuales*, memoria, consciencia o

aprendizaje consciente e inconsciente a lo largo del tiempo, tienen una relevancia relativa, y un no del todo conocido rol e interinfluencia mutua.

Makensen (2004) sugirió que todos estos indicios auditivos interactúan conjuntamente, por lo que para ser suficientemente preciso, todo experimento tendería que incluir tantas como fuese posible.

Han (1992, en Makensen) sugirió, que movimientos de la fuente no influyen en la resolución de inversiones delante-detrás una vez que la posición y la fuente hayan sido identificadas y memorizadas. En este experimento con dos altavoces, frontal y trasero, los sujetos memorizaban la posición de alguno de ellos identificándolo visualmente. Posteriormente, en un test a ciegas, fueron incapaces de distinguir trasero y delantero, incluso con movimiento de estos, lo que implica que la información memorizada fue mas poderosa que el resto de los indicios.

Esto esta relacionado con el hecho de que tendemos a elegir como localización de la fuente a la localización racional visiblemente más cercana, lo que Gardner llamó **efecto de proximidad de la imagen** (*the proximity image effect*, Gardner 1967,1972) especialmente para ambientes poco reverberantes.

Blauert relacionó la sensación de movimiento con el órgano vestibular, y Begault considera que los indicios de tipo cognitivo son una importante parte del mecanismo de la audición, como puede ser el **efecto de desplazamiento doppler** (*doppler shift effect*), que explica cómo detectamos fuentes en movimiento aproximándose o alejándose de nosotros, una vez que el sistema cognitivo aprende como funciona este efecto.

Se puede hablar también de **indicios ambientales**. El hecho es, que las reflexiones tempranas en una sala no anecoica puede actuar solucionando algunas ambigüedades del cono de confusión una vez que aprendemos a interpretar sonidos secuenciados similares cuyos frentes de onda alcanzan cada oído con diferentes direcciones de incidencia. Sin embargo, en algunos casos la información ambiental puede aportar todavía más confusión, como en el caso de que el campo reverberante sea excesivo.

Las diferencias interaurales dependen también de la distancia a la fuente. Para distancias cortas el frente de onda de la señal acústica puede considerarse como esférico, mientras que a medida que las distancias se hacen mayores, el frente de onda comienza a comportarse como aproximadamente plano. Estas diferencias en las características de la señal acústica entrante en función de la distancia a la fuente, han sido definidas bajo el término de ***paralaje auditivo*** (*auditory parallax*) por Blauert (1996).

Dado que las grabaciones hechas para bases de datos HRTDF son típicamente a 1 metro de distancia, grabaciones binaurales de campo con cabeza de maniquí o con micros son necesarias para comprobar este tipo de fenómeno.

2.2. HERRAMIENTAS DE ESPACIALIZACIÓN DE AUDIO.

2.2.1. Herramientas de espacialización clásicas.

Existen diversas herramientas de posproducción de audio que permiten el procesamiento de señales con diversos propósitos estéticos o narrativos.

La estereofonía fue la primera aproximación con funciones de espacializar fuentes de sonido creando un valor narrativo (además de una mejora en la inteligibilidad). La separación de fuentes en el espacio virtual de la imagen estéreo permite un mejor reconocimiento y menor enmascaramiento entre las fuentes. Una serie de métodos fueron desarrollados para enriquecer a percepción sonora mediante esta vía, haciendo la imagen global mas amplia e inmersiva.

Una de las técnicas tradicionalmente utilizadas para localizar fuentes monofónicas en una imagen estéreo es la **panoramización de nivel** (derivada del clásico potenciómetro panoramizador o *panpot*, típico de las mesas de mezclas de audio). Esta técnica utiliza diferencias interaurales (estereofónicas) de nivel entre los dos canales de salida estéreo para localizar la fuente monofónica a lo largo de la línea recta que une los dos altavoces estéreo (en caso de auriculares esta línea une ambos oídos). Esta técnica puede ser considerada como un modo artificial y simplificado de crear ILD, pero de forma homogénea en frecuencia. A pesar de esto, el método suele funcionar bien debido que la mayoría de los sonidos naturales tienen suficiente contenido por encima de la región de corte entre ITD e ILD (1-1.5 KHz), y el oído es sensible a los indicios ILD para lateralización a lo largo de la mayoría del rango de frecuencias audibles desde más o menos 200 Hz (Blauert, 1996).

La otra técnica utilizada para localizar espacialmente una señal monofónica es la **traslación de fase** (*phase shift*), que consiste en aplicar retardos a la señal en un oído con relación al otro, haciendo uso del fenómeno,

denominado de diversas formas según autores, como *efecto precedencia* (Wallach, 1939), *efecto Haas* (en honor a Haas) o *ley del primer frente de onda* (Blauert, 1996). Aplicando un retardo artificial entre canales el efecto auditivo percibido es el representado en la Figura 2.6 :

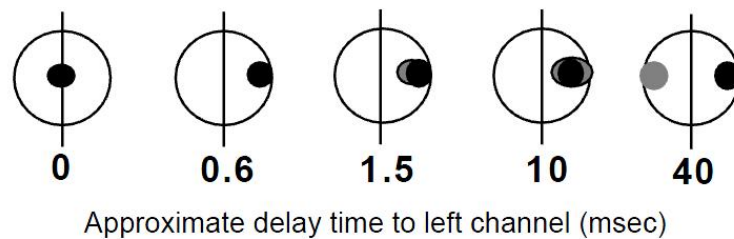


FIGURA 2.6. EFECTO PERCEPTUAL AL INCREMENTAR ITD ENTRE 0 Y 40 MILISEGUNDOS. (BEGAULT)

- Para retardos temporales desde 0 a 0,6 milisegundos, la imagen estéreo se traslada a lo largo de eje que una los altavoces (entre los dos oídos en caso de escucha con auriculares) como predice la teoría dúplex.
- Hasta 1.5 milisegundos se produce un ligero aumento en la anchura espacial percibida de la imagen estereofónica.
- Incrementando el retardo sigue incrementándose la anchura y la coloración del sonido, hasta que a retardos mayores de 10 ms el “centro de gravedad” de la fuente comienza a moverse de nuevo hacia el centro de la imagen virtual.
- A partir de retardos de aproximadamente 40 ms, por encima del umbral de eco dependiendo del tipo de fuente, el evento auditivo se empieza a separar en dos imágenes puntuales.

Ambas técnicas mencionadas utilizan los principales indicios binaurales utilizados en la escucha espacial (ITDs e ILDs) para simular audición espacial simple a partir de señales monofónicas. Podemos ver estos métodos como una aproximación, ya no pueden reemplazar totalmente el efecto de las diferencias

interaurales de tiempo y de nivel que dependen de características físicas y fisiológicas propias de cada individuo.

Por lo tanto, podemos pensar que estas aproximaciones clásicas “elementales” pueden ser mejoradas, usando otras propiedades añadidas de la escucha humana binaural, como las espectrales o las relativas al movimiento.

2.2.2. Procesado HRTF

El camino entre una fuente sonora y el oído interno, se ve afectado por una serie de obstáculos que dependen de la constitución física de cada individuo, y que pueden ser consideradas como un filtrado espectral denominado *función de transferencia derivada de la cabeza* (*head-related transfer function* o **HRTF**) que se define como el ratio entre la transformada de Fourier de la presión sonora que llega al oído y la equivalente transformada en el punto localizado en el centro de la cabeza del oyente con ese oyente ausente.

Se puede medir la HRTF como conjunto de respuestas impulsivas capturadas mediante pequeños micrófonos de prueba situados a la entrada del canal auditivo del sujeto, o simulando éste mediante una cabeza de maniquí, obteniendo de esta forma la HRIR (*respuesta impulsiva derivada de la cabeza*), que se puede ver como la transformada de Fourier en el tiempo de la HRTF. Tomando medidas para un número discreto de puntos que cubra el espacio esférico tridimensional, y teniendo en cuenta la resolución espacial del oído, tendríamos un conjunto de respuestas impulsivas. Si aplicamos una *convolución* entre estas y cualquier fuente monofónica sin información espacial asociada, podemos transformar esa fuente en una fuente virtual tridimensional, localizada teóricamente en cada punto del espacio que queramos.

2.2.3. HRTFs no individualizadas. Confusiones.

En la mayoría de los casos, la toma de medidas de respuestas impulsivas individuales de tipo HRTF no es práctica, debido a la alta tecnología de equipamiento necesaria, la cantidad de tiempo que se necesita y la localización especial en donde debe ser realizada. Esto hace que se busque una solución de compromiso, realizando promedios de medidas antropométricas y registrando muestras de HRTF para un grupo de individuos promedio o construyendo un dummy head a partir de ellas

Como se mencionó previamente, los indicios espectrales ayudan a desambiguar el cono de confusión, pero son altamente dependientes de las características físicas antropométricas de cada sujeto, derivando en importantes diferencias en la respuesta en frecuencia individual para cada uno.

La utilización de HRTF no individualizadas, debido a estas diferencias individuales conlleva inexactitudes en la representación de sonidos filtrados son principalmente las conocidas en la literatura como **localización dentro de la cabeza** (*inside-the-head locations*) también denominadas **internalizaciones**, y las **confusiones delante-detrás** (*front-back confusions*) también denominadas **inversiones**.

Aunque las confusiones delante-detrás pueden ocurrir en la escucha cotidiana, la proporción se incrementa para la escucha con auriculares, especialmente con HRTFs no individualizadas, la Tabla 2.3 muestra un resumen histórico de resultados recogidos por Begault (2000), para presentaciones en campo libre y auriculares.

Study	Free-field reversals	Virtual source reversals	Ratio B-F to F-B reversals
Burger (1958); one-octave noise bands	20%		
Laws (1974); 0° target only	0	35%	n/a
Oldfield and Parker (1984a, b)	3.4% (12.5% monaural listening; 26% pinnae occluded)	n/a	
Wightman and Kistler (1989b); user's own HRTF	6%	11%	2:1
Makous and Middlebrooks (1990)	6%	n/a	
Asano <i>et al.</i> (1990); 2 subjects	n/a	subject 1: 1.9% with own HRTF; 0.9% with other subject's HRTF subject 2: 7.4% with own HRTF; 14.8% with other subject's HRTF	
Wenzel <i>et al.</i> (1993)	19%	31%	free-field: 8.5:1 virtual: 4.0:1
Begault and Wenzel (1993); speech	n/a	37%	3:1

TABLA 2.3. PORCENTAJE DE CONFUSIONES DELANTE-DETRÁS PARA ESCUCHA EN CAMPO LIBRE, AURICULARES (FUENTES VIRTUALES) Y RATIO DE CONFUSIONES DELANTE-DETRÁS Y DETRÁS-DELANTE. LAS FUENTES SON RUIDO DE BANDA ANCHA EXCEPTO EN BURGER (RUIDO DE BANDA DE OCTAVA) Y WENZEL (VOZ). (EN BEGAULT, 2000).

El porcentaje de inversiones de delante a detrás es mayor que la de detrás a delante, especialmente cuanto mas ancho es el espectro de la fuente, con peores resultados en el caso de la voz.

Otras experiencias también sugieren que el porcentaje de confusiones es también mayor en el plano medio, para sonidos provenientes directamente del centro, aunque ésta es el área con mayor precisión en la percepción y es esta fuera del cono de confusión. El hecho es que en el plano medio la posición de la

fuente es simétrica respecto de ambos oídos, con lo que las ITD y las ILD no son útiles.

Los resultados de Asano (1990), en la *tabla 2.3*, tiene otra peculiaridad. Para el sujeto 1, el número de inversiones es menor con la HRTF de otro sujeto que con la suya propia. Respecto a esto, Begault (2000) sugiere que existen diferencias en las HRTF individuales que hacen pensar en “buenos y malos oyentes”. La cuestión subyacente es si es posible para el sistema cognitivo reaprender el uso de una HRTF de un “buen oyente” una vez reemplazada la propia.

El fenómeno de la *localización dentro de la cabeza* (también denominado internalización o lateralización según Begault), se refiere al caso especial en que la percepción espacial sitúa la fuente sonora dentro de la cabeza, hecho que normalmente se produce sobre auriculares. El hecho es que este fenómeno parece depender de las características de la fuente sonora y la manera en la cual ésta interactúa con la HRTF. Para trenes de ruido rosa impulsivo la externalización es relativamente fácil de conseguir, al contrario que por ejemplo, señales de voz (Wenzel et al, 1993). Algunos experimentos previos sugieren que la externalización puede ser alcanzada con la ayuda de indicios dinámicos e incluso con referencias visuales (Blauert, 1996).

Aplicación de HRTFs no individualizadas.

La aplicación de procesado HRTF en productos audiovisuales comerciales tiene que resolver el problema de la producción en masa para conseguir un mínimo estándar fiable. Respecto a esto se han seguido diferentes estrategias.

Seeber & Fastl (2003), desarrollaron un método rápido para la selección subjetiva de HRTF previamente a la experiencia audiovisual, que permite a un usuario, a través de escucha, acomodarse a la HRTF más próxima a la suya a partir de una base de datos dada, mediante una elección en dos pasos de refinamiento.

La teoría de Blauert de las **bandas reforzadas** (*boosted bands*) se puede aplicar mediante el análisis de la relación entre los máximos en la HRTF y el comportamiento predecible para las bandas direccionales. Un promediado de HRTF se puede originar también a partir del análisis tanto de características físicas como espectrales. El tamaño de la cabeza y pina se puede medir para un gran número de individuos y luego promediarse, o también examinar el espectro de un gran número de respuestas impulsivas a través de sus transformadas de Fourier y diseñar un “promedio espectral”. El promediado de HRTF y posterior elección de una HRTF promedio de acuerdo con su contenido espectral ha sido obtenido satisfactoriamente por So et al. (2010), reduciendo de forma efectiva el ratio de confusiones.

Bases de datos HRTF

Dado que el proceso de medida de un conjunto de HRTFs es relativamente costoso en tiempo y recursos. Una serie de proyectos han realizado estas medidas en forma de bases de datos de HRTF de libre distribución para propósitos de investigación. Estas bases de datos han sido concebidas para adaptarse a una población promedio.

- La base de datos MIT KEMAR (Gardner & Martin, 1994). Consiste en dos series de medidas (para dos tamaños distintos de pinas artificiales) de una cabeza de maniquí simétrica modelo KEMAR. La resolución máxima es 5° en el plano horizontal.
- La base de datos CIPIC (Algazi et al., 2001). Consiste en medidas de 45 sujetos a 25 azimuts diferentes y 50 elevaciones, en incrementos angulares de 5° aproximadamente. La base de datos también contiene medidas antropométricas de cada sujeto.
- La base de datos del IRCAM (2003). La configuración consta de 10 ángulos de elevación empezando en -45° y finalizando en +90°, en incrementos verticales de 15° y variables en la rotación sobre el eje medio (en vertical), desde 24 muestras a 0° hasta 1 sola muestra para 90° de elevación.

2.3. TECNOLOGIAS BASADAS EN FILTRADO HRTF

Se puede usar sonido envolvente en gran cantidad de aplicaciones. Las más vanguardistas desde el punto de vista tecnológico son las militares, dado el hecho de que la navegación aérea se ve mejorada añadiendo audio con contenido tridimensional al visual. Las aplicaciones medicas y la industria del entretenimiento son las más extendidas.

La industria audiovisual se caracteriza por una evolución continua. El sonido 3D tiene una aplicación obvia en videojuegos, cine, o comunicaciones móviles.

Uno de los inconvenientes del procesado HRTF es su alta demanda en cuanto a potencia de procesado. Esto ha llevado al desarrollo de tecnologías híbridas que suelen aplicar procesados más sencillos a los objetos sonoros.

EAX, de Creative Labs (Creative, en línea), dispone de la tecnología de simulación más avanzada para videojuegos con características como obstrucción de objetos, transformación del ambiente mediante *morphing*, etc. Es una tecnología derivada de **Aureal**, basada en procesado HRTF, y que hace uso de DirectSound3D, el componente aural de DirectX, cuya API incluye soporte para DirectSound3D mediante C++ para efecto doppler, atenuación con la distancia, etc. (Creative, 2001).

En la industria del cine, **Dolby Headphone**, basada en la tecnología precursora *Lake technologies*, incorpora procesado HRTF basado en DSP en un sistema que puede reproducir 7.1 canales surround en auriculares. El chip decodificador Dolby Headphone puede estar integrado en el reproductor de DVD, la TV, en un procesador de sobremesa o portátil, o incluso incorporado en unos auriculares “activos”. Normalmente están implementados con una serie de preprogramaciones o *presets* con diversas configuraciones acústicas para varios tipos de recinto (casa, sala de cine, etc.).

Las opciones “pasivas” han sido también implementadas a modo de bandas sonoras pre-codificadas para DVD, el primer DVD lanzado al mercado

que permitía esta opción fue Pearl Harbor: (“Pearl Harbor. 60th Anniversary commemorative Edition”, 2001).

SRS headphone (recurso en línea), Yamaha **silent cinema** y Smyth research son otras tecnologías basadas mas o menos en la aplicación de procesado HRTF.

La conversión de un flujo AC3 con 5.1 canales en binaural es posible de forma simplificada tomando el ángulo teórico para cada altavoz del sistema 5.1, y filtrando la HRTF para esa ángulo y posteriormente mezclando las señales resultantes en una mezcla estéreo de dos canales, transformada ya en binaural. Esta técnica ha sido probada satisfactoriamente como aproximación incluso más inversiva que la mezcla original AC3 5.1 (So et al, 2006).

Sin embargo, en una producción audiovisual, además de procesar el producto final, existe la posibilidad de aplicar el proceso de conversión a binaural en la etapa de diseño sonoro para cada sonido individualmente, permitiendo una especialización específica en cada caso, teniendo en cuenta las características de la muestra de audio y obteniendo probablemente mejores resultados finales.

Consideraciones sobre reproducción de sonido binaural

Como hemos comentado, el material sonoro con información binaural es muy sensible a las características sonoras que portan los indicios espectrales. Por lo tanto es necesario para la reproducción de dicho material eliminar en la medida de lo posible, las propias.

Esta condición se cumple en la escucha a través de auriculares, en la que los oídos reciben señales sonoras separadas, y funcionará mejor cuanto más fiable sea la respuesta en frecuencia de éstos y el aislamiento respecto de ruidos externos. De esta manera conseguimos, por así decirlo, reemplazar nuestros oídos por aquellos con los que a sido registrada la información binaural (Figura 2.7).

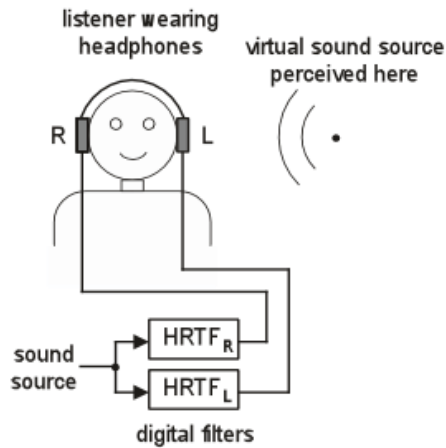


FIGURA 2.7. ESQUEMA DE LA CONFIGURACION IDÓNEA PARA ESCUCHA BINAURAL

La opción más novedosa e interesante en este sentido aplicada a la reproducción cinematográfica es la que utiliza el sistema **IMAX**, a través del sistema **PSE** (*Personal Sound Environment*). Es un sistema personal de sonido, diseñado para complementar el impacto visual de las películas en 3D, utilizando escucha binaural mediante la utilización de unos “cascos” equipados con dos **nearphones** (pequeños altavoces situados próximos a los oídos) (*Figura 2.8*). Permite situar sonidos en cualquier punto de la sala y cada espectador percibe una sensación similar, sin importar en que punto de la sala se encuentre, evitando el fenómeno de *sweet spot*. Suele equipar los cascos con gafas LCS para visión 3D. También se usa para que cada espectador seleccione el idioma deseado, ya que los *nearphones* no bloquean la entrada al oído posibilitando la utilización conjunta con los altavoces de sala.



FIGURA 2.8. AURICULARES IMAX PSE SIN GAFAS. Fot.: Sonics Associates

Reproducción a través de altavoces.

En reproducción a través de altavoces (estéreo, por ejemplo) una gran parte de la señal reproducida por el altavoz izquierdo alcanza el oído derecho y viceversa (*Figura 2.9*), produciéndose una interferencia cruzada denominada **diafonía** o *crosstalk*, que degrada por completo la escucha binaural.

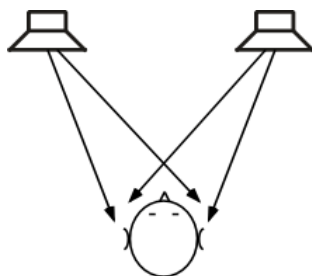


FIGURA 2.9. DIAFONÍA (CROSSTALK) PRODUCIDA POR DOS ALTAVOCES.

Existen diferentes métodos para minimizar este efecto cruzado, utilizando filtros complejos basados en líneas de retardo o incluso medidas de HRTF en para una posición concreta. La condición que debe cumplirse en la escucha es, por tanto, que el oyente se sitúe en una posición centrada respecto a los dos altavoces (*sweet spot*), que limita en gran medida la utilización de esta técnica, reportándose elevadas tasas de error de tipo frontal-posterior y de elevación (Wave Arts, 2007). Las confusiones más comunes suelen ser las de tipo de detrás hacia delante, lo que tendría como resultado que en una panoramización de 180 grados desde delante hacia un lado y luego hacia atrás, el resultado percibido sería en realidad de delante hacia el lado y hacia delante de nuevo.

Salvo para casos especiales como reproducción a través de computadora, donde se supone al oyente sentado y centrado en una posición bastante fija, este tipo de recursos no es efectivo, limitando la reproducción binaural al caso de auriculares.

DISEÑO EXPERIMENTAL

2. DISEÑO EXPERIMENTAL

2.1. OBJETIVOS

Desde el punto de vista de la posproducción de audio, es interesante conocer la efectividad de diversos métodos existentes para generar audio con características espaciales aprovechando las propiedades de la escucha binaural. También el estudio del comportamiento de distintos tipos de sonido de morfología variada, permite una elección más adecuada de éstos en función del objetivo buscado.

Además, como ya se ha mencionado, numerosos experimentos han mostrado que los movimientos relativos cabeza-fuente solucionan en cierta medida las inexactitudes en la representación de eventos auditivos de forma virtual. Se intentará simular fuentes en movimiento aprovechando la facilidad que proporcionan las herramientas de posproducción de audio, para observar su comportamiento.

Como los resultados son altamente dependientes de características individuales, se diseñará un test perceptivo cuyo propósito es obtener resultados sobre una muestra de sujetos. Se tratará de testear la mejora en los errores de localización para una serie de ángulos de localización a través de señales reales con diferentes características espectrales y usando como métodos de filtrado el entorno *matlab*, el software comercial proporcionado por *wave arts* y *logic audio* y el uso de micrófonos que permiten grabación binaural directa.

Hay numerosas consideraciones a tener en cuenta en el diseño del test, dada la complejidad en la interrelación de los diversos indicios utilizados por el sistema auditivo para la localización espacial.

2.2. ESTÍMULOS

2.2.1. Elección del material sonoro

Ya se ha mencionado la importancia del contenido frecuencial de las señales para su localización, debido a la elevada dependencia respecto de ITD e ILD y sobre todo de los indicios espectrales como sugieren, entre otras, la teoría dúplex o la teoría de las bandas direccionales de Blauert (1996).

Las características espectrales de la HRTF, en parte individualmente y en parte como promedio, ayudan a resolver ciertos conflictos entre ITDs e ILDs, y las debidas al cono de confusión. Como el mayor obstáculo en la síntesis con HRTF no individualizadas, son las diferencias interindividuales, es de suponer que para distintos tipos de sonidos el comportamiento será relativamente diferente.

Las características temporales son también relevantes, dado que el *indicio por envolvente de ITD* también toma información de los transitorios en intensidad de las señales. El incremento inicial en la envolvente de una señal se ha demostrado que reduce la precisión en la localización, mientras que para sonidos sin transitorios, con ataques abruptos, la localización mejora de forma continua con la densidad espectral de la fuente (Hartman, 1982).

Sonido sintetizado.

Normalmente, los test psicoacústicos utilizan como señal de referencia ruido blanco gaussiano de banda ancha en forma de impulsos, salvo para averiguaciones concretas acerca del comportamiento espectral donde se suele utilizar ruido blanco filtrado en bandas estratégicas.

Para la realización de este test se ha optado por elegir **ruido rosa** en lugar de ruido blanco. El ruido rosa es un ruido de tipo aleatorio que se diferencia del ruido blanco en la distribución espectral de energía. En el ruido rosa, la energía

se distribuye uniformemente según bandas de octava de manera que su distribución en frecuencia es constante en una escala logarítmica. El oído humano utiliza ese tipo de escala a través de bandas de octava como patrón de integración de señales en frecuencia, por lo que la percepción frecuencial será más fiable para bajas frecuencias.

Sonido ecológico

El análisis del comportamiento de señales del mundo real es claramente interesante desde el punto de vista de la posproducción de audio. Los sonidos ecológicos son más complejos de estudiar, ya que están muy asociados memorísticamente a imágenes y otros indicios de tipo racional que probablemente influyen en la localización sonora.

Al margen de este hecho, las características dinámicas y frecuenciales permiten elaborar una categorización de éstos de cara a su comportamiento experimental.

Por ejemplo, se puede esperar que en sonidos de **amplio espectro** y similares a ruido aleatorio, como podría ser ruido de mar, un río, lluvia o un avión, su comportamiento sea similar al del ruido blanco o rosa, y que se obtengan mejores resultados.

Por otro lado, sonidos de **banda estrecha** como un mosquito o un tono telefónico, deberían ser más conflictivos en términos de localización.

Algunos sonidos de tipo **impulsivo** de ataque rápido, como un disparo, un golpe o una puerta cerrándose, en tanto en cuanto son similares a una señal impulso Delta típica, con transitorios muy abruptos y contenido frecuencial continuo, se espera que se comporten mejor desde el punto de vista de la localización.

La **voz** es un caso particular dada su importancia en aplicaciones audiovisuales. En este tipo de señales, la mayor parte de la energía se concentra por debajo de los 7Khz, y la mayor parte del contenido que aporta información

sobre la fuente se sitúa en la zona indeterminada de corte entre ITDs e ILDs (1-3 KHz). Resultados de Begault y Gardner (en Begault, 2000), muestran que existe una alta variación en los errores de localización entre individuos especialmente para estas señales, aunque la presentación vía auriculares mejora ligeramente al caso de campo libre.

Para este experimento se ha elegido una muestra de *mar*, por ser la más similar a ruido rosa; una grabación *de puerta cerrándose*, como sonido impulsivo y dos muestras de voz, una masculina y otra femenina, que además tienen contenido fónico distinto.

3.2.2. Obtención del material sonoro.

El ruido rosa se ha obtenido a través de síntesis mediante el software de procesado *Sony soundforge*, de una muestra monofónica que se ha editado para una duración de 1 segundo con transitorios inicial y final de 50 ms para que pudiera ser comparable con la muestra *mar*.

Como muestra *mar* se ha aprovechado material grabado con los micros soundman, desechando uno de los canales para quedarnos con una señal monofónica. Se ha aplicado un transitorio inicial y final de 50 ms para evitar artificios.

La señal de voz femenina es una locución monofónica de la frase : “como me gusta el sonido diegético” de 1,8 segundos de duración, grabada en el estudio de radio de la EPSG de Gandía, en un ambiente que se puede considerar suficientemente anecoico. La frase resulta interesante como test, ya que tiene contenido vocálico y consonántico variado.

Grabación binaural. Micrófonos de campo Soundman OKM

Algunas técnicas de grabación fueron desarrolladas para propósitos audiovisuales tratando de simplificar el caso de las medidas reales de HRTFs. Por ejemplo, el soundman OKM utiliza micrófonos de coste reducido insertados en una carcasa de tipo auricular de botón, permitiendo un sistema de grabación portátil que registra al menos algunas de las propiedades tipo ITD e ILD y una aproximación a las propiedades espectrales. Una comparativa con respecto a una el sistema de grabación binaural de tipo cabeza dummy de Neumann no mostro excesivas diferencias (SAE-Middlesex Univ., 2009). Hay que tener en cuenta que todas las grabaciones que se hagan con estos micros llevaran impresas las características binaurales del individuo que los utilice, con lo que puede que no sean tan generalizables que las realizadas con dummies diseñados siguiendo criterios de promedio de medidas antropomórficas para uso general.



FIGURA 3.1. DETALLE DE MICROFONOS DE GRABACIÓN DE CAMPO SOUNDMAN OKM.

Para poder comparar muestras de material monofónico filtrado y material registrado directamente con información binaural, se han realizado dos grabaciones de campo estéreo utilizando los micrófonos Soundman, la primera para obtener una señal de voz que pudiéramos comparar con la voz monofónica, y la segunda un sonido de puerta cerrándose, que será la señal de contenido impulsivo que utilizemos.

La limitación que imponen las bases de datos HRTF es que las muestras están tomadas a un metro de distancia, con lo que se seguirá ese criterio para las grabaciones.

En el caso de la señal de voz, se han utilizado dos sujetos, situados de pie con las cabezas a la misma altura, uno de ellos con los micrófonos binaurales incorporados en la cavidad auditiva, y el otro situándose a un metro de distancia proyectando la misma frase en posiciones correspondientes a diferentes ángulos de interés. La frase proyectada es “babebibobu”, que tiene contenido vocálico diverso y una consonante bilabial con bastante contenido impulsivo. Una característica a tener en cuenta de esta señal es que la toma se ha realizado en el entorno en una sala de tamaño medio-largo con suelo de madera. Esto aporta características ligeramente reverberantes que pueden influenciar el resultado del test de alguna manera.

Se ha realizado una toma microfónica binaural también para el caso del sonido impulsivo, una puerta cerrándose, en este caso se ha realizado también una toma natural en un ambiente medianamente reverberante. Para la toma, el sujeto que porta los micrófonos binaurales se ha situado también a un metro de distancia y ha rotado su posición para obtener muestras de distintos ángulos de incidencia.

3.2.3. Elección del tipo de procesado

Existen aplicaciones software basadas en HRTF para edición y mezcla de audio para los principales programas de postproducción: *Binaural panner* para Apple Logic Audio (Apple, 2009) y *Panorama* (Wave Arts, 2007) instalable como plugin RTA y audio suite, dentro del paquete Power suite de Wave arts, siendo operativo tanto para Protools como para logic Audio.

Como punto intermedio tenemos el binaural panner desarrollado por la universidad UC Davis (UC Davis, recurso en línea), de libre distribución, ejecutable sobre el entorno Max/MSP, que permite aplicar filtrado HRTF de una forma relativamente sencilla manteniendo amplias posibilidades de control del

proceso de filtrado, pudiendo, por ejemplo, elegir entre cualquiera de las medidas realizadas a sujetos de la base de datos CIPIC.

A continuación se detalla información sobre el plugin *panorama 5* de *Wave Arts* (Wave Arts Inc., 2007). Se obvia entrar en detalle para el *binaural panner* de *logic audio* ya que no proporciona información tan detallada además de no aportar información acerca de su algoritmo de procesado ni la base de datos que utiliza. Este último provee una interfaz más simplificada y menos parámetros, pero suficiente para el propósito buscado. Se remite al manual de usuario de Apple logic audio 9 (Apple Inc., 2009), para más detalles.

PANORAMA 5

Se trata de un procesador de acústica virtual que utiliza tecnología de modelado acústico para simular la propagación de sonido en espacios reales. Utiliza algoritmos de síntesis basados en HRTF para recrear las características binaurales que nos permiten localizar los sonidos en el espacio, incluyendo procesados específicos para:

- Modelado de la distancia a la fuente.
- Modelado de cancelación cruzada interaural (*crosstalk cancelation*), que permite optimizar el procesado binaural para escucha mediante altavoces en lugar de auriculares.
- Efecto Doppler.
- Utilización de la base de datos HRTF medidos sobre el dummy head de tipo KEMAR realizada por el MIT y una muestra sobre individuos humanos realizada por el IRCAM.

Esta aplicación permite además simular características acústicas adicionales como son:

- Modelado de reflexiones tempranas (*early room reflections*)
- Reverberación difusa.
- Modelado de reflexión en superficies específicas.

En el siguiente diagrama de flujo (*Figura 3.2*) se representa el algoritmo utilizado por panorama para procesar las distintas características acústicas que permiten reproducir una señal sonora en una ubicación específica del espacio virtual.

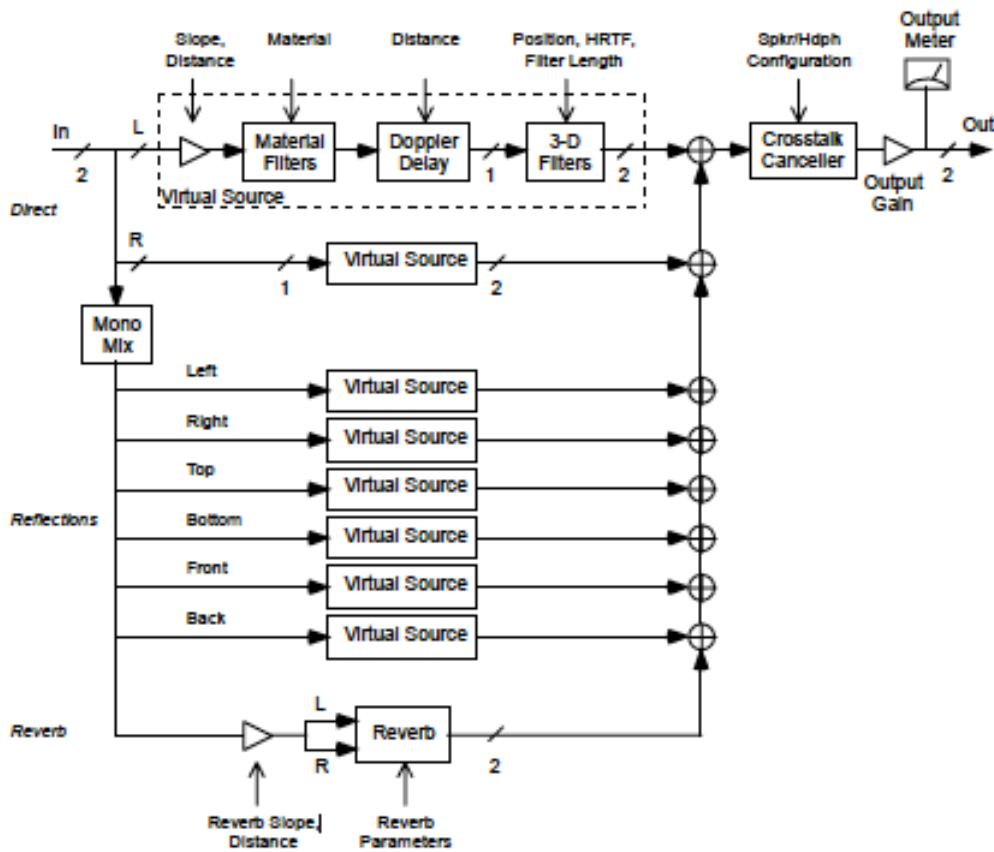


FIGURA 3.2. DIAGRAMA DE FLUJO DE PANORAMA 5.

A partir de una señal monofónica (denominada *virtual source*), o una señal estereofónica dividida en dos señales monofónicas se aplica un control de ganancia para simular la atenuación en función de la distancia. Seguidamente se aplica (en su caso), un filtro que simulan reflexiones sobre algún material. Después se sitúa una línea de retardo que simula la propagación en medio aéreo, incluyendo efecto doppler en caso de que se requiera, y por ultimo se aplica el filtro HRTF (3D en la figura), que añade el efecto de torso, cabeza y orejas, que resulta evidentemente en una señal estéreo con las características del entorno.

En el caso de que la señal de entrada sea estéreo, los dos canales resultantes se suman.

La simulación de reflexiones tempranas y reverberación se recrea añadiendo líneas adicionales.

Finalmente se puede eliminar la cancelación cruzada interaural, añadiendo un filtro para cada canal estéreo final. Se trata de un filtro basado en medidas de HRTFs reales. La señal resultante mejora la separación de canales en un promedio de 20 dB en el rango de 100Hz a 6 KHz para la mayoría de oyentes. (Wave Arts Inc., 2007)

La interfaz de usuario de panorama 5 permite la ubicación de fuentes a través de valores numéricos o una ventana gráfica que se puede mostrar en forma lineal o polar configuraciones de azimut y elevación. Las líneas de reflexiones y delay pueden activarse o desactivarse en función de la conveniencia o no de simular entornos reales. Existen controles para la ganancia de entrada y de salida, simulación de frente de onda plano o esférico y longitud del filtro empleado (permite elegir una solución de compromiso entre calidad/potencia de procesado).



FIGURA 3.3. INTERFAZ DE PANORAMA 5.

Filtrado mediante HRTF

A partir de los sonidos elegidos, se procederá a obtener una señal estéreo binaural mediante la aplicación de filtros HRIR a las señales originales. La introducción de características binaurales será más efectiva si partimos de señales monofónicas. En este caso, tenemos ruido rosa y una locución de voz femenina originalmente monofónicas, pero en el caso de la grabación de voz estéreo binaural que será procesada, se ha elegido una toma lo más representativa posible de la serie (el canal derecho para la toma correspondiente a 90°) y se ha convertido a monofónica. El mismo procedimiento se ha seguido para el sonido de puerta. De esta manera se puede aplicar un filtrado binaural, en este caso artificial, para compararlo con el caso “natural”.

Selección de las HRIR y filtrado con *matlab*.

La elección de la base de datos MIT KEMAR implica filtros tomados íntegramente a partir de una cabeza dummy. De hecho, las grabaciones fueron realizadas solo para uno de los hemisferios laterales (0° a 180°), asumiendo que la cabeza artificial era simétrica. Esto permite un testado promedio pero no necesariamente adecuado a las realidades individuales.

El uso de la base de datos IRCAM permite testear condiciones para diferentes sujetos, pero su elección implica cierta complejidad. Una simple inspección de trenes de ruido rosa filtrados con esta base de datos confirma la variabilidad de los resultados (IRCAM, recurso en línea). Dependiendo del filtro elegido llega a resultar imposible el reconocimiento de trayectorias angulares en dirección horaria o anti horaria, incluso para trayectorias relativamente largas. Además, esta base de datos tiene la particularidad de que la densidad de muestreo es variable, ya que en la zona central-posterior las muestras han sido tomadas en intervalos de 15° en base a que la precisión en la localización es menor en esta región, de manera que limitaría las posibilidades de filtrado para el caso de simulación de movimiento.

Se han tomado por tanto, las muestras de la base de datos del MIT, y se ha utilizado para realizar el filtrado el toolkit *Roomsim*, ejecutable en el entorno de programación matlab, y que incorpora una función que permite convolucionar dos señales en formato wav de forma sencilla y almacenar el archivo resultante en el mismo formato.

La base de datos de HRIR contiene muestras a las que se ha aplicado ecualización de campo difuso, para adecuar el entorno de grabación con el entorno de escucha por auriculares, evitando el error debido a la casi inexistencia propagación entre la fuente (muestras de cabeza dummy) y los auriculares.

Filtrado mediante software

Se ha realizado el procesado de cada muestra para dos tipos de software comercial: la versión 5 el plugin *panorama* de Wave Arts y el *binaural panner* incluido en Apple logic Audio 9. El entorno y los parámetros modificables son similares. Aunque el plugin de logic no incluye procesado acústico adicional, para este experimento solo nos interesa el caso de fuente directa lo más anecoica posible.

Una vez añadido el plugin y ajustado los parámetros, se ha exportado y se ha hecho una normalización sonido por sonido, ya al ser pasajes de distintas características dinámicas, aunque se igualen los niveles características de sonoridad para cada uno son distintas.



FIGURA 3.4. DETALLE DEL ENTORNO GRÁFICO DE PANORAMA 5 Y BINAURAL PANNER.

Señales de test

Para cada sonido se han elegido una serie de localizaciones a un metro de distancia variando en azimut, es decir, en el plano horizontal, en las que es interesante comparar resultados. (ver *Tabla 3.2*). Las direcciones situadas en elevación se han obviado. Las direcciones de más interés en el son la frontal delantera y trasera (0° y 180°) para comparar confusiones entre esas dos posiciones, las direcciones oblicuas frontal derecha e izquierda 45° y 315° , y sus simétricas respecto del plano frontal para comparar el efecto de posibles confusiones del mismo tipo, y por ultimo las direcciones en ángulo recto 90° y 270° . También se incluyen muestras monofónicas que se supondrán situadas en el centro (dentro de nuestra cabeza), para comparar resultados de internalización-externalización con las señales ubicadas a 1 metro.

Dado que se pretende comparar hasta 4 sonidos y 4 opciones de filtrado, el numero de muestras que resultaría podría ser impracticable para un test perceptivo corriendo el riesgo de provoca fatiga auditiva y dificultando la tarea de los sujetos. Por tanto no se han incluido muestras para todas las posiciones y para todos los tipos de procesado.

Además hay que tener en cuenta que solo se incluyen dos sonidos obtenidos en grabación binaural de campo, la voz masculina y la puerta.

En la tabla 3.1 se muestra la combinación de muestras que se hay incluido en el test, donde la voz masculina es denotada como ambiental y la voz femenina como seca.

	Matlab Roomsim			Protools Panorama5			Logic Audio Binaural panner			Grabación Soundman		
Rosa	0°		45°	centro		0°	0°		45°			
	90°	135°	45°	45°	135°	180°	135°		180°			
Voz ambiental	0°		180°	centro		0°	180°		0°		180°	
	225°	270°	315°	225°	270°	315°	225°	270°	315°	225°	270°	315°
Voz seca	0°		45°	centro		0°	0°		45°			
	135°		180°	45°	135°	180°	135°		180°			
Puerta	0°		45°	centro		0°	0°		45°		0°	
	90°	135°	180°	45°	135°	180°	135°		180°		135°	180°
mar	0°		45°									

TABLA 3.1. CONFIGURACIÓN DE MUESTRAS UTILIZADAS

En el siguiente gráfico se muestra el numero total de muestras resultante para cada posición.

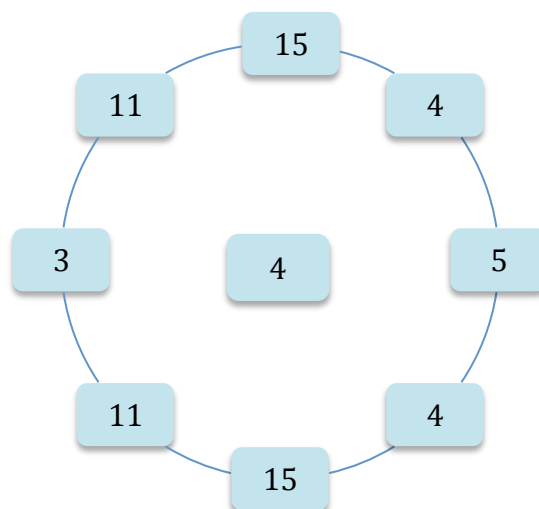


FIGURA 3.5. NUMERO DE MUESTRAS EN CADA POSICIÓN, VISTO SOBRE EL EJE HORIZONTAL

En el siguiente esquema (*Tabla 3.2*). se resume gráficamente la distribución espacial de las muestras para los distintos casos.

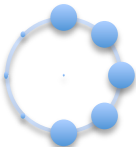
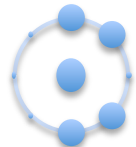
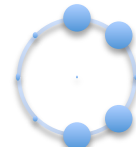
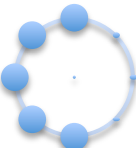

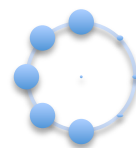
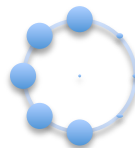
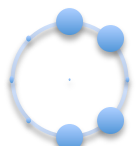
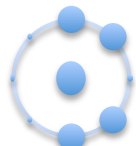

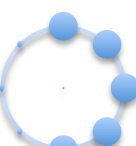
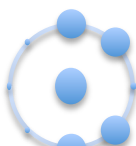
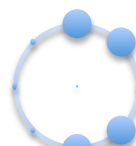
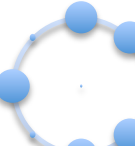
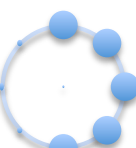
	Matlab Roomsim	Protools Panorama5	Logic Audio Binaural panner	Grabación Soundman
Rosa				
Voz ambiental				
Voz limpia				
Puerta				
mar				

TABLA 3.2. DISTRIBUCIÓN DE LAS MUESTRAS UTILIZADAS EN CADA CASO

3.2.4. Simulación de movimientos de cabeza.

El filtrado HRTF permite simular fuentes en movimiento tomando muestras consecutivas a lo largo de un desplazamiento angular determinado. Para simular un desplazamiento continuo de la fuente a partir de muestras discretas es necesario realizar interpolación. La interpolación consiste en promediar coeficientes de los filtros HRTF para crear un filtro aproximado en una posición intermedia entre posiciones y requiere un procesado complejo. La simulación de movimiento que se ha empleado aquí, ha sido obtenida de forma simple a través de software mediante *panorama 5*, cuyos detalles de interpolación se desconocen. Se empleara una automatización del parámetro *azimut* para conseguir una variación continua a lo largo del tiempo cuyos resultados son bastante satisfactorios.



FIGURA 3.5. DETALLE DE AUTOMATIZACIÓN CON PANORAMA 5.

De acuerdo con Thurlow (1967), el promedio de máxima rotación angular registrado para movimientos involuntarios de la cabeza 29° para altas frecuencias y 42° para bajas frecuencias.

Se han simulado movimientos de longitudes de arco de entre 5° y 30° en las posiciones central-delantera y central-trasera, dado que esta es la región más sensible a confusiones delante-detrás a la vez que tiene mejor localización borrosa y mejor precisión fija y angular (al menos en el caso central-anterior). También se han simulado para las zonas entorno a 45° y 135° .

Se han simulado esos movimientos para giros horarios y anti horarios, para observar si la aproximación o alejamiento a la zona conflictiva tiene alguna influencia.

3.3. PROCEDIMIENTO

3.3.1. Sujetos

Se ha realizado el test a 11 individuos, en un rango de edades de entre 25 y 64 años, distribuidos entre ambos sexos, y mayoritariamente sin experiencia previa en test de este tipo ni relación con el audio profesional, y que no hay reportado ningún problema auditivo previo.

3.3.2. Condiciones de escucha.

El test debe realizarse bajo las condiciones de presentación binaural, por lo tanto es necesario utilizar auriculares y minimizar el ruido de fondo que pueda superponerse sobre las señales de test. Los auriculares de tipo circumaural son los más adecuados en este sentido. Aunque se carece de datos cuantitativos, si se aprecia que la calidad de los auriculares, especialmente respecto de la respuesta en frecuencia influye de manera determinante sobre la nitidez de la apreciación binaural. Se ha intentado que todos los sujetos realicen el test con los mismos auriculares. En este caso se han utilizado unos auriculares sennheiser HD 280 pro, indicados para monitorización de estudio y respuesta en frecuencia es prácticamente plana para el rango de frecuencias audibles (figura). El valor de aislamiento que proporcionan respecto de ruido externo hasta 32 dB y su respuesta en frecuencia se muestra en la figura.

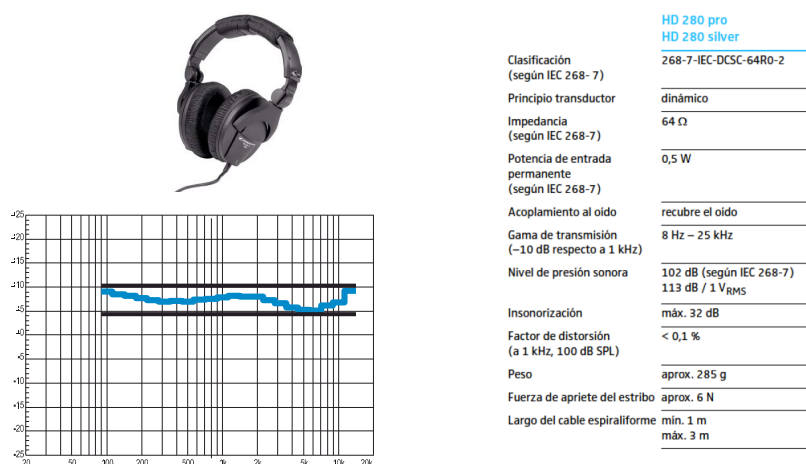


FIGURA 3.6 CARACTERÍSTICAS DE LOS AURICULARES SENNHEISER HD 280 (SENNHEISER)

3.3.3. Tarea

Para la presentación del test, se ha utilizado un formulario pdf interactivo en el que se coloca embebido el material de audio, junto con una grafica de referencia para identificar 8 regiones del espacio en que se divide una circunferencia (Figura 3.7) que representa al plano horizontal visto en planta, en el centro de la cual se sitúa el oyente. Las muestras empleadas corresponden a la posición centrada de cada región. En la posición centro (0), se sitúa la muestra monofónica. En cada división se indica un código numérico que se pide transcribir en cada recuadro de respuesta (Figura 3.8) con el fin de recopilarlo posteriormente para su procesado.



FIGURA 3.7. GRÁFICA DE REFERENCIA EN EL TEST.

El test se divide en dos partes. En la primera se pide identificar la aparente dirección de procedencia de un sonido, y ubicarlo en uno de los octantes en los que se divide una circunferencia (Figura 3.7).

La segunda parte incluye además sonidos en movimiento, por lo que una segunda tarea será indicar si aparentemente el la fuente se mueve o no (Figura 3.8).

Las muestras se han localizado de forma aleatoria y se permite escuchar cada muestra varias veces.

En la ultima parte del test se pide a los sujetos que indiquen cuantas veces en promedio han escuchado cada sonido.

El tiempo calculado necesario para realizar el test es entre 20 y 30 minutos.

Los resultados del test se recopilan a través de la aplicación para formularios de *adobe*, que permite exportarlos, entre otros, en formato *excel*.

Se ha añadido en el ANEXO II el formulario completo del test. El formulario interactivo es operativo en la versión *pdf*, **incluido en la versión pdf del trabajo entregada en CD**, en el que se pueden En el CD se adjunta también el formulario *pdf* de respuestas donde, están recogidos los resultados brutos del cuestionario del test perceptivo. de este trabajo. Para que el material sonoro pueda reproducirse correctamente, es necesaria la versión 9 o superior de *Acrobat Reader*.

Test 2.

Identifica la dirección de donde proviene el sonido como en el test 1 e indica además si crees que el sonido es estático o se mueve.



	<input type="radio"/> Se mueve	<input type="radio"/> No se mueve
	<input type="radio"/> Se mueve	<input type="radio"/> No se mueve
	<input type="radio"/> Se mueve	<input type="radio"/> No se mueve
	<input type="radio"/> Se mueve	<input type="radio"/> No se mueve
	<input type="radio"/> Se mueve	<input type="radio"/> No se mueve
	<input type="radio"/> Se mueve	<input type="radio"/> No se mueve
	<input type="radio"/> Se mueve	<input type="radio"/> No se mueve
	<input type="radio"/> Se mueve	<input type="radio"/> No se mueve
	<input type="radio"/> Se mueve	<input type="radio"/> No se mueve

FIGURA 3.8. DETALLE DEL CUESTIONARIO

3.4. RESULTADOS

A partir de las respuestas del test, se han obtenido parámetros de interés para evaluar el comportamiento de cada caso: aciertos globales según posición, influencia del tipo de procesado y de material sonoro e influencia del movimiento.

También se ha realizado un test T (T-test) o test de Student, que permite evaluar la significatividad de la diferencia de dos grupos, siempre que la comparación entre los grupos se haga para elementos que correspondan a las mismas condiciones. A partir de este test podemos obtener datos fiables de parámetros concretos que nos interesa estudiar como son la relación entre aciertos dentro y fuera y delante y detrás.

3.4.1. ACIERTOS / POSICIÓN.

Empezaremos analizando los resultados obtenidos para el grado de acierto global en la ubicación de la fuente posición por posición, sin tener en cuenta ni el tipo de filtrado ni el material sonoro utilizado. En la serie de gráficas de tipo histograma (Figura 3.9-3.42) se representa para cada muestra recreada, el porcentaje de respuestas (eje de ordenadas) en cada una de las posibles zonas (eje de abscisas). Las 8 zonas han sido codificadas en estas gráficas de la misma manera que en el cuestionario del test (de 0 a 8). La denominación utilizada en las gráficas para referirse a cada una de las muestras es *posición real*, referida a la región en la que se ha recreado la fuente.

Posición central (0):

Para la muestra de carácter monofónico, que fue recreada en la posición codificada como 0 (centro), aproximadamente un tercio de los sujetos la identifican correctamente en esa posición, y aproximadamente la cuarta parte la identifican con la posición frontal (0°) o la posición trasera (180°). En la gráfica (Figura 3.9).

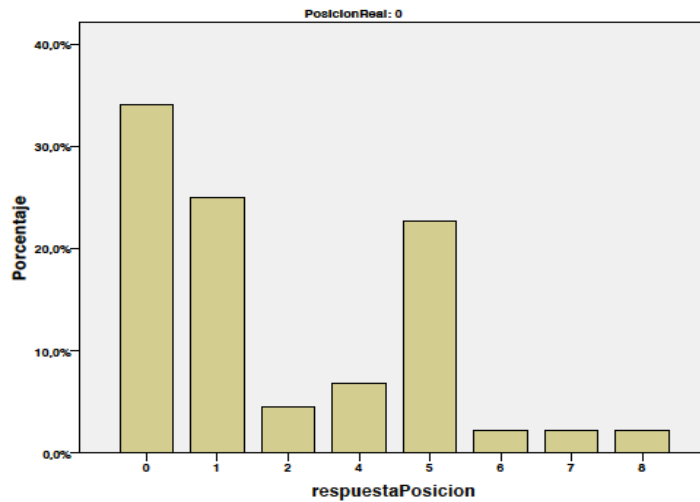


FIGURA 3.9. DISTRIBUCIÓN DE LAS RESPUESTAS PARA LA MUESTRA EN 0 (CENTRO)

Posiciones frontal (1) y posterior (5):

Para los casos en que la fuente está situada directamente delante o detrás (Figuras 3.10 y 3.11) se observa que el grado de acierto menor que para el caso anterior, siendo un 10% mayor para el caso trasero respecto del delantero.

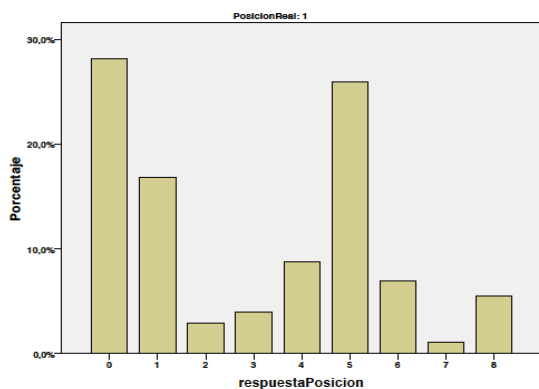


FIGURA 3.10. RESPUESTAS PARA LA MUESTRA EN LA POSICIÓN 1 (DELANTE).

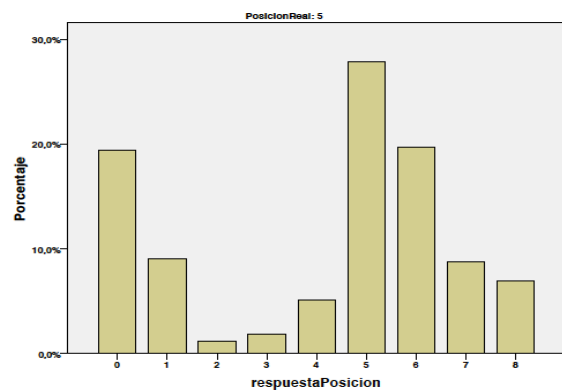


FIGURA 3.11. RESPUESTAS PARA LA MUESTRA EN 5 (DETRÁS)

Para el caso frontal, el grado de acierto es muy pobre (menos del 20%) y el mayor número de confusiones se produce hacia la posición central y la trasera, siendo las respuestas dadas para estas dos posiciones erróneas un 10% mayores que para la posición real). Se observa que para el caso trasero el grado de acierto se acerca al 30%, y que las confusiones se reparten entre más posiciones (posterior-izquierda y central). De estos resultados se puede extraer que el grado de confusión delante-detrás es mayor que el de confusiones detrás-delante, y que el grado de confusión hacia el centro es también mayor para la posición frontal.

Posiciones frontal-derecha(2)/ posterior-derecha(4) y frontal izquierda(8) / posterior izquierda(6):

El grado de confusión delante-detrás también afecta a las posiciones delantera-derecha (45°) hacia la trasera-derecha (135°) y las posiciones delantera-izquierda (315°) sobre la trasera izquierda (225°), como se aprecia en las gráficas de las *Figuras 3.12 a 3.15*.

Se aprecia una tasa de acierto es bastante pobre (en torno al 20% para ambos casos) en las posiciones delanteras registrándose casi la mitad de las respuestas desplazadas hacia la región contigua hacia atrás (90° y 270°).

Para estas regiones se aprecia un grado mayor de confusiones hacia atrás que para los casos anteriores, siendo de entorno al 70% sobre el total y más del doble del que se produce en dirección inversa.

Se aprecia una tasa de acierto considerable de entorno al 50% para las posiciones traseras en ambos casos, produciéndose un error hacia delante sobre todo hacia la región contigua (90° y 270°), pero considerablemente menor que para los casos simétricos hacia adelante.

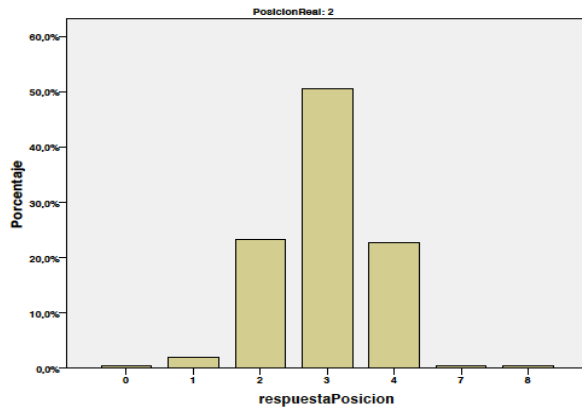


FIGURA 3.12. DERECHA-DELANTE (45°)

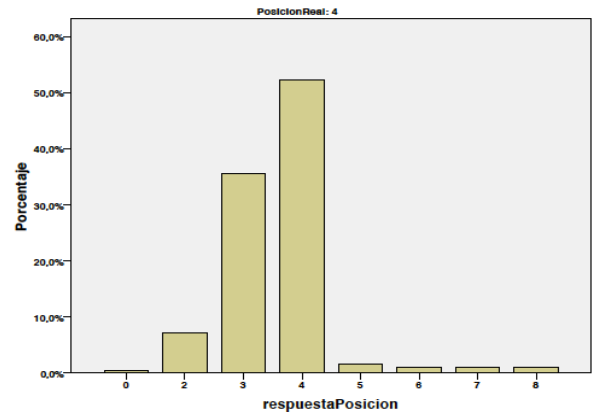


FIGURA 3.13. DERECHA-DETRÁS (135°)

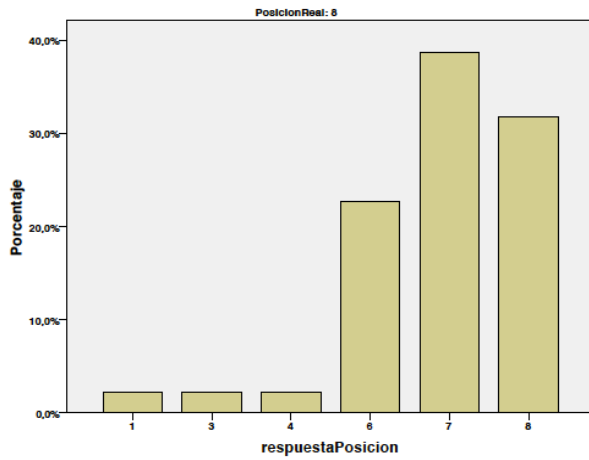


FIGURA 3.14. IZQUIERDA-DELANTE (315°)

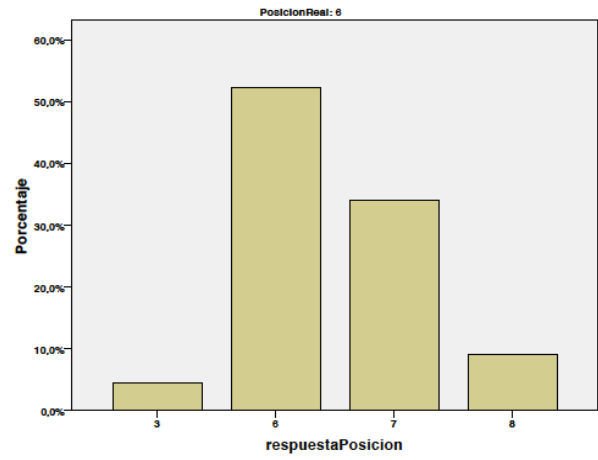


FIGURA 3.15. IZQUIERDA-DETRÁS (225°)

Posiciones derecha (3) e izquierda (7):

Finalmente para las posiciones directamente laterales, correspondientes a posiciones alineadas con ambos oídos, se refleja un buen grado de acierto, llegando al 60% (Figuras 3.16, 3.17).

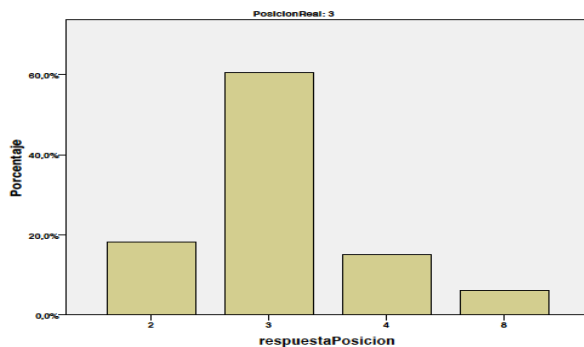


FIGURA 3.16. RESPUESTAS PARA LA MUESTRA EN 3 (DCHA)

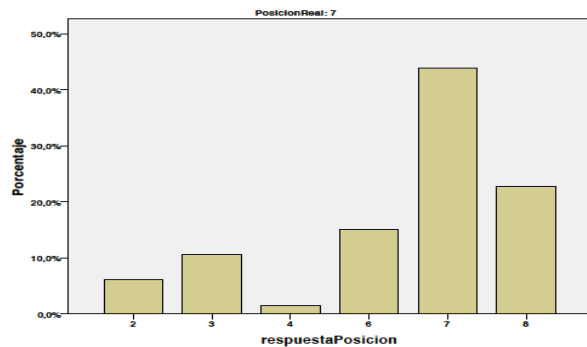


FIGURA 3.17. RESPUESTAS PARA LA MUESTRA EN 7 (IZQ)

3.4.3. ACIERTO / POSICIÓN/ TIPO DE MATERIAL

Un primer análisis de la influencia del tipo de material sonoro en los resultados indica que el comportamiento tiende a ser similar para todos los sonidos, con ciertas excepciones interesantes que se mencionan a continuación :

Posición central (Figuras 3.18, 3.19,3.20):

- Para **ruido rosa** se observa que el numero de respuestas es igual para la posición central que para la delantera.
- La **puerta** fue ubicada mas a menudo en la posición trasera que en su posición real (mas de la mitad de las respuestas).
- La **voz femenina (seca)** fue ubicada con acierto más de la mitad de las veces, lo que indica una mayor facilidad para internalizarla.
- La voz **masculina (ambiental)** prácticamente no se desvía del promedio.

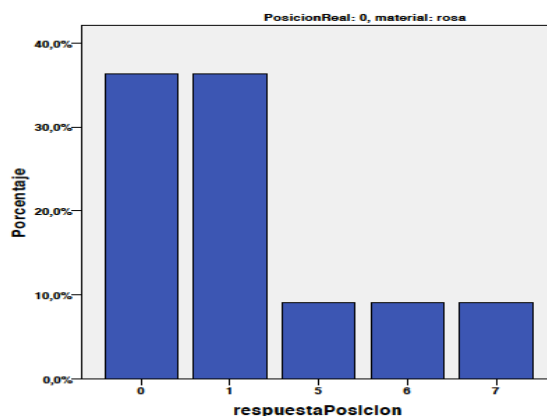


FIGURA 3.18. RUIDO ROSA RECREADO EN EL EL CENTRO (0)

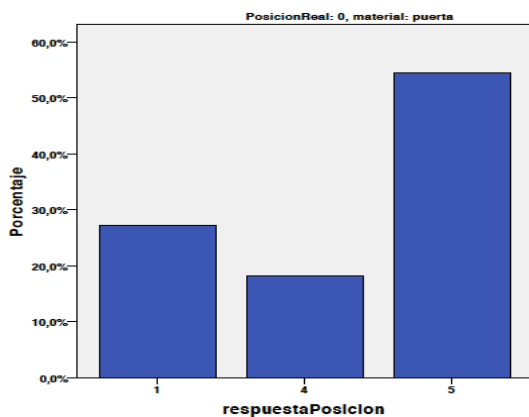


FIGURA 3.19. PUERTA EN EL CENTRO

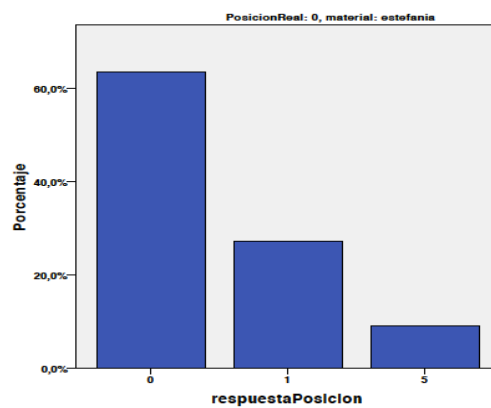


FIGURA 3.20. VOZ FEMENINA EN EL CENTRO

Posición frontal, Figuras (3.21, 3.22):

- La **puerta**, al igual que para la posición central, tiende a ubicarse en la zona trasera con más frecuencia, incluso mayor que para la posición real.
- El sonido de **mar** tiende a localizarse en el centro, al igual que el ruido rosa ver gráfica (Figura 3.36, pg. 72, en el apartado influencia del movimiento), más a menudo que en su posición, sin embargo, a diferencia de éste, apenas presenta confusión hacia atrás.
- La **voz femenina** (se puede ver en la gráfica de la Figura 3.41, junto con movimiento, pg. 73), presenta un excesivo error hacia la posición centro (65%), lo que confirma la clara tendencia a internalizarla.

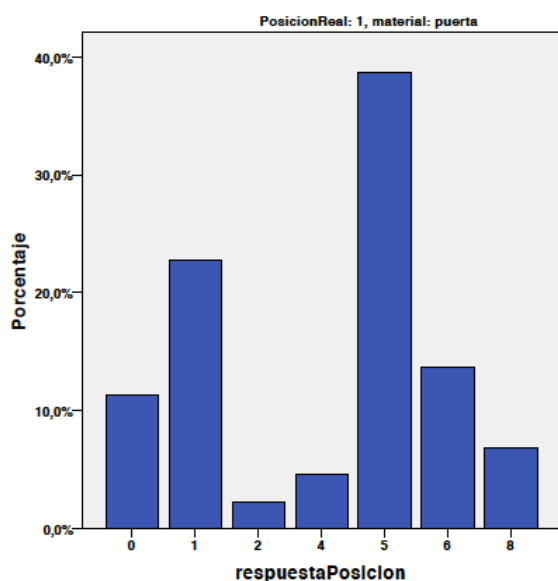


FIGURA 3.21. PUERTA DELANTE

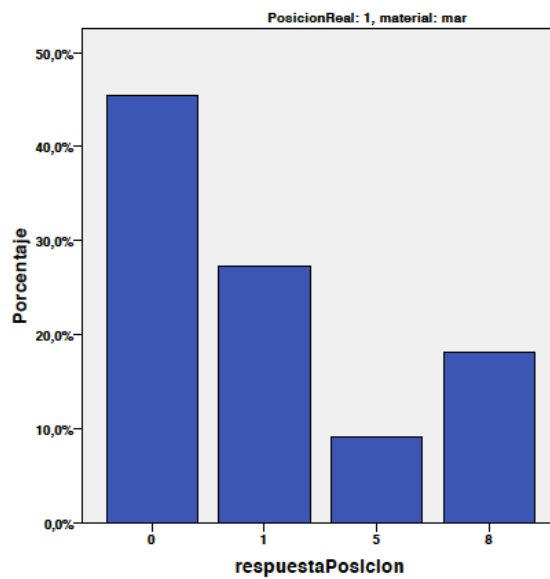


FIGURA 3.22. MAR DELANTE

Posición trasera (Figuras 3.23, 3.24):

- La **puerta** se localiza en esta posición con mucho más acierto que las demás.
- El **mar** se localiza algo mejor en esta posición, y además apenas se confunde con la posición central, pero sí con la delantera.
- La **voz femenina**, vuelve a confundirse con el centro más que las demás, gráfica se muestra en la *Figura 3.41*.
- El ruido **rosa** también tiene más resultados en el centro que detrás (ver *Figura 3.40*).

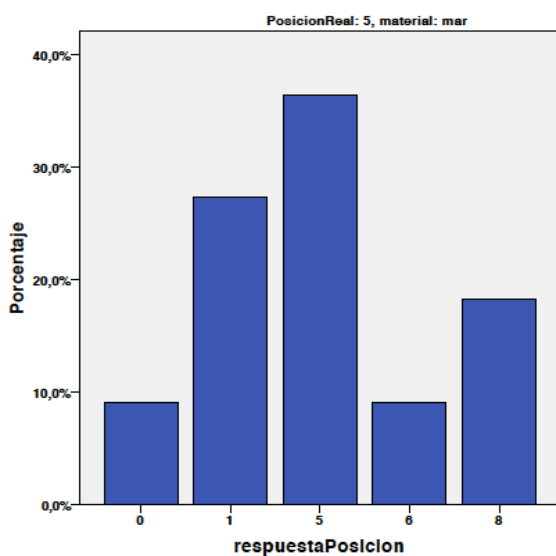


FIGURA 3.23. MAR DETRÁS

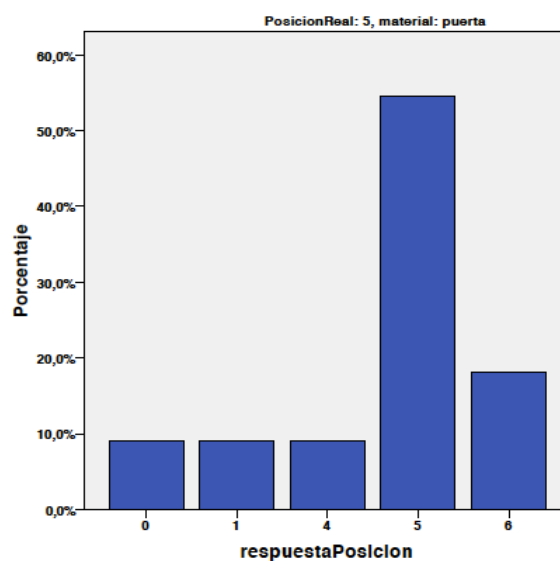


FIGURA 3.24. PUERTA DETRÁS

Posición frontal-derecha.

En esta posición la localización para la muestra **puerta** y **mar** (Figuras 3.25, 3.26), se confunde la mayoría de las veces con la posición trasera-derecha, especialmente para la puerta, que se localiza el doble de veces hacia atrás, confirmando la tendencia. Sin embargo, **ruido rosa** y las **voces** siguen una tendencia a que predomine la componente lateral sobre la frontal-trasera, localizándose más veces en ángulo recto.

En el caso simétrico hacia la izquierda, la única muestra estudiada es la de la voz masculina, con resultados similares.

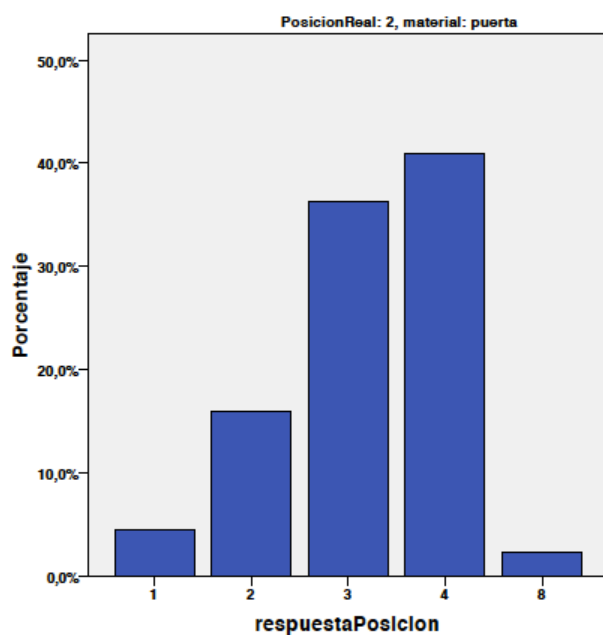


FIGURA 3.25. PUERTA EN 45°

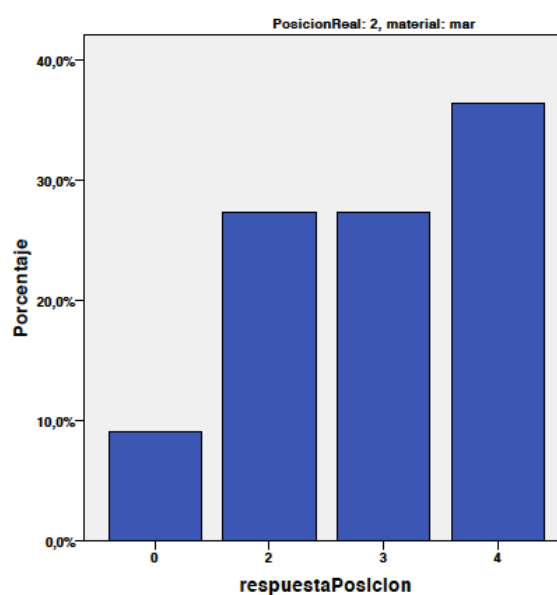


FIGURA 3.26. MAR EN 45°

Posición trasera derecha (4):

Destaca para esta localización el alto grado de acierto para el material **puerta** (Figura 3.28), que ya se mostraba tendente a localizarse hacia atrás cuando era presentada en posiciones delanteras. Para el sonido de **mar** (Figura 3.29), se produce un hecho peculiar, ya que presenta más respuestas en la posición directamente lateral que la propia, de hecho el grado de acierto no mejora respecto del caso anterior en estando situada a 45° se la desplazaba a 135°.

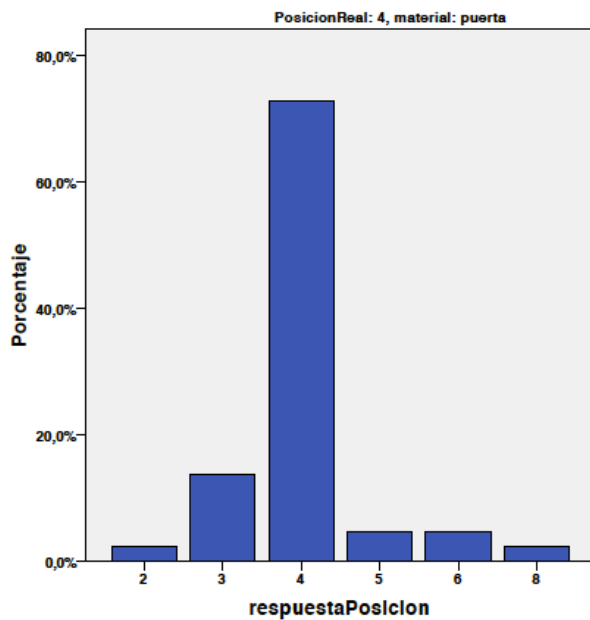


FIGURA 3.28. PUERTA EN 135°

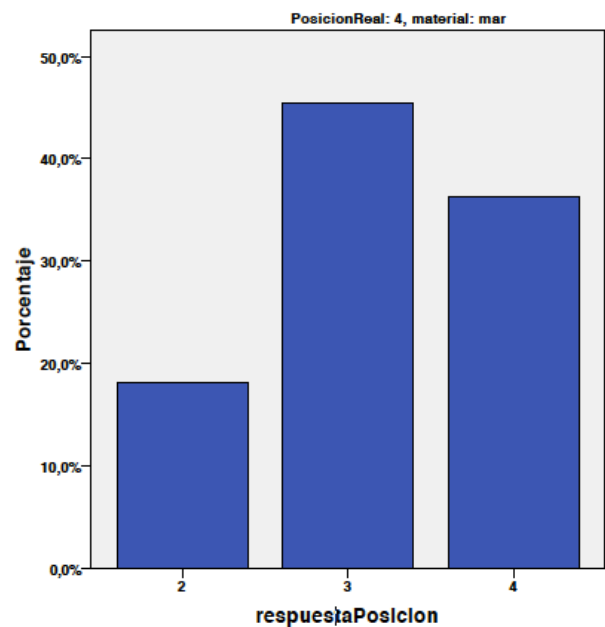


FIGURA 3.29. PUERTA EN 135°

3.4.2. Influencia del tipo de material sonoro.

ANÁLISIS ESTADÍSTICO MEDIANTE TEST-T.

Para estudiar la influencia de los distintos materiales y los tipos de procesado sobre el resultado, se ha efectuado un TEST-T (paired-samples t-test) de muestras pareadas en el que se compara la media de aciertos cada par de condiciones (par de materiales o par de procesados). Las medias se calculan para una serie de variables de interés:

- **Acierto dentro-fuera:** grado de acierto en la posición cero (centro) y el resto. Permite extraer información acerca de la fiabilidad de cada caso para localizar interna y externamente.
- **Respuesta dentro-fuera:** Medida de las respuestas realizadas para cada grupo en el centro y el resto de las posiciones. Permite extraer información de la tendencia a internalizar o externalizar con cada grupo.
- **Acierto delante-detrás:** grado de acierto en las posiciones delanteras ($-315^\circ, 0^\circ, 45^\circ$), y las traseras ($135^\circ, 180^\circ, 225^\circ$). Mide la fiabilidad de cada caso para localizar correctamente delante o detrás.
- **Respuesta delante-detrás:** Medida de las respuestas realizadas para cada grupo hacia delante y detrás. Permite extraer información acerca de la tendencia a responder delante o detrás para cada caso.

Cada comparación de par de datos se realiza para las muestras que pueden ser comparables, es decir en los que el mismo sonido se testea en la misma posición para ambos pares de datos, por lo que la variable bajo estudio puede obtener distintos valores si el número de muestras es distinto.

Las Variables acierto se miden en tanto por ciento de aciertos, mientras que las variables respuesta entre +1 y -1 siendo +1 el valor para la primera respuesta posible (dentro, delante) y -1 el valor para la segunda (fuera, detrás).

Una vez calculado el promedio, un valor de la variable próximo a +1 indica que el promedio de respuestas se aproxima más a dentro o delante, y viceversa.

La significatividad indica que la muestra de resultados para el conjunto de sujetos es suficientemente representativa como para que los datos sean fiables.

Influencia del material sonoro

El test-T indica los siguientes resultados significativos (Tabla completa en ANEXO I):

Acierto dentro-fuera:

- Hay diferencias significativas entre la media de acierto dentro-fuera para la condición *voz ambiental* (67%) y la condición *voz seca* (48%), con un resultado del test ($t=2,3$; $gl=65$; $p=0,02$)¹.
- No hay diferencias significativas *voz femenina (seca)* (73%) y *puerta* (87%); ($t=-3,2$; $gl=142$; $p = 0,002$).
- No hay diferencias significativas entre *puerta* (84%) y *voz ambiental* (74%), con ($t=-2$; $gl=109$; $p=0,005$).
- No hay diferencias significativas entre *ruido rosa* (80%) y la *voz seca* (73%); ($t=-2$; $gl=142$ $p=0,05$); ni entre éste y *puerta* (87%); ($t=2$; $gl=142$; $p=0,005$).

De los resultados se extrae que la voz masculina (ambiental), obtiene mejores resultados de internalización y externalización, en parte puede ser debido a que parte de las muestras son registros estereofónicos tomados con los micrófonos Soundman, por lo que puede haber influencia de factores ambientales relacionados con la sala donde fueron tomados que aporten información espacial

¹ t : resultado del test; gl : grados de libertad; p : significatividad.
La diferencia es "significativa" si $p < 0,05$ y "muy significativa" si $p < 0,001$.

En el siguiente gráfico (*Figura 3.30*) pueden verse de modo más general los resultados de la comparación del acierto con un intervalo de confianza del 95% (eje de ordenadas), agrupados por materiales sonoros (eje de abscisas). En general los resultados son muy similares, salvo la tendencia a que la voz seca se localice algo peor, aunque en general la tasa de acierto es alta. La voz grabada en estudio, al tratarse de una voz neutra sin características ambientales, muestra una tendencia a la internalización, como se ha observado en la comparativa general de aciertos por posición por tipo de material sonoro.

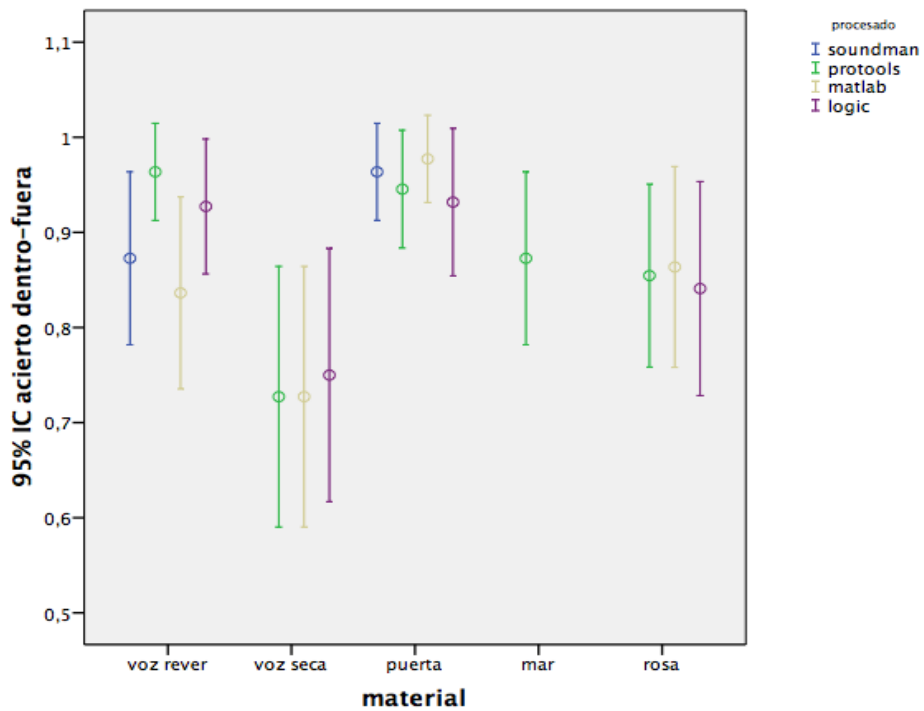


FIGURA 3.30. COMPARACIÓN DE ACIERTO DENTRO FUERA POR TIPO DE MATERIALES

Respuesta dentro-fuera (dentro=0; fuera=1).

- Hay diferencias muy significativas en la tendencia a localizar dentro o fuera entre *voz ambiental (masculina)*(0,71) y *voz seca* (0,44); ($t=3,7$; $gl=65$; $p < 0,001$).
- Hay diferencias muy significativas en la tendencia entre *voz ambiental* (0,76) y *puerta* (0,94); ($t=-3,6$; $gl=109$ $p < 0,001$).
- Hay diferencias muy significativas entre *voz seca* (0,71) y *puerta* (0,95); ($t=-5,8$; $gl=142$ $p < 0,001$).
- No hay diferencias significativas entre *voz seca* (0,71) y *ruido rosa* (0,83); ($t=-3,1$; $gl=142$; $p=0,02$).
- No hay diferencias muy significativas entre *puerta* (0,95) y *ruido rosa* (0,83); ($t=3,7$; $gl=142$ $p < 0,001$).

Los resultados constatan que las muestras de puerta se localizan externamente con mucha más frecuencia que el resto, y confirma la tendencia antes indicada hacia la internalización de la voz neutra.

Acierto delante-detrás:

- Solo hay una comparación con significatividad alta para esta variable entre el acierto delante-detrás para la *voz seca* (34%) y la *puerta* (48%); ($t=-2,6$; $gl=142$; $p=0,01$).

En la grafica se muestra el acierto desglosado para cada grupo de posiciones(*Figura 3.31*), las tres de delante, las tres del medio (incluyendo el centro y los lados) y las tres de detrás. Se ve una tendencia a que el acierto sea mayor para los sonidos que están detrás que para los que están delante, especialmente para la puerta. Aunque el intervalo de confianza es demasiado alto

para cuantificar la diferencia. Se aprecia que el acierto para los sonidos delanteros es bastante más pobre.

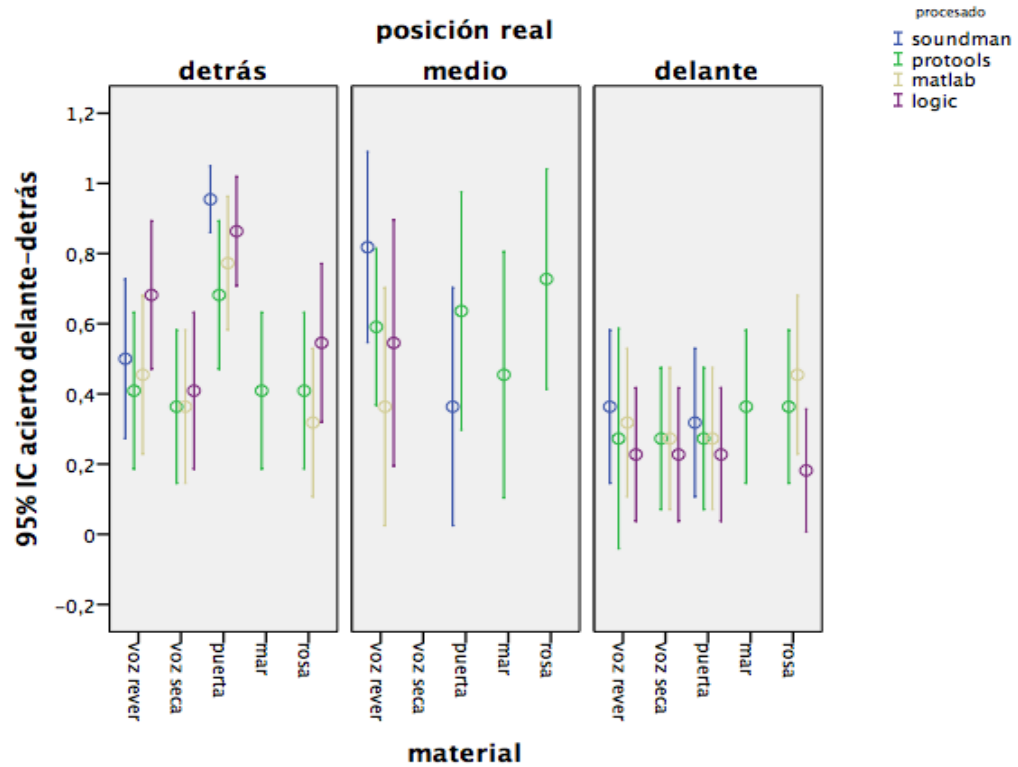


FIGURA 3.31. ACIERTO DELANTE-DETRÁS AGRUPADO POR MATERIALES. LAS TRES DIVISIONES (DELANTE, MEDIO, DETRÁS) SE REFIEREN A LA ZONA DONDE SE PRESENTABA EL SONIDO

Respuesta delante-detrás (delante=1; detrás=-1):

- Hay diferencias muy significativas entre *voz seca* (-0,01) en *puerta* (-0,47); ($t=6,9$; $gl=141$; $p< 0,001$).
- Hay diferencias muy significativas entre *puerta* (-0,48) y ruido rosa (-0,08); ($t=-4,8$; $gl=142$; $p< 0,001$).
- Hay diferencias significativas entre *puerta* (-0,41) y *mar* (0,02); ($t=-3$; $gl=43$; $p=0,003$).
- Hay diferencias significativas entre la tendencia de *voz ambiental* (-0,12) y la de *puerta* (-0,4) a localizarse detrás ($t=2,8$; $gl=109$; $p=0,005$).
- No hay diferencias significativas entre *voz ambiental* (-0,2) y *voz seca* (0,02); en la tendencia a ser percibida detrás ($t=-2$; $gl=65$; $p=0,04$).

Los resultados confirman la tendencia a que la puerta se localice más frecuentemente en las posiciones traseras, mientras el resto se mantienen en valores neutrales.

En la gráfica (Figura 3.32) se observa que en general todas las muestras, salvo quizá el mar y el ruido rosa tienden a localizarse detrás, lo que indica una tendencia a la confusión delante-detrás. En el único caso que se destaca claramente de los demás, la puerta, se observa que el ratio de confusión de las posiciones traseras con las frontales es alto, teniendo en cuenta que se testean el mismo numero de muestras delante que detrás.

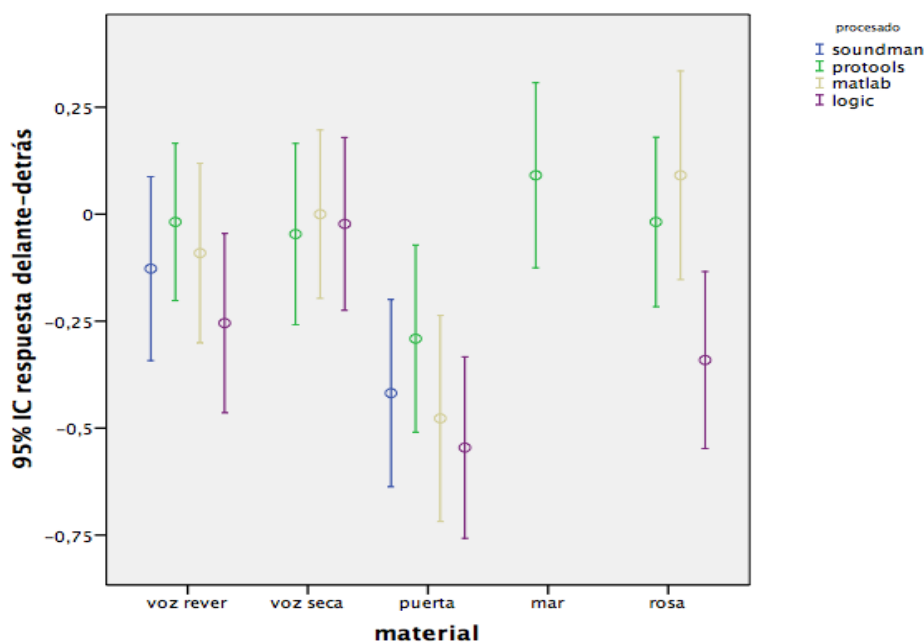


FIGURA 3.32. RESPUESTA DELANTE-DETRÁS SEGÚN MATERIAL

3.4.3. Influencia del tipo de procesado.

El test T de muestras relacionadas que indica diferencias significativas (ANEXO I):

Acierto dentro-fuera:

- La diferencia es muy significativa entre el material grabado en los micros *Soundman* (18%) y el filtrado hecho con matlab (91%); ($t=-6,2$; $gl=21$; $p<0,001$).
- Es significativamente menor para las muestras *soundman* (36%) que para el filtrado con binaural panner de *logic* (82%); ($t=-2,9$; $gl=21$; $p=0,009$).

Dado que las muestras comparables entre *Soundman*, *matlab* y *logic* no incluían ninguna muestra en el centro, los malos resultados de para las muestras *Soundman* son atribuibles a errores de internalización para el material registrado con esos micros.

Respuesta delante-detrás (delante=1; detrás=-1):

- La diferencia entre la tendencia de esta variable es significativa entre las muestras *Soundman* (0,14) y las de *logic* (0,41); ($t=-2$; $gl=21$; $p=0,02$).
- No hay diferencias significativas en la tendencia entre *Matlab* (-0,12) y *logic* (-0,29); ($t=-2$; $gl=174$; $p=0,009$).
- Tampoco es significativa entre *protools* (-0,12) y *logic* (-0,29); ($t=-2$; $gl=186$; $p=0,005$)

Al margen de variaciones puntuales, los resultados obtenidos del test son en general bastante poco significativos y no se puede extraer ninguna tendencia general. Dicho de otro modo, no parece que ninguna de las opciones de procesado implique diferencias destacables sobre las demás.

Esto se aprecia si observamos las graficas para acierto dentro-fuera (*Figura 3.30, pg. 66*). En general, se aprecia una influencia más o menos clara del tipo de material sonoro sobre la localización, pero no así del tipo de procesado utilizado para situar esa localización.

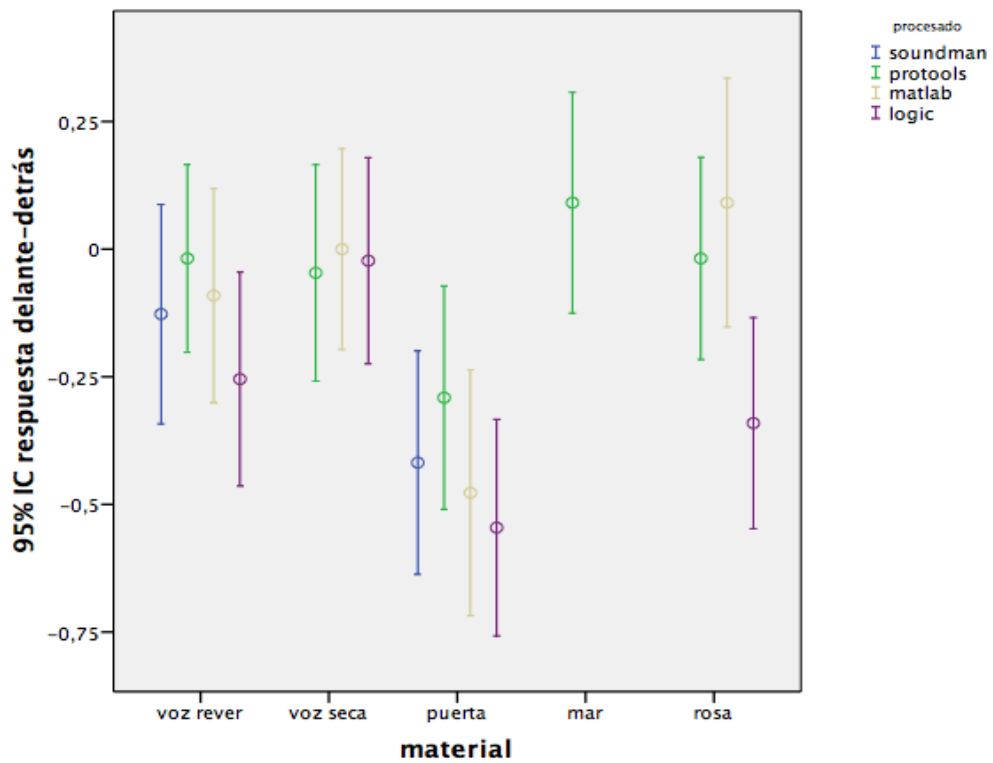


FIGURA 3.33. RESPUESTA DELANTE-DETRÁS AGRUPADA POR MATERIALES. Y TIPOS DE PROCESADO.

3.4.4. Influencia del movimiento

Posición central delantera.

En la grafica (*Figura 3.34*) se observa que cuando se introduce un ligero movimiento a la muestra de la posición frontal (barras en verde), el ratio de errores hacia el centro disminuye respecto al caso estático (en azul), sin embargo también lo hace el de respuestas dadas hacia la posición correcta.

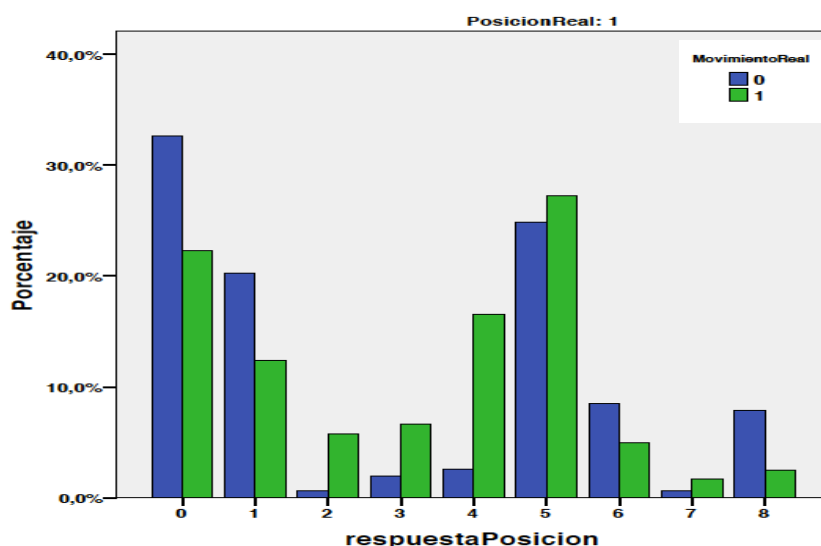


FIGURA 3.34. PORCENTAJE DE RESPUESTAS EN CADA POSICIÓN CUANDO SE SIMULABA MOVIMIENTO EN TORNO A LA POSICIÓN FRONTAL (MOVIMIENTO=1), COMPARADO CON EL CASO ESTÁTICO (MOVIMIENTO=0)

En general, se produce un aumento de confusiones hacia atrás. Los resultados son bastante coincidentes para el ruido rosa (*Figura 3.36*), pero más exagerados en el caso de la voz (*Figura 3.35*). Vemos en la figura como la voz seca, que en el caso estático tenía un alto grado de internalización, éste disminuye drásticamente, tendiendo a situarse hacia la posición central trasera.

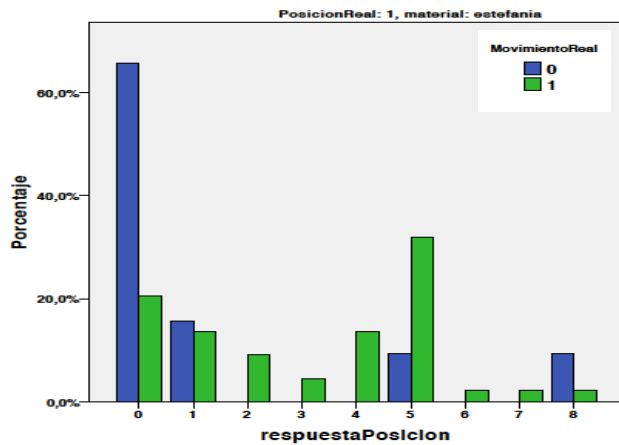


FIGURA 3.35. COMPARACIÓN DE RESPUESTAS POR POSICIÓN PARA MUESTRAS DE VOZ FEMENINA EN LA ZONA FRONTAL, CON MOVIMIENTO (1) Y SIN ÉL (0)

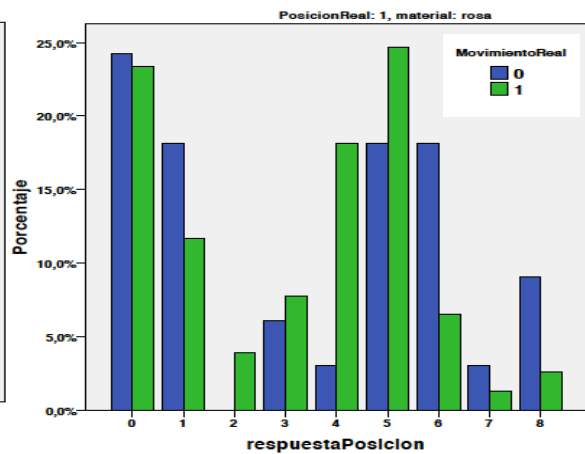


FIGURA 3.36. COMPARACIÓN DE RESPUESTAS POR POSICIÓN PARA MUESTRAS DE RUIDO ROSA EN LA ZONA FRONTAL, CON MOVIMIENTO (1) Y SIN ÉL (0)

Posición Frontal-derecha.

La tendencia hacia atrás se manifiesta también en esta posición. En las gráficas (3.37 y 3.38) se observa que la tendencia ya acusada para el caso estático de ubicar los sonidos originalmente a 45° en 90° se acentúa todavía más si estos sonidos son móviles. La tendencia es más acusada para ruido rosa (Figura 3.38), casi el doble de casos de error respecto al caso estático. En la voz (Figura 3.37) la tendencia es a desplazar un poco más el sonido (hacia 135°).

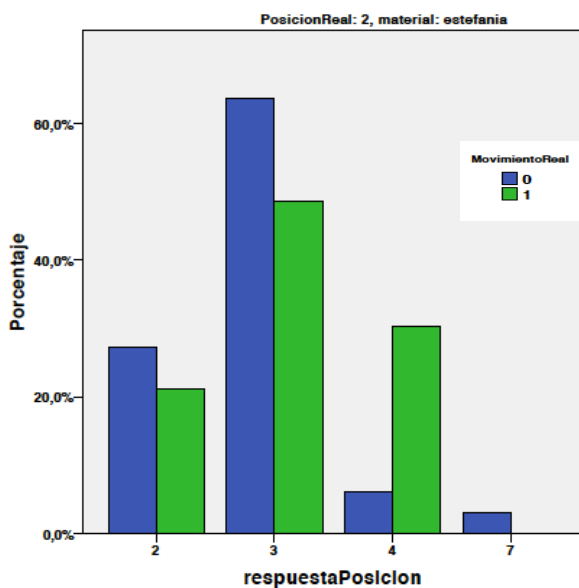


FIGURA 3.37. COMPARACIÓN DE RESPUESTAS POR POSICIÓN PARA MUESTRAS DE VOZ FEMENINA EN LA ZONA FRONTAL, CON MOVIMIENTO (1) Y SIN ÉL (0)

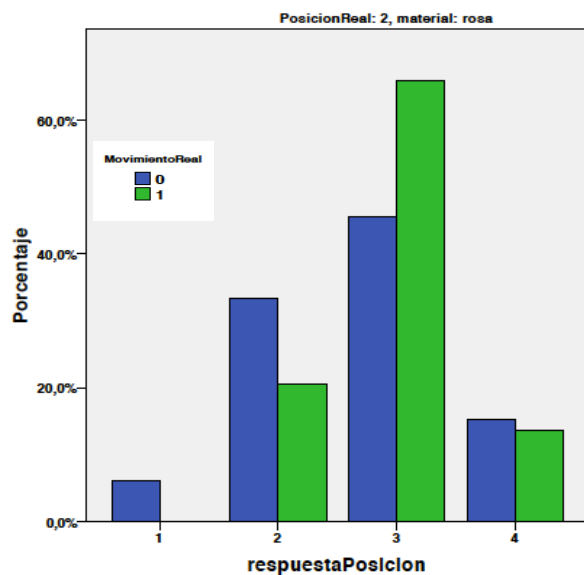


FIGURA 3.38. COMPARACIÓN DE RESPUESTAS POR POSICIÓN PARA MUESTRAS DE RUIDO ROSA EN LA ZONA FRONTAL, CON MOVIMIENTO (1) Y SIN ÉL (0)

Posición trasera derecha.

Los resultados apenas indican si acaso un ligero descenso de los errores hacia delante y mejores ratios de precisión, sobre todo para la voz (Figuras 3.39 y 3.40).

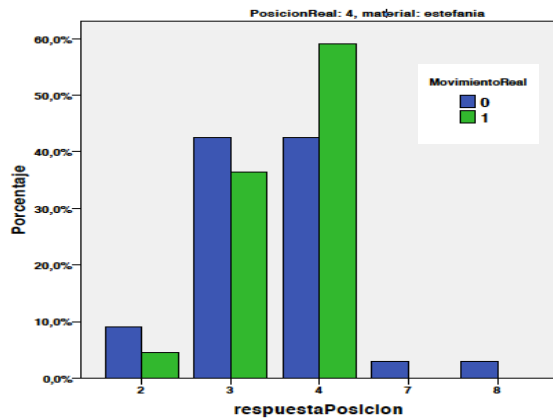


FIGURA 3.39. COMPARACIÓN DE RESPUESTAS POR POSICIÓN PARA MUESTRAS DE VOZ FEMENINA EN LA ZONA FRONTAL, CON MOVIMIENTO (1) Y SIN ÉL (0)

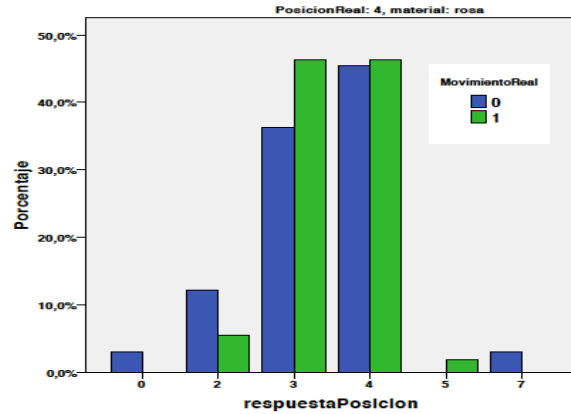


FIGURA 3.40. COMPARACIÓN DE RESPUESTAS POR POSICIÓN PARA MUESTRAS DE RUIDO ROSA EN LA ZONA FRONTAL, CON MOVIMIENTO (1) Y SIN ÉL (0)

Posición central trasera.

Para esta posición la mejora sí es considerable. Se reduce tanto el grado de internalización como el de errores detrás-delante, tanto para la voz (Figura 4.41) como para el ruido rosa (Figura 3.42), sin embargo, los errores no se corrigen del todo, ya que la tendencia es a ubicarlos en la posición trasera-izquierda, situada hacia la dirección del movimiento.

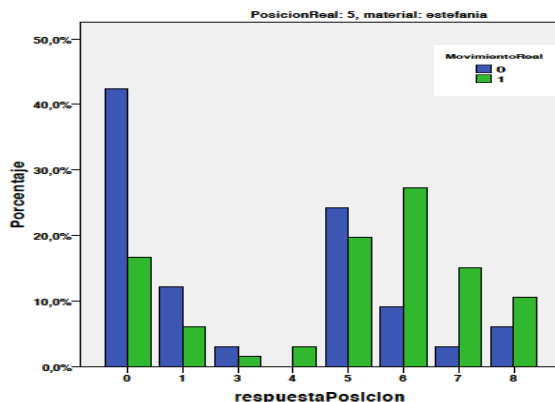


FIGURA 3.41. COMPARACIÓN DE RESPUESTAS POR POSICIÓN PARA MUESTRAS DE VOZ EN LA ZONA FRONTAL, CON MOVIMIENTO (1) Y SIN ÉL (0)

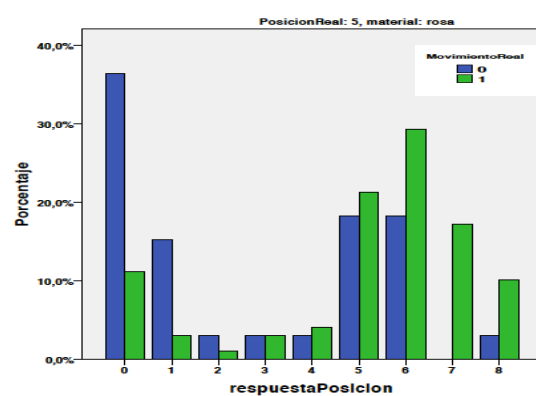


FIGURA 3.42. COMPARACIÓN DE RESPUESTAS POR POSICIÓN PARA MUESTRAS DE RUIDO ROSA EN LA ZONA FRONTAL, CON MOVIMIENTO (1) Y SIN ÉL (0)

Detección del movimiento

Se ha realizado un test-T para comparar el grado de detección de movimiento para los distintos movimientos angulares. Aunque todos las muestras del test 2, incluían movimiento simulado, la consciencia de éste no se hacía presente en todos los casos. En general, solo ha habido unanimidad en la respuesta para giros de 10° de arco (horario y anti horario) en la parte trasera en donde no se apreciaba movimiento, pero solo con las muestras de voz; y para giros de entre 20° y 30° donde siempre se detectaba movimiento.

De los resultados del test (ver tabla en el anexo) se extrae que los porcentajes de detección de movimiento son siempre:

- Menores al 20% para 5°.
- Menores al 35% para 10°.
- Menores al 55% para 20°.
- Mayores al 75% para 30°.

Por pares de datos, hay una diferencia significativa en la detección de movimiento en los casos:

- Entre 5° y 30° de un 72%.
- Entre 5° y 20° de un 45%.
- Entre 10° y 30° de un 67%
- Entre 10° y 20° de un 28%.
- Entre el giro de 30° desde y hacia posiciones centradas de un 14% a favor del último.

Se puede apreciar que no siempre se detecta movimiento cuando es suficientemente corto y casi siempre para movimientos suficientemente largos, y que hay una ligera tendencia a detectar más fácilmente el movimiento cuando se dirige desde un ángulo lateral hacia delante o detrás, que a la inversa. Esto es coherente con el hecho de que las posiciones directamente delante y directamente detrás disponen de menos indicios de localización.

DISCUSIÓN

4. DISCUSIÓN

El objetivo del diseño experimental planteado era evaluar la efectividad de la síntesis binaural como método de postproducción de sonido espacial. El principal hándicap que presentaba este método es la ineffectividad a la hora de representar correctamente la información espacial necesaria para una interpretación satisfactoria por parte del sistema auditivo, lo que deriva en confusiones en la localización del sonido recreado, o incluso la falta de localización espacial.

Errores en la localización.

En general los resultados de localización no son muy buenos, alcanzándose los mejores resultados para los sonidos situados regiones laterales en torno 90° y 270° , posiciones alineadas directamente con el eje aural, donde efectivamente, las diferencias interaurales de tiempo e intensidad son mayores. No obstante, el grado de acierto no sobrepasa el 60%.

Se aprecia que predominan los errores hacia regiones colindantes, salvo para los casos 0° y 180° donde dominan los errores de delante a detrás y de detrás a delante. En general, como era de esperar, se confirma una tendencia a la confusión hacia atrás.

Para estas regiones, los resultados indican un grado de acierto bastante pobre, sobre todo para el caso frontal (en torno a un 10% más). Esta posición se confunde la mayoría de veces con en el centro o detrás, mientras que los errores para el caso trasero se reparten más sobre regiones contiguas en la parte posterior y también en el centro.

Sin embargo los errores asociados al centro son también más numerosos hacia 0° y 180° , lo que indica que en general, es difícil distinguir las muestras monofónicas de las que están situadas situadas en el plano medio. En efecto, las características binaurales de éstas son similares, dada la ausencia de indicios de

tipo ITD e ILD. Aun así se observa una tendencia ligeramente mayor a la internalización que a la externalización.

En los casos intermedios (posiciones oblicuas 45° , 135° , 235° , y 270°), también dominan las confusiones detrás-delante y detrás-delante, con el matiz de que en el caso de las regiones delanteras el grado de acierto es peor que en las traseras y la confusión es mayor hacia las detrás que a la inversa, lo cual concuerda con los datos de la tabla 2.3, recopilados por Begault (2000). Se aprecia también que domina la componente lateral sobre la frontal-trasera, sobre todo para los sonidos delanteros, que tienden a lateralizarse (desplazarse hacia la región 90° - 270°).

En las regiones situadas en el plano medio, con menores diferencias interaurales, el fenómeno de confusión delante-detrás es más acusado. Esta tendencia también se da para localizaciones que teóricamente tienen más información de ese tipo, pero están situadas en la zona del cono de confusión (*Figuras 3.12 a 3.15*). Estas regiones son sensibles a la ausencia de indicios espectrales .

Influencia del tipo de procesado.

El estudio de esta condición era interesante dada la alta sensibilidad al filtrado con HRTF no individualizadas.

Según los resultados obtenidos, ninguno de los métodos es claramente más efectivo que el resto, pero el alto índice de errores hace pensar en búsqueda de estrategias de mejora.

Desde un punto de vista práctico, el software de Wave Arts ofrece más posibilidades aportando procesado de tipo acústica de recintos, que puede aportar más realismo al procesado sonoro de cara a su efectividad en la especialización.

Se observa que la grabación de campo con micrófonos obtiene resultados similares a los del procesado de material neutro, aunque se ha limitado su uso a

condiciones impuestas por el filtrado, el estudio de grabaciones que utilicen la influencia del paralaje puede ser interesante. Aun así se debe tener en cuenta que el material grabado contendrá la información espectral fina aportada por el sujeto que realiza la grabación, con lo que el material resultante será mejor localizado por sujetos con rasgos físicos similares.

Influencia del movimiento.

La introducción de movimiento en las muestras parece producir en general una mejora en la tasa de errores de largo recorrido, delante-detrás, pero aumenta los de pequeño recorrido, entre posiciones contiguas.

Parece tener un efecto positivo en la externalización de la voz (un 40%), pero no para el ruido rosa. Para la posición frontal sin embargo el acierto no mejora, aumentando incluso la confusión hacia atrás.

Sin embargo para la posición trasera, los resultados de externalización mejoran para ambos sonidos (un 25%), casi anulándose además la confusión de detrás hacia adelante.

Para los ángulos oblicuos sucede algo parecido, en los delanteros, el movimiento empeora las confusiones hacia detrás, mientras que para los traseros mejora levemente las confusiones hacia delante.

Estos resultados irregulares pueden deberse a la elección de movimientos excesivamente largos, ya que según comentarios generalizados, el movimiento tendía a percibirse mucho más largo de lo que realmente era, lo que probablemente inducía confusiones.

Se aprecia una necesidad de estudiar el efecto del movimiento en detalle, ya que en principio, los movimientos cortos de menos de 10° parecen pasar bastante desapercibidos y podrían ser una técnica interesante a tener en cuenta.

Influencia del tipo de material sonoro.

La muestra de tipo impulsivo *puerta* tiene el mayor grado de confusión hacia el exterior y de acierto en las posiciones traseras, pero también mayor tendencia a confundirse con esa posición. De acuerdo con comentarios realizados por algunos sujetos al finalizar el test, este sonido tendía a imantarse hacia detrás. Puede haber una causa racional en esto. Al escuchar una puerta y no tener ninguna referencia visualmente próxima, el sistema cognitivo puede tender a interpretar que por lógica debe estar situada detrás. Además, el fenómeno cotidiano de una puerta cerrándose, es algo que generalmente sucede de espaldas a nosotros. Se trataría de un fenómeno parecido al planteado por Gardner (1967,1972) que apreció indicios racionales determinantes en la decisión *proximity image effect*.

La voz, especialmente si es anecoica (seca), tiene el mayor grado de acierto en la posición central, y también de confusión del exterior hacia el centro, lo que parece indicar que es más fácilmente internalizable. Este material sonoro es el más difícil de ubicar en general según los comentarios posteriores al test. Esto también podría explicarse de un modo racional, ya que estamos acostumbrados a oír nuestra propia voz internalizada sin características añadidas por entorno. La voz ambiental, en consonancia con esto, obtiene mejores ratios de localización en el exterior y de confusión hacia el exterior.

El ruido rosa y mar tienden a errores de internalización. Aunque son sonidos de características similares, aquí no se aprecia que el contenido simbólico del sonido real aporte algún grado de externalización.

Consideraciones adicionales.

Hay que notar que el diseño experimental propuesto permite para cada muestra respuestas relativas a 8 regiones del plano horizontal por lo que no permite evaluar con precisión el grado de errores delante-detrás para pares específicos (como 0°-180° o 45°-135°) al producirse dispersión en las respuestas

hacia más regiones, lo que también influye en el bajo grado de acierto en general. Sin embargo permite apreciar las tendencias de forma más realista en las confusiones.

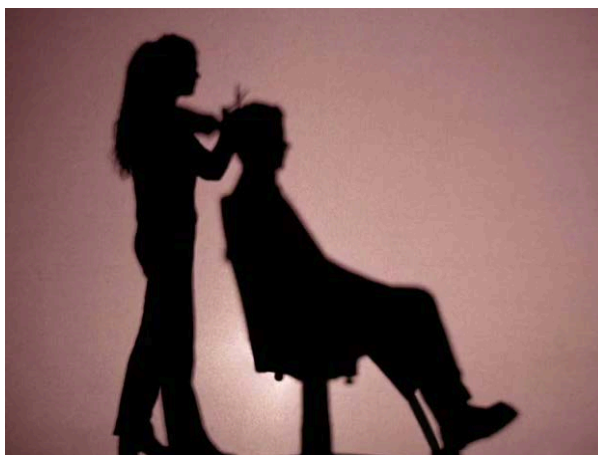
En general se confirma una tendencia de confusión hacia atrás. Este hecho podría explicarse a través de la hipótesis de los *indicios visuales*. En general cuando oímos un sonido y no lo vemos, nuestro cerebro racionalmente lo ubica detrás. En el caso de sonidos ecológicos, esta tendencia podría acentuarse por el hecho de que cualquier sonido real suele llevar una imagen asociada memorísticamente. En las instrucciones indicadas para la realización del test no se especificaba la condición test a ciegas, sin embargo, varios de los sujetos indicaron que en caso de duda, cerrar los ojos al escuchar ciertas muestras ayudaba a la decisión. La condición ojos abiertos - ojos cerrados no ha sido estudiada en este test, pero podría aportar resultados interesantes. El estudio de la influencia perceptiva de imágenes conjuntamente con sonido es una posibilidad claramente interesante a la hora de identificar ciertas tendencias de tipo cognitivo.

Hacia una espacialización binaural

Los resultados, no siendo concluyentes, indican que existe una cierta fiabilidad a la hora de usar herramientas de posproducción de audio para recrear una espacialización aceptable en la escucha binaural. En todo caso, es necesario una estrategia de diseño sonoro que nos permita sacarle el máximo partido.

De los resultados del diseño experimental, podemos extraer que los sonidos impulsivos tienden a destacar sobre los demás en cuanto a realismo. Los sonidos impulsivos se caracterizan por tener un espectro frecuencial homogéneo y amplio, y transitorios de ataque corto. Sonidos como golpes, disparos o sonidos de motor, ya son utilizados de forma efectiva como golpe de efecto inmersivo en las producciones estéreo multicanal cinematográficas.

El producto más popular de diseño binaural, el clásico “*The virtual Barber shop*”, que utiliza precisamente este tipo de trucos para conseguir un efecto inmersivo más convincente.



**“VIRTUAL BARBER SHOP”. PINCHAR EN LA FIGURA PARA REPRODUCIR.
ES NECESARIO ADOBE ACROBAT 9 O SUPERIOR.**

En efecto, el sonido de tijeras o la maquina cortadora de pelo, son sonidos de tipo impulsivo que proporcionan un resultado convincente, mientras la voz o la radio de fondo, no resultan creíbles en todo momento. Este producto utiliza un recurso interesante, el paralaje auditivo. Ya se ha comentado que las diferencias interaurales son más acusadas para frentes de onda esféricos, por lo que la utilización de grabación de campo o cabezas de maniquí, permite exagerar la espacialización sin que por otro lado sea apreciable, ya que en principio, la información de la distancia a la fuente no se transmite de forma muy eficiente de forma binaural, permitiendo “falsear” la espacialización.

Como aplicación a la música audiovisual, se intuye que el sonido de notas de piano, será también por el mismo motivo mas eficaz que por ejemplo los instrumentos de cuerda o de viento, lo cual parece necesario tener en cuenta en un hipotético diseño musical binaural.

Respecto a los diálogos, está bastante establecida la idea de que la percepción de la voz está altamente influida por la referencia visual del interlocutor, lo se suele conocer como *efecto ventriloquismo* (Muñoz y Recuero, 2008) o *vococentrismo* (Chion, 1990): el sistema perceptivo utiliza información

visual correlacionada para corregir una localización irracional de una voz, haciendo inútil cualquier espacialización. Esto llevó a que los diálogos en una obra cinematográfica se sitúen de forma sistemática en el centro y se graben de forma monofónica, haciendo inútil el planteamiento original de Blumlein, que en su día presentó su sistema estéreo (BBC, 2008), como la forma de registrar a el movimiento de los actores sobre la pantalla.

El hecho destacable es que la mezcla binaural parece que podría permitir, aunque de una forma elemental el posible manejo de la voz entre el interior (la voz interior o voz pensamiento) y el exterior (voz narración o diálogos) , de una forma más realista que el estéreo.

En todo caso, el hecho de que la localización no sea eficiente en recreaciones binaurales, no supone una limitación drástica. Las mezclas que se realizan para sonido multicanal, solo suelen usar efectos localizables en determinados pasajes de acción, ya que resultaría contraproducente en base al conocido *efecto bastidor* o "*exit sign effect*" (Holman, 1990), que explica cómo al reconocer un sonido detrás nuestra, tendemos a girar la cabeza y encontrarnos, en el caso de una sala de cine, la luz del letrero de salida, única referencia visual posible. Sin embargo, suele ser común que recurran de forma más sistemática a sonidos ambientales como mar, tráfico, etc., que también son un recurso inmersivo, que aunque menos apreciable, es interesante desde el punto de vista estético.

¿Escucha estéreo o binaural?

En vista de la aparente preponderancia de los indicios visuales, manifestada en el comentado *efecto ventriloquismo*, y que Chion plantea de una forma más general como *imantación del sonido por la imagen*, cabe preguntarse si la escucha binaural ofrece alguna aportación respecto del estéreo convencional.

Los dispositivos móviles, donde el uso de auriculares está extendido, podrían hacer suplir carencias en tamaño de las resoluciones de imagen con contenido sonoro espacial, invirtiendo la preponderancia de la imagen respecto del sonido, aunque haría falta determinar si se produciría un efecto rechazo por incoherencia sonido-imagen, o si el efecto imantación se mantendría. El ámbito de los videojuegos también parece idóneo, ya que se asemeja al caso de oyente no móvil, y donde resulta interesante conseguir efectos inmersivos.

El desarrollo de nuevas posibilidades de reproducción, como las salas de cine *IMAX dome*, donde una pantalla esférica excede el ángulo de visión pero permitiendo que dirijamos la mirada sobrepasando la limitación del efecto ventana clásico, unida a sistemas de auriculares *nearphone* tipo IMAX PSE, y su combinación con la estereoscopía con el formato IMAX sólido, pueden ser opciones válidas para la escucha binaural.

Independientemente de si la mezcla binaural aporta algo interesante a la mezcla de sonido, sí es evidente que la calidad se ve mejorada respecto al estéreo, al conseguirse mejor separación de elementos sonoros.

La retrocompatibilidad del estéreo respecto al mono, era problemática. Desfases indeseados de las señales en la reproducción, podían resultar en la cancelación de una parte importante de de la señal, quedándonos en el caso de una contrafase, solo con las partes diferentes entre los dos canales, es decir sin la parte mono. Este no es el caso de la reproducción binaural, ya que la condición e escucha se ciñe precisamente, a presentar a ambos oídos señales totalmente separadas en el espacio. Eso hace posible una convivencia entre ambos formatos, planteándose la posibilidad de una escucha binaural que podría aportar un valor añadido sobre la estéreo de forma complementaria.

5. CONCLUSIONES.

La audición binaural es un campo de estudio de interés en el campo de la postproducción de sonido. La escucha binaural es un recurso que permite, utilizar características de la localización espacial humana para conseguir una experiencia inmersiva utilizable en aplicaciones audiovisuales mejorando el estéreo tradicional y aportando mejoras en la calidad y las posibilidades creativas, además disponer de posibilidades simplicidad y movilidad.

Sin embargo, una aproximación a la cuestión en profundidad, pone de manifiesto que la eficacia de los métodos de espacialización binaural es limitada, debido a dificultades para transmitir la información espacial de forma efectiva y generalizable. El análisis en profundidad de los fenómenos implicados en la percepción binaural humana mediante experimentación de tipo perceptivo puede mejorar la comprensión del fenómeno, y es especialmente interesante desde la perspectiva del diseño de sonido audiovisual, mediante la utilización de elementos de uso común en esta área, grabación, software, y morfología de los sonidos.

Los resultados del diseño experimental planteado en este trabajo no indican diferencias significativas entre la efectividad de las herramientas de postproducción de uso común estudiadas, y la utilización de grabaciones de campo realizadas con microfonía binaural tiene una efectividad comparablemente limitada.

Sin embargo, sí parece apreciarse una tendencia a que el material sonoro utilizado influya en los resultados. La comparación sonidos con distintas características frecuenciales y dinámicas, parece indicar que en el resultado no es sensible a la estructura frecuencial, pero sí a la dinámica y temporal, en beneficio de sonidos de tipo impulsivo. La simulación de distintos tipos de movimiento en las fuentes empleadas, también parece influir en los resultados, pero de forma dispar y no concluyente.

El estudio detallado del comportamiento de material sonoro que se emplea en postproducción es un campo de estudio a seguir claramente interesante, que puede aportar nuevas estrategias de diseño y especialización de sonido.

REFERENCIAS

6. REFERENCIAS

Bibliografía

- Begault, D. R. (2000). *3-D Sound for Virtual Reality and Multimedia*. Boston: Academic Press. [Online: http://humanfactors.arc.nasa.gov/publications/Begault_2000_3d_Sound_Multimedia.pdf]
- Blauert, J. (1996). *Spatial Hearing: The psychophysics of human sound localization*. Cambridge, MA: MIT Press.
- Blauert, J. (2005). *Communication acoustics*. Berlin: Springer-Verlag.
- Chion, M. (1990). *La audiovision*. Paris: Éditions Nathan.
- Mackensen, P. (2004). *Auditive localization. Head movements, an additional cue in Localization*. PhD dissertation. Berlin: Technischen Universität Berlin, Germany. Jan. 2004. [http://edocs.tu-berlin.de/diss/2004/mackensen_philip.pdf]
- Leung Ngan Ming. (2001). *Optimizing the performance of Virtual Headphone-based Surround Sound (VHSS) Systems by manipulating the spectra of non-individualized HRTFs*. PhD Thesis. Hong kong: The Hong Kong University of Science and Technology.
- Owsinski, B. (1999). *The mixing engineer's handbook*. CA: Mixbooks
- Zwicker and Fastl. (2007). *Psycoacoustic: Facts and models*. Berlin: Springer Verlag.

Artículos científicos

- Algazi, V.R., Duda R.O., Thompson D. M., and Avendano C. (2001). "The CIPIC HRTF database". Davis, CA. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. 99-102.
- Algazi, V.R. and Duda, R. O. (2008). "Effective Use of Psychoacoustics in Motion Tracked Binaural audio ". *Proc. Tenth IEEE International Symposium on Multimedia*. Berkeley, CA. p. 739-746, p. 562-567.
- Algazi, V.R. and Duda, R.O. (2011). "Headphone-based spatial sound". *IEEE Signal Processing Magazine*, Vol. 28, No. 1, pp. 33-42.
- Blumlein, A. (1933). "Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems". *GB patent 394325*.
- Boer, K.D. and Urk, V. (1941). "Some particulars of directional hearing". *Phillips techn. Rev.* 6. p. 359-364.
- Cetta, P. (2003). "Modelos de localización especial del sonido y su implementación en tiempo real". FACM-UCA.
- Féron, F.X., Frissen, I., Boissinot, J. and Guastavino, C. (2010). "Upper limits of auditory rotational motion perception", *J. Acoust Soc Am.* 128(6):370314.
- Gardner, M.B (1967). "Proximity Image Effect in Sound Localization". *J. Acoust. Soc. Am.* p. 163.
- Gardner, M.B. and Gardner, R.S. (1972). "Problem of localization in the median plane: Effect of pinnae cavity occlusion". *J. Acoust. Soc. Am.*, 53(2): 400-408.
- Han, H. (1992). "On the relation between directional bands and head movements". In 92. *AES Convention, Vienna*.
- Hartmann, W. M. (1983). "Localization of sound in rooms". *Journal of the Acoustical Society of America*, 74, 1380-1391.
- Holman, T. (1990). "New Factors in Sound for Cinema and Television". *AES 80th convention. Session-Paper# C-I-2*.
- Middlebrooks J. and Green, D. (1989). "Directional dependence of interaural envelope Delays". *J. Acoust. Soc. Am.*, 87:2149-2162.
- Motokuni I, Kazuhiro I. and Morimoto, M. (2007). "Individual differences in directional

- bands in median plane localization". *Applied Acoustics*, 68, 909-915.
- Muñoz, R. y Recuero. M (2008). "Umbral absoluto de percepción de coherencia de origen entre fuente sonora y visual". *FIA 2008-A160*
- SAE Institute in association with University of Middlesex. (2009). "Comparison between Soundman OKM II Studio Classic and Neumann Art Head KU81i in technical and timbral aspects". *Bachelor-Thesis*. Berlin.
- Seeber, B. U. and Fastl, H. (2003). "Subjective selection of non-individual head-related transfer functions, Boston, MA. *Proceedings of the 2003 International Conference on Auditory Display, Boston, MA.1-4*.
- So,R.H.Y., Leung, N.M, Braasch,J. and Leung K.L. (2006). "A low cost, non-Individualized surround sound system based upon head related transfer functions: An ergonomics study and prototype development". *Applied Ergonomics*. 37 695–707.
- So, R. H. Y. , Ngan, B. , Horner, A., Braasch, J. , Blauert, J. and Leung, K. L. (2010). "Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study", *Ergonomics*. 53: 6, 767 — 781.
- Strybel, T. Z., Manligas, C. L., and Perrott, D. R. (1992). "Minimum Audible Movement Angle as a Function of the Azimuth and Elevation of the Source" *Human Factors*, 34, 267–275.
- Thurlow,W. R., Mangels,J. W. and Runge,P. S.(1967). "Head movements during sound localization" *J. Acoust. Soc. Am.vol.42, no. 2, pp. 489-493*.
- Wallach, H. (1939). "On sound localization," *J. Acoust. Soc. America, vol. 10, no. 4, pp. 270–274*.
- Wallach, H. (1949). "The role of head movements and vestibular and visual cues in sound localization". *Journal of experimental psicology, Vol 2, N° 4*.
- Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). "Localization using non-individualized head-related transfer functions". *Journal of the Acoustical Society of America*, 94,111–123.
- Wenzel, E. M. (1995) "the relative contribution of interaural time and magnitude cues to dynamic sound localization", *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustic., NY. 15-18*.
- Wersényi,G.(2009). "Effect of emulated head-tracking for reducing localization errors

in virtual audio simulation”. *IEEE transactions on audio, speech, and language processing*, vol. 17, no. 2, 247.

Wightman, F.L. and Kistler, D. (1992). “The dominant role of low-frequency interaural time differences in sound localization,” *J. Acoust. Soc. Amer.*, vol. 91, no. 3, pp. 1648–1661.

Wightman, F.L. and Kistler, D. (1999). “Resolution of front–back ambiguity in spatial hearing by listener and source movement” *J. Acoust. Soc. Ame.*, vol. 105, no. 5, pp. 2841–2853.

Manuales

- Apple Inc. (2009). *Logic Pro 9 User manual* [Online resource: <http://documentation.apple.com/en/logicpro/>].
- Creative. (2001). *Environmental Audio Extensions: EAX 2.0. Version 1.3.* [<http://developer.creative.com/games/files/EAX20SDK.exe>]
- Gardner, B. and Martin, K. (1994). *HRTF Measurements of a KEMAR Dummy Head Microphone*. MIT Media Lab Perceptual Computing Technical Report #280
- Mutanen, J.(2003). *I3DL2 and Creative EAX*.
- Wave Arts Inc. (2007). *Wave Arts Power suite 5 manual*. [Online resource: [http://wavearts.com/uploads/docs/Power Suite 5 Manual.pdf](http://wavearts.com/uploads/docs/Power_Suite_5_Manual.pdf)].

Recursos en linea

- Campbell, D. R. (2007). *Roomsim V3.4 User Guide*. [online resource: http://www.mathworks.com/matlabcentral/forums/5184/1/content/Roomsim_3p4p1_small/User%20Guide%20WEBver/Roomsim%20User%20Guide%20v3p4.htm]
- Creative. (*Homepage of Creative EAX*. [<http://eax.creative.com>]
- Creative. *EAX*. [<http://www.creative.com/soundblaster/technology/eax/welcome.aspx>]
- Dolby Labs.[<http://www.dolby.com/professional/technology/games/dolby-headphone.html>]
- BBC. (2008). “Blumlein early stereo recordings restored”. [online: <http://news.bbc.co.uk/2/hi/technology/7537782.stm>]
- IRCAM. [http://recherche.ircam.fr/equipes/salles/listen/system_protocol.html]
- RAE. [online: www.rae.es/.html]
- Smith research. [<http://www.smyth-research.com>]
- SRS. [<http://www.srslabs.com>]
- UC Davis. Binaural tools. [http://www.ece.ucdavis.edu/binaural/binaural_tools.html]
- Yamaha. *Silent cinema*. [http://www.yamaha.co.jp/english/product/av/guide/technologies/silent_cinema_dsp.html].

Obras audiovisuales.

- Pearl Harbor. 60th Anniversary Commemorative Edition.(2001).Dir.Michael Bay.
DVD. Touchtone-buenavista home entertainment.

Virtual Barber shop. (Desconocida). Cetera algorithm. [Online:
www.youtube.com/watch?v=pgeFdOayeaw].

ANEXOS

ANEXO I
TABLAS DE RESULTADOS DE TEST-T

ANEXO II. TABLAS DE RESULTADOS DE TEST-T.

Acierto dentro-fuera

Estadísticos de muestras relacionadas

		Media	N	Desviación típ.	Error típ. de la media
Par 1	voz reverb	,67	66	,475	,058
	voz seca	,48	66	,504	,062
Par 2	voz reverb	,74	110	,443	,042
	puerta	,84	110	,372	,035
Par 3	voz reverb	,91	22	,294	,063
	mar	,95	22	,213	,045
Par 4	voz reverb	,71	77	,455	,052
	rosa	,69	77	,466	,053
Par 5	voz seca	,73	143	,447	,037
	puerta	,87	143	,333	,028
Par 6	voz seca	,73	44	,451	,068
	mar	,84	44	,370	,056
Par 7	voz seca	,73	143	,447	,037
	rosa	,80	143	,398	,033
Par 8	puerta	,93	44	,255	,038
	mar	,84	44	,370	,056
Par 9	puerta	,87	143	,333	,028
	rosa	,80	143	,398	,033
Par 10	mar	,87	55	,336	,045
	rosa	,85	55	,356	,048

Acierto dentro-fuera

Prueba de muestras relacionadas

		Diferencias relacionadas							
		Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 1	voz reverb - voz seca	,182	,630	,078	,027	,337	2,344	65	,022
Par 2	voz reverb - puerta	-,100	,524	,050	-,199	-,001	-2,003	109	,048
Par 3	voz reverb - mar	-,045	,213	,045	-,140	,049	-1,000	21	,329
Par 4	voz reverb - rosa	,026	,606	,069	-,112	,164	,376	76	,708
Par 5	voz seca - puerta	-,147	,543	,045	-,237	-,057	-3,232	142	,002
Par 6	voz seca - mar	-,114	,538	,081	-,277	,050	-1,402	43	,168
Par 7	voz seca - rosa	-,077	,461	,039	-,153	-,001	-1,996	142	,048
Par 8	puerta - mar	,091	,421	,064	-,037	,219	1,431	43	,160
Par 9	puerta - rosa	,070	,422	,035	,000	,140	1,981	142	,050
Par 10	mar - rosa	,018	,408	,055	-,092	,128	,331	54	,742

**Respuesta dentro-fuera
Estadísticos de muestras relacionadas^a**

		Media	N	Desviación típ.	Error típ. de la media
Par 1	voz reverb	,71	66	,456	,056
	voz seca	,44	66	,500	,062
Par 2	voz reverb	,76	110	,427	,041
	puerta	,94	110	,245	,023
Par 3	voz reverb	,91	22	,294	,063
	mar	,95	22	,213	,045
Par 4	voz reverb	,75	77	,434	,049
	rosa	,73	77	,448	,051
Par 5	voz seca	,71	143	,457	,038
	puerta	,95	143	,217	,018
Par 6	voz seca	,73	44	,451	,068
	mar	,84	44	,370	,056
Par 7	voz seca	,71	143	,457	,038
	rosa	,83	143	,381	,032
Par 8	puerta	,93	44	,255	,038
	mar	,84	44	,370	,056
Par 9	puerta	,95	143	,217	,018
	rosa	,83	143	,381	,032
Par 10	mar	,87	55	,336	,045
	rosa	,85	55	,356	,048

**Respuesta dentro-fuera
Prueba de muestras relacionadas^a**

		Diferencias relacionadas							
		Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 1	voz reverb - voz seca	,273	,596	,073	,126	,419	3,717	65	,000
Par 2	voz reverb - puerta	-,173	,504	,048	-,268	-,077	3,592	109	,000
Par 3	voz reverb - mar	-,045	,213	,045	-,140	,049	1,000	21	,329
Par 4	voz reverb - rosa	,026	,606	,069	-,112	,164	,376	76	,708
Par 5	voz seca - puerta	-,245	,507	,042	-,328	-,161	5,778	142	,000
Par 6	voz seca - mar	-,114	,538	,081	-,277	,050	1,402	43	,168
Par 7	voz seca - rosa	-,119	,452	,038	-,194	-,044	3,147	142	,002
Par 8	puerta - mar	,091	,421	,064	-,037	,219	1,431	43	,160
Par 9	puerta - rosa	,126	,409	,034	,058	,193	3,682	142	,000
Par 10	mar - rosa	,018	,408	,055	-,092	,128	,331	54	,742

**Diferencia delante-detrás
Estadísticos de muestras relacionadas**

		Media	N	Desviación típ.	Error típ. de la media
Par 1	voz reverb	,03	66	1,136	,140
	voz seca	-,18	66	,975	,120
Par 2	voz reverb	,02	110	1,092	,104
	puerta	,30	110	1,010	,096
Par 3	voz reverb	-,32	22	,894	,191
	mar	-,68	22	,894	,191
Par 4	voz reverb	,05	77	1,075	,122
	rosa	-,05	77	1,111	,127
Par 5	voz seca	,00	142	,989	,083
	puerta	,46	142	1,056	,089
Par 6	voz seca	,02	43	1,058	,161
	mar	-,05	43	1,214	,185
Par 7	voz seca	,00	142	,989	,083
	rosa	,06	142	1,099	,092
Par 8	puerta	,41	44	1,187	,179
	mar	-,02	44	1,210	,182
Par 9	puerta	,48	143	1,060	,089
	rosa	,08	143	1,107	,093
Par 10	mar	-,09	55	1,127	,152
	rosa	,02	55	1,027	,139

**Diferencia delante-detrás
Prueba de muestras relacionadas**

		Diferencias relacionadas							
		Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 1	voz reverb - voz seca	,212	,832	,102	,008	,417	2,070	65	,042
Par 2	voz reverb - puerta	-,282	1,042	,099	-,479	-,085	-2,838	109	,005
Par 3	voz reverb - mar	,364	1,093	,233	-,121	,848	1,560	21	,134
Par 4	voz reverb - rosa	,104	,940	,107	-,109	,317	,970	76	,335
Par 5	voz seca - puerta	-,465	,805	,068	-,598	-,331	-6,882	141	,000
Par 6	voz seca - mar	,070	1,009	,154	-,241	,380	,453	42	,653
Par 7	voz seca - rosa	-,063	,844	,071	-,203	,077	-,895	141	,372
Par 8	puerta - mar	,432	,925	,139	,151	,713	3,096	43	,003
Par 9	puerta - rosa	,399	,987	,083	,235	,562	4,830	142	,000
Par 10	mar - rosa	-,109	,936	,126	-,362	,144	-,864	54	,391

Respuesta delante-detrás Estadísticos de muestras relacionadas

		Media	N	Desviación típ.	Error típ. de la media
Par 1	voz reverb	-,20	66	,808	,099
	voz seca	,02	66	,644	,079
Par 2	voz reverb	-,12	110	,775	,074
	puerta	-,40	110	,826	,079
Par 3	voz reverb	-,18	22	,733	,156
	mar	,18	22	,853	,182
Par 4	voz reverb	-,19	77	,779	,089
	rosa	-,09	77	,747	,085
Par 5	voz seca	-,01	142	,658	,055
	puerta	-,47	142	,778	,065
Par 6	voz seca	-,05	43	,688	,105
	mar	,02	43	,831	,127
Par 7	voz seca	-,01	142	,658	,055
	rosa	-,07	142	,769	,064
Par 8	puerta	-,41	44	,816	,123
	mar	,02	44	,821	,124
Par 9	puerta	-,48	143	,777	,065
	rosa	-,08	143	,770	,064
Par 10	mar	,09	55	,800	,108
	rosa	-,02	55	,733	,099

Respuesta delante-detrás Prueba de muestras relacionadas

		Diferencias relacionadas							
		Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 1	voz reverb - voz seca	-,212	,832	,102	-,417	-,008	-2,070	65	,042
Par 2	voz reverb - puerta	,282	1,042	,099	,085	,479	2,838	109	,005
Par 3	voz reverb - mar	-,364	1,093	,233	-,848	,121	-1,560	21	,134
Par 4	voz reverb - rosa	-,104	,940	,107	-,317	,109	-,970	76	,335
Par 5	voz seca - puerta	,465	,805	,068	,331	,598	6,882	141	,000
Par 6	voz seca - mar	-,070	1,009	,154	-,380	,241	-,453	42	,653
Par 7	voz seca - rosa	,063	,844	,071	-,077	,203	,895	141	,372
Par 8	puerta - mar	-,432	,925	,139	-,713	-,151	-3,096	43	,003
Par 9	puerta - rosa	-,399	,987	,083	-,562	-,235	-4,830	142	,000
Par 10	mar - rosa	,109	,936	,126	-,144	,362	,864	54	,391

Estadísticos de muestras relacionadas

Variable			Media	N	Desviación típ.	Error típ. de la media	
acierto delante- detrás	Par 1	soundman	,23	22	,429	,091	
		matlab	,55	22	,510	,109	
	Par 2	soundman	,41	22	,503	,107	
		protools	,36	22	,492	,105	
	Par 3	soundman	,41	22	,503	,107	
		logic	,45	22	,510	,109	
	Par 4	matlab	,40	176	,491	,037	
		protools	,40	176	,492	,037	
	Par 5	matlab	,40	176	,491	,037	
		logic	,44	176	,498	,038	
	Par 6	protools	,40	187	,491	,036	
		logic	,43	187	,496	,036	
acierto dentro-fuera	Par 1	soundman	,18	22	,395	,084	
		matlab	,91	22	,294	,063	
	Par 2	soundman	,36	22	,492	,105	
		protools	,59	22	,503	,107	
	Par 3	soundman	,36	22	,492	,105	
		logic	,82	22	,395	,084	
	Par 4	matlab	,86	176	,350	,026	
		protools	,86	176	,350	,026	
	Par 5	matlab	,86	176	,350	,026	
		logic	,88	176	,332	,025	
	Par 6	protools	,85	187	,358	,026	
		logic	,87	187	,341	,025	
diferencia delante- detrás	Par 1	soundman	,14	22	,889	,190	
		matlab	-,32	22	,894	,191	
	Par 2	soundman	-,14	22	,774	,165	
		protools	,14	22	1,167	,249	
	Par 3	soundman	-,14	22	,774	,165	
		logic	,41	22	1,182	,252	
	Par 4	matlab	,05	175	1,089	,082	
		protools	,06	175	1,087	,082	
	Par 5	matlab	,05	175	1,089	,082	
		logic	,22	175	1,012	,077	
	Par 6	protools	,12	187	1,096	,080	
		logic	,29	187	1,028	,075	
diferencia dentro- fuera	Par 1	soundman	-0,818181818	22	0,394771017	0,084165464	
		matlab	0,090909091	22	0,294244943	0,062733233	
	Par 2	soundman	-,64	22	,492	,105	
		protools	,41	22	,503	,107	
	Par 3	soundman	-,64	22	,492	,105	
		logic	,18	22	,395	,084	
	Par 4	matlab	,14	176	,350	,026	
		protools	,14	176	,350	,026	
	Par 6	protools	,15	187	,358	,026	
		logic	,13	187	,341	,025	
	respuesta delante- detrás	Par 1	soundman	-,14	22	,889	,190
			matlab	-,18	22	,733	,156
Par 2		soundman	,14	22	,774	,165	
		protools	-,14	22	,774	,165	
Par 3		soundman	,14	22	,774	,165	
		logic	-,41	22	,796	,170	
Par 4		matlab	-,12	175	,760	,057	
		protools	-,13	175	,778	,059	
Par 5		matlab	-,12	175	,760	,057	
		logic	-,29	175	,728	,055	
Par 6		protools	-,12	187	,781	,057	
		logic	-,29	187	,727	,053	
respuesta dentro- fuera	Par 1	soundman	,82	22	,395	,084	
		matlab	,91	22	,294	,063	
	Par 2	soundman	,64	22	,492	,105	
		protools	,59	22	,503	,107	
	Par 3	soundman	,64	22	,492	,105	
		logic	,82	22	,395	,084	
	Par 4	matlab	,86	176	,350	,026	
		protools	,86	176	,350	,026	
	Par 5	matlab	,86	176	,350	,026	
		logic	,88	176	,332	,025	
	Par 6	protools	,85	187	,358	,026	
		logic	,87	187	,341	,025	

Prueba de muestras relacionadas*										
Variable			Diferencias relacionadas							
			Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia		t	gl	Sig. (bilateral)
						Inferior	Superior			
acierto delante-detrás	Par 1	soundman - matlab	-,318	,780	,166	-,664	,028	-1,914	21	,069
	Par 2	soundman - protools	,045	,844	,180	-,329	,420	,253	21	,803
	Par 3	soundman - logic	-,045	,722	,154	-,366	,275	-,295	21	,771
	Par 4	matlab - protools	-,006	,619	,047	-,098	,086	-,122	175	,903
	Par 5	matlab - logic	-,045	,584	,044	-,132	,041	-1,033	175	,303
	Par 6	protools - logic	-,027	,553	,040	-,107	,053	-,661	186	,509
acierto dentro-fuera	Par 1	soundman - matlab	-,727	,550	,117	-,971	-,483	-6,197	21	,000
	Par 2	soundman - protools	-,227	,869	,185	-,613	,158	-1,226	21	,234
	Par 3	soundman - logic	-,455	,739	,157	-,782	-,127	-2,887	21	,009
	Par 4	matlab - protools	,000	,355	,027	-,053	,053	,000	175	1,000
	Par 5	matlab - logic	-,017	,362	,027	-,071	,037	-,624	175	,533
	Par 6	protools - logic	-,016	,319	,023	-,062	,030	-,687	186	,493
diferencia abs. Delante-detrás diferencia absoluta posición diferencia dentro-fuera	Par 1	soundman - matlab	,455	1,184	,252	-,071	,980	1,800	21	,086
	Par 2	soundman - protools	-,273	1,279	,273	-,840	,294	-1,000	21	,329
	Par 3	soundman - logic	-,545	1,299	,277	-1,122	,031	-1,969	21	,062
	Par 4	matlab - protools	-,006	,894	,068	-,139	,128	-,085	174	,933
	Par 5	matlab - logic	-,171	,854	,065	-,299	-,044	-2,656	174	,009
	Par 6	protools - logic	-,171	,818	,060	-,289	-,053	-2,860	186	,005
diferencia abs. Izquierda derecha diferencia absoluta posición	Par 1	soundman - matlab	-,909	,426	,091	-1,098	-,720	-10,000	21	,000
	Par 2	soundman - protools	-,1045	,486	,104	-1,261	-,830	-10,095	21	,000
	Par 3	soundman - logic	-,818	,501	,107	-1,040	-,596	-7,659	21	,000
	Par 4	matlab - protools	,000	,355	,027	-,053	,053	,000	175	1,000
	Par 5	matlab - logic	,017	,362	,027	-,037	,071	,624	175	,533
	Par 6	protools - logic	,016	,319	,023	-,030	,062	,687	186	,493
	Par 4	matlab - protools	,006	,509	,038	-,070	,082	,149	174	,882
	Par 5	matlab - logic	,000	,557	,042	-,083	,083	,000	174	1,000
	Par 6	protools - logic	,000	,539	,039	-,078	,078	,000	186	1,000
diferencia posición respuesta delante-detrás	Par 4	matlab - protools	-,107	2,087	,176	-,456	,242	-,607	139	,545
	Par 5	matlab - logic	,121	1,973	,166	-,208	,449	,726	140	,469
	Par 6	protools - logic	,053	1,656	,135	-,213	,319	,393	150	,695
	Par 1	soundman - matlab	,045	1,214	,259	-,493	,584	,176	21	,862
	Par 2	soundman - protools	,273	,985	,210	-,164	,709	1,299	21	,208
	Par 3	soundman - logic	,545	1,011	,215	,097	,994	2,531	21	,019
diferencia posición respuesta dentro-fuera	Par 4	matlab - protools	,006	,894	,068	-,128	,139	,085	174	,933
	Par 5	matlab - logic	,171	,854	,065	,044	,299	2,656	174	,009
	Par 6	protools - logic	,171	,818	,060	,053	,289	2,860	186	,005

Variable = Respuesta	Movimiento				
Estadísticos de muestras relacionadas*					
		Media	N	Desviación típ.	Error típ. de la media
Par 8	Mov 5	,18	22	,395	,084
	Mov -10	,32	22	,477	,102
Par 9	Mov 5	0,18	22,00	0,39	0,08
	Mov +10	0,41	22,00	0,50	0,11
Par 10	Mov 5	0,09	11,00	0,30	0,09
	Mov -20	0,55	11,00	0,52	0,16
Par 11	Mov 5	0,18	22,00	0,39	0,08
	Mov +20	0,59	22,00	0,50	0,11
Par 12	Mov 5	0,18	22,00	0,39	0,08
	Mov -30	0,91	22,00	0,29	0,06
Par 13	Mov 5	0,18	22,00	0,39	0,08
	Mov +30	0,77	22,00	0,43	0,09
Par 14	Mov -10	,19	43	,394	,060
	Mov +10	,26	43	,441	,067
Par 15	Mov -10	,19	21	,402	,088
	Mov -20	,48	21	,512	,112
Par 16	Mov -10	,21	33	,415	,072
	Mov +20	,58	33	,502	,087
Par 17	Mov -10	,16	64	,366	,046
	Mov -30	,83	64	,380	,048
Par 18	Mov -10	,21	33	,415	,072
	Mov +30	,73	33	,452	,079
Par 19	Mov +10	,55	11	,522	,157
	Mov -20	,55	11	,522	,157
Par 20	Mov +10	,36	33	,489	,085
	Mov +20	,73	33	,452	,079
Par 21	Mov +10	,23	66	,422	,052
	Mov -30	,77	66	,422	,052
Par 22	Mov +10	,30	44	,462	,070
	Mov +30	,73	44	,451	,068
Par 23	Mov -20	,55	11	,522	,157
	Mov +20	,64	11	,505	,152
Par 24	Mov -20	,50	22	,512	,109
	Mov -30	,77	22	,429	,091
Par 25	Mov -20	,55	11	,522	,157
	Mov +30	,64	11	,505	,152
Par 26	Mov -30	,85	55	,356	,048
	Mov +30	,71	55	,458	,062

Prueba de muestras relacionadas*									
		Diferencias relacionadas							
		Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 8	Mov 5 - Mov -10	-0,14	0,47	0,10	-0,34	0,07	-1,37	21,00	0,19
Par 9	Mov 5 - Mov +10	-0,23	0,61	0,13	-0,50	0,04	-1,74	21,00	0,10
Par 10	Mov 5 - Mov -20	-,455	,522	,157	-,805	-,104	-2,887	10	,016
Par 11	Mov 5 - Mov +20	-,409	,503	,107	-,632	-,186	-3,813	21	,001
Par 12	Mov 5 - Mov -30	-,727	,456	,097	-,929	-,525	-7,483	21	,000
Par 13	Mov 5 - Mov +30	-,591	,590	,126	-,853	-,329	-4,695	21	,000
Par 14	Mov -10 - Mov +10	-,070	,507	,077	-,226	,086	-,903	42	,372
Par 15	Mov -10 - Mov -20	-,286	,561	,122	-,541	-,031	-2,335	20	,030
Par 16	Mov -10 - Mov +20	-,364	,549	,096	-,558	-,169	-3,807	32	,001
Par 17	Mov -10 - Mov -30	-,672	,473	,059	-,790	-,554	11,358	63	,000
Par 18	Mov -10 - Mov +30	-,515	,566	,098	-,716	-,315	-5,231	32	,000
Par 19	Mov +10 - Mov -20	,000	,775	,234	-,520	,520	,000	10	1,000
Par 20	Mov +10 - Mov +20	-,364	,653	,114	-,595	-,132	-3,200	32	,003
Par 21	Mov +10 - Mov -30	-,545	,560	,069	-,683	-,408	-7,917	65	,000
Par 22	Mov +10 - Mov +30	-,432	,587	,088	-,610	-,253	-4,883	43	,000
Par 23	Mov -20 - Mov +20	-,091	,539	,163	-,453	,271	-,559	10	,588
Par 24	Mov -20 - Mov -30	-,273	,550	,117	-,517	-,029	-2,324	21	,030
Par 25	Mov -20 - Mov +30	-,091	,539	,163	-,453	,271	-,559	10	,588
Par 26	Mov -30 - Mov +30	,145	,488	,066	,014	,277	2,213	54	,031

ANEXO II. FORMULARIO PDF INTERACTIVO.

**CONTIENE EL TEST PERCEPTIVO CON LAS
MUESTRAS DE AUDIO.**

REQUIERE ADOBE ACROBAT 9 O SUPERIOR.

Test de Percepción: Localización binaural.

Disponemos de una serie de sonidos: ruido rosa, mar, voz y una puerta, que están localizados en varias direcciones del espacio en el plano horizontal, es decir, alrededor de nuestra cabeza y a la altura de los oídos.

El objetivo del test es identificar la región donde aparentemente están situados los sonidos. Para ello se utiliza como referencia una gráfica, donde el plano horizontal (visto desde arriba) se divide en 8 áreas y una región central, donde se supone que nos encontramos.



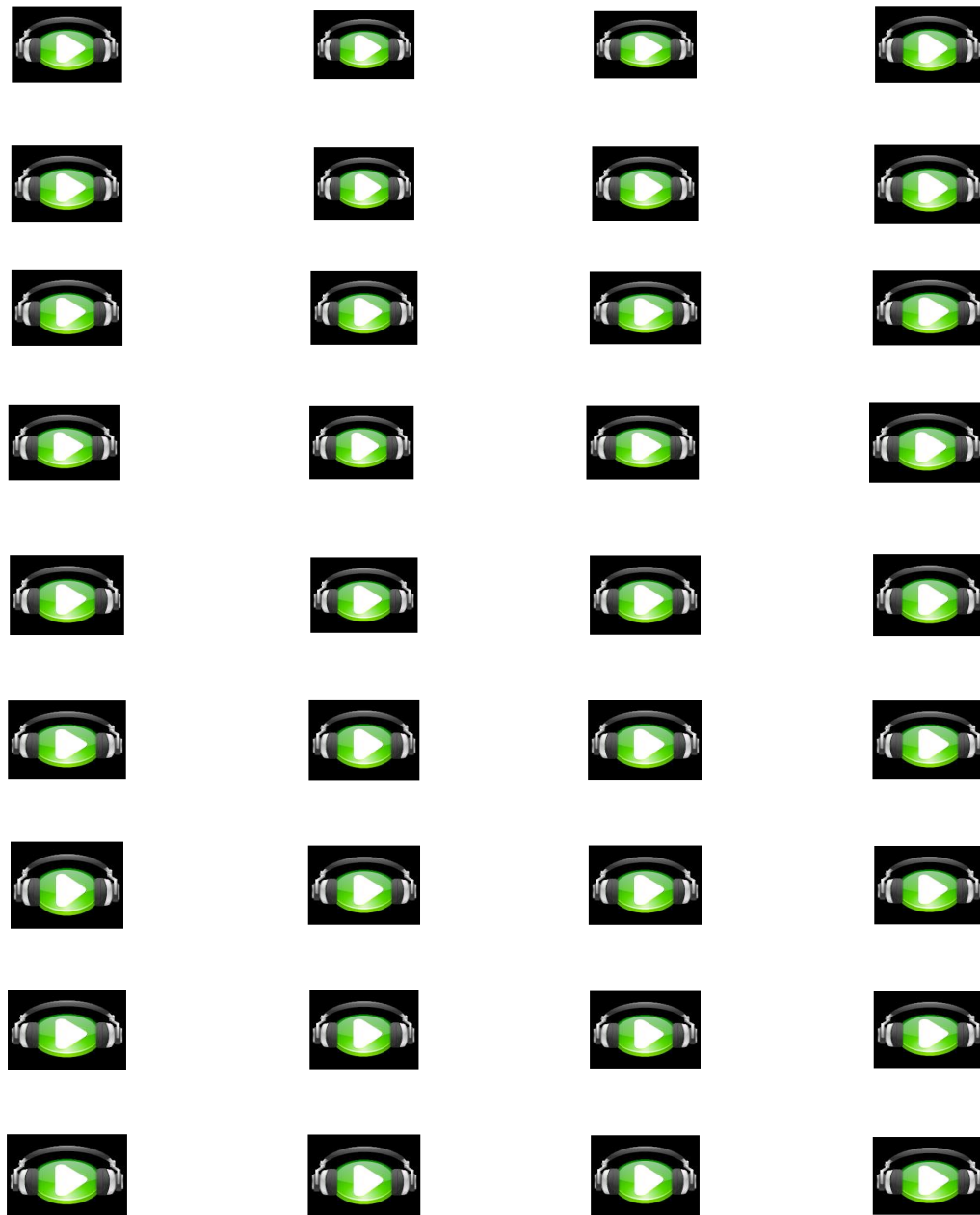
Instrucciones: La escucha debe hacerse con auriculares.

Los sonidos pueden escucharse más de una vez si es necesario.

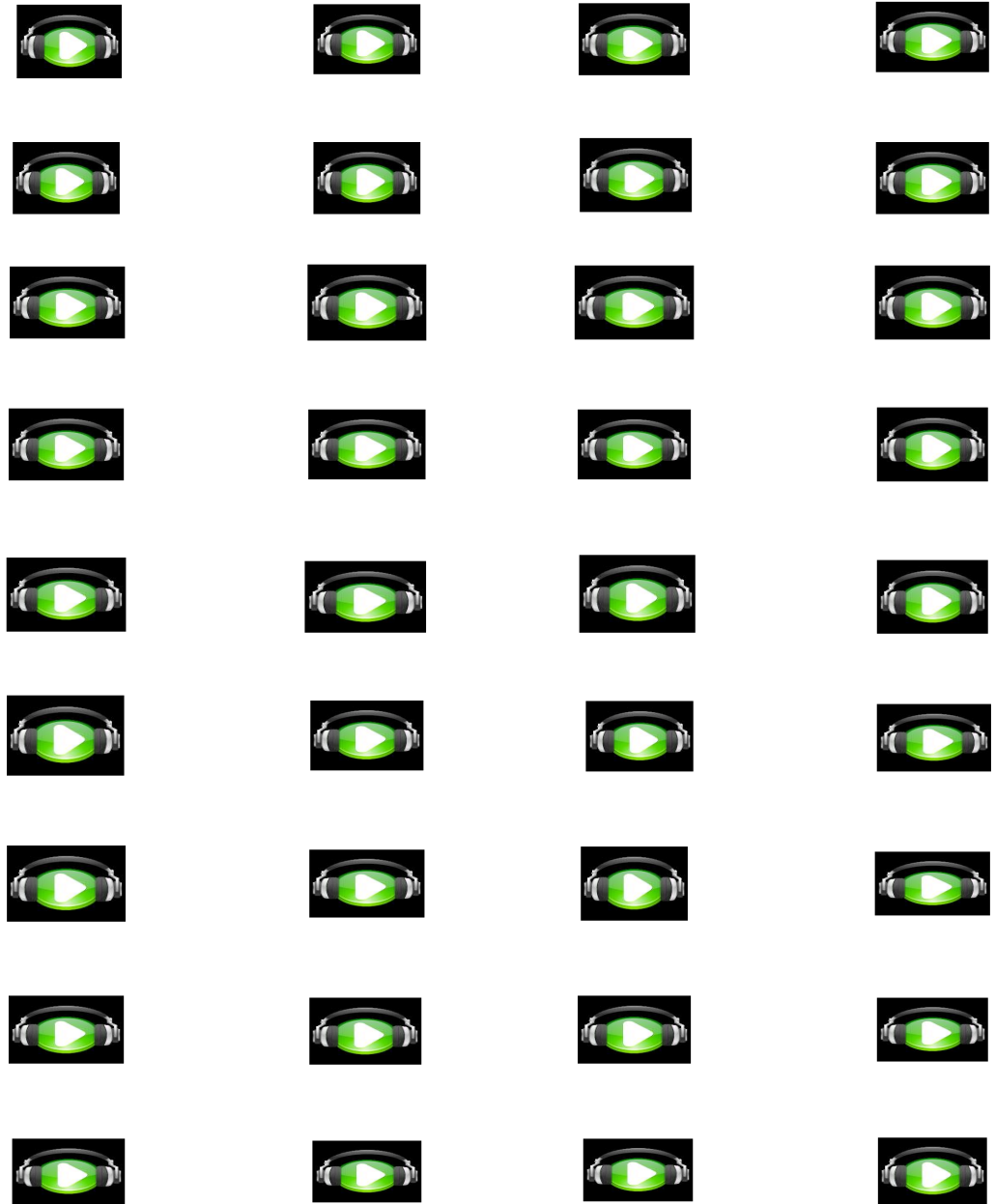
Se necesita [Adobe reader 9 o superior](#).

Test 1.

Pincha en cada icono para escuchar los sonidos.
Identifica la dirección de donde proviene el sonido
según la figura e introduce el valor numérico
correspondiente en la pestaña desplegable.

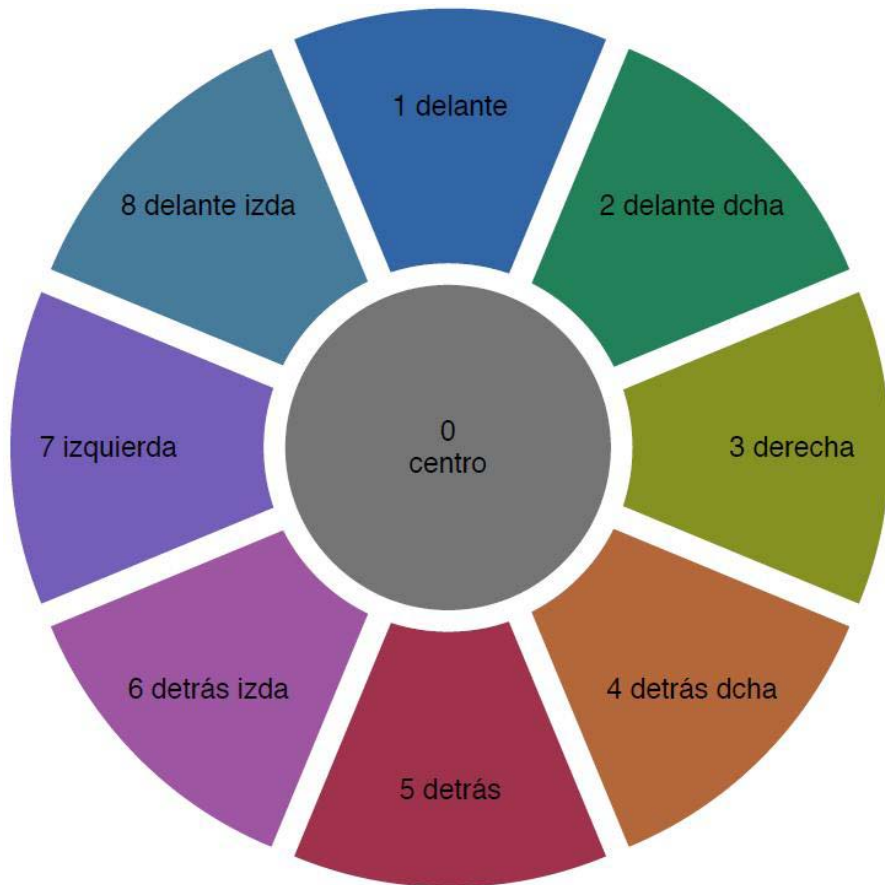


Test 1 (continuación)



Test 2.

Identifica la dirección de donde proviene el sonido como en el test 1 e indica además si crees que el sonido es estático o se mueve.



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve

Test 2 (continuación)



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve



Se mueve
No se mueve

¡FIN!

Solo un par de cosas más:

Escribe tu nombre:

Pon aquí el modelo de auriculares que has utilizado

Una Dos Tres Más

¿Cuántas veces, más o menos, has escuchado cada sonido?

Recuerda guardar el documento antes de cerrar.

¡Graciñas!