

UNIVERSIDAD POLITECNICA DE VALENCIA

ESCUELA POLITECNICA SUPERIOR DE GANDIA

---



UNIVERSIDAD  
POLITECNICA  
DE VALENCIA



ESCUELA POLITECNICA  
SUPERIOR DE GANDIA

**“Procesador de sonido y estudio de  
métodos de interpolación de la  
localización de fuentes basados en  
HRTFs para la generación de audio 3D”**

***TESIS DE MASTER***

Autor:  
**Karmelo Usategi de la Peña**

Director/es:  
***D. Miguel Ferrer Contreras***

***GANDIA, 2010***

# Procesador de sonido y estudio de métodos de interpolación de la localización de fuentes basados en HRTFs para la generación de audio 3D

*Autor:* Karmelo Usategi de la Peña

*Director:* Miguel Ferrer Contreras

*Resumen — Debido a la creciente investigación entorno a la escucha espacial y el sonido holofónico, la comunidad científica está invirtiendo cada vez más recursos y tiempo en la investigación de las Head Related Transfer Function (HRTF) y su relación con la localización de fuentes en el espacio.*

*Como ya está establecido en la bibliografía disponible sobre el tema, las HRTF contienen toda la información necesaria para la creación de la localización de fuentes en un entorno virtual 3D. Por lo tanto, si dichas HRTFs son presentadas a un oyente, mediante una señal monoaural filtrada, éste percibe la localización de la fuente tal y como se localizaría en un entorno real.*

*En el presente trabajo, ya que las HRTFs son medidas siempre en posiciones discretas del espacio, se pretende analizar distintos métodos de interpolación de HRTFs compatibles con la implementación a tiempo real de las mismas para generar la localización virtual de las fuentes en todo el espacio 3D, y así poder establecer tanto objetivamente como subjetivamente la importancia relativa de interpolar o no interpolar las señales de salida de las HRTFs.*

*Además, se pretende crear un interfaz que permita al usuario situar fuentes en cualquier punto del espacio tridimensional de forma dinámica y seleccionar el método de interpolación para la generación de fuentes virtuales en puntos donde no se disponen las HRTF's, de entre las diferentes estrategias desarrolladas. Se pretende así, que dicha interfaz realice el filtrado de la señal deseada por una HRTF determinada en tiempo real mediante la utilidad Playrec que da acceso a tiempo real a las tarjetas de sonido sobre la plataforma MatLab.*

*Abstract — Due to the constant increase of the research on the spatial hearing and holophonic sound, the science community is spending more and more money and time on the research of Head Related Transfer Functions (HRTF) and it's relation with the sound source localization.*

*As it can be found in the bibliography available around the topic, the HRTF included all the necessary information for the creation of virtual sound localization in 3D space. So if they are presented to a subject, via filtering with a monoaural signal, the subject feel the source localization like it was located on a real world situation.*

*In the present work, due to the fact that HRTFs are measured in discrete positions around the entire 3D space, we try to analyse some different HRTF interpolation methods, all of them compatible with its real time implementation, to create the sound source localization in the entire 3D space, and establish the relative importance of interpolating or not output signals from HRTF subjectively and objectively.*

*Apart from that, we also pretend to create an interface that allows the user to locate a virtual sound source in any point of the entire 3D space, giving the user the possibility to choose dynamically between different type of signals and different type of interpolation methods that have been created. We pretend to build a real time signal filtering of HRTF using Playrec utility that provides real time access to PC soundcards on MatLab.*

**INDICE**

<b>0. Objeto y motivación.....</b>	<b>3</b>
<b>1. Escucha espacial y localización de fuentes.....</b>	<b>3</b>
I.1. Determinación del ángulo lateral.....	4
I.2. Determinación de la elevación.....	5
I.3. Ambigüedad en la localización.....	6
I.4. Head Related Transfer Function (HRTF) y Head Related Impulse Response (HRIR).....	7
<b>2. Espacialización del sonido (sound spatialization) y holofonía.....</b>	<b>9</b>
II.1. Grabación con maniquí o dummy.....	9
II.2. Sistemas basados en el filtrado por HRTF.....	10
<b>3. Método experimental.....</b>	<b>12</b>
III.1. Descripción de los medios utilizados.....	12
III.1.1. Biblioteca de HRTFs ofrecida por el IRCAM.....	13
III.1.2. Utilidad Playrec.....	15
III.2. Generación del campo sonoro a demanda.....	16
III.2.1. Interfaz gráfica generada mediante la herramienta GUIDE de MatLab.....	16
III.2.2. Código asociado a la interfaz gráfica.....	19
III.2.3. Código asociado a Playrec.....	22
III.3. Métodos de interpolación desarrollados.....	25
III.3.1. Aproximación por HRTF más cercana: Interpolación por umbrales.....	26
III.3.2. Ponderación lineal inversamente proporcional a la distancia.....	27
III.3.3. Ponderación cuadrática de las señales de salida.....	29
III.4. Interfaces gráficas generadas: Capacidades y limitaciones.....	31
III.5. Test psicoacústico para la toma de información.....	35
<b>4. Resultados.....</b>	<b>37</b>
IV.1. Comparación numérica de los métodos de interpolación.....	38
IV.2. Comparación psicoacústica de los métodos de interpolación.....	42
<b>5. Conclusiones y líneas futuras.....</b>	<b>49</b>
<b>AGRADECIMIENTOS.....</b>	<b>50</b>
<b>BIBLIOGRAFÍA.....</b>	<b>51</b>

## 0. Objeto y motivación

En el presente proyecto se pretende desarrollar una herramienta capaz de filtrar pistas de audio de modo que el oyente pueda situarlas virtualmente en ‘cualquier’ punto del espacio tridimensional dinámicamente. Se pretende que las posiciones virtuales de las fuentes puedan variar a lo largo de la reproducción de las pistas (en tiempo real) para lo cual se utilizan las Head Related Transfer Functions (HRTF).

Puesto que las bases de datos de las HRTF existentes sólo ofrecen dichas respuestas en algunos puntos discretos de todo el espacio, se hace hincapié en el estudio de distintos métodos de interpolación de dichas respuestas para colocar virtualmente las fuentes en puntos donde no se disponen de HRTFs y observar su resultado en distintos sujetos, intentando observar la sensación subjetiva.

Para realizar este trabajo se ha hecho uso de las bibliotecas de HRTFs y HRIRs existentes (Head Related Transfer Functions y Head Related Impulse Responses) de acceso público, creadas por diferentes laboratorios de diferentes países como EEUU, Francia o Japón y la notable investigación generada durante los últimos años en torno a las HRTFs y la localización de fuentes en el espacio, así como del campo sonoro en 3D mediante la reproducción binaural, de corte más reciente, o la escucha espacial (spatial hearing).

Así, para alcanzar el objetivo, se ha utilizado Matlab como entorno de programación y la librería PlayRec para la interacción en tiempo real con la tarjeta de sonido del PC desde Matlab.

## 1. Escucha espacial y localización de fuentes

Para poder determinar la ubicación de una fuente sonora en el espacio, el sistema auditivo utiliza un conjunto de pistas que es necesario situar dentro de un sistema de coordenadas para poder estudiarlas. Como la cabeza puede aproximarse con una esfera, la comunidad científica utiliza como origen el centro de la cabeza y el acimut, la elevación y la distancia como variables que definen cualquier punto del espacio alrededor de esta. La forma más usual de definir dichas coordenadas es el sistema de coordenadas vertical-polar. Con esta elección, las superficies de acimut constante son planos a través del eje Z y las superficies de elevación constante son conos concéntricos alrededor del eje Z [1]. Para clarificarlo:

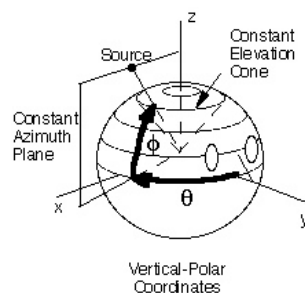


Fig.1. Detalle de la definición del sistema de coordenadas vertical-polar.

Alrededor de hace 100 años, John Strutt, más conocido como Lord Rayleigh fue pionero en la investigación de la escucha espacial y desarrolló una teoría que aún se utiliza a día de hoy, en la cual se presentan dos sistemas principales para la determinación de la teoría de localización de fuentes, que son las Diferencias Interaurales de Tiempo (ITD) y las Diferencias Interaurales de Intensidad (IID) [2][3].

A continuación se describen los aspectos más importantes para la localización de las fuentes tanto en el plano horizontal como en el vertical. No se hará mención ninguna a los mecanismos de diferenciación de la distancia del sujeto a la fuente ya que en este trabajo no se tiene en cuenta dicha variable y se suponen todas las fuentes virtuales a la misma distancia del sujeto, puesto que el oído humano es mejor diferenciando el acimut que la elevación y a su vez es mejor diferenciando la elevación de la fuente que la distancia del sujeto a la misma [3].

### 1.1. Determinación del ángulo lateral

Las Diferencias Interaurales de Intensidad (IID) y las Diferencias Interaurales de Tiempo (ITD) se dan a partir de las diferentes distancias que deben recorrer las ondas para llegar a cada uno de los oídos, y por lo tanto, de las diferentes atenuaciones que sufren en el camino, incluida la sombra acústica producida por la cabeza, y los retrasos entre las señales captadas por cada uno de los oídos. Es por tanto, un proceso que se realiza con una escucha binaural.

Como es sabido, la cabeza es un obstáculo efectivo únicamente para las altas frecuencias (por encima de los 1000Hz), concretamente para aquellas que tienen longitudes de onda menores a la mitad del radio de la cabeza [3], por lo que se ha comprobado que la cabeza se comporta como un filtro acústico paso bajo y además, que no es posible localizar una fuente en el plano medio del espacio únicamente mediante Diferencias Interaurales de Intensidad (IID).

Por ello, Lord Rayleigh propone en su teoría las Diferencias Interaurales de Tiempo (ITD), que para sonidos senoidales sería la diferencia de fase o tiempo entre las ondas sonoras que llegan a cada uno de los oídos, debido a la orientación de la cabeza y la longitud de onda de los sonidos captados [2].

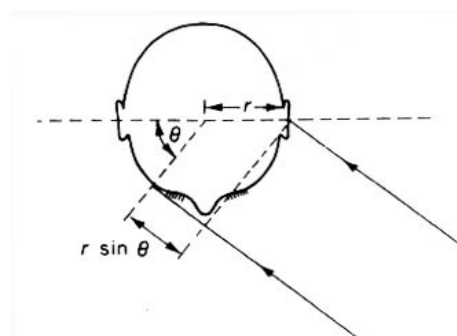


Fig.2. Diferencia Interaural de Tiempo en función del acimut.

Si, por ejemplo, la distancia entre los oídos es de aproximadamente 20cm, supone sólo el 3% de la longitud de onda para una frecuencia de 50Hz, lo que hace muy difícil que una fracción tan

pequeña se pueda detectar, mientras que para una frecuencia de 800Hz la separación de los oídos es aproximadamente de media longitud onda, pudiendo dar lugar a ciclos enteros de diferencia de fase entre los sonidos que llegan a cada uno de los oídos, con lo que el sistema auditivo no tiene forma de determinar qué ciclo corresponde a cada uno de los oídos.

Por tanto, para frecuencias altas, resulta ambiguo juzgar la dirección de la fuente sonora basándose sólo en la diferencia de fase entre el sonido percibido por los dos oídos y se puede concluir que la localización auditiva por diferencia de fase (ITD) es definitiva en una banda de frecuencias comprendidas entre los 200 y los 800 Hz [3].

Todo lo expuesto anteriormente está limitado a tonos puros, y asumiendo también que el observador evita girar la cabeza con el fin de encontrar la localización de la fuente en acimut. Generalmente estas hipótesis no se dan en la realidad, y por tanto no se puede determinar la dirección de llegada de un sonido ordinario con el alto grado de precisión que se obtienen en los cálculos con tonos puros. La naturaleza compleja de los sonidos, indica que el espectro de frecuencia se extiende sobre la mayor parte del rango audible, por lo que la diferencia entre niveles de intensidad y fases pueden ser utilizadas simultáneamente para localizar el origen del sonido.

En resumen, para frecuencias altas (por encima de los 1000 Hz), la localización se debe fundamentalmente a la intensidad, mientras que para las bajas frecuencias (menos de 800 Hz) la localización que domina es la debida a la diferencia de fases o tiempo de retardos en ambos oídos. En frecuencias medias se presenta una indeterminación que los seres humanos resuelven por medio de una localización tanto por fase como por nivel de intensidad, es decir que utilizamos ambas de forma simultánea y combinada. Esta teoría fue denominada "teoría duplex" por Lord Rayleigh en 1907 [2].

A pesar de la simplicidad y el éxito de la teoría duplex, la teoría solo explica la percepción del acimut dando lugar a posiciones espaciales de fuentes que están a la misma distancia del oyente y generan los mismos valores de IID o ITD. Esta ambigüedad fue detectada por von Hornbostel y Wertheimer alrededor del 1920, definiendo el 'cono de confusión' generado en campo lejano por los puntos que contienen el mismo ITD e IID, cuyas localizaciones eran indistinguibles desde el punto de vista de la teoría duplex.

Este problema se acentúa en el plano medio que separa las dos orejas y corre verticalmente sobre la cabeza ya que en cualquier punto sobre este plano, los valores de IID e ITD son cero haciendo que la comparación interaural sea ineficiente. Debido a que los oyentes son capaces de localizar fuentes en este plano, diferentes científicos han sugerido que la habilidad de localizar sonidos en el plano medio es debida a la capacidad de escucha monoaural y la coloración espectral producida por el torso, la cabeza y el oído externo o pinna [3][4][5].

## *1.2. Determinación de la elevación*

Mientras que la determinación del ángulo lateral, como ya se ha comentado, es un proceso binaural, la determinación de la elevación es monoaural y se deriva del hecho de que nuestra oreja actúa

como una antena acústica cuyas cavidades resonantes y geometría amplifican algunas frecuencias y atenúan otras. Es más, su respuesta frecuencial depende de la dirección de llegada del sonido [6][3].

La siguiente figura muestra la medida de la respuesta frecuencial de dos direcciones de llegada diferentes. En cada caso observamos que existen dos caminos desde la fuente hasta el canal auditivo: un camino directo y un camino más largo que sigue una reflexión en la oreja. Para frecuencias moderadamente bajas, la oreja esencialmente recolecta energía sonora adicional, y las señales de los dos caminos llegan en fase. Sin embargo, para frecuencias elevadas, la señal retardada no está en fase con la señal directa y se producen interferencias destructivas. La mayor interferencia ocurre cuando la diferencia de longitud de los caminos ( $d$ ) es de media longitud de onda, es decir, cuando  $f = c/2d$ . En el ejemplo, esto produce una "muesca" en el espectro recibido sobre los 10 kHz. Para valores típicos de  $d$ , la frecuencia de la muesca está en un rango entre 6 y 16 kHz.

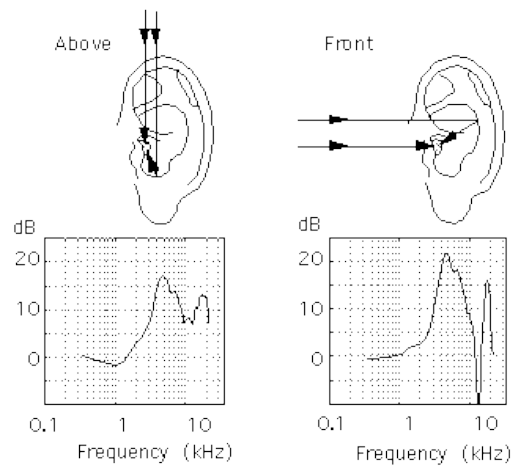


Fig.3. Filtrado de la señal emitida en el dominio frecuencial dependiente del ángulo de elevación.

Como la oreja es un reflector mucho mejor para sonidos que llegan por el frente que por arriba, la muesca resultante es mucho más pronunciada para fuentes frontales que fuentes situadas sobre la cabeza. Además, la diferencia de longitud entre los caminos varía con los cambios en el ángulo de elevación, por lo que la frecuencia de la muesca varía con la elevación. Aunque aún hay disputas sobre que características son porcentualmente más importantes, está bien establecido que la oreja es determinada como la fuente primaria para la detección de la elevación [3].

### 1.3. Ambigüedad en la localización

Existen casos de ambigüedad en la localización de la fuente, como por ejemplo en casos donde los sonidos proceden de dos fuentes situadas de forma simétrica y a la misma distancia de la cabeza. En este caso, los sonidos llegan al oído izquierdo con la misma diferencia de tiempo con respecto al derecho que de modo contrario. Afortunadamente la forma asimétrica del pabellón auditivo ayuda a aclarar cualquier tipo de duda porque el sonido se modifica sustancialmente antes de llegar al tímpano dependiendo de la dirección de llegada.

Otro caso se da en recintos cerrados, donde existen reflexiones sonoras ya que el oyente recibe diferentes señales, unas provenientes directamente de la fuente sonora y otras que llegan después de reflejarse, una o más veces, en las superficies interiores del recinto o en los objetos que contiene el mismo. Estas señales alcanzan los oídos del oyente con diversos retardos de tiempo y desde direcciones diferentes. Como el camino más corto entre dos puntos es la línea recta, el sonido que llega primero es el directo, que es el utilizado para localizar la fuente sonora por medio de la escucha binaural [7][8]. Posteriormente llegan las reflexiones, que han perdido parte de su potencia en los choques con las superficies reflejantes, por lo que contienen menos energía que el sonido directo. Estos sonidos reflejados proporcionan el ambiente acústico o reverberación.

Hay varios tipos de efectos temporales, entre ellos dos casos extremos:

- Sonidos reflejados que llegan inmediatamente después del directo y que contribuyen a enriquecer la tonalidad del sonido original.
- Sonidos que llegan mucho más tarde y que se perciben como ecos.

Entre estas dos situaciones, existen una gran cantidad de posibles casos intermedios que fomentan la ambigüedad en la localización.

Cuando el sonido es continuo, el oído recibe una información directa y otra reflejada, que son idénticas en tono y cuyas intensidades varían dependiendo de las condiciones del recinto. Por esta razón es muy difícil localizar en salas normales tonos puros cuando estos se están produciendo. En cambio, si estos tonos se interrumpen momentáneamente puede determinarse su origen, ya que los primeros ciclos llegan con un cierto adelanto al oído con respecto a las reflexiones, que inevitablemente se retrasarán [8].

Otro posible caso de ambigüedad se produce cuando la fuente sonora está detrás o delante del oyente. En este caso, los sonidos llegan al oyente con igual intensidad y fase a ambos oídos y es imposible discernir su origen. Se podría solucionar el problema moviendo la cabeza con el objetivo de establecer unas diferencias temporales que nos ayuden en la localización, aunque la escucha binaural mediante auriculares no permita dicho movimiento, porque, todo el espacio acústico se movería con el movimiento de la cabeza.

Además se ha comprobado también que el Efecto de Precedencia o Efecto Haas, influye drásticamente en la localización del sonido que se puede generar cuando un oyente que “ve” dos altavoces situados a la misma distancia bajo un ángulo de  $45^\circ$ , que percibe dos emisiones idénticas (cada una de ellas procedente de un altavoz), que no se diferencian entre ellas más que por el intervalo temporal que separa su llegada al oído [7][9]. Los experimentos, además, han llevado a sugerir que el efecto de precedencia tiene su explicación en procesos cognitivos que se desarrollan en niveles relativamente altos [3][7].

#### *1.4. Head Related Transfer Function (HRTF) y Head Related Impulse Response (HRIR)*

Mientras las relaciones entre los ITD e IID y la localización espacial eran razonablemente predecibles, la comunidad científica encontró problemas para relacionar la percepción espacial de



los sonidos con su variación en el contenido espectral mediante expresiones matemáticas. Por ello, el primer paso hacia la comprensión de los cambios de contenido espectral con la escucha direccional fue la realización de modelos físicos [10], además de medidas y experimentos empíricos [11], y más recientemente simulaciones computacionales [12], para la obtención de la variación del contenido espectral dependiente de la dirección obtenida directamente de la oreja física de una persona.

Así, estas medidas fueron denominadas Head Related Transfer Functions (HRTF) o Funciones de Transferencia Relativas a la Cabeza y resumían el filtrado acústico dependiente de la dirección producido por el torso, la cabeza y pinna sobre un sonido emitido en campo libre. De esta manera, los científicos esperaron primero obtener la respuesta en frecuencia de un oído y así analizar las variaciones de la percepción respecto a la dirección de llegada del sonido.

Formalmente, las HRTF se definen como la respuesta en frecuencia en campo lejano de un oído (izquierdo o derecho) específico de cada individuo, medido desde un punto concreto del campo libre a un punto concreto del oído del individuo, normalmente para un radio fijado de distancia desde la cabeza del oyente. Estas medidas se realizan para diferentes variaciones del punto de emisión en grados o radianes, del plano horizontal (acimut) y el plano frontal (elevación) [11].

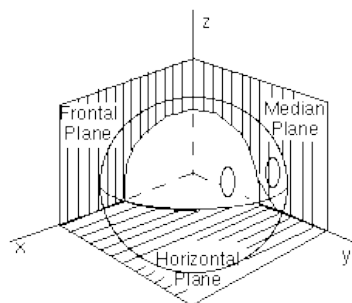


Fig.4. Definición de los planos de escucha.

Por lo tanto, la HRTF es una función de cuatro variables (si no contamos la distancia de la cabeza a la fuente, tres): tres coordenadas espaciales y la frecuencia. En coordenadas esféricas, para distancias mayores de un metro, se dice que la fuente está en el campo lejano y la HRTF decae inversamente con la distancia. La mayoría de las medidas de la HRTF se realizan en el campo lejano, lo que reduce la HRTF a una función del acimut, la elevación y la frecuencia [13]. Cabe mencionar que mediante la Transformada de Fourier Inversa se consigue la expresión análoga en el dominio temporal denominada Respuesta al Impulso Relativa a la Cabeza o Head Related Impulse Response (HRIR).

Las HRIR son usualmente definidas como filtros de Respuesta al Impulso Finita o FIR y suman la información de ambos sistemas definidos por Lord Rayleigh en la 'Teoría Duplex', IID e ITD de forma que los retardos de tiempo (o la información de ITD) son codificados en el espectro en la fase del filtro y la información de las diferencias de nivel es codificada en el módulo en función de la frecuencia del filtro [14].

## 2. Espacialización del sonido (sound spatialization) y holofonía

Teniendo en cuenta todos los factores que influyen en la escucha espacial, y teniendo en cuenta que las HRTFs han demostrado ser los filtros acústicos necesarios para la localización de fuentes en el espacio tridimensional, se han creado, a raíz de estos avances, las técnicas de espacialización del sonido y la holofonía que se entienden como las técnicas empleadas para la recreación tridimensional de una fuente sonora.

Las primeras grabaciones que consiguen este efecto fueron realizadas por el investigador italo-argentino Hugo Zucarelli, quien para ello utilizó un maniquí o dummy que reproducía fielmente una cabeza y torso humano con las mismas o similares características acústicas que el cuerpo humano. El principal inconveniente de estas grabaciones era la falta de interactividad, ya que la posición del dummy en el momento de la grabación respecto a la fuente sonora era la determinante de la sensación que percibiría el oyente. Si éste moviera la cabeza mientras escucha la grabación, la fuente parecería provenir siempre de la misma dirección, llegando incluso a anularse el efecto holofónico [15].

Sin embargo, con el desarrollo de la computación han aparecido sistemas informáticos capaces de reproducir el efecto de la cabeza y torso humanos sobre el sonido que percibimos, recreando el efecto artificialmente. Son los llamados sistemas basados en HRTF (Head Related Transfer Function) que son la base del trabajo presente. Gran parte de la investigación actual se centra en el análisis y determinación de estas funciones, y de su caracterización mediante un conjunto limitado parámetros [16].

En el siguiente subapartado se introducen las distintas técnicas utilizadas a lo largo de los años por la comunidad científica para crear la espacialización del sonido.

### *II.1. Grabación con maniquí o dummy*

Esta técnica consiste en emplear una reproducción fidedigna de la cabeza, orejas y torso humanos para la grabación de un ambiente sonoro. El maniquí contiene dos pequeños micrófonos en el interior de sus oídos que captan la señal cuando ésta ya se ha visto afectada por los materiales que componen el muñeco. Estos materiales recrean fielmente los tejidos del cuerpo humano así como sus propiedades acústicas y, en particular, como se ve afectado el sonido a las diferentes frecuencias cuando los atraviesa o se refleja en ellos [17].

Las grabaciones estéreo convencionales consiguen una cierta sensación de dirección en el plano horizontal, sirviéndose únicamente de la diferencia temporal y de intensidad entre los estímulos recibidos por uno y otro oído (el ITD y el ILD). Sin embargo, con esto sólo se consigue situar, utilizando altavoces, la imagen sonora a lo largo del segmento que los une y nunca fuera de estos (más alejada o hacia un lado) [3]. En el caso de que se usen auriculares, la imagen se forma en el interior de la propia cabeza y el sonido se sitúa más a la izquierda o derecha respectivamente.

Hay varios métodos de realizar estas grabaciones holofónicas, que registran información direccional:

1. Mediante el uso de un maniquí o modelo artificial, como se ha explicado hasta ahora [17].
2. A través de un voluntario que lleve un par de micrófonos especialmente diseñados para colocarse en el interior del oído, para tener así un auténtico modelo real del efecto del cuerpo sobre el sonido. Éste método tiene la principal ventaja de la movilidad del voluntario, pudiéndose realizar grabaciones que simulen situaciones más reales y que incluyen movimiento, como un paseo por un parque o la grabación de un diálogo en una escena de una película en la que el espectador ocupe la posición de uno de los actores. Sin embargo, este sistema tiene también desventajas, como la compatibilidad limitada entre el voluntario que realiza la grabación y quien finalmente la escucha, debido a las diferencias anatómicas de sus cabezas y pabellones auditivos [18].
3. Por último, se puede simular electrónicamente el efecto de la cabeza para incorporar la información sobre la dirección al sonido. Estos sistemas emplean las llamadas HRTF asociadas a cada dirección de origen del sonido [12].

### *II.3. Sistemas basados por filtrado por HRTF*

En base a cualquiera de los métodos de registro de información direccional expuestos sobre estas líneas y con el avance en las técnicas de computación y la informática, han surgido sistemas enfocados a la aplicación de las funciones características de la cabeza, HRTF, a fuentes de sonido para lograr así el efecto de auralización. Si bien existen multitud de sistemas, todos ellos siguen un mismo principio básico de funcionamiento [13][17]:

- Se dispone de una fuente de sonido monoaural. En general, sencillamente una grabación mono del sonido a procesar.
- Se determina la posición relativa en que se quiere situar dicha fuente respecto del oyente.
- Se procesa el sonido según la HRTF indicada para esa dirección de procedencia, convolucionando la señal de entrada con la HRIR correspondiente para cada oído. La distancia se establece, normalmente, atenuando el sonido en función de la lejanía o proximidad de la fuente. Esto se hace así debido al elevado coste experimental que conllevaría emplear funciones HRTF dependientes no sólo de dos variables (el acimut y la elevación), sino además de la distancia.

En el siguiente diagrama de bloques se representa esquemáticamente como pueden llevarse a cabo este tipo de sistemas [18]:

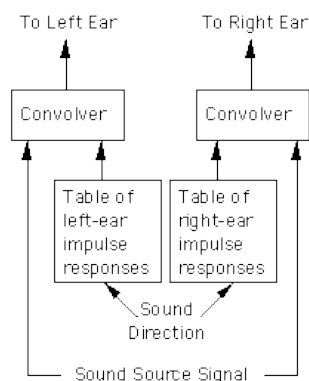


Fig.5. Diagrama de bloques genérico de generación de sonido binaural mediante el filtrado por HRTF.

Como vemos, el sistema tiene almacenada una tabla con las respuestas al impulso (HRIR) de cada oído para cada dirección. En realidad, se puede utilizar una única tabla de respuestas al impulso y obtener la otra calculando su simétrica. Es decir, si suponemos la cabeza y torso humanos (y por tanto sus propiedades acústicas) simétricos respecto al plano medio, la HRTF correspondiente a una dirección para el oído derecho, por ejemplo, será equivalente a la del oído izquierdo para la dirección de procedencia simétrica respecto a ese plano medio, esto es, con la misma elevación pero el acimut cambiado de signo. Así, se puede evitar tener almacenada información redundante en nuestro sistema.

Una vez escogida la HRIR a aplicar, la señal (independiente para cada oído) pasa a un bloque convolucionador donde se aplica la HRIR y llega a la salida.

Es importante notar que este esquema sólo nos permite procesar una fuente sonora al mismo tiempo. Si quisiéramos reproducir dos o más al mismo tiempo, tendríamos que repetir el esquema una vez por cada nueva fuente, y sumar los resultados de cada uno combinándolos.

El número de HRIR almacenadas en las tablas (la 'resolución' en el acimut y la elevación) puede reducirse empleando técnicas de interpolación apropiadas entre puntos del espacio, que es lo que se pretende en el presente estudio, además de hacer hincapié en la sensación subjetiva producida en el oyente por cada una de las interpolaciones propuestas.

Actualmente se han desarrollado una gran variedad de aplicaciones informáticas enfocadas a la *auralización* acústica. Aunque aún es un campo joven, las importantes aplicaciones que se le vaticinan en el campo del entretenimiento y las comunicaciones han fomentado su desarrollo. Como ya se ha dicho, todas ellas comparten el mismo principio común: el tratamiento de la señal acústica a auralizar mediante la función HRTF apropiada [13][17].

Estas HRTFs pueden consistir, sencillamente, en un conjunto de pequeños archivos de audio que contienen la grabación obtenida por el micrófono en el oído del voluntario o del maniquí cuando se reproduce en un altavoz, situado en la dirección de procedencia a analizar, una breve secuencia de ruido pseudoaleatorio semejante estadísticamente a un delta de Dirac. Su duración es muy breve, de unos pocos milisegundos, conteniendo normalmente unas 512 muestras. Así, esa grabación obtenida se asemeja a la respuesta al impulso del voluntario, pues recoge las modificaciones sufridas por su cabeza, orejas, etc. sobre un impulso o delta de Dirac. Es decir, en esos pequeños

archivos .WAV tenemos la colección de HRIR asociadas a una persona o maniquí en particular [17] [18].

En otras ocasiones, como en el caso de la base de datos del IRCAM [19], las HRTF se suministran en forma de tabla de valores en un archivo .mat interpretable desde el entorno MATLAB. Esta base de datos es la utilizada en el presente proyecto para generar las diferentes localizaciones de las fuentes a gusto del consumidor.

De esta forma, y haciendo uso de los recursos que se tienen al alcance, se pretende crear un sistema basado en el filtrado por HRTFs que se detalla a continuación en el que se pretende que se puedan localizar las fuentes sonoras en cualquier punto del espacio deseado por el usuario del sistema.

### **3. Método experimental**

En el siguiente apartado se describirá el método experimental utilizado para el desarrollo de la aplicación y sus limitaciones. Así, este apartado se dividirá en diferentes subsecciones. La primera de ellas irá destinada a la descripción del conjunto de HRTFs (ofrecidos por el laboratorio IRCAM) utilizados para la generación de la aplicación además de la introducción de la utilidad 'Playrec' y sus características. En segundo lugar, se describirá el proceso mediante el cual se ha generado la aplicación para la generación del sonido proveniente de cualquier punto equidistante del espacio. A continuación, se describen los distintos métodos de interpolación ideados para la creación del campo sonoro, seguidamente, se presentan las interfaces desarrolladas para el usuario y se comentan sus limitaciones y finalmente, se describirá el 'modus operandi' ideado para analizar las sensaciones subjetivas de los sujetos usuarios del sistema.

#### *III.1. Descripción de los medios utilizados*

En este primer apartado del método experimental se describirán, como ya se ha dicho, los medios utilizados para la recreación del campo sonoro tridimensional a demanda del usuario. Para ello, se procederá primero a la introducción de la biblioteca de HRTFs o HRIRs utilizada para este menester ofrecida en la red, por cortesía de Oliver Warusfel y el Equipo de Acústica de Salas del Instituto de Investigación y Coordinación de la Música/Acústica de París (Institut de Recherche et Coordination Acoustique/Musique, IRCAM) [19] y tras ello, se procederá a la descripción de la utilidad Playrec desarrollada para Octave o MatLab la cual ofrece un acceso a las tarjetas de sonido de cualquier PC en tiempo real utilizando las librerías de entrada/salida de audio de libre distribución 'PortAudio' la cual puede utilizarse en diferentes plataformas como Windows, Macintosh y Unix.

### III.1.1 Biblioteca de HRTFs ofrecida por el IRCAM

Este instituto, ofrece un conjunto de HRIRs medidas en cámara anecoica, en los oídos de un conjunto de 51 sujetos. Se decide tomar esta biblioteca de todas las existentes en la red por la simplicidad que supone tener los ficheros en el formato de datos de MatLab para su posterior uso. La única desventaja de esta biblioteca es que no contiene la respuesta de un maniquí especialmente diseñado para este tipo de medidas por lo que se ha optado por utilizar la HRTF que más se ajustaba al autor.

Las medidas son tomadas en una cámara anecoica (de dimensiones  $8.1 \times 6.2 \times 6.45 = 324 \text{ m}^3$ ) cubierta por cuñas de fibra de vidrio de 1.1m de longitud que absorbe las ondas por encima de los 75Hz. La configuración cuenta con una silla de oficina adherida a un plato giratorio (B&K 9640, conectado a un controlador B&K 5960) y una estructura que se encarga de colocar la fuente en el ángulo de elevación deseado. El control de la elevación, acimut y colocación correcta de la cabeza del usuario se realizó mediante un software dedicado a cada uno de los menesteres.



Fig.6. Detalle del método de medida de las HRTF.

El conjunto de elementos electroacústicos utilizados para la toma de medidas es por un lado, una fuente TANNOY System600 alimentada por un amplificador YAMAHA cuya respuesta en frecuencia y directividad son medidas mediante un sistema de micrófono omnidireccional (B&K 4149) y un amplificador alimentado con una señal de barrido logarítmico. La cápsula del micrófono se colocó lo más cerca posible de la situación de la cabeza del sujeto durante la prueba ofreciendo los resultados mostrados en las siguientes gráficas:

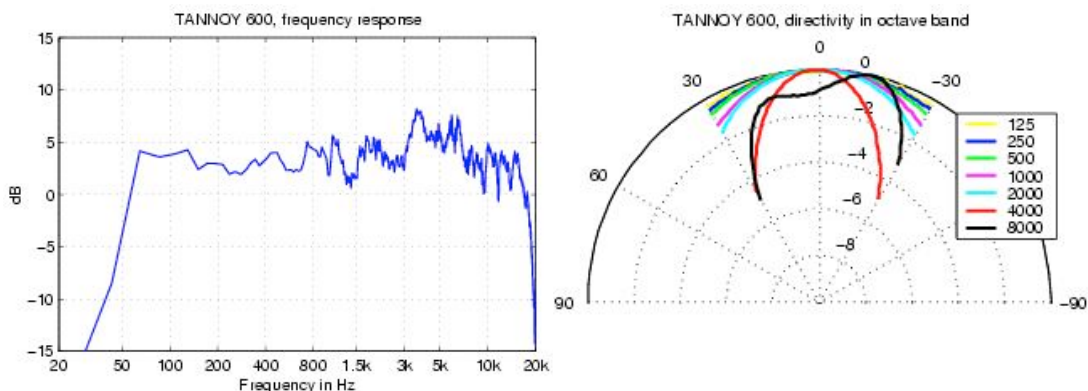


Fig.7.Respuesta en frecuencia y directividad de la fuente Tannoy System600.

Y por otro lado, los micrófonos utilizados para captar la señal en las orejas de los sujetos. En este caso optan por utilizar un par de mini-micrófonos Knowles FG3329 y un previo construido para el experimento con una ganancia de 40dB. La respuesta en frecuencia de dichos micrófonos, suponiendo ambas iguales por el laboratorio mentado es:

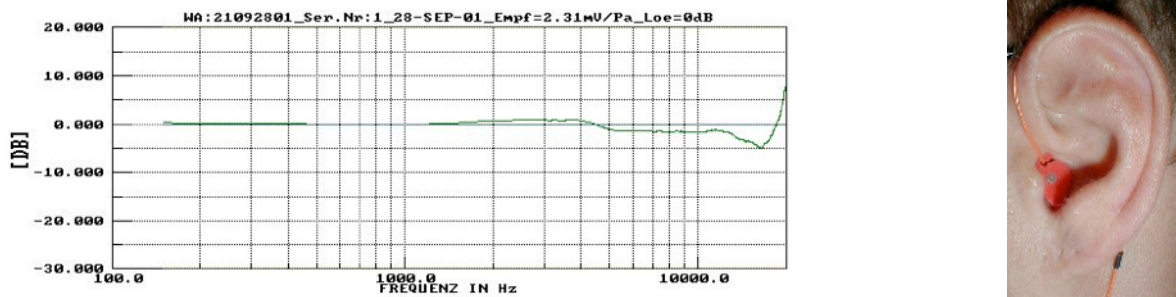


Fig.8. Respuesta en frecuencia de los mini-micrófonos Knowles FG3329 y colocación del receptor.

Estos dispositivos son colocados en la oreja de los sujetos mediante el método *blocking-meatus*, lo que supone que la parte frontal de la cápsula se sitúa en la entrada del canal auditivo (con ayuda de un molde de silicona realizado para cada sujeto) bloqueando la entrada de sonido en el mismo para evitar resonancias, siendo esta la manera óptima de hacer estas medidas según distintos autores [20][21].

El protocolo de medida para cada uno de los sujetos desarrollado por este laboratorio comprende un total de 187 medidas por cada sujeto, que consiste en 10 ángulos de elevación que van desde  $-45^\circ$  a  $+90^\circ$  en saltos de  $15^\circ$  de resolución vertical. Los saltos de rotación o acimut varían según el ángulo de elevación desde 24 medidas a una única en  $+90^\circ$  de elevación. Dicho protocolo se resume en la tabla 1.

Además de la base de datos, el IRCAM ofrece ciertos códigos de ayuda implementados sobre MatLab como por ejemplo las sentencias necesarias para la llamada a las HRIRs o el código necesario para compensar el retardo de propagación y partir la función de transferencia en filtros de fase mínima y componentes paso todo.

Elevación (Grados)	Incremento de acimut (Grados)	Puntos por elevación
-45	15	24
-30	15	24
-15	15	24
0	15	24
15	15	24
30	15	24
45	15	24
60	30	12
75	60	6
90	360	1

Tabla.1. Resumen del protocolo de medida de las HRTF.

### III.1.2 Utilidad Playrec

La utilidad Playrec fue desarrollada por Robert Humphrey inicialmente como parte de un proyecto desarrollado para la Universidad de York, en el Reino Unido, en 2006. El proyecto fue denominado ‘Automatic Speaker Location Detection for use in Ambisonic Systems’ [22].

Esta utilidad, observada como una biblioteca computacional en la que se basa MatLab, ha de ser compilada previamente a poder ser utilizada. Esta compilación está correctamente descrita para cada uno de los posibles escenarios y S.O. posibles pero su compilación, por ejemplo, en sistemas basados en 64 bits no está desarrollado todavía, por lo que no hace uso de la potencia computacional actual.

Playrec, ofrece ciertas características que se describen en su página web y se resumen a continuación incluyendo las limitaciones que ha ofrecido durante el desarrollo de la aplicación en tiempo real.

Así las características resumidas son:

- Acceso a la tarjeta de sonido sin bloqueos: *‘Todas las muestras son almacenadas en un bucle por lo que MatLab puede continuar con otros procesos mientras se produce la reproducción y la grabación del sonido’.*

Siendo cierto que todas las muestras son almacenadas, este almacenamiento se produce por bloques de datos (denominados ‘páginas’) que se van poniendo en cola, para ser procesados. Esto supone que cada una de las páginas son puestas en cola y numeradas y el usuario va accediendo y haciendo uso de ellas. Las páginas pueden ser de tres tipos distintos, de grabación, de reproducción o de grabación/reproducción según se quiera hacer uso de las mismas, siendo su tamaño determinable por el usuario en el código.

Aquí se presenta una de las limitaciones de la utilidad ya que cuanto mayor es el tamaño de la página escogido, más sencillo es el control de los datos, pero mayor es la latencia del sistema propuesto. Es decir, que existe la necesidad de encontrar un consenso entre la comodidad del tratamiento de datos por parte del usuario y la latencia exigida para la aplicación diseñada.

- Grabación y reproducción continua sin cortes: *‘Todas las nuevas muestras de salida son automáticamente añadidas a cualquier muestra restante. Esto hace posible que el procesado de las muestras solo se vea limitado por la capacidad de computacional del ordenador utilizado. Debido a ello, la grabación sin pérdida de muestras puede ser fácilmente alcanzada’.*

Frente a la grabación sin pérdida de muestras que ofrece el desarrollador, Robert Humphrey, vuelve ser clave el tamaño de la página de datos ya que cuanto mayor es el tamaño de la página, mayor capacidad computacional es necesaria para que el tiempo de proceso de cálculo en la página de muestras actual no sea mayor que el de adquisición de la nueva página necesaria. Por lo tanto se da que si el bloque de muestras es muy pequeño, su tiempo de adquisición es muy corto comparado con el tiempo de cálculo y se producen pérdidas y como al final, en cualquier procesado a tiempo real no interesa que exista latencia, el bloque de datos debe ser corto.



Este hecho, demuestra que el desarrollo de esta utilidad está bastante limitado y resulta ser un código bastante optimizado pensado para otro tipo de sistemas.

- Configuración mínima: *Se adjunta una serie de funciones para obtener una lista de los dispositivos de audio disponibles en los que únicamente es necesario introducir como variable el número identificador del dispositivo a utilizar y la frecuencia de muestreo requerida. El buffer puede establecerse y manejarse dinámicamente.*
- Número de muestras en el cálculo no impuesta: *No hay ningún límite impuesto para restringir el número mínimo o máximo de las muestras que pueden ser pasadas a la utilidad en un momento dado, las únicas restricciones son aquellos debido a limitaciones en la memoria disponible o la capacidad de procesamiento computacional.*

La máquina utilizada es un ASUS F3sv, que cuenta con un procesador Intel Core 2 Duo T7500 @ 2.20 GHz, una memoria de caché L2 de 2047MB sobre una placa AsusTek SantaRosa. La API de audio en este utilizada en este caso es ASIO sobre una tarjeta de audio Realtek HD comercial. Esta capacidad a día de hoy no es la capacidad máxima computacional adquirible a nivel comercial, pero es una máquina suficientemente potente.

Las características descritas por el autor son más, pero se refieren a la opciones de la utilidad respecto a los distintos sistemas operativos y distintas API que soporta.

### *III.2. Generación del campo sonoro a demanda*

Como ya se ha mencionado, en esta segunda sección se procederá a la descripción de los pasos seguidos para generar la interfaz gráfica mediante la herramienta GUIDE de MatLab que interacciona con el usuario generando el sonido como si proviniese de cualquier punto del espacio deseado por el mismo. Se comentarán además las limitaciones de la misma y los problemas encontrados a lo largo de su desarrollo.

Siguiendo la tónica de apartados anteriores, primero se introducirá la interfaz desarrollada, desgranando cada uno de sus componentes y su funcionamiento. Posteriormente, se detallará el método utilizado para la generación de los campos sonoros así como los problemas encontrados durante el desarrollo, la toma de datos de las estructuras de las HRTF o la generación de la señal de entrada.

#### *III.2.1. Interfaz gráfica mediante la herramienta GUIDE de MatLab*

Inicialmente, se pretende crear una interfaz que pueda interactuar con el usuario con relativa facilidad y que esta además, genere hasta una nueva petición por parte del usuario, el campo sonoro procedente de una fuente situada virtualmente donde el usuario demandó, dotando de flexibilidad y una mayor interactividad al programa.

Para ello, se decide que el usuario presionará con el ratón, en una imagen similar a la situada bajo estas líneas, la colocación en el plano horizontal de la fuente mientras que la colocación en

elevación se hará mediante el teclado. Aún a sabiendas de que muchos investigadores han creado la sensación de lejanía en la fuente en incluso la externalización de la fuente haciendo uso de caídas de nivel para la primera de las variables y sumando al sonido cierta reverberación conocida creando sensación de espacialidad correspondiente a la segunda, en el presente trabajo se suponen, por simplicidad, todas las fuentes a la misma distancia y sin crear la sensación de externalización. Aprovechando este hecho, se decide que la inserción del ángulo de elevación dentro del campo de un *Edit Text* dentro de la interfaz generada será el detonante para que el usuario pueda interactuar con la imagen y situar la fuente en acimut y además sirva como ‘botón’ de actualización del resto de parámetros. Se pretende así que el usuario no tenga límites en colocar la fuente con una resolución de  $1^\circ$  en acimut en la interfaz, mientras que en elevación el usuario quedará limitado a situar la fuente entre  $-45^\circ$  y  $+90^\circ$  de elevación, con la misma resolución, debido a que, como ya se ha visto, la biblioteca de HRTFs no contiene medidas realizadas por debajo de los  $-45^\circ$ .

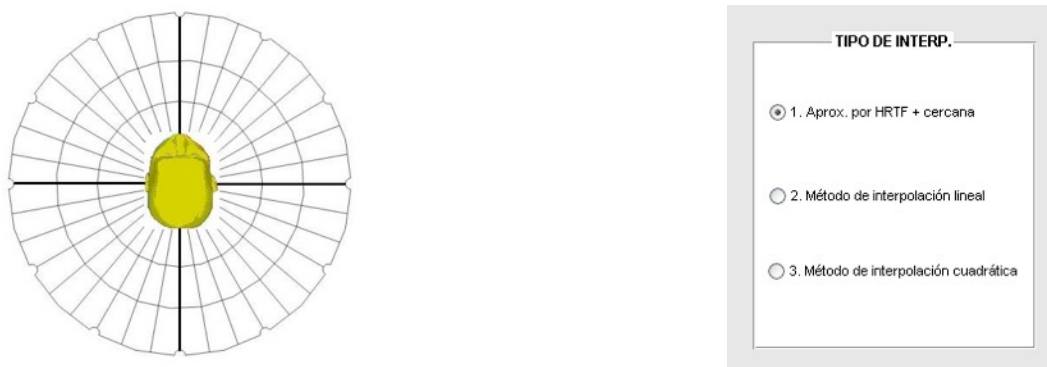


Fig.9. Detalle del plano horizontal y el panel de botones incluidos en la interfaz gráfica generada.

Obviamente, y como ya se ha comentado, para cada ángulo de elevación deseado, el usuario podrá situar la fuente virtual en cualquier punto del plano horizontal únicamente clicando con el ratón la posición adecuada en la imagen. Con el fin de facilitar este proceso, se opta por mostrar una imagen de una cabeza vista desde la parte superior y centrada, rodeada de un círculo con las posiciones marcadas cada  $30^\circ$  de acimut similar a la de la figura 9. Con este fin, se pretende que el usuario clique con el ratón en los bordes del círculo, evitando así generar la sensación de poder colocar la fuente teniendo en cuenta la distancia de la cabeza a la fuente, y facilitando así la colocación de la fuente en el espacio. La imagen del plano horizontal mentada, se carga dentro de un *panel* para así tener la capacidad de poder clicar en la misma, y que los datos generados por esa acción puedan ser utilizados a lo largo del código.

Además de la interacción ‘espacial’ con el usuario, también se decide establecer un panel de botones o *button group* (cuya imagen se puede observar sobre estas líneas, en la fig. 14) que permita al usuario seleccionar entre distintos tipos de interpolación que se detallarán en apartados posteriores con el fin de obtener las diferencias y sensaciones subjetivas asociadas a cada uno de ellos. Dicho panel es accesible al usuario mientras se reproduce cualquier sonido deseado y la última fuente generada no deja de sonar nunca, hasta que se altera de nuevo cualquiera de las variables o se introduce un nuevo valor en el ángulo de elevación, favoreciendo también la

concentración del usuario a las diferencias entre los distintos sonidos generados. De la misma forma, se da la posibilidad al usuario de que pueda variar el tipo de sonido generado entre un ruido modulado con una frecuencia de 20Hz, un ruido aleatorio (con densidad espectral de potencia constante), un tono puro al cual se le podrá cambiar la frecuencia dinámicamente y un archivo de audio (un pasaje de vientos metálicos) predefinido obtenido del Denon Anechoic Orchestral Music Recording disponible en la red [23]. La elección de este tipo de sonidos viene dada por la necesidad de monitorizar los resultados y la observación de que el sistema funciona correctamente.

Con todo lo anterior se puede resumir que el funcionamiento de la interfaz gráfica sigue el siguiente diagrama de bloques:

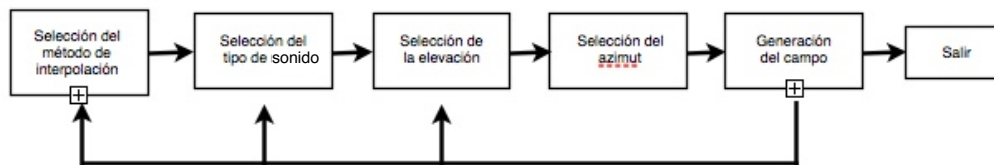


Fig.10. Diagrama de bloques del funcionamiento de la interfaz.

Todo lo anteriormente mencionado, esto es, el control de los bucles, se produce en el *Callback* del texto editable (*Edit Text*) correspondiente a la introducción de la elevación. Así, cada una de las funciones asociadas a los diferentes botones pasan la información necesaria a ésta, y se programa el control de las variables y los bucles.

Posteriormente a la creación de este diagrama de bloques, se decide introducir la posibilidad de crear la misma aplicación, dando la posibilidad de la escucha de otra fuente adicional y así, abrir el estudio a nuevas posibilidades de localización multifuente.

Para ello y tomando como base el diagrama presentado, se procede a la inserción de dos botones (establecer y restablecer) auxiliares que permiten establecer la fuente que se está reproduciendo actualmente para, sobre esa localización y fuente generada, establecer una nueva fuente donde el usuario lo considere oportuno, otorgando la posibilidad de alterar dinámicamente como en el caso anterior, las variables de la misma (tipo de interpolación, de sonido, ángulo de elevación y acimut). Obviamente, el botón restablecer únicamente se encarga de eliminar la fuente fijada por el usuario para que este pueda fijar una nueva. La posibilidad de reproducir más de una fuente situada virtualmente en posiciones diferentes ha sido restringida como mucho a dos fuentes en dos puntos diferentes. Esta limitación viene impuesta por la complejidad de la interfaz generada y el difícil control de los bucles necesarios para su control además de la multiplicación de los elementos presentes en la misma. Es obvio que esta utilidad también puede utilizarse para reproducir una única fuente como la anterior, resumiendo ambas posibilidades en esta última. De esta forma, se genera el siguiente diagrama de bloques:

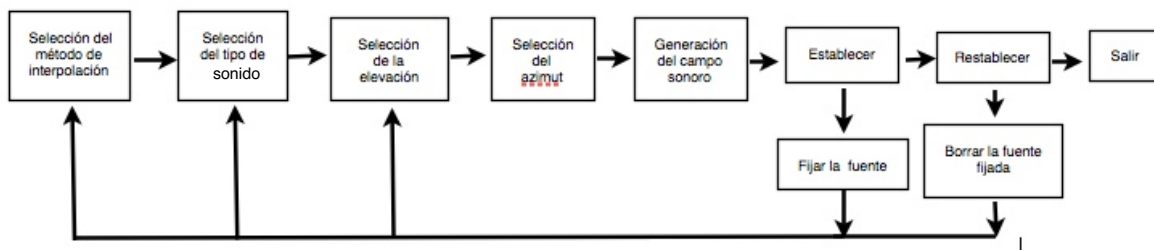


Fig.11. Diagrama de bloques del funcionamiento de la interfaz.

Por último, se decide incluir en la interfaz un botón (*button*) que permita salir de la ejecución del programa. Así, la interfaz generada cuenta con esta apariencia:

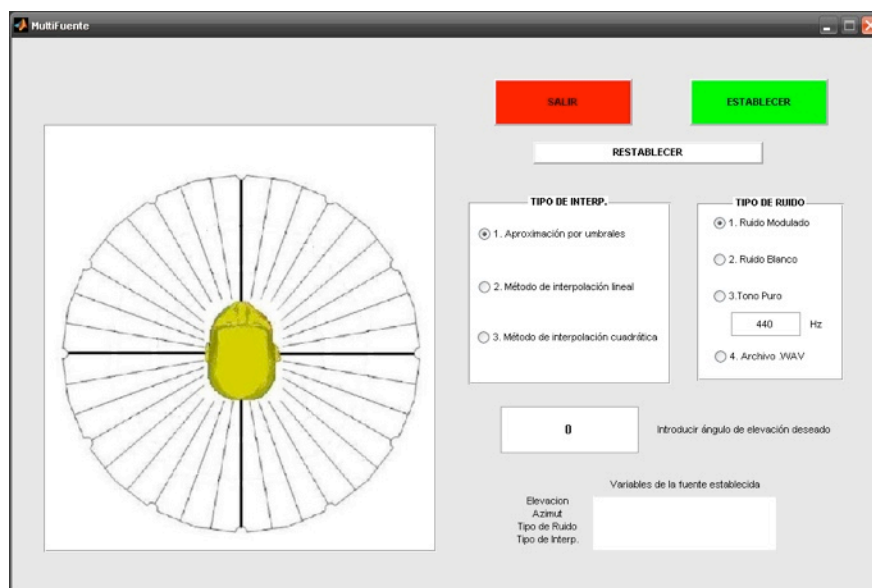


Fig.12. Aspecto de la interfaz gráfica generada.

### III.2.2. Código asociado a la interfaz gráfica

Como ya se ha mencionado, el detonante para que todo el sistema funcione y se actualice es la introducción del ángulo de elevación deseado por parte de usuario, lo que permite seleccionar un ángulo deseado en la pantalla y así generar el campo sonoro debido, además de actualizar todos los botones que se hayan pulsado. Así, se observa una de las limitaciones de la aplicación, y es que se pueden alterar todas las variables durante la ejecución y escuchar su efecto, pero no se puede visualizar el estado del mismo, hasta que se introduce de nuevo un valor en el campo de la elevación. Este hecho, se ha determinado que no impide el correcto funcionamiento del programa, ya que para cambiar la localización de la fuente, primero, como bien lo muestran los diagramas de bloques presentados, se ha de alterar, o al menos pulsar la tecla *enter*, dentro del campo correspondiente a la elevación y esto, cuando el usuario se familiariza con la interfaz, se realiza continuamente.

Así, en la *Opening\_Fcn* de la interfaz creada, se procede únicamente a establecer la imagen del plano horizontal en el *panel* creado para tal efecto siguiendo un código sencillo que está disponible tanto en MatLab Central como en la red.

Una vez inicializada la interfaz y debido a que, como ya hemos dicho, el detonante es la introducción del ángulo de elevación en el *Edit Text* generado para ello, el siguiente bloque del código se presenta dentro de la función *Callback* del *Edit Text* establecido. Así dentro de *eleva\_Callback* se genera un código que lo primero que hace es esperar a que se pulse cualquier botón del ratón para empezar a correr y así tomar el punto actual en el que el ratón ha sido clicado mediante la instrucción *gca* (*get current axes*) para después almacenar sus coordenadas *x* e *y*.

Tras la obtención las coordenadas, se comprueba si dichas coordenadas están dentro de los límites establecidos por la imagen o el *axes* generado para cargarla tomando el campo *XLim* e *YLim* del eje de gráficas o *axes* actual. Si la coordenada establecida por el usuario estuviese fuera de estos límites, el usuario sería informado en la misma interfaz.

Una vez obtenidas las coordenadas establecidas por el usuario, y habiendo identificado el punto del centro de la imagen previamente, o el centro de la cabeza en la imagen, se procede al cálculo del ángulo de acimut necesario para la generación del campo sonoro deseado. Para este menester se procede primero a la identificación del cuadrante seleccionado por el usuario y al cálculo del ángulo relativo que identifica al ángulo seleccionado por el usuario mediante la teoría de triángulos regulares de Pitágoras. Una vez se obtiene el ángulo relativo, este es sumado al ángulo establecido por el cuadrante seleccionado. Así, al ángulo clicado en el primer cuadrante se le suma  $0^\circ$ , al ángulo clicado en el segundo  $90^\circ$  y así sucesivamente.

Después de calcular el ángulo de acimut demandado por el usuario, se procede a la obtención del ángulo de elevación introducido por teclado mediante la instrucción *get* y se presenta la información de acimut como título de la imagen del plano horizontal.

Como ya se ha mencionado, ya se puede trabajar en la generación del campo sonoro deseado debido a la obtención de los dos parámetros necesarios para ello, la elevación y el acimut. Así, se procede primero a determinar si el botón *salir* ha sido pulsado para no generar ningún campo o si por el contrario el usuario ha pulsado cualquiera de los botones que permiten seleccionar entre los distintos métodos de interpolación o los distintos tipos de generación del sonido o señal a filtrar. Si el usuario pulsa el botón *salir*, el programa limpia todas las variables y cierra el programa, si en cambio selecciona un método de interpolación y un tipo de sonido, este llama a las funciones desarrolladas para la generación de la señal y posteriormente su señal de salida es reproducida mediante el comando *sound* indefinidamente, o hasta que el usuario realice un nuevo cambio de las variables, en cuyo caso se reproduce la nueva señal generada en donde las variables son alteradas y se espera a la siguiente ejecución del bucle de reproducción para que sus efectos sean reproducidos. De esta forma, lo que se hace es reproducir una señal cíclicamente y cuando se detecta un cambio en el interfaz, se cambian las condiciones para generar la señal que se reproduce en el siguiente

ciclo. Con esto lo que se persigue es que la aplicación procese y presente los resultados en *pseudo-tiempo real*.

Una vez establecido el comportamiento general de la interfaz, se ha generado un bucle *while* teniendo en cuenta que se haya pulsado el botón de establecimiento de fuente, y si es así, generando la salida sumada de ambas fuentes. Para la reproducción de la señal de salida generada se ha utilizado el comando *soundsc* que procesa la señal a reproducir para que sus valores queden dentro del rango dinámico permitido para que no saturé, [-1,1].

Para la generación de los distintos métodos de interpolación se decide implementar diferentes funciones dependiendo de las necesidades de cada una (primero en tiempo real mediante Playrec y después mediante MatLab), pero se observa que todas tienen la misma forma de llamada a las HRTFs contenidas en un fichero tipo *struct*. Como cada una de las interpolaciones desarrolladas se presentarán en apartados posteriores, en este se describirá la forma de acceder a la información necesaria dentro del *struct* de datos. Es conveniente hacer hincapié en que este código no está presente en el archivo de inicialización y funcionamiento de la interfaz si no que está contenido en cada una de las funciones de interpolación desarrolladas.

Como ya se ha expuesto anteriormente, las HRTFs están disponibles en formato *struct* para cada uno de los 51 sujetos que realizaron el experimento en el IRCAM [19]. Este *struct*, se divide en otros dos *struct* al realizar la carga del archivo, correspondiente al contenido de cada uno de los oídos del sujeto. Así se generan dentro de cada uno de estos, 5 campos correspondientes a la elevación y acimut (guardadas en un vector columna), el tipo de filtro (FIR), la frecuencia de muestreo con la que se hizo la medida (*sampling\_hz*), y la matriz de contenido de cada una de las posiciones de fuente (denominado *content\_m*) almacenada por filas. Los vectores columna correspondientes a la elevación y acimut comparten los índices que definen cual de los filtros o HRTFs de 512 puntos de longitud se debe tomar para colocar la fuente en la posición debida, es decir, que para la obtención del filtro adecuado, se realiza una búsqueda de la elevación y el acimut en sus respectivos vectores columna de donde se extrae el índice de la matriz de contenido (*content\_m*) correspondiente a los ángulos demandados por el usuario.

Los ángulos de elevación y acimut se almacenan de manera que para cada ángulo elevación, partiendo de  $-45^\circ$  hasta  $90^\circ$ , están almacenados todos los saltos en acimut correspondientes de  $0^\circ$  a  $360^\circ$  (véase el protocolo de medida presentado en la Tabla 1). De esta manera, se decide primero establecer en un vector auxiliar los índices que corresponden a los cambios de elevación, realizando una búsqueda en el vector de acimut correspondiente a todos los valores de  $0^\circ$ . Así sabiendo cuantos índices del vector de elevación comparten el mismo valor, se determina el número de HRTFs disponibles en ese ángulo de elevación (obviamente, haciendo la resta entre los índices almacenados en el vector auxiliar mentado) y el incremento de acimut y el número de HRTFs en cada uno de ellos, haciendo la división entre  $360^\circ$  (correspondiente a todo el círculo). Una vez obtenidos estos datos, se divide el ángulo de acimut introducido por el usuario por el incremento de acimut establecido para el ángulo de elevación demandado y se redondea a la baja, obteniendo así el índice relativo (dentro del ángulo de elevación seleccionado) de la HRTF

escogida. Así sumando ambos índices (el relativo y el correspondiente a los cambios de elevación) se obtiene el índice absoluto con el que se accede al vector de contenidos (*content\_m*) de las HRTF. A continuación se programan las excepciones necesarias en los casos límite como son entre 345° y 0° donde los índices no son consecutivos o el caso especial de 0°.

### III.2.3. Código asociado a Playrec

Siendo la idea original del proyecto la de conseguir una interfaz que situase virtualmente y en tiempo real la posición de una fuente generada también a tiempo real, se intentó desarrollar el filtrado utilizando la utilidad Playrec descrita en capítulos anteriores.

Concretamente la idea inicial de la aplicación a generar, era la de conseguir un código que captase las ondas incidentes sobre un micrófono y a su vez que el código filtrase dicha señal capturada, para situarla en cualquier punto del espacio y reproducir dicha ubicación de la fuente de vuelta al usuario en tiempo real. Esta idea podría tener como variación el hecho de que el sonido en vez de ser capturado por el micrófono fuese directamente obtenido de la tarjeta de sonido, pero ello no supondría mayor problema.

De esta forma, se decide optar por una función programada por el creador de la utilidad y ofrecida en la página web de Playrec, que consiste en un *loopback* o *bypass* en el que la señal que es capturada por el micrófono es directamente emitida al usuario sin ningún tipo de procesado de la señal.

Haciendo uso de la función mentada, se cuenta con la parte del código de adquisición de la señal y emisión de la misma ya programado, lo único que es necesario es programar un filtrado eficiente en la misma. Así, primero, se parte del desarrollo de un algoritmo para el filtrado monocanal en tiempo real de la señal captada por el micrófono de cara a observar si es posible realizar el filtrado en tiempo real monocanal de una señal cualquiera.

A este código desarrollado, es necesario pasarle las variables del identificador del dispositivo de grabación y reproducción, el número de canales de grabación y reproducción además de la frecuencia de muestreo y la frecuencia de corte del filtro. Como el número de coeficientes del filtro es del mismo tamaño que el tamaño de la ‘pagina’ de datos (*pageSize*) utilizando la instrucción `filter` para el filtrado de la señal de entrada, se consigue un filtrado ‘eficiente’ y sin pérdida de muestras.

Una vez desarrollado el filtrado monocanal se procede a identificar las diferencias existentes entre las exigencias del filtrado paso bajo monocanal desarrollado y las exigencias del filtrado multicanal a desarrollar. De esta forma, se observa que:

1. La información de las HRIR de las que se dispone, son respuestas impulsivas en el dominio temporal y vienen definidas con los coeficientes de un filtro de respuesta finita, FIR, como el filtro generado en el caso anterior.

2. Como se sabe que hay opción de usar la instrucción `filter` o `conv` para realizar el filtrado, pero su eficiencia computacional es baja, es necesario construir una nueva estrategia para el mismo. Se decide realizar el filtrado en el dominio frecuencial ya que ofrece una mayor eficiencia computacional que la instrucción `conv`.
3. El procesado y filtrado de la señal de entrada es multicanal y por lo tanto, la señal de salida en ambos oídos (estéreo) no comparte la misma información.
4. Son necesarios cuatro filtros distintos y cuatro filtrados distintos por lo tanto, son necesarios todos los cálculos de filtrado de la señal de entrada necesarios en el filtrado paso bajo implementado previamente, por cuadruplicado, para poder realizar la interpolación de las señales localizadas en los dos puntos más próximos al deseado por el usuario.
5. Es necesario un cálculo de los coeficientes de interpolación (pesos) en función del ángulo de acimut deseado por el usuario para realizar la mezcla entre las señales.

Teniendo identificadas las necesidades, se procede a la implementación del código para el filtrado por las HRTF que se desea implementar sin alterar de ninguna forma la estructura y código referente a la grabación y reproducción de las señales, ni el código correspondiente a la inicialización y puesta en funcionamiento de Playrec, el cual cuenta con estructura compleja de programación.

En este punto, es necesario mencionar que a la señal de entrada, no se realiza ningún tipo de filtrado para la eliminación de ruido o componentes no deseadas por lo que la calidad del sonido generado esta directamente relacionada con la calidad del micrófono utilizado para la captación del sonido aunque este fuese un procesado que podría sumarse al necesario *a posteriori*.

El código inicialmente implementado contempla un tamaño de página de datos de 512 muestras, siendo este también el tamaño de cada uno de los filtros HRIR incluidos en la biblioteca. En el código se contempla que la Transformada de Fourier (*fft*) se realiza con el doble de muestras que el tamaño de página, tanto la del filtro como la de la señal temporal de entrada, cuyo bloque de datos también es de 512 muestras. Es decir, que tanto la señal temporal como la HRIR utilizada tienen 512 muestras cada una y son transformadas con el doble de muestras para asegurar que la convolución circular calculada coincida con la lineal, y por lo tanto con el resultado de realizar el filtrado.

En este momento, entra en juego una de las limitaciones de la utilidad y quizá la más importante, que es el hecho de que trabaja con páginas de datos y procesos que únicamente pueden ser de tres tipos: grabación, reproducción y grabación/reproducción y tienen un tamaño definido por el usuario.

Esta limitación supone que si utilizamos de forma intercalada, obviamente cuando sea necesario, la páginas de grabación y reproducción por separado, podemos optar a programar las páginas de grabación con un tamaño distinto al de reproducción que es justamente el resultado de un filtrado en el dominio temporal, pero no podemos tener acceso continuo a las muestras grabadas y reproducidas, ya que se deben grabar, procesar y tras ello reproducir debiendo establecer el proceso de grabación y reproducción en páginas diferentes.



Así la opción más obvia conduce a la utilización de las página de grabación/reproducción en la que su mayor limitación es que la cantidad de muestras de grabación y de reproducción son las mismas, por lo que cualquier procesado que implique una alteración de duración como es el filtrado, supone una pérdida de muestras que es necesario evitar.

Intentando solucionar este problema, se ha optado por realizar el filtrado por bloques mediante el método de solapamiento y suma, de forma que la mitad del bloque de muestras de salida se reproduce (las primeras 512 muestras del bloque de muestras de salida) mientras que la otra mitad se almacena en un vector para ser sumada a la primera mitad del bloque de muestras de salida siguiente. El proceso se resume en la siguiente imagen:

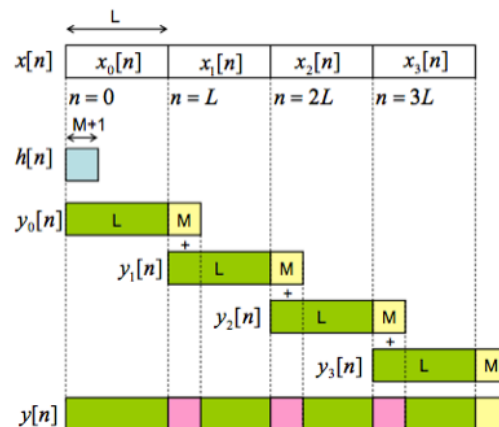


Fig.13. Detalle del filtrado por solapamiento y suma.

De esta forma, siendo  $x[n]$  la señal de entrada que es dividida en bloques de  $L$  muestras, 512 en este caso, y  $h[n]$  siendo el filtro de  $M+1$  muestras, en este caso 512 muestras también, se obtiene la señal de salida  $y[n]$  cuya duración es de  $L+M$  muestras. Así, se implementa un filtrado teóricamente eficiente de la señal captada por el micrófono, que localiza dicho sonido en el espacio tridimensional mediante el filtrado por HRTFs, sin ninguna pérdida teórica de muestras.

Además, se comprueba que *Playrec* es capaz de realizar el filtrado tanto en el dominio temporal, como en el dominio frecuencial, realizando el filtrado de la señal captada por un micrófono por una HRTF de una localización escogida al azar. Se opta por realizar el filtrado para la localización de  $-90^\circ$  para poder apreciar correctamente los resultados del filtrado.

En este punto, se plantea un nuevo problema a resolver en el planteamiento de la interfaz gráfica presentada, y es que mientras *Playrec* está ejecutándose, no existe forma alguna de alterar sus variables sin que sea dentro del bucle de toma, procesado y reproducción de las muestras, para que se pueda apreciar la variación de la localización mientras se está ejecutando. En este punto, parece que la única opción viable resulta la de incluir todo el código asociado al cálculo de variables como azimut, y elevación, búsqueda de su correspondiente HRTF, filtrado, cálculo de los pesos de cada una de las interpolaciones, además de las llamadas a cada uno de los elementos de la interfaz gráfica, dentro del código de ejecución de *Playrec*.

### III.3. Métodos de interpolación desarrollados

Como bien se menciona en la introducción de este apartado de la memoria, a continuación se describen los diferentes métodos de interpolación desarrollados. Como ya se ha presentado anteriormente, esta aplicación no realiza la interpolación de las HRTFs como tal sino que interpola las señales de salida de estos filtros para crear la sensación de localizar las fuentes en cualquier punto del espacio permitido.

Cabe mencionar, antes de cualquier descripción de cualquier método de interpolación desarrollado, que los métodos empleados en el presente proyecto son métodos orientados a la implementación de la aplicación en tiempo real, por lo que los métodos de interpolación de alto coste computacional se han descartado por no cumplir con este requisito. De esta forma, la *interpolación por división en subbandas y alineamiento de las portadoras* [23] o *sinc interpolator* [24] no se ha hecho además de incluirlas por exigir hasta 8 filtrados en frecuencia por cada uno de los canales de audio a reproducir, lo que dificulta en un alto grado, hasta mayor eficiencia computacional, su implementación orientada al tiempo real. Ambos métodos de interpolación de HRTFs han resultado ser los más eficientes incluso en casos donde se han programado con bibliotecas que a priori, no cumplen el criterio del *aliasing espacial* de Nyquist, como es el caso de la *interpolación por división en subbandas y alineamiento de las portadoras*. Partiendo de esa base, lo que pretende este estudio es determinar de forma subjetiva, el grado de precisión que los métodos de interpolación deben tener para generar un campo sonoro tridimensional a demanda del usuario y eficiente en el aspecto subjetivo de la escucha.

Como bien establecieron Lord Rayleigh, [2] y posteriormente, más firmemente, Blauert [3], el oído tiene una menor capacidad para discernir las diferencias de elevación que las de acimut y a su vez tiene una menor capacidad de distinción de la distancia del oyente a la fuente que la elevación. Debido a este hecho, los métodos de interpolación desarrollados solo contemplan la interpolación de las salidas de las HRTFs en el plano horizontal pero son fácilmente extrapolables a la elevación.

En todos los métodos de interpolación que se desarrollan a continuación, y en cualquier algoritmo que pretenda reconstruir las señales provenientes de cualquier punto de un círculo, es necesario cumplir el criterio de *aliasing espacial*, muy similar al *aliasing frecuencial* o *aliasing temporal* más usual en el tratamiento de señales. En el criterio definido por Ajdler y sus compañeros [24] que se cita a lo largo del trabajo como criterio de Nyquist en pos de la no obtención de *aliasing espacial* en la generación de campos sonoros en un círculo, tiene como razón de ser la incertidumbre que se produce en un frente de ondas cuando se muestrea en puntos separados del orden de la longitud de onda de la señal. Así, suponiendo una señal con una longitud de onda muy pequeña que se 'muestrea espacialmente' en dos puntos (en este caso los micrófonos que se usan para medir las HRTF's), si los micrófonos está muy separados, las señales con longitudes de onda pequeñas se captaran en fases destructivas o constructivas (según la longitud de onda) creando efectos indeseados como los similares a un filtrado en peine (cuando se suman en

contrafase) o coloración de la señal (si las fuentes, además se mueven, se puede apreciar un efecto de modulación). Por lo tanto, si no se muestrea espacialmente con la distancia adecuada, la señal que representemos a partir del muestreo espacial no será capaz de representar fielmente a la señal original [24] y ése es un requisito a cumplir necesariamente.

$$|l_\theta| \leq |\omega| \frac{d}{2c} \approx |\omega| \frac{0.09}{c}. \quad (1)$$

En la fórmula superior se considera la separación fija, (la distancia entre los oídos de un adulto, 18cm aprox.) y aproximan el coseno del ángulo de incidencia a la separación de las HRTF's,  $\omega$  es la frecuencia angular máxima de la señal generada.

### III.3.1. Aproximación por la HRTF más cercana: Interpolación por umbrales

A continuación se procede a la descripción del primero de los métodos de interpolación de señales generado. Obviamente, como su propio nombre indica en este caso no se realiza ningún tipo de interpolación ya que se trata de reproducir la señal filtrada por la HRTF más cercana al ángulo demandado por el usuario.

De esta forma, lo que se ha realizado ha sido la medida de la distancia desde el ángulo demandado por el usuario ( $\Theta_i$ ) a los ángulos más cercanos (se entiende como el anterior y posterior) de los que se dispone de HRTF ( $\Theta_i$ ) y ( $\Theta_{i+1}$ ) mediante la sencilla regla expresada a continuación:

Si el valor absoluto de la resta (es decir si el ángulo contenido,  $\alpha$  o  $\beta$ ) entre el ángulo de acimut deseado ( $\Theta_i$ ) y el ángulo anterior ( $\Theta_j$ ) es menor que el valor de la resta entre el ángulo deseado ( $\Theta_i$ ) y el ángulo posterior ( $\Theta_{j+1}$ ) se selecciona la HRTF correspondiente al ángulo anterior ( $\Theta_j$ ) y viceversa. Con la imagen siguiente se intenta expresar lo comentado:

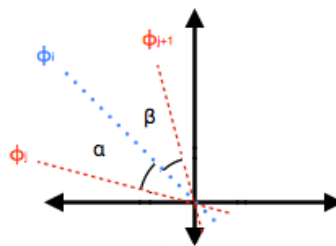


Fig.14. Detalle del método de elección de HRTF.

Con el siguiente método se trata de establecer de forma objetiva y subjetiva la importancia de interpolar las señales con alguno de los métodos existentes, o en el presente trabajo, con los que se presentan a continuación, para crear un campo sonoro proveniente de cualquier punto del espacio. Además se pretende observar en los resultados hasta que punto es necesario interpolar las señales de salida de las HRTFs en pos de obtener un campo sonoro en cualquier punto de un círculo.

### III.3.2. Ponderación lineal inversamente proporcional a la distancia

En este apartado, se procede a la descripción de un método de interpolación basado en la bibliografía existente al respecto [26], en la que la localización de la señal se realiza ponderando las señales de salida obtenidas mediante el filtrado por las HRTFs más cercanas al ángulo demandado por el usuario. De esta forma, lo que se pretende es establecer una regla que permita dotar de mayor importancia a la salida de la HRTF que más cerca se encuentre del ángulo determinado por el usuario, y dotar de menos a la que más lejos está, siguiendo una regla de proporcionalidad lineal comprendida entre 0 y 1.

Así, se pretende desarrollar un método de interpolación de las señales de salida de las HRTFs similar al desarrollado por Hartung, Braasch, y Sterbing en su paper ‘Comparison of different methods for the interpolation of Head-Related Transfer Functions’ [25], previamente desarrollado por otros investigadores.

Una vez definido el objetivo, primero se procede al cálculo de la distancia, aproximando dicha distancia por la cuerda que une los ángulos, o puntos más próximos donde se dispone de HRTF, extremos. Es decir, que la aproximación de la distancia se realiza mediante la aproximación de la longitud de la cuerda angular (es decir, aproximamos el arco por su tangente) siendo sólo válido, como ya es sabido, para ángulos muy pequeños. Tendiendo en cuenta que las HRTFs, como bien se ha mencionado en el método experimental, están separadas entre sí  $15^\circ$ , se decide aproximar el arco por su tangente obteniendo buenos resultados.

Así de esta forma, se define la regla de interpolación de las señales de salida de los filtros HRTF mediante una ponderación inversamente proporcional a la distancia entre los puntos ( $D_1$  y  $D_2$ ) sobre el arco del círculo de 1m creado (dibujado en color verde en la fig. 15) uniendo los puntos extremos en los que se dispone de HRTFs medidas. El ángulo de acimut deseado es denominado ( $\Theta_j$ ), el ángulo anterior ( $\Theta_j$ ) y el ángulo posterior de las que se dispone de HRTF para el filtrado es denominado ( $\Theta_{j+1}$ ).

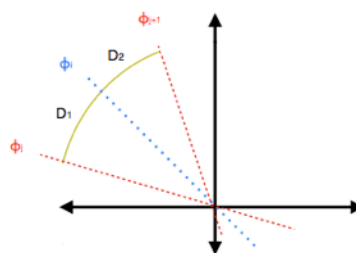


Fig.15. Detalle de la toma de variables para el cálculo de los pesos.

De esta forma, el ratio de mezcla, o los pesos adscritos a cada una de las señales de salida disponibles en los ángulos anterior y posterior ( $\Theta_j$  y  $\Theta_{j+1}$ ) se definen como:

$$\begin{aligned} ratio(\theta_j) &= \frac{|\theta_i - \theta_j|}{|\theta_{j+1} - \theta_j|} \\ ratio(\theta_{j+1}) &= \frac{|\theta_{j+1} - \theta_i|}{|\theta_{j+1} - \theta_j|} \end{aligned} \quad (2)$$

Partiendo de las siguientes expresiones (3), que son normalizadas y simplificadas para obtener las expresiones situadas sobre estas líneas (2), se puede observar que se derivan de la medida de la longitud del arco entre el ángulos de acimut deseado ( $\theta_i$ ), los ángulos anterior ( $\theta_j$ ) y posterior ( $\theta_{j+1}$ ).

$$\begin{aligned} ratio(\theta_j) &= \frac{2\pi}{360} \cdot (\theta_i - \theta_j) \\ ratio(\theta_{j+1}) &= \frac{2\pi}{360} \cdot (\theta_{j+1} - \theta_i) \end{aligned} \quad (3)$$

Obviamente, para que los pesos sean inversamente proporcionales a la distancia aproximada por el arco se le debe de restar el valor unidad a los pesos definidos para obtener la señal de salida ( $y$ ), como bien se muestra en la siguiente expresión que se define para cada uno de los canales estéreo procesados:

$$y_{izq}(\theta_i) = (1 - ratio(\theta_j)) \cdot y_{izq}(\theta_j) + (1 - ratio(\theta_{j+1})) \cdot y_{izq}(\theta_{j+1}) \quad (4)$$

De esta forma, la salida de la HRTF más cercana al ángulo deseado, obtiene un mayor peso en la realización de la composición del sonido, que la que está más alejada, otorgando así pesos equitativos según la necesidad.

Es cierto que para los ángulos extremos, no ha sido necesaria la introducción de excepciones debido a que el método desarrollado para la mezcla de las salidas, tiene en cuenta que el usuario pueda pinchar en un ángulo de acimut ( $\theta_i$ ) del que se disponga de la HRTF, en cuyo caso, al ser la distancia mínima entre la salida de la HRTF deseada y la salida de la HRTF que se dispone, el peso otorgado a la salida de la HRTF disponible es máximo haciendo que nada de la HRTF más alejada se mezcle.

A continuación, se presentan los pesos que irán adscritos a cada una de las señales de salida de las HRTFs más cercanas disponibles al ángulo de localización de la fuente demandado por el usuario. De esta forma, y como ya es sabido que se cuenta con HRTFs cada  $15^\circ$ , se procede a introducir los pesos definidos por este método de interpolación para dicho rango angular en el que no se dispone de HRTFs para su filtrado por la señal monoaural. Suponiendo que  $sound_n$  es la señal filtrada por la HRTF en  $0^\circ$  y  $sound_m$  es la señal filtrada por la HRTF en  $15^\circ$ :

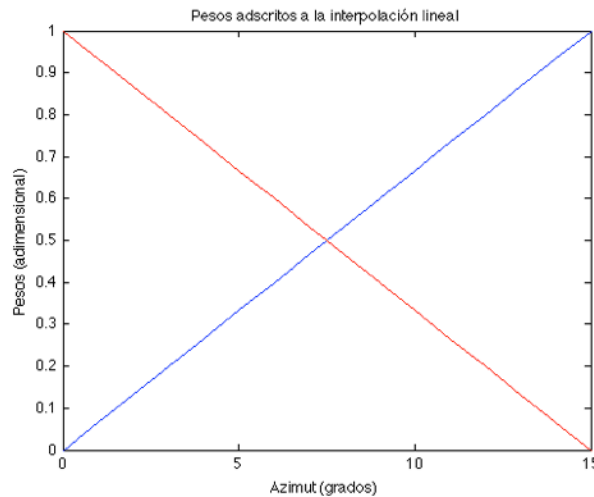


Fig.16. Pesos obtenidos para cada una de las separaciones entre los ángulos de los que se posee de HRTF.

Es necesario mencionar que en este caso, no ha sido necesario procesar el rango dinámico de la señal resultante, ya que al ser una suma, como se muestra en la imagen superior y en la expresión matemática (3), los niveles no sobrepasan el rango dinámico de -1 y 1 al ser como mucho multiplicadas las salidas en los puntos, por el valor máximo de cualquiera de los pesos generados. Obviamente la curva azul (recordando que los pesos están sin invertir) es la correspondiente a *sound\_n* y la roja es la correspondiente a *sound\_m*.

### III.3.3. Ponderación cuadrática de las señales de salida

En medio del desarrollo del método de interpolación anterior, se genera la idea de poder dotar de mayor peso aún a la salida de la HRTFs más cercanas al ángulo seleccionado por el usuario y dotar de menos (proporcionalmente hablando, respecto al método de interpolación anterior) a la salida del ángulo más alejado, pretendiendo además, que este método no altere (al igual que el anterior) el rango dinámico de la señal resultante, por lo que resulta necesario establecer el peso de mezcla de las señales en el caso en el que el usuario demanda el punto medio entre los dos ángulos más cercanos de los que se dispone de HRTF.

De esta forma, y persiguiendo que los pesos en los puntos más cercanos sean mayores que los ofrecidos por la interpolación lineal y tratando de observar si al tender en un mayor grado hacia las HRTFs disponibles, la interpolación ofrece mejores resultados, se implementa este método haciendo uso de un sistema básico de ecuaciones.

Así, y sabiendo que el método debe seguir una expresión matemática genérica, se opta por la expresión cuadrática siguiente:

$$y = k + mx^2 \quad (5)$$

Por lo tanto, sabiendo que  $y$  debe valer 0 en  $\theta_j$  y sabiendo también que  $y$  debe valer 0.5 en el punto medio entre  $\theta_j$  y  $\theta_{j+1}$  se obtienen los valores de  $k$  y  $m$  dependientes del valor del ángulo central  $\Delta\theta$  y de la separación angular entre  $\theta_j$  o  $\theta_{j+1}$ . Las expresiones son las siguientes:

$$k = - \left( \frac{(\Delta\theta - (\theta_{j+1} - \theta_j))^2}{2 \cdot ((2 \cdot (\theta_{j+1} - \theta_j) \cdot \Delta\theta) - \Delta\theta^2)} \right) \quad (6)$$

$$m = \frac{1}{2 \cdot ((2 \cdot (\theta_{j+1} - \theta_j) \cdot \Delta\theta) - \Delta\theta^2)}$$

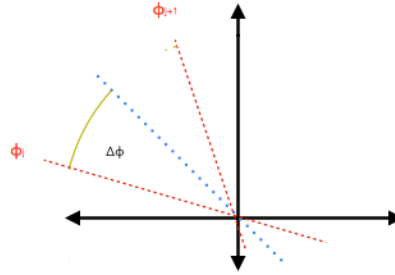


Fig.17. Detalle de la toma de variables para el cálculo de los pesos.

De esta forma, el ratio de mezcla, o los pesos adscritos a cada una de las señales de salida disponibles en los ángulos anterior y posterior ( $\theta_j$  y  $\theta_{j+1}$ ) suponiendo que el ángulo deseado es  $\theta$ , tiene la siguiente forma:

$$\begin{aligned} ratio(\theta_j) &= 1 - \left( - \left( \frac{(\Delta\theta - (\theta_{j+1} - \theta_j))^2}{2 \cdot ((2 \cdot (\theta_{j+1} - \theta_j) \cdot \Delta\theta) - \Delta\theta^2)} \right) + \left( \frac{1}{2 \cdot ((2 \cdot (\theta_{j+1} - \theta_j) \cdot \Delta\theta) - \Delta\theta^2)} \right) \cdot \theta^2 \right) \\ ratio(\theta_{j+1}) &= \left( - \left( \frac{(\Delta\theta - (\theta_{j+1} - \theta_j))^2}{2 \cdot ((2 \cdot (\theta_{j+1} - \theta_j) \cdot \Delta\theta) - \Delta\theta^2)} \right) + \left( \frac{1}{2 \cdot ((2 \cdot (\theta_{j+1} - \theta_j) \cdot \Delta\theta) - \Delta\theta^2)} \right) \cdot \theta^2 \right) \end{aligned} \quad (7)$$

Obviamente, como los pesos adscritos a cada una de las señales disponibles, son distintos pero complementarios, por lo que en el momento de realizar la suma de señales para obtener  $y$ , se aplica a cada una de las señales sus propios pesos sin necesidad de que sean invertidos:

$$y_{izq}(\theta_i) = ratio(\theta_j) \cdot y_{izq}(\theta_j) + ratio(\theta_{j+1}) \cdot y_{izq}(\theta_{j+1}) \quad (8)$$

De esta forma y siguiendo los pasos descritos en líneas anteriores, se obtiene el conjunto de pesos que irán adscritos a cada una de las señales filtradas por las HRTFs más cercanas al ángulo demandado por el usuario. Nuevamente como solo se disponen de HRTFs cada  $15^\circ$ , a continuación se presentan los pesos generados para cada uno de los ángulos de separación entre cualesquiera HRTFs más cercanas disponibles.

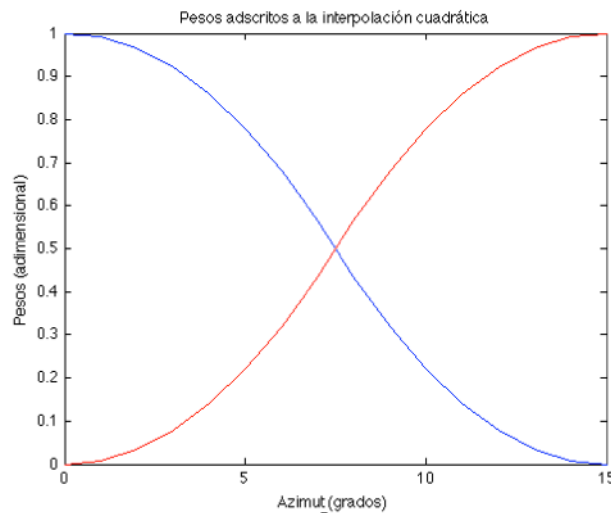


Fig.18. Pesos obtenidos para cada una de las separaciones entre los ángulos de los que se posee de HRTF.

Obviamente, y debido a que se ha decidido que sea de esta forma, los pesos generados no alteran en ningún caso el rango dinámico de la señal resultante, haciendo de este un método novedoso de interpolación, debido a la creación propia de las expresiones y a que la literatura no ofrece casos en los que se haya optado por una interpolación cuadrática en vez de lineal. Si bien es cierto, que el presente método es mucho menos riguroso que el lineal en cuanto a justificación científica, ambos comparten la misma idea intuitiva de ofrecer un mayor peso, el método cuadrático más aún, a la señal de la HRTF más cercana al ángulo demandado por el usuario.

#### III.4. Interfaces gráficas generadas: Capacidades y limitaciones

Como bien se detalla en la introducción del método experimental, a continuación se procede a presentar la interfaz generada y para ello, se procederá primero a detallar el funcionamiento de las mismas, detallando lo que es capaz de hacer, para después ahondar en las limitaciones que tiene.

Primero, presentaremos la interfaz generada para la localización multifuente, enumerando los distintos controles que permiten el uso de la interfaz, para después detallar su funcionamiento y capacidades.

Atendiendo a la imagen inferior (fig.19), se observa el botón 'Salir' numerado con el número 5 en la imagen, que es el encargado de dejar que la interfaz se deje de ejecutar y se cierre. Debido a que en la ejecución de programa, hay un bucle, el cual se encarga del cálculo de las señales de salida interpoladas y su reproducción, se crea el problema en la ejecución del botón salir. Este problema consiste en que cuando se ha decidido dejar de utilizar la interfaz generada, y se aprieta el botón, los controles de la interfaz dejan de funcionar, pero el programa sigue reproduciendo sonidos. Concretamente realiza la reproducción de todos los sonidos que se han generado durante la ejecución siguiendo el orden inverso (es decir del último al primero), hasta que han sido todos reproducidos.



Esta ‘limitación’ en la ejecución de la interfaz se ha intentado subsanar mediante distintas técnicas sin éxito alguno. Debido a esto, se pretende identificar el error del código que genera dichas instrucciones concluyendo en que la reproducción de los sonidos en el orden inverso al generado por el usuario se debe al hecho de que al pulsar cualquiera del resto de controles para cambiar cualquiera de las variables del sonido generado, el bucle de generación de sonidos, sigue generando audio mientras dichas instrucciones se ejecutan (y el sonido se reproduce), lo que genera que haya instrucciones correspondientes a la reproducción de audio que se quedan en una ‘cola’ para su posterior ejecución. Este efecto indeseado no se ha sabido rectificar adecuadamente, y debido a que el usuario puede quitarse los auriculares al pulsar el botón ‘Salir’ sin que se aprecie dicho efecto, no se la ha decidido dar mayor importancia.

A continuación, se procede a describir las capacidades y limitaciones de los paneles numerados como 3 y 4 en los que se dota al usuario de la interfaz, de la libertad de seleccionar el tipo de interpolación deseado y el tipo de sonido a generar para ser filtrado por cualquiera de las HRTFs correspondientes. Ambos paneles, son directamente accesibles por el usuario durante la reproducción de cualquier sonido deseado, lo que dota a la interfaz de mayor flexibilidad, y sus cambios son perceptibles únicamente esperando a que el sonido en actual reproducción termine.

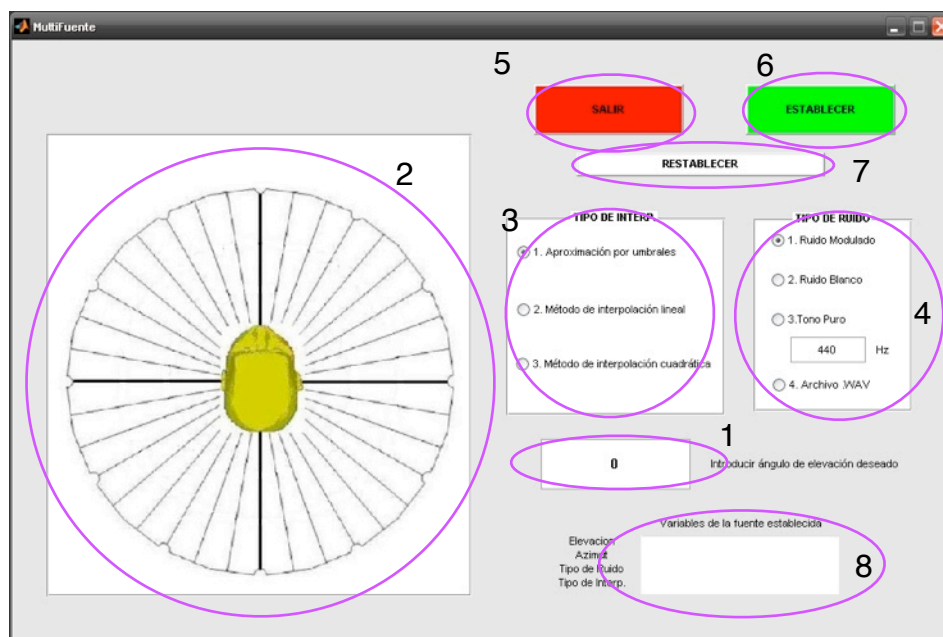


Fig.19. Detalle de la interfaz generada.

De esta forma se consigue que el usuario pueda realizar y apreciar dichos cambios realizados en pseudo tiempo real. Uno de los problemas que se pueden identificar en el uso de la aplicación, es que su información, es actualizada únicamente cuando se decide cambiar la posición de la fuente en el espacio, más concretamente, cuando se altera (o no se altera pero se pulsa *enter*) el contenido del campo de inserción de la información de la elevación, por lo que a pesar de que la ejecución de la interfaz es en pseudo tiempo real, la visualización de los cambios generados por el usuario solo se actualiza cada vez que el usuario decide cambiar la posición de la fuente. Este efecto no sucedería

si no se utilizase la instrucción *'sound'*, ya que dicha instrucción necesita toda la señal a reproducir antes de ejecutarse y no acaba hasta que se reproduce toda.

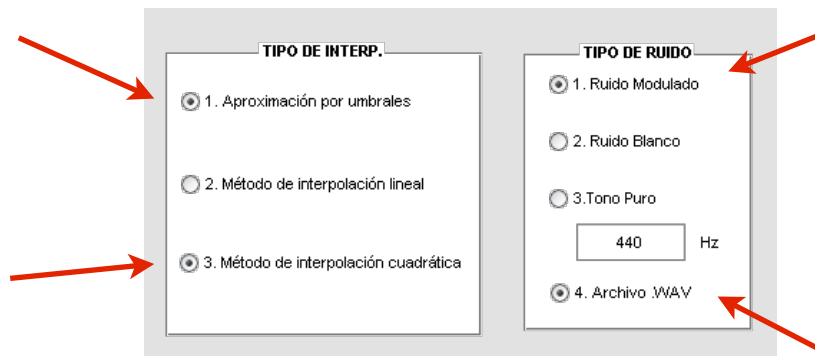


Fig.20. Detalle de la limitación en la ejecución de la interfaz.

Esto genera, durante la utilización de la interfaz, por ejemplo en el caso del panel de selección del tipo de sonido, que si el usuario decide intercambiar los tipos de sonido sin alterar la localización, los botones de selección de cada uno de los tipos de sonido, se quedan ‘activados’ dando lugar a una posible confusión visual de cual es el sonido que actualmente se está reproduciendo (véase fig.20). Este hecho no es demasiado importante en el panel de selección del tipo de sonido, porque con la información auditiva recibida, se puede determinar claramente cual de ellos está realmente activado, pero no es tan sencillo en el caso del panel de selección de interpolación.

Como ya se ha mencionado, en el panel de selección del tipo de interpolación, el error sigue siendo el mismo, pero esta vez el efecto es bastante más notorio en el aspecto en que la información recibida por vía auditiva no es tan concluyente como lo es en el caso de la selección del tipo de sonido. Por ello, y debido a que el control de la ejecución del programa no ofrece otra solución, ya que en teoría, cada vez que se pulsa uno de los botones, este debería ser activado y el resto desactivados, se opta por que sea el usuario el que marque y a su vez desmarque el tipo de interpolación deseado durante la ejecución. Es decir, se plantea como solución, que el usuario decida y active el tipo de interpolación deseado, pero que antes de que decida activar otro tipo de interpolación, desactive la apariencia del botón actualmente activado y posteriormente active el método de interpolación deseado, para que la utilización de la interfaz no lleve a errores de interpretación de las sensaciones generadas si se diese el caso.

Una vez concluida la descripción de la capacidad de las partes 3, 4 y 5 de la interfaz presentada, se procede a la explicación de la forma de funcionamiento, capacidad y limitaciones de las partes 1 y 2 de la interfaz presente en la fig. 19.

Las partes mencionadas, tienen una ejecución ligada entre ellas de forma que para que se pueda seleccionar un punto en la parte 2 de la interfaz, correspondiente a la selección del ángulo de acimut deseado, primero es necesario establecer (mediante el pulsado de la tecla *enter*) el ángulo de elevación deseado en la parte de la interfaz identificada con el número 1 en la fig. 19. Esto puede

parecer a simple vista, que el usuario debe realizar la selección de ambas variables de una forma rápida (cuando se está familiarizado con la interfaz, se hace), para que la reproducción de los sonidos no cese en ningún momento. Es cierto que, si el usuario establece el ángulo de elevación en su correspondiente casilla y no procede a la selección de un ángulo de acimut en el tiempo que tarda en reproducirse el último sonido generado, la reproducción se detiene hasta la introducción del ángulo de acimut deseado.

Este hecho puede resultar un problema desde que se describe que es necesario fijar un valor de elevación en su campo correspondiente para el establecimiento del acimut, pero el funcionamiento se plantea de forma inversa. En las instrucciones presentadas al usuario se pide primero que decida de forma personal, interior, donde quiere situar la fuente para acto seguido introducir el ángulo de elevación en su campo, establecer el ratón en el ángulo de acimut deseado, establecer el ángulo de elevación introducido pulsando la tecla *enter* y seguidamente clicar en la imagen para que se establezca el ángulo de acimut deseado.

Tras el uso continuado de la interfaz, el usuario ha demostrado realizar este proceso sin apenas tener conciencia de las limitaciones de la interfaz y sobretodo, sin tener sensación de que la reproducción de los sonidos generados es discontinua. Esta sensación percibida, es altamente satisfactoria si en el empleo de la interfaz se siguen las consejos mencionados.

Es necesario mencionar, que para que la interfaz comience a funcionar, únicamente es necesario establecer los ángulos de elevación y acimut debido a que se establece el primer método de interpolación y el primer tipo de sonido de sus correspondientes opciones de la lista, para ofrecer un comienzo de la reproducción de sonidos, en la que no haya errores por la falta de establecimiento de las variables que entran en juego.

Como se puede apreciar en color verde y numerado con el 6 en la figura 19, se encuentra el botón establecer. La función de este botón es fijar la fuente que actualmente se esta reproduciendo, almacenando sus variables en la casilla de texto estático blanca numerada con el 8, para que al establecer una nueva fuente, véase alterar la localización, tipo de sonido o método de interpolación, esta nueva fuente sea sumada a la ya establecida y almacenada en el objeto mentado.

La activación del botón establecer, se puede realizar en cualquier punto de la ejecución del programa, siempre que se esté ejecutando la reproducción de alguna fuente ya localizada y simulada. Pero volvemos a encontrar que los cambios realizados solo se visualizan cuando se establece de nuevo la información de elevación, de la misma forma que ocurría con el resto de la información a presentar, a pesar de que, como en los casos anteriores, los cambios se reproduzcan en pseudo tiempo real.

En cuanto al botón denominado ‘restablecer’ en la interfaz descrita, presenta un comportamiento similar en cuanto a limitaciones de refiere que el botón ‘establecer’ anteriormente explicado, pero su comportamiento es opuesto. Es decir, que este botón se encarga de eliminar la fuente anteriormente establecida por el botón ‘establecer’, además de limpiar la tabla de texto que almacena sus variables sin que se deje de emitir el sonido generado por la última de las fuentes situadas antes de ser presionado.

### III.5. Test psicoacústico para la toma de información

Como ya se ha detallado en la introducción de este apartado referente al método experimental utilizado en este trabajo, a continuación se procede a la descripción del pequeño test psicoacústico implementado, intentando obtener conclusiones subjetivas de los métodos de interpolación propuestos.

De esta forma primero se decide establecer el método de elección de la respuesta y tras ello construir el test con las variables que se estimen oportunas. El método de elección propuesto para determinar por el sujeto sometido al test si una de las opciones es mejor valorada que las otras ha sido el de ABX Test en que se pregunta al sujeto cual de las dos opciones propuestas se parecen más (o crea una sensación similar) a la señal original que se presenta.

Una vez se decide que tipo de test psicoacústico que se va a implementar para una mejor obtención y extracción de información se decide el criterio a establecer para determinar si una respuesta es más valorada que otras y por lo tanto, establecer cual es el método de interpolación que mejores resultados ofrece. J.S. Bradley [26] por ejemplo, y siendo su criterio de determinación del valor investigado el utilizado en el presente trabajo, determinó el *jnd* del parámetro  $C_{50}$ , haciendo que los sujetos escuchasen pares de campos sonoros (en este caso, un discurso oral) de variadas diferencias en los valores de dicho parámetro, para determinar el *jnd* como el valor de la diferencia escuchada por el 50% de los sujetos que realizaron la prueba.

Además de ese criterio se utilizará también un criterio estadístico que nos permite medir la aleatoriedad o azar con el que los sujetos responden a un determinado conjunto de preguntas del test que compartan alguna de sus características. El método Chi-cuadrado permite examinar si un conjunto de datos mantienen una determinada tendencia o si por el contrario las preguntas son respondidas al azar, lo que demostrará una preferencia en los sujetos por una configuración dada. El grado de significación elegido para confirmar o rechazar una configuración frente a otra ha sido  $p < 0,05$ .

Por lo tanto, en el presente proyecto se tratará de establecer cual de los métodos es mejor en su labor de interpolación de señales primero estableciendo el/los métodos que ofrecen un porcentaje de respuestas mayor al 50% y se analizará la aleatoriedad de las respuestas de los sujetos mediante la ‘Chi-Cuadrado de Pearson’ y, tras ello, haciendo uso de la consistencia con la que se ha creado el test, se determinará, si es posible, el método que mejores resultados ofrece según el caso de interpolación propuesto.

Además, se ha decidido introducir el factor repetitividad en el test, introduciendo una pregunta trampa en cada una de las páginas (fig. 21) del test creado para poder valorar la fiabilidad del sujeto en la realización del test y evitar la aleatoriedad en las respuestas.

El test psicoacústico se ha implementado de la siguiente manera:

- La señal a comparar, denominada ‘señal original’ en el test, es la señal de salida obtenida filtrando los tres ‘sonidos’ presentados con la HRTF correspondiente al ángulo de 30°.

- Esta señal original, deberá ser comparada con señales interpoladas que simulan o pretenden simular la misma localización de la fuente en el espacio que la señal original.
- Esta interpolación se ha realizado intentando simular diferentes situaciones en las interpolaciones implementadas en la aplicación desarrollada. De esta forma, se presentan señales interpoladas filtradas por las HRTFs de  $0^\circ$  y  $60^\circ$ , de  $15^\circ$  y  $45^\circ$ , de  $0^\circ$  y  $45^\circ$  y de  $15^\circ$  y  $60^\circ$ . Como se puede observar, estas interpolaciones contienen tanto casos ‘centrados’ en lo que la localización deseada está en el punto medio de los ángulos de las HRTFs utilizadas para interpolar, como casos asimétricos donde la proporción no es la mitad, para observar el comportamiento en un mayor número de casos de todos los posibles sin aumentar mucho el rango angular utilizado para la interpolación (en este caso entre  $0^\circ$  y  $60^\circ$ ) y sin multiplicar el número de pruebas a presentar a los sujetos.
- Debido a lo comentado sobre el rango angular utilizado, es necesario mencionar que las señales utilizadas en el test deben cumplir el criterio de Nyquist para evitar el efecto del *aliasing espacial*. De esta forma, siguiendo la expresión (1) y sabiendo que la separación angular máxima entre las HRTFs utilizadas para interpolar es de  $60^\circ$ , la frecuencia máxima a utilizar en las señales emitidas en el test, para que el campo sonoro se cree correctamente es de 1819Hz.
- Como consecuencia de la restricción en la frecuencia de las señales, se prevé necesario utilizar un conjunto de señales que incluye un tono puro de 500Hz, que cumple el criterio de Nyquist, un ruido aleatorio de densidad espectral constante filtrado a 1500Hz que también cumple el criterio, y un tono puro fuera del cumplimiento del criterio de Nyquist de 2kHz para establecer diferencias si la hubiere.
- Por último se decide formar el test con cierta consistencia de forma que las respuestas de los sujetos ayuden a discriminar los resultados del mismo. De esta forma, se presentan los tres métodos de interpolación desarrollados por pares, siguiendo la regla de que si uno es mejor que otro, por la comparación entre los tres, el resultado con más votos sea el mejor de los tres.
- No se contempla la indecisión por parte del sujeto, por ello, no se propone la opción de que el par de interpolaciones propuestas sean igual de buenas en la localización de la fuente. Para contrarrestar la posible dificultad de poder decidir entre una de las dos opciones, se da al sujeto la opción de repetir la escucha de la señal original y el par de señales interpoladas tantas veces como desee.

Una vez sentados los estatutos del test psicoacústico, se procede a crear en MatLab una interfaz gráfica simple que ayude a la recolección de los datos ofrecidos por los sujetos sometidos al test de un modo computacional y automático. Así se implementa la siguiente interfaz que se desgrana en el párrafo siguiente.

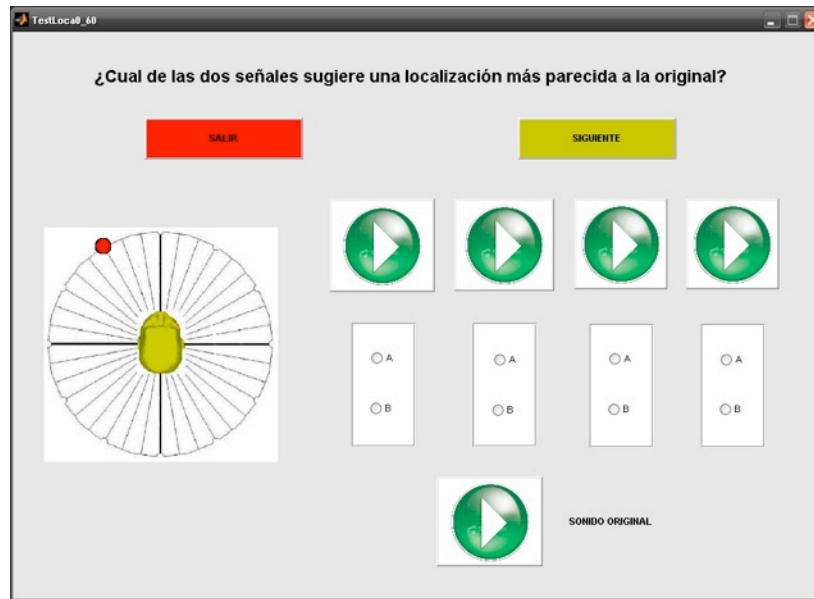


Fig.21. Detalle de la interfaz para el test psicoacústico generada.

La interfaz presentada cuenta con ciertos controles, pero su funcionamiento es básico. El botón ‘Play’ asociado a la señal original reproduce la localización de la fuente en  $30^\circ$  haciendo uso de la HRTF medida en ese punto acimutal. De la misma forma, cada uno de los botones ‘Play’ asociado a cada panel de dos botones, reproduce la señal interpolada mediante los tres métodos de interpolación descritos para cada par de ángulos utilizados para interpolar, y también para cada tipo de señal de los tres propuestos. Cada panel de botones asociado a cada uno de los botones de reproducción servirá para que el usuario determine cual de las dos señales está mejor interpolada. De esta forma, cuando el usuario haya respondido a la pregunta formulada en cada uno de los paneles de cada página del test, al pulsar el botón ‘siguiente’, se procederá al almacenamiento de sus respuestas para su posterior análisis.

#### 4. Resultados

Una vez se han implementado las interpolaciones necesarias para las interfaces generadas, resulta ético calcular cual es el mejor método de interpolación de todos los implementados. Esta acción en el presente proyecto se puede llevar a cabo de dos formas completamente distintas. Por un lado, se puede programar un pequeño test psicoacústico en el que un número determinado de sujetos determinen cual es el método de interpolación que mejores resultados ofrece en cuanto a la localización de las fuentes. Pero por otro lado se puede realizar un análisis matemático de los resultados, comparando la señales interpoladas en un número determinado de ángulos, con las señales generadas con las HRTFs existentes en esos ángulos. Por lo tanto, en este apartado de resultados, primero se procederá al análisis de los resultados ofrecidos por las interpolaciones de una forma numérica y tras ello, se hará uso del pequeño test psicoacústico implementado para

intentar obtener resultados subjetivos de cual es el mejor método de interpolación de los desarrollados.

#### IV.1. Comparación numérica de los métodos de interpolación

En el siguiente apartado se procederá al análisis de los métodos de interpolación desarrollados, a través de cálculos numéricos que pretenden establecer la diferencia entre las señales de salida interpoladas con cada uno de los métodos, y la señal original en ese mismo punto.

Más concretamente, se ha decidido establecer mediante el MSE (*Mean Square Error*) las diferencias entre las señales interpoladas en determinados ángulos a través de las HRTFs disponibles en dos ángulos conocidos y extremos y las señales obtenidas al ser filtradas por las HRTFs disponibles en esos mismos ángulos que han sido interpolados.

De esta forma, se ha decidido realizar la interpolación de los ángulos existentes cada  $15^\circ$  dentro del rango de  $[-60^\circ, +60^\circ]$ , siendo  $0^\circ$  una fuente situada de frente al sujeto y a la altura de sus oídos, utilizando las HRTFs disponibles de  $-60^\circ$ ,  $0^\circ$  y  $60^\circ$  y comparando las señales interpoladas con las señales obtenidas de las HRTFs existentes en cada uno de los puntos de interpolación establecidos. Intentando explicar lo anterior se crea esta imagen, en las que los puntos verdes son las HRTFs reales utilizadas para realizar la interpolación en los puntos morados en lo que se dispone de HRTFs medidas, para realizar la comparación entre las señales interpoladas y las señales medidas.

Como ya se ha mentado, el método elegido para comparar estas señales ha sido el Error Cuadrático Medio o *Mean Square Error*, que se es normalmente denominado MSE. El MSE es normalmente expresado en unidades logarítmicas para la comparación muestra a muestra de las señales implicadas en el cálculo, utilizada por la mayor parte de investigadores que han servido de orientación y base para el presente proyecto [13] [25] [26].

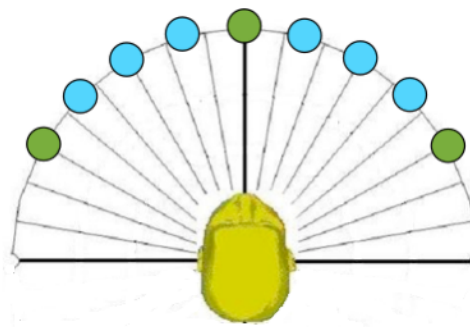


Fig.22. Detalle de los puntos donde se ha dispuesto de HRTFs y puntos donde se han interpolado las localizaciones.

El MSE se define para cada uno de los oídos como la división entre la suma acumulativa de la resta al cuadrado entre la señal original ( $y$ ) y la señal interpolada ( $y_{int}$ ) en ese punto de acimut ( $\theta_i$ ) y

la suma acumulativa de la señal original ( $y$ ) al cuadrado en  $\theta_i$ , ambas sumas realizadas muestra a muestra. Notar que este cálculo puede hacerse tanto en el dominio temporal como en el frecuencia:

$$MSE(\theta) = 10 \log \left( \frac{\sum_{i=1}^N (y(\theta, i) - y_{int}(\theta, i))^2}{\sum_{i=1}^N y(\theta, i)^2} \right) \quad (9)$$

Realizando este cálculo para cada una de las señales en todo el rango angular propuesto para el análisis, se obtiene la magnitud decibélica de la comparación muestra a muestra de la señales originales (suponiendo estas medidas con las HRTFs reales) con la señales interpoladas para cada uno de los oídos por separado. Los resultados obtenidos se presentan en las siguientes figuras:

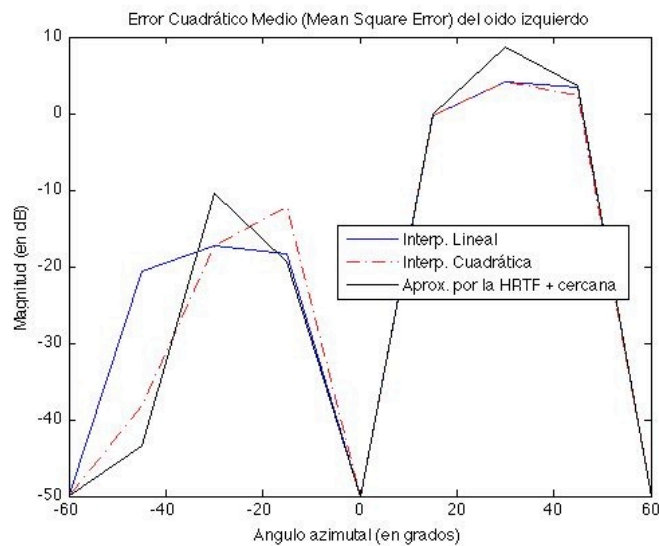


Fig.23. MSE asociado al oído izquierdo.

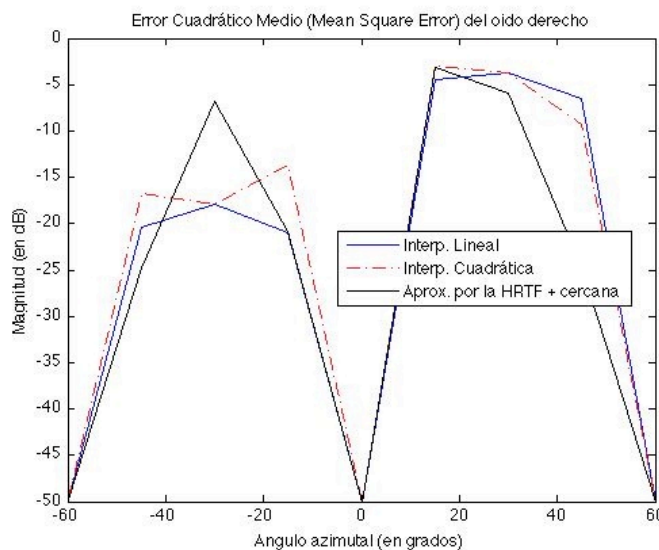


Fig.24. MSE asociado al oído derecho.



Ya que lo que se está intentando medir es la eficiencia de las interpolación en la localización de fuentes, no parece tener mucho sentido analizar los resultados obtenidos en oídos diferentes, cuando la sensación de localización es creada mediante un proceso cognitivo binaural. Debido a ello, se realiza un promediado logarítmico de los MSE obtenidos para cada uno de los oídos, intentando así aunar la respuesta de ambos y darle el sentido pretendido. Así, de esta forma, el resultado obtenido promediando las respuestas de ambos oídos se presenta en la siguiente figura.

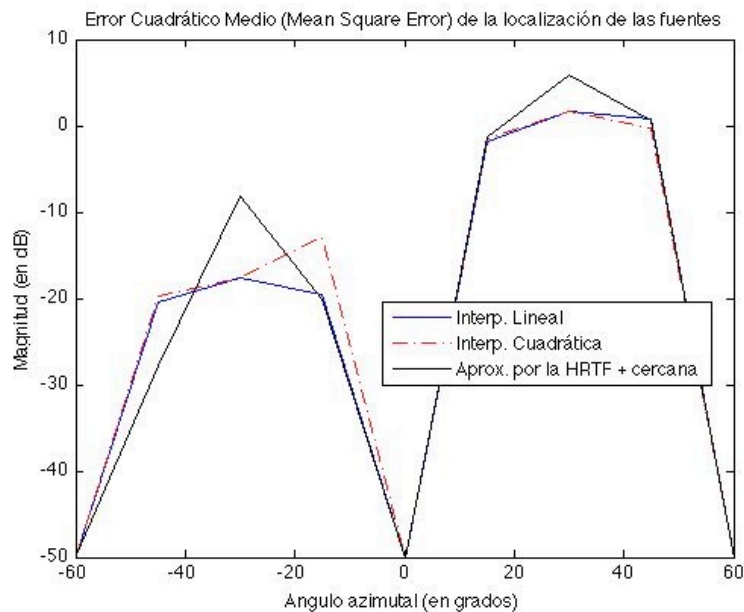


Fig.25. MSE de la localización de las fuentes virtualmente situadas.

Si observamos la distribución de ambas gráficas (fig. 23 y 24) se puede observar un efecto curioso en el que las señales que son interpoladas simulando una fuente cada  $15^\circ$  dentro del rango  $[-60^\circ, 0^\circ]$  ofrecen entre 10 y 20dB menos de error que las señales interpoladas cada  $15^\circ$  dentro del rango  $[0^\circ, +60^\circ]$ . Dicho efecto podría ser debido a las asimetrías presentes en la fisionomía del ser humano o sujeto que fue utilizado para realizar las medidas de las HRTF utilizadas en el presente proyecto, ya que como bien se dice en el método experimental, las HRTF son obtenidas mediante medidas realizadas a sujetos reales y no a maniqués contruidos para tal efecto.

Independientemente de ese efecto, se pretende analizar con esta figura, cual de los tres métodos de interpolación dibujados ofrece mejores resultados siendo el MSE del método de interpolación por acercamiento a la HRTF más cercana el pintado en negro, el MSE del método de interpolación inversamente proporcional a la distancia el pintado en azul y el MSE del método de interpolación cuadrático el pintado en línea discontinua en rojo.

Una vez sabiendo identificar cada una de las curvas, la primera conclusión clara a extraer de la figura, es que es mejor interpolar la señales de salida de las HRTFs disponibles que no hacerlo, como se puede observar claramente en el MSE obtenido para en rango  $[0^\circ, +60^\circ]$ , en el punto medio ( $-30^\circ$ ), y en el MSE obtenido para el rango  $[-60^\circ, 0^\circ]$ .

Pero por otro lado y más concretamente, se puede observar también, en este caso en todo el rango angular analizado, que la diferencia entre interpolar y no interpolar las señales en los puntos más cercanos al punto al cual se dispone de HRTF para interpolar, (es decir en los valores de  $[-45^\circ, -15^\circ, 15^\circ, 45^\circ]$  ya que se utilizan para interpolar las de  $-60^\circ, 0^\circ$  y  $60^\circ$ ) es muy pequeña. De hecho, es sólo de apenas 1-2 dB en la parte de  $0^\circ$  a  $+60^\circ$  e incluso es casi mejor no interpolar en ciertos casos, como se puede observar en la parte de  $-60^\circ$  a  $0^\circ$ .

De la misma forma, de la misma figura (fig. 25), se puede extraer una clara conclusión observando que la interpolación inversamente proporcional a la distancia (en azul), ofrece en todos los casos un error menor o prácticamente igual que la interpolación cuadrática (en rojo) desarrollada. Únicamente, se puede observar en el punto de  $+45^\circ$ , que la interpolación cuadrática ofrece un mejor resultado en la interpolación que el método lineal pero además, también se puede observar que en el ángulo  $-15^\circ$ , el efecto es inverso, y además mucho más pronunciado, y la interpolación cuadrática ofrece un peor resultado que la lineal. Así, la primera conclusión observada establece que la interpolación cuadrática ofrece similares resultados, en el mejor de los casos, que la interpolación lineal presentada.

Si observamos los valores de MSE en cada uno de los oídos por separado, por otro lado, se puede observar que en el punto de  $-45^\circ$  en el oído izquierdo en la fig. 23 y en el punto simétrico ( $+45^\circ$ ) en el oído derecho en la fig. 24 los valores de MSE correspondientes a la curva de la interpolación por la HRTF más cercana (negra) ofrecen un valor de MSE mucho menor que cualquiera de los otros métodos de interpolación. Por lo que se puede deducir que los métodos de interpolación propuestos ofrecen distintos resultados según el canal, respectivo a cada uno de los oídos, en el que se realice.

Debido a ello, se puede ser un poco más riguroso y proponer que en líneas futuras del presente trabajo se propongan algoritmos adaptativos de interpolación de las señales filtradas por HRTF, realizado el proceso independientemente para cada canal (oído) con el fin de minimizar más aún las diferencias entre las señales interpoladas y las señales obtenidas en esos mismos puntos mediante HRTFs medidas experimentalmente, mediante magnitudes que cuantifiquen el error entre las señales como el MSE.

Es necesario mencionar en este punto, que existe un método previamente desarrollado por otros científicos denominado *interpolación por splines esféricas* [25] en el que lo que hace el algoritmo es buscar entre todas las HRTFs disponibles en cada uno de los puntos donde esta información existe, para ver cual se parece en un mayor grado a la señal que se quiere generar y con ella, interpolar la localización de la fuente en ese punto. Es necesario mencionar también que este método de interpolación no hace distinciones entre cada uno de los oídos, siendo la búsqueda de la HRTF más parecida realizada para ambos oídos y no independientemente para cada uno de ellos y que además es necesario contar con un número de HRTFs tal que se pueda establecer una separación angular mayor entre ellas, para comparar las señales interpoladas con HRTFs existentes en dicho punto para hacer la selección.

#### IV.2. Comparación psicoacústica de los métodos de interpolación

Para realizar la comparación subjetiva de los diferentes métodos de interpolación desarrollados, se ha realizado la prueba del test psicoacústico a un total de 20 sujetos, indistintamente hombres y mujeres, con edades comprendidas entre los 20 y 40 años y sin problemas de audición conocidos.

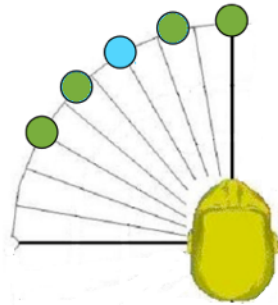


Fig.26. Detalle de los puntos donde se ha dispuesto de HRTFs y puntos donde se han interpolado las localizaciones.

Como ya se ha mencionado, el test ha sido implementado, de forma que el sujeto deba comparar subjetivamente las señales interpoladas simulando una localización de la fuente de  $-30^\circ$  en azimut y  $0^\circ$  en elevación con la misma señal filtrada por la HRTF disponible en este mismo punto. Este proceso se ha realizado con 4 pares de HRTFs que son  $0^\circ - 60^\circ$ ,  $0^\circ - 45^\circ$ ,  $15^\circ - 45^\circ$  y  $15^\circ - 60^\circ$ , de esta forma los puntos verdes en la figura 32 son las localizaciones de las HRTFs utilizadas para interpolar y los puntos azules, la localización interpolada.

Una vez definido esto, para realizar una comparación con los valores objetivos, se ha procedido al cálculo del MSE (9) de cada una de las señales interpoladas con la señal original, para así disponer de una base numérica para la comparación de los resultados. De esta forma, y siguiendo la expresión (9), se obtienen los siguientes valores de MSE recogidos en la siguiente tabla.

Señales interpoladas (HRTFs utilizadas, método interp.)	MSE (IZQ.) (dB)	MSE (DCHA.) (dB)	MSE (dB)
0 y 60 (500Hz) --> Umbrales	-4,56	-2,92	-3,67
0 y 60 (500Hz) --> Lineal	-17,24	-17,88	-17,55
0 y 60 (500Hz) --> Cuadrática	-17,24	-17,88	-17,55
0 y 60 (R.M.) --> Umbrales	2,43	9,06	6,90
0 y 60 (R.M.) --> Lineal	1,20	4,90	3,43
0 y 60 (R.M.) --> Cuadrática	1,20	4,90	3,43
0 y 60 (2kHz) --> Umbrales	3,68	6,67	5,43
0 y 60 (2kHz) --> Lineal	-3,14	3,19	1,09
0 y 60 (2kHz) --> Cuadrática	-3,14	3,19	1,09
15 y 45 (500Hz) --> Umbrales	-11,40	-9,17	-10,14
15 y 45 (500Hz) --> Lineal	-28,61	-29,31	-28,94
15 y 45 (500Hz) --> Cuadrática	-28,61	-29,31	-28,94
15 y 45 (R.M.) --> Umbrales	2,48	3,77	3,18
15 y 45 (R.M.) --> Lineal	0,92	0,16	0,56
15 y 45 (R.M.) --> Cuadrática	0,92	0,16	0,56

15 y 45 (2kHz) --> Umbrales	-5,37	0,27	-1,69
15 y 45 (2kHz) --> Lineal	-18,70	-11,09	-13,40
15 y 45 (2kHz) --> Cuadrática	-18,70	-11,09	-13,40
0 y 45 (500Hz) --> Umbrales	-13,59	-10,46	-11,75
0 y 45 (500Hz) --> Lineal	-22,67	-24,01	-23,29
0 y 45 (500Hz) --> Cuadrática	-24,30	-25,59	-24,90
0 y 45 (R.M.) --> Umbrales	2,63	0,05	1,53
0 y 45 (R.M.) --> Lineal	1,74	2,63	2,21
0 y 45 (R.M.) --> Cuadrática	1,77	2,40	2,10
0 y 45 (2kHz) --> Umbrales	-5,20	-1,69	-3,10
0 y 45 (2kHz) --> Lineal	-6,93	-3,30	-4,74
0 y 45 (2kHz) --> Cuadrática	-7,29	-3,60	-5,06
15 y 60 (500Hz) --> Umbrales	-11,40	-9,17	-10,14
15 y 60 (500Hz) --> Lineal	-23,06	-23,52	-23,28
15 y 60 (500Hz) --> Cuadrática	-21,70	-22,15	-21,92
15 y 60 (R.M.) --> Umbrales	2,57	3,73	3,19
15 y 60 (R.M.) --> Lineal	1,20	1,72	1,47
15 y 60 (R.M.) --> Cuadrática	1,26	1,83	1,56
15 y 60 (2kHz) --> Umbrales	-5,37	0,27	-1,69
15 y 60 (2kHz) --> Lineal	-13,77	-3,10	-5,75
15 y 60 (2kHz) --> Cuadrática	-13,12	-3,05	-5,66

Tabla.2. Resumen de los valores MSE en dB de cada una de las señales utilizadas en el test.

Obviamente, en los casos de interpolación en ángulos centrales (como es el caso de  $0^\circ$  y  $60^\circ$  y  $15^\circ$  y  $45^\circ$ ), el MSE no ofrece mayor información respecto a cual de los métodos es mejor interpolador, porque las señales son creadas de la misma manera y los pesos, como ya se ha visto, son iguales para ambos métodos de interpolación, aunque sí se puede observar que en todos los casos, como ya se ha mentado en el apartado anterior, cualquiera de los dos métodos de interpolación ofrece un MSE mucho menor que el ofrecido por el método de aproximación por umbrales. También se puede apreciar atendiendo a los valores de la tabla, que el error cuadrático medio es mucho mayor en los casos en los que se interpola una señal de banda ancha, debido a que al realizar la composición de la señales, se produce una pérdida de ciertas componentes frecuenciales (sobretudo a altas frecuencias) similar a la que se produce por el efecto peine.

Este efecto indeseado sobre la mezcla de las señales filtradas por las HRTFs correspondientes puede llegar a empañar los resultados del test psicoacústico que se presentan a continuación ya que a pesar de que la pregunta formulada en el test no deja lugar a dudas, el oyente puede fijarse más en que la composición espectral de las señales sea más parecida, en vez de responder a la pregunta del test de cual de las señales sugiere una localización más parecida de la fuente a la señal original.

Si se sigue observando los valores presentados en la tabla se puede observar que los valores de MSE correspondientes a las señales interpoladas linealmente y cuadráticamente no ofrecen grandes variaciones unas respecto de las otras por lo que parece ser, como ya se ha comentado en el apartado anterior que ambos métodos ofrecen resultados similares entorno a la interpolación de las señales de salida de las HRTF.

Además, y como se puede observar marcado en amarillo en la tabla, se da un único caso en el que el método de aproximación por umbrales ofrece un valor menor de MSE que los otros dos métodos de interpolación, lo que se ajusta a lo comentado anteriormente ya que se produce en la interpolación de señales de banda ancha como es el ruido modulado.

Una vez establecidos los valores de MSE correspondientes a las señales interpoladas utilizadas para el test, se procede a la presentación de los resultados que el test ha ofrecido tras ser realizado al conjunto de 20 sujetos que se ha presentado previamente. Todos los sujetos tras la realización del test concluyen que resulta difícil en muchos casos decidir entre los pares de señales preguntados. Los resultados serán analizados de tres formas distintas:

1. Mediante un análisis global de los resultados. Sin hacer distinciones, ni del tipo de señal ni de las parejas de HRTFs utilizadas para obtener las señales interpoladas, es decir, tomándose todas las pruebas realizadas para el cómputo.
2. Mediante el análisis dependiente de la localización de las HRTFs disponibles para la realización de la interpolación. Es decir, agrupando las señales que comparten los pares de HRTFs utilizados para realizar la interpolación.
3. Mediante el análisis dependiente de los resultados obtenidos para las diferentes señales que se han utilizado en la realización del test.

Observando la tabla 3 correspondiente a los datos globales del test, es decir del primero de los métodos propuestos, se puede observar claramente, que la gran mayoría (85%) de los usuarios sometidos al test prefieren la recreación de la localización de la señal mediante alguno de los dos métodos de interpolación presentados, que la presentación de la señal filtrada por la HRTF más cercana al punto demandado por el usuario. Además atendiendo a la misma tabla, se puede observar la consistencia con la que se ha construido el test, ya que para ambos casos de interpolación presentados se tiene que un 85% de los sujetos prefieren la señal presentada por alguno de los métodos de interpolación que la emisión de la HRTF más cercana, pero en la pregunta de si prefieren más el método lineal o el cuadrático, la respuesta no está tan clara quedando casi en la indeterminación por parte de los sujetos examinados, ya que los porcentajes son de 55.4% y 44.6% respectivamente.

Ahora, si atendemos a la tendencia que puedan demostrar las respuesta de los sujetos de manera global mediante la Chi-Cuadrado, se reconoce lo que se ha mencionado antes acerca de que todos los usuarios prefieren cualquiera de los dos métodos de interpolación propuestos frente a la aproximación de la HRTF más cercana, pero no hay una tendencia definida que diga que la probabilidad de que un 95% de la población se decida por una u otra.

De esta manera, queda reflejado también que los resultados numéricos obtenidos mediante el métodos de Mínimos Cuadrados (MSE) ya que en su análisis se ha demostrado que entre las señales interpoladas la diferencia del MSE era muy parecida ya que había diferencias de apenas 3dB entre ellas, mientras que entre cualquiera de las señales interpoladas y las aproximadas por las HRTFs más cercanas ofrecen entre sí diferencias de alrededor de 15dB.

	Lineal vs. Cuadrática	Lineal vs. Cuadrática	Lineal vs. Umbrales	Lineal vs. Umbrales	Cuadrática vs. Umbrales	Cuadrática vs. Umbrales
<i>Método</i>	<i>Lineal</i>	<i>Cuadrática</i>	<i>Lineal</i>	<i>Umbral</i>	<i>Cuadrática</i>	<i>Umbral</i>
0° y 60° (500Hz)	9	11	18	2	18	2
0° y 60° (RM)	10	10	20	0	19	1
0° y 60° (2kHz)	9	11	20	0	19	1
15° y 45° (500Hz)	12	8	16	4	15	5
15° y 45° (RM)	9	11	17	3	16	4
15° y 45° (2kHz)	13	7	17	3	17	3
15° y 60° (500Hz)	13	7	19	1	18	2
15° y 60° (RM)	9	11	17	3	17	3
15° y 60° (2kHz)	12	8	16	4	18	2
0 y 45 (500Hz)	12	8	18	2	18	2
0 y 45 (RM)	9	11	14	6	14	6
0 y 45 (2000Hz)	16	4	13	7	17	3
<b>TOTAL (sujetos)</b>	<b>133</b>	<b>107</b>	<b>205</b>	<b>35</b>	<b>206</b>	<b>34</b>
<b>Porcentaje (%)</b>	<b>55,4</b>	<b>44,6</b>	<b>85,4</b>	<b>14,6</b>	<b>85,8</b>	<b>14,2</b>
<b>JiCuadrado</b>	<b>2,82</b>		<b>120,42</b>		<b>123,27</b>	
		<b>IND.</b>		<b>Lineal (95%)</b>		<b>Cuadrática (95%)</b>

Tabla.3. Resumen de las respuesta global del test de los sujetos.

Si optamos, en cambio, por analizar los resultados de los sujetos en función de la localización (o los grados) de los pares de HRTFs utilizados para la creación de las señales interpoladas, se procederá primero al análisis de los resultados por separado de los casos ‘centrados’, es decir aquellos en los que las señales han sido obtenidas por pares de HRTFs de 15° - 45° y 0° - 60°, y tras ello, en los en los que las señales son obtenidas con los pares de HRTF de 15° - 60° y 0° y 45°.

Así los casos ‘centrados’ mentados en el párrafo anterior se resumen en la tabla 4, en las que los resultados son similares a los ya descritos en el análisis global de los resultados realizado. Por otro lado, se observa que, como puede resultar natural, los porcentajes disminuyen cuando se utiliza el par de HRTF más cercanas a la localización destino (30°) que es el par 15° - 45°. Estos porcentajes, siguen definiendo una tendencia clara por parte de los sujetos examinados a decantarse por cualquiera de los métodos de interpolación frente a la aproximación a la HRTF más cercana como sería la de 15° en este caso, pero también define un comportamiento que supondría que si se pudiese utilizar una biblioteca de HRTFs definida en una malla tal que el sujeto no pudiese tener constancia del cambio de localización, menor al *jnd* (*just noticeable difference*) de la localización espacial de una fuente, la técnica de aproximación por la HRTF más cercana al ángulo demandado por el usuario sería efectiva.

	Lineal vs. Cuadrática	Lineal vs. Cuadrática	Lineal vs. Umbrales	Lineal vs. Umbrales	Cuadrática vs. Umbrales	Cuadrática vs. Umbrales
<i>Método</i>	<i>Lineal</i>	<i>Cuadrática</i>	<i>Lineal</i>	<i>Umbral</i>	<i>Cuadrática</i>	<i>Umbral</i>
15° y 45° (500Hz)	12	8	16	4	15	5
15° y 45° (RM)	9	11	17	3	16	4
15° y 45° (2kHz)	13	7	17	3	17	3
<b>TOTAL (sujetos)</b>	<b>34</b>	<b>26</b>	<b>50</b>	<b>10</b>	<b>48</b>	<b>12</b>
<b>Porcentaje (%)</b>	<b>56,7</b>	<b>43,3</b>	<b>83,3</b>	<b>16,7</b>	<b>80,0</b>	<b>20,0</b>
<b>JiCuadrado</b>	<b>1,07</b>		<b>26,67</b>		<b>21,6</b>	
		<b>IND.</b>		<b>Lineal (95%)</b>		<b>Cuadrática (95%)</b>

	Lineal vs. Cuadrática	Lineal vs. Cuadrática	Lineal vs. Umbrales	Lineal vs. Umbrales	Cuadrática vs. Umbrales	Cuadrática vs. Umbrales
<i>Método</i>	<i>Lineal</i>	<i>Cuadrática</i>	<i>Lineal</i>	<i>Umbral</i>	<i>Cuadrática</i>	<i>Umbral</i>
0° y 60° (500Hz)	9	11	18	2	18	2
0° y 60° (RM)	10	10	20	0	19	1
0° y 60° (2kHz)	9	11	20	0	19	1
<b>TOTAL (sujetos)</b>	<b>28</b>	<b>32</b>	<b>58</b>	<b>2</b>	<b>56</b>	<b>4</b>
<b>Porcentaje (%)</b>	<b>46,7</b>	<b>53,3</b>	<b>96,7</b>	<b>3,3</b>	<b>93,3</b>	<b>6,7</b>
<b>JiCuadrado</b>	<b>0,27</b>		<b>52,27</b>		<b>45,07</b>	
	<b>IND.</b>		<b>Lineal (95%)</b>		<b>Cuadrática (95%)</b>	

Tabla.4. Resumen de las respuestas de los sujetos en función de la localización de las HRTF ‘centradas’ utilizadas.

De la misma forma, analizando los casos ‘no-centrados’ se observa también una disminución del porcentaje de sujetos que prefieren las señales interpoladas frente a las aproximadas a las HRTF más cercanas y sigue en este caso también la tendencia definida por el método ‘Chi-Cuadrado’. Además, también como ya se ha dicho, aunque el porcentaje de sujetos que prefiere el método de interpolación lineal frente al cuadrático es algo mayor (61,7% frente a 38,3%), no supone una tendencia clara por parte de la mayoría de la población, por lo que no cumple ambos criterios de forma concluyente.

	Lineal vs. Cuadrática	Lineal vs. Cuadrática	Lineal vs. Umbrales	Lineal vs. Umbrales	Cuadrática vs. Umbrales	Cuadrática vs. Umbrales
<i>Método</i>	<i>Lineal</i>	<i>Cuadrática</i>	<i>Lineal</i>	<i>Umbral</i>	<i>Cuadrática</i>	<i>Umbral</i>
15° y 60° (500Hz)	13	7	19	1	18	2
15° y 60° (RM)	9	11	17	3	17	3
15° y 60° (2kHz)	12	8	16	4	18	2
<b>TOTAL (sujetos)</b>	<b>34</b>	<b>26</b>	<b>52</b>	<b>8</b>	<b>53</b>	<b>7</b>
<b>Porcentaje (%)</b>	<b>56,7</b>	<b>43,3</b>	<b>86,7</b>	<b>13,3</b>	<b>88,3</b>	<b>11,7</b>
<b>JiCuadrado</b>	<b>1,07</b>		<b>32,27</b>		<b>35,27</b>	
	<b>IND.</b>		<b>Lineal (95%)</b>		<b>Cuadrática (95%)</b>	

	Lineal vs. Cuadrática	Lineal vs. Cuadrática	Lineal vs. Umbrales	Lineal vs. Umbrales	Cuadrática vs. Umbrales	Cuadrática vs. Umbrales
<i>Método</i>	<i>Lineal</i>	<i>Cuadrática</i>	<i>Lineal</i>	<i>Umbral</i>	<i>Cuadrática</i>	<i>Umbral</i>
15° y 45° (500Hz)	12	8	18	2	18	2
15° y 45° (RM)	9	11	14	6	14	6
15° y 45° (2kHz)	16	4	13	7	17	3
<b>TOTAL (sujetos)</b>	<b>37</b>	<b>23</b>	<b>45</b>	<b>15</b>	<b>49</b>	<b>11</b>
<b>Porcentaje (%)</b>	<b>61,7</b>	<b>38,3</b>	<b>75,0</b>	<b>25,0</b>	<b>81,7</b>	<b>18,3</b>
<b>JiCuadrado</b>	<b>3,27</b>		<b>15,00</b>		<b>24,07</b>	
	<b>IND.</b>		<b>Lineal (95%)</b>		<b>Cuadrática (95%)</b>	

Tabla.5. Resumen de las respuestas de los sujetos en función de la localización de las HRTF ‘no-centradas’ utilizadas.

De esta forma, como se puede observar que el análisis de los resultados del test en función de la localización de las HRTF utilizadas para la interpolación no ofrecen ninguna conclusión desconocida se procede al análisis de las respuestas en función de la señal de test utilizada, haciendo una comparación de los resultados obtenidos con el tono puro de 500Hz, con el ruido de banda ancha filtrado paso bajo a 1.5kHz (cumpliendo el criterio de *aliasing espacial*) y un tono puro de 2kHz fuera del criterio de Nyquist.

Como los resultados no ofrecen ninguna conclusión que se deba añadir respecto a las ya mencionadas se omite la presentación de las tablas de resultados.

De esta forma, solo queda el análisis a realizar de aquellas ‘preguntas trampa’ que se han introducido en la construcción del test. Como recuerdo cabe decir que las preguntas trampa proceden de preguntas en el test en el que el sujeto era sometido a la comparación de un par de señales en las que una de las señales era la señal original a comparar. Obviamente, el resultado debería ser que todos los sujetos debieran reconocer la señal original de las señales interpoladas o aproximadas por las HRTF más cercanas.

Como no se puede disponer del mismo número de resultados, ni en las mismas condiciones que para las preguntas de interés se decide hacer un examen del conjunto de respuestas ofrecidas por las preguntas trampa sin hacer distinciones entre los datos en función de la localización de las HRTF utilizadas para interpolar ni el tipo de señal de test utilizada. Así, se presentan los resultados en la tabla siguiente.

	<b>Cuerda vs Real</b>	<b>Cuerda vs Real</b>	<b>Cuadrática vs Real</b>	<b>Cuadrática vs Real</b>	<b>Umbral vs Real</b>	<b>Umbral vs Real</b>
<b>Método</b>	<b>Real</b>	<b>Lineal</b>	<b>Real</b>	<b>Cuadrática</b>	<b>Real</b>	<b>Umbral</b>
0° y 60° (500Hz)	18	2			20	0
0° y 60° (RM)						
0° y 60° (2kHz)			16	4		
15° y 45° (500Hz)	16	4	20	0		
15° y 45° (RM)					16	4
15° y 45° (2kHz)					19	1
15° y 60° (500Hz)			17	3		
15° y 60° (RM)						
15° y 60° (2kHz)	17	3				
0 y 45 (500Hz)			18	2	19	1
0 y 45 (RM)						
0 y 45 (2000Hz)	10	10				
<b>TOTAL (sujetos)</b>	<b>61</b>	<b>19</b>	<b>71</b>	<b>9</b>	<b>74</b>	<b>6</b>
<b>Porcentaje (%)</b>	<b>76,3</b>	<b>23,8</b>	<b>88,8</b>	<b>11,3</b>	<b>92,5</b>	<b>7,5</b>
<b>JiCuadrado</b>	<b>22,05</b>		<b>48,05</b>		<b>57,80</b>	
	<b>Lineal (95%)</b>		<b>Lineal (95%)</b>		<b>Cuadrática (95%)</b>	

Tabla.6. Resumen de las respuestas de los sujetos en las ‘preguntas trampa’ del test.

Como se puede observar en la tabla, existe un porcentaje de sujetos, en el peor de los casos de 24%, en el que los sujetos afirman que la señal localizada mediante cualquiera de los métodos de interpolación o por aproximación a la HRTF más cercana, ofrecen una mejor localización de la fuente que la señal filtrada por la HRTF disponible en el punto interpolado de -30°. Además, a pesar de que la tendencia por parte de los sujetos a identificar la señal original como la respuesta correcta está bien definida, se puede reconocer que esta tendencia está mucho más definida para los casos de pregunta trampa realizada mediante las señales interpoladas por el método de interpolación cuadrática o por la aproximación de la señal a la HRTF más cercana, siendo alrededor de el 90% de los sujetos encuestados los que prefieren la señal original.

Si como se ha realizado anteriormente, optamos por el análisis de los resultados en función de la localización de las HRTF utilizadas para generar las señales del test (tabla 8), se puede observar que la tendencia se sigue manteniendo excepto en el caso en el que se utilizan los pares de señales de 0° - 45°, en el que se observa que en el caso de la señal de 2kHz, los sujetos no prefieren una ni otra, si no que no hay determinación por ninguna de las dos, quedando los porcentajes exactamente



al 50% de los sujetos que prefieren la señal interpolada de forma lineal frente a la real (o original), en su comparación con la misma señal original.

Si atendemos ahora, en busca del dato numérico en el cálculo del MSE que permita justificar este hecho en la tabla 8, no se observa que el error de la señal interpolada linealmente tenga un valor especialmente pequeño para que genere esa indeterminación en las respuestas entre la señal original y la señal interpolada, pero lo que sí es cierto es que es la prueba de un tono puro de 2kHz, una señal que no es fácilmente localizable en el espacio, primero por tratarse de un tono puro y segundo por tratarse de un tono puro de una frecuencia cuya longitud de onda es de 17cm, una prueba complicada para cualquier oído humano sin previo entrenamiento en la materia de localización de fuentes.

Este hecho no supone un cambio de tendencia respecto a lo ya comentado ya que se habla de un caso muy particular frente al extenso grupo de datos analizados, debido a ello se supone este hecho, como un caso aislado que cuya tendencia no tiene porque ser mantenida al aumentar el número de sujetos sometidos a la prueba.

También resulta necesario mencionar por otro lado, que los mayores porcentajes de sujetos que prefieren la señal original frente a las señales interpoladas por cualquiera de los métodos se da en los casos en los que la comparación se realiza sobre las señales de Ruido Modulado. Este efecto es debido claramente al hecho diferenciador que contienen las señales en el aspecto en que cuando realizamos la mezcla de dos señales para la interpolación de una señal determinada, el proceso introduce un filtrado en frecuencia debido a la suma destructiva y constructiva de determinadas componentes frecuenciales contenidas en el espectro, debido a ello, es mucho más fácil para los sujetos identificar cual de las dos señales presentadas se trata de la original.

	Cuerda vs Real	Cuerda vs Real	Cuadrática vs Real	Cuadrática vs Real	Umbral vs Real	Umbral vs Real
<i>Método</i>	Real	Lineal	Real	Cuadrática	Real	Umbral
0° y 60° (500Hz)	18	2				
0° y 60° (RM)					20	0
0° y 60° (2kHz)			16	4		
<b>TOTAL (sujetos)</b>	<b>18</b>	<b>2</b>	<b>16</b>	<b>4</b>	<b>20</b>	<b>0</b>
<b>Porcentaje (%)</b>	<b>90,0</b>	<b>10,0</b>	<b>80,0</b>	<b>20,0</b>	<b>100,0</b>	<b>0,0</b>
<b>JiCuadrado</b>	<b>12,80</b>		<b>7,20</b>		<b>20,00</b>	
	<b>Real (95%)</b>		<b>Real (95%)</b>		<b>Real (95%)</b>	
	Cuerda vs Real	Cuerda vs Real	Cuadrática vs Real	Cuadrática vs Real	Umbral vs Real	Umbral vs Real
<i>Método</i>	Real	Lineal	Real	Cuadrática	Real	Umbral
15° y 45° (500Hz)	16	4				
15° y 45° (RM)			20	0		
15° y 45° (2kHz)					16	4
<b>TOTAL (sujetos)</b>	<b>16</b>	<b>4</b>	<b>20</b>	<b>0</b>	<b>16</b>	<b>4</b>
<b>Porcentaje (%)</b>	<b>80,0</b>	<b>20,0</b>	<b>100,0</b>	<b>0,0</b>	<b>80,0</b>	<b>20,0</b>
<b>JiCuadrado</b>	<b>7,20</b>		<b>20,00</b>		<b>7,20</b>	
	<b>Real (95%)</b>		<b>Real (95%)</b>		<b>Real (95%)</b>	

Tabla.7. Resumen de las respuestas de los sujetos a la ‘pregunta trampa’ en función de la localización de las HRTF ‘centradas’ utilizadas.

	Cuerda vs Real	Cuerda vs Real	Cuadrática vs Real	Cuadrática vs Real	Umbral vs Real	Umbral vs Real
<i>Método</i>	Real	Lineal	Real	Cuadrática	Real	Umbral
15° y 60° (500Hz)			17	3	19	1
15° y 60° (RM)						
15° y 60° (2kHz)	17	3				
<b>TOTAL (sujetos)</b>	<b>17</b>	<b>3</b>	<b>17</b>	<b>3</b>	<b>19</b>	<b>1</b>
<b>Porcentaje (%)</b>	<b>85,0</b>	<b>15,0</b>	<b>85,0</b>	<b>15,0</b>	<b>95,0</b>	<b>5,0</b>
<b>JiCuadrado</b>	<b>9,80</b>		<b>9,80</b>		<b>16,20</b>	
	<b>Real (95%)</b>		<b>Real (95%)</b>		<b>Real (95%)</b>	

	Cuerda vs Real	Cuerda vs Real	Cuadrática vs Real	Cuadrática vs Real	Umbral vs Real	Umbral vs Real
<i>Método</i>	Real	Lineal	Real	Cuadrática	Real	Umbral
0° y 45° (500Hz)			18	2	19	1
0° y 45° (RM)						
0° y 45° (2kHz)	10	10				
<b>TOTAL (sujetos)</b>	<b>10</b>	<b>10</b>	<b>18</b>	<b>2</b>	<b>19</b>	<b>1</b>
<b>Porcentaje (%)</b>	<b>50,0</b>	<b>50,0</b>	<b>90,0</b>	<b>10,0</b>	<b>95,0</b>	<b>5,0</b>
<b>JiCuadrado</b>	<b>0,00</b>		<b>12,80</b>		<b>16,20</b>	
	<b>IND.</b>		<b>Real (95%)</b>		<b>Real (95%)</b>	

Tabla.8. Resumen de las respuestas de los sujetos a la ‘pregunta trampa’ en función de la localización de las HRTF ‘no-centradas’ utilizadas.

## 5. Conclusiones y líneas futuras

A continuación se procede al resumen de todas las conclusiones extraídas del presente trabajo que se han ido redactando a lo largo de esta memoria:

1. Se comprueba la **necesidad de cumplir el criterio de Nyquist** en pos de evitar el *aliasing espacial*, ya que este criterio asegura que en cualquier punto de un círculo el campo sonoro generado físicamente (el sonido presentado) será exactamente igual al original generado por unas fuentes determinadas en esa localización y para ese punto de escucha.
2. Se comprueba tanto numéricamente como subjetivamente que en cualquier caso, centrado o no centrado, es **preferida** por la mayoría de los oyentes **la generación de la localización de forma virtual utilizando cualquiera de los dos métodos de interpolación presentados**, en vez de la aproximación por la HRTF más cercana disponible. Es decir, se comprueba que es necesario generar un método de interpolación para generar en el sujeto la sensación de localización de fuente de forma óptima en cualquier punto de todo el círculo.
3. Debido a que los resultados obtenidos entre los MSE obtenidos para las señales interpoladas linealmente y los obtenidos para las señales interpoladas cuadráticamente son muy parecidos, y a que no se ha concluido subjetivamente si uno de los dos es preferido por la mayoría de la población, **se concluye que es suficiente la interpolación lineal** y no es necesario aumentar la demanda de cálculo de la interpolación cuadrática porque sus resultados no lo merecen.
4. Los algoritmos de interpolación presentados son una forma sencilla y relativamente eficiente, teniendo en cuenta el *aliasing espacial*, de obtener sonidos localizados en cualquier punto del espacio tridimensional. Está claro que debido al efecto del *aliasing espacial* **dicha eficiencia**

**depende de la definición angular de la biblioteca de HRTFs** hasta tal punto que si fuese una biblioteca muy extensa, el método de aproximación a la HRTF más cercana sería también válido.

5. Se comprueba que con la utilidad *Playrec* es posible el filtrado por una determinada HRTF de una señal captada por un micrófono y su reproducción en tiempo real, pero no se ha comprobado la interacción dinámica del filtrado con la interfaz gráfica generada.

Como apartado final al presente trabajo se enumeran a continuación las vías que se podrían seguir para continuar profundizando en el trabajo presentado:

1. Debido a la limitación de tiempo, no se ha podido implementar dicha aplicación en tiempo real, en su totalidad. Un reto dirigido al conocimiento de la herramienta GUIDE de MatLab sería la generación de una interfaz que interactuase a tiempo real con el usuario de la aplicación. Además de sobre Matlab, que no resulta ser el entorno más apropiado para este tipo de aplicaciones, se podría pensar en su implementación por ejemplo en C++ sobre una DSP.
2. Se plantea la posibilidad de realizar un test psicoacústico más exhaustivo, no solo compuesto por pruebas con ruidos abstractos, si no realizado con sonidos naturales que el humano está acostumbrado a utilizar, y jugando con sonidos con distinta carga espectral ya que es bien conocido que la localización de los sonidos depende de la frecuencia de los mismos. Así se obtendría resultados subjetivos más concretos y de mayor rigor científico.
3. Debido a que cada una de las HRTFs proveniente de un sujeto determinado están ajustadas a sus características físicas, estas HRTF pueden tender a no crear la sensación de localización correctamente en determinado sujetos con determinadas propiedades físicas. Otra de las líneas de futuro sería la de intentar hacer la aplicación proponiendo un conjunto de HRTFs provenientes de sujetos con distintas características antropométricas, para que usuario pudiese elegir la que mejor se ajustase a sus características físicas. Es decir, la idea respondería a la necesidad de una estandarización de los rasgos físicos de los sujetos utilizados para la obtención del conjunto de HRTFs.

## AGRADECIMIENTOS

Me gustaría agradecer especialmente el apoyo moral y académico ofrecido por mi director de Tesina Miguel Ferrer Contreras y agradecer enormemente también a tod@s l@s compañer@s del Master, la realización del test psicoacústico implementado. Además, dar las gracias a compañeros del Master que me han ayudado de una u otra forma en la realización de esta Tesina, ellos saben quienes son.

**BIBLIOGRAFÍA**

- [1] Recuero, M. *Ingeniería Acústica*. Capítulo 11. Editorial Paraninfo: 1994.
- [2] Rayleigh, L. *On our perception of sound direction*. Philosophical Magazine 13: 1907.
- [3] Blauert, J. *Spatial Hearing*. The MIT Press, Cambridge: 1997. Revised Edition.
- [4] Moore, B.J.C. *An introduction to the psychology of hearing*. Academic Press, London: 1989.
- [5] Batteau, D. *The role of pinna in human localization*. Proc. R. Soc. London Bi Biol. pp. 158-180: 1967.
- [6] Handel, S. *Listening*. The MIT Press, Cambridge: 1989.
- [7] Litovsky, R.Y., Colburn, S.H., Yost, W.A., Guzman, S.J. *The precedence effect*. J.A.S.A., 106. 1633-1654, 1999.
- [8] Zurek, P.M. *The Precedence Effect and its possible role in the avoidance of interaural ambiguities*. J.A.S.A., 67. 952-964, 1980.
- [9] Freyman, R. L., Clifton, R. K., Litovsky, R. Y. *Dynamic processes in the precedence effect*. J.A.S.A., 90. 874-884, 1991.
- [10] Shaw, E.A.G. *The External Ear: Handbook of Sensory Physiology VI/1: Auditory System, Anatomy Physiology*. Springer-Verlag, New York: 1974.
- [11] Wightman, Frederic L. and Kistler, Doris J. *Headphone simulation of free-field listening. I: Stimulus synthesis*. J.A.S.A., 85(2): February 1989.
- [12] Kahana, Yuvi et al. *Numerical Modelling of the Transfer Functions of a Dummy-Head and of the External Ear*. 16th Audio Engineering Society (AES) International Conference on Spatial Sound Reproduction, Rovaniemi, Finland: 1999.
- [13] Oppenheim, A. V., Schaffer, R. W. *Discrete-Time Signal Processing*. Prentice Hall, Englewood Cliffs, New Jersey: 1989.
- [14] Cheng, C. I., Wakefield, G. H. *Introduction to Head-Related Transfer Functions (invited tutorial)*. University of Michigan, Department of Electrical Engineering and Computer Science. U.S.A.
- [15] Fontana, S., Farina, A., Grenier, Y. *Binaural for popular music: Case of study*. 13th International Conference on Auditory Display, Montréal, Canada, June 26-29, 2007.
- [16] Zhong, X. L., Xie, B. S. *Progress on the research of Head-Related Transfer Function*. Audio Eng, 12: 44-46, 2004.
- [17] Gardner, B., Martin, K. *HRTF Measurements of a KEMAR Dummy-Head Microphone*. MIT Media Lab. Perceptual Computing – Technical Report 280, May, 1994.
- [18] Algazi, V. R., Duda, R. O., Thompson, D. M. *The Cipic HRTF Database*. New Platz, New York, October 2001.
- [19] Warusfel O. *Listen HRTF Database*. Room Acoustics Team, IRCAM Institute, Paris. 2003.
- [20] Riederer K. *Repeatability analysis of head-related transfer function measurements*. 105th Convention of Audio Engineering Society, preprint paper 4846. 1998.
- [21] Hammershoi D., Moller H. *Sound transmission to and within the human ear canal*. J.A.S.A., 100. 408-427. 1996.
- [22] Humphrey R. A., *Automatic loudspeaker location detection for use in Ambisonic systems*. MEng Thesis in Electronic Engineering, University of York and KTH. 2006.
- [23] Denon Records PG-6006. *Anechoic orchestral music recording*. Audio CD, 1988.
- [24] Ajdler T., Faller C., Sbaiz L., M. Vetterli. *Interpolation of Head-Related Transfer Functions Considering Acoustics*. 118th Convention of the Audio Engineering Society, J.A.E.S. vol. 53, p. 662. 2005.
- [25] Hartung K., Braasch J., Sterbing S. J. *Comparison of different methods for the interpolation of head-related transfer functions*. Conference of the Audio Engineering Society, pp. 319-329. 1999.
- [26] J.S. Bradley, R. Reich and S.G. Norcross, *A just noticeable difference in C50 for speech*. Applied Acoustics 58, 99-108. 1999.