

UNIVERSIDAD POLITÉCNICA DE VALENCIA

ESCUELA POLITÉCNICA SUPERIOR DE GANDÍA

MÁSTER EN POSTPRODUCCIÓN DIGITAL

---



UNIVERSIDAD  
POLITECNICA  
DE VALENCIA



ESCUELA POLITÉCNICA  
SUPERIOR DE GANDÍA

**“Estudio teórico-práctico de las  
posibilidades y limitaciones de las  
técnicas binaurales en la postproducción  
cinematográfica”**

**TRABAJO FINAL DE MÁSTER**

*Autor: Bartosz Roszko*

*Director: Blas Payri*

*Gandía, julio de 2013*

## RESUMEN

El presente trabajo de fin de master se enfoca en las posibilidades de aplicación del sonido binaural en las películas como método alternativo del sonido envolvente. Se estudia el proceso de la sonorización del material de vídeo con la técnica binaural, desde las grabaciones hasta la postproducción. Se presentan las técnicas disponibles del procesado del audio, usando tanto el sonido del plano capturado con los micrófonos binaurales OKM Soundman como el registrado luego en el estudio de grabación. Se muestran los problemas que pueden ocurrir durante el proceso de grabación (posición incorrecta de la persona que lleva los micrófonos binaurales, su respiración) y postproducción (corrección del sonido que suena demasiado cerca del oído con el uso de procesamiento M/S, externalización a través de añadir los primeros rebotes, corrección de los niveles de diferentes sonidos grabados a la vez con el uso del compresor, reducción de la reverberación con expander). Las soluciones propuestas son un resultado del conocimiento de la localización de sonido y además de los experimentos realizados por el autor. Cada una de las cinco escenas grabadas con la cámara se muestra en cinco versiones que presentan métodos alternativos del procesamiento de sonido (la grabación bruta de los micrófonos OKM, esa misma grabación procesada en Logic Pro, el sonido grabado en el estudio de postproducción y procesado con Binaural Panner de Logic Pro, Panorama 5 de Wave Arts o mezclado al estéreo). Después de la finalización del tratamiento, todas las escenas son evaluadas cuantitativamente por los oyentes y además cualitativamente por los ingenieros de sonido. Las conclusiones principales son que la mayoría de la gente se concentra solamente en el movimiento derecha-izquierda y no delante-detrás. Se observa la tendencia de indicar las grabaciones de OKM y Panorama 5 como preferidos, aunque en algunos casos los oyentes eligen el estéreo.

## **ABSTRACT**

The following master's thesis focuses on the possibilities of application of binaural sound in the movies as an alternative method of surround sound. The process of sound design using binaural technique for video is studied in this thesis, in the range from the recording to the postproduction. The available techniques of sound processing are presented using the sound captured with binaural microphones as well as the sound registered later in the recording studio. The problems which might occur during the recording (incorrect position of the person who has the binaural microphones put on, his/her breath present on the recording) and the postproduction (correction of the sound which sounds too close to the ear using M/S processing, externalization by adding early reflections, correction of the levels of different sounds recorded at the same time with the use of a compressor, reverb reduction with an expander). The proposed solutions are the result of both applying the knowledge about sound localization and conducting experiments by the author. Each of the five scenes recorded with a camera are presented in five versions reflecting different types of processing (raw recording from the OKM microphones, the same one, but processed in Logic Pro, the sound registered in the postproduction studio and then processed using Binaural Panner available in Logic Pro, Panorama 5 by Wave Arts or mixed to stereo). After the processing all the scenes are evaluated quantitatively by the audience, and also in terms of quality by sound engineers. The main conclusions show that most people focus only on the right-left movements and not the front-back ones. Moreover, there's a tendency to point out the OKM and Panorama 5 recordings as the preferred ones, although in some cases the audience favors the stereo version.

# ÍNDICE

RESUMEN.....	2
ABSTRACT .....	3
1 INTRODUCCIÓN AL SONIDO BINAURAL .....	6
1.1 Definición.....	6
1.2 Estado del arte.....	6
1.3 Ventajas y deventajas del sonido binaural.....	7
2 ELECCIÓN DEL TEMA Y LOS OBJETIVOS DEL TRABAJO .....	9
3 MECANISMOS DE LOCALIZACIÓN DE SONIDO POR EL HOMBRE .....	9
3.1 ITD.....	9
3.2 ILD.....	10
3.3 HRTF.....	10
3.4 Otros mecanismos de localización.....	11
4 TECNICAS DE GRABACIONES BINAURALES.....	13
4.1 Micrófonos binaurales.....	13
4.2 Dummy head.....	13
5 IMPORTANCIA DE LOS AURICULARES EN LA ESCUCHA BINAURAL .....	14
6 REALIZACIÓN DE GRABACIONES .....	15
7 USO DE HERRAMIENTAS SOFTWARE .....	19
7.1 Especificación de las herramientas usadas.....	19
7.2 Preparación de trabajo.....	19
7.3 Mezcla estéreo.....	20
7.4 Tratamiento de las grabaciones binaurales.....	21

7.4.1	Los obstáculos encontrados. ....	21
7.4.2	Mejora de calidad de las grabaciones. ....	23
7.5	Mezcla con el uso de la herramienta Binaural Panner. ....	27
7.6	El uso de la herramienta Panorama 5.....	30
8	EVALUACIÓN PERCEPTIVA.....	32
8.1	Análisis de datos cuantitativos. ....	32
8.1.1	Congruencia.....	33
8.1.2	Calidad.....	35
8.2	Análisis cualitativo. ....	37
9	CONCLUSIONES .....	39
10	BIBLIOGRAFÍA .....	40

# **1 INTRODUCCIÓN AL SONIDO BINAURAL**

## **1.1 Definición.**

La fuente del sonido puede ser un simple sistema de altavoces, mono, estéreo, un sistema multicanal (cuadrofonía, 5.1, 6.1, 7.1, etc.) y también auriculares. Respecto a la manera de recepción se distingue la escucha monoaural y binaural. En el presente trabajo, por *sonido binaural*, se entenderá el sonido escuchado a través de auriculares, recibido con ambos oídos, grabado o procesado con la intención de reproducir fielmente el espacio acústico y la localización de las fuentes.

## **1.2 Estado del arte.**

A pesar de muchas imperfecciones del sonido binaural (sobre todo en la correcta localización de la fuente sonora) éste es el punto de interés de numerosos centros de ciencia y de especialistas del campo audiovisual. Esa posición se debe a las posibilidades potenciales que en los últimos años nos han abierto los propios avances de la tecnología multimedia. El objetivo de las grabaciones binaurales, es proporcionar a los oyentes la impresión de encontrarse en un espacio acústico particular (por ejemplo una sala de concierto, un plano de una película) de la manera más realista posible, incluso, más que con el típico sistema de estéreo o 5.1.

La primera transmisión con un prototipo de sonido binaural fue realizada en el año 1881 de Opera Garnier a través de un sistema llamado Théâtrophone, pero un desarrollo más profundo dio inicio hace cerca de medio siglo. Gracias a Internet y a que los auriculares empezaron a ser fácilmente accesibles las grabaciones se hicieron más populares. Se comenzaron a observar tendencias de construir prototipos propios de micrófonos y de publicar el material registrado en la red global. Aunque la mayoría de las grabaciones que se pueden encontrar consisten en solo audio, existen también una gran cantidad de vídeos con sonido binaural. Sin embargo, hay que tener en cuenta que las grabaciones presentes tienen más frecuentemente el carácter experimental.

Uno de los usos profesionales del sonido binaural en el arte es el disco *Street Hassle* de Lou Reed del año 1978 que se considera como el primer álbum pop grabado en técnica binaural (Nusser, 1978).

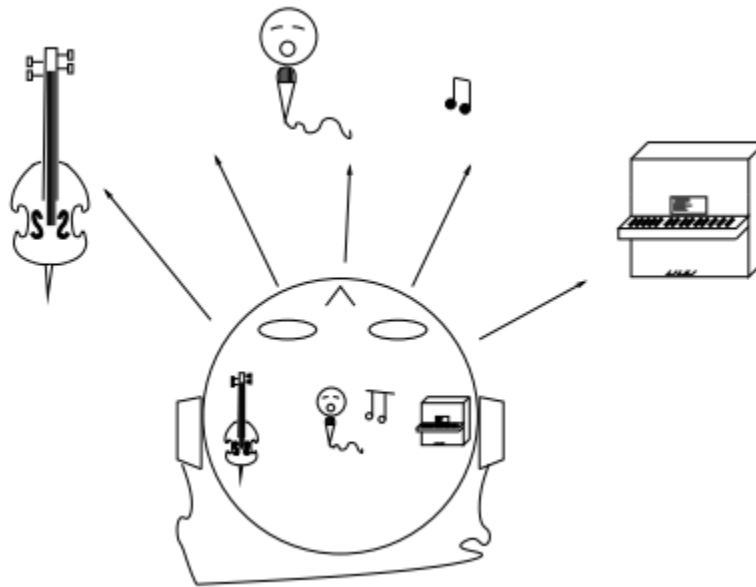
Otro uso de la técnica binaural puede ser en audiocuentos como por ejemplo el proyecto *Mind Theatre* donde se ha creado el ambiente sonoro con el fin de colocar el oyente en el centro de acción.

En conjunto con el arte visual el sistema descrito está presente casi únicamente en las películas relacionadas con música como grabaciones de conciertos (no han sido tomadas en consideración las películas que sirven solamente para presentar el sonido binaural), pero aparecen unas tendencias de usarlo en películas con argumento (por ejemplo el cortometraje *Interior* de Zachary Beckler que va a estar realizado en el año 2013 sólo con el sonido binaural). Otro ejemplo del uso de sonido binaural con vídeo puede ser el proyecto *Soundscapes* de Sony, donde el oyente podrá situarse en la escena de un concierto. Gracias a eso será posible identificarse con los músicos y su punto de vista.

### **1.3 Ventajas y deventajas del sonido binaural.**

Una evidente ventaja de la solución descrita es el coste del sistema de reproducción (auriculares) relativamente bajo. Al contrario que con los sistemas de altavoces tradicionales, no se requiere la esucha en *sweet spot*, es decir un lugar precisamente definido (en realidad, el oyente binaural se pasea con su *sweet spot*, ya que la escucha por auriculares hace que el *sweet spot* varíe con el oyente). El inconveniente que puede tener una correcta percepción del espacio acústico en el caso de los sistemas de altavoces múltiples es su poca comodidad, ya que un leve movimiento de la cabeza de tan solo unos pocos centímetros, puede provocar una errónea localización de la fuente sonora, incluso la percepción del sonido de un único altavoz (efecto de Haas). En el caso de grabaciones binaurales este problema no existe. Otra ventaja es posibilidad de localización de sonido no sólo en el plano horizontal, sino también el vertical.

A pesar de haber realizado numerosos estudios a lo largo de la historia sobre el sonido binaural, sus posibilidades teóricas nunca han sido explotadas plenamente. El motivo principal son las diferencias fisiológicas de cada persona, que serán comentadas más adelante. Las consecuencias de los problemas técnicos son sobre todo, la incorrecta localización de la fuente virtual de sonido, como por ejemplo confusiones delante-detrás y localización de sonido dentro de la cabeza (como a menudo en el caso de escuchar grabaciones estéreo con auriculares) (Figura 1.1).



*Figura 1.1. El efecto de localización dentro de la cabeza durante la escucha de grabación estéreo con los auriculares (Liitola, 2006).*

Otra desventaja de este tipo de escucha, es el movimiento de la escena acústica con respecto al propio movimiento de la cabeza (en configuración típica, es decir sin *head-tracking*) que puede ser recibido como antinatural.

Además hay que tener muy en cuenta la fatiga que pueden sufrir los oídos durante una larga sesión con este tipo de auriculares. No obstante la experiencia del autor demuestra que la fatiga es más grande en el caso de grabaciones estéreo que de las mismas pero procesados al sistema binaural. La localización dentro de la cabeza de grabaciones estéreo es en este caso decisivo y mucho menos natural.



## **2 ELECCIÓN DEL TEMA Y LOS OBJETIVOS DEL TRABAJO**

Teniendo en cuenta las consideraciones indicadas más arriba, y las tendencias al uso del sonido binaural dentro del campo audiovisual, las posibilidades han sido estudiadas para ser aplicadas en películas, gracias al desarrollo de herramientas de postproducción (como en este caso Logic Pro). La elección del tema ha sido motivada por la poca información existente en este campo. Mientras se han realizado muchos estudios sobre el sonido binaural y muchas publicaciones que tratan sobre el sonido en película, tras una larga y extensa búsqueda no se ha encontrado ninguna publicación que trate de la postproducción con el uso de sonido binaural de manera compleja. A pesar de las desventajas mencionadas, las posibilidades que el sonido binaural está abriendo en las películas parecen ser atractivas y dignas de estudiar.

En este trabajo se explorará una metodología de rodaje y postproducción del sonido binaural, para estudiar los efectos y limitaciones que tiene sobre los diferentes elementos de la banda sonora, y sobre todo la relación entre movimiento en la imagen y en el sonido. Se realizará el diseño del sonido de una serie de secuencias audiovisuales-tipo y se validará perceptivamente con un grupo de sujetos espectadores. Además se estudiará interferencias del procesado del sonido (procesadores de dinámica, reverberación, cambios de base estéreo) en la correcta localización de la fuente binaural.

## **3 MECANISMOS DE LOCALIZACIÓN DE SONIDO POR EL HOMBRE**

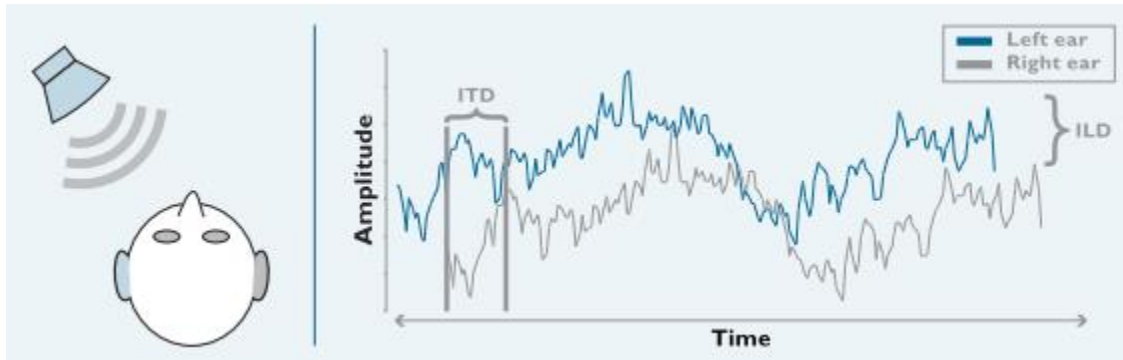
### **3.1 ITD.**

El hombre localiza la fuente del sonido basándose de varios parámetros, uno de los más importantes es la diferencia interaural de tiempo (*Interaural Time Difference*). Describe la diferencia entre la llegada de onda acústica al uno y al otro oído. Para las fuentes físicas su valor máximo, que corresponde al sonido que viene de un lado, llega

hasta unos 0,6 ms. Este parámetro es significativo en el margen de frecuencias por debajo de 1,5 kHz (Şek, 2000).

### 3.2 ILD.

ILD (*Interaural Level Difference*), es decir, la diferencia interaural de nivel es el resultado de absorber la energía acústica por la cabeza. Con respecto a la difracción de las frecuencias bajas, este parámetro es significativo sólo para las frecuencias por encima de 1,5 kHz (Stern, 2006). Su valor llega hasta unos 20 dB para los sonidos laterales. En la *Figura 2.1* se ha mostrado gráficamente los parámetros descritos.



*Figura. 3.1. Los parámetros ITD e ILD para la fuente de sonido como en la imagen (Litovsky).*

### 3.3 HRTF.

Otro importante parámetro analizado por el cerebro en el proceso de localización de sonido es la función de transferencia derivada de la cabeza HRTF (*Head Related Transfer Function*) que define las diferencias en el espectro de frecuencias del sonido que llega a los dos oídos. HRTF es definido como la relación entre la presión acústica junto al tímpano y la presión acústica en el punto correspondiente al centro de la cabeza pero sin oyente (Begault, 2000). Este parámetro es parecido entre personas diferentes en el margen de frecuencias por debajo de 4 kHz, por encima de este valor las diferencias individuales son muy notables como se puede observar en la gráfica siguiente.

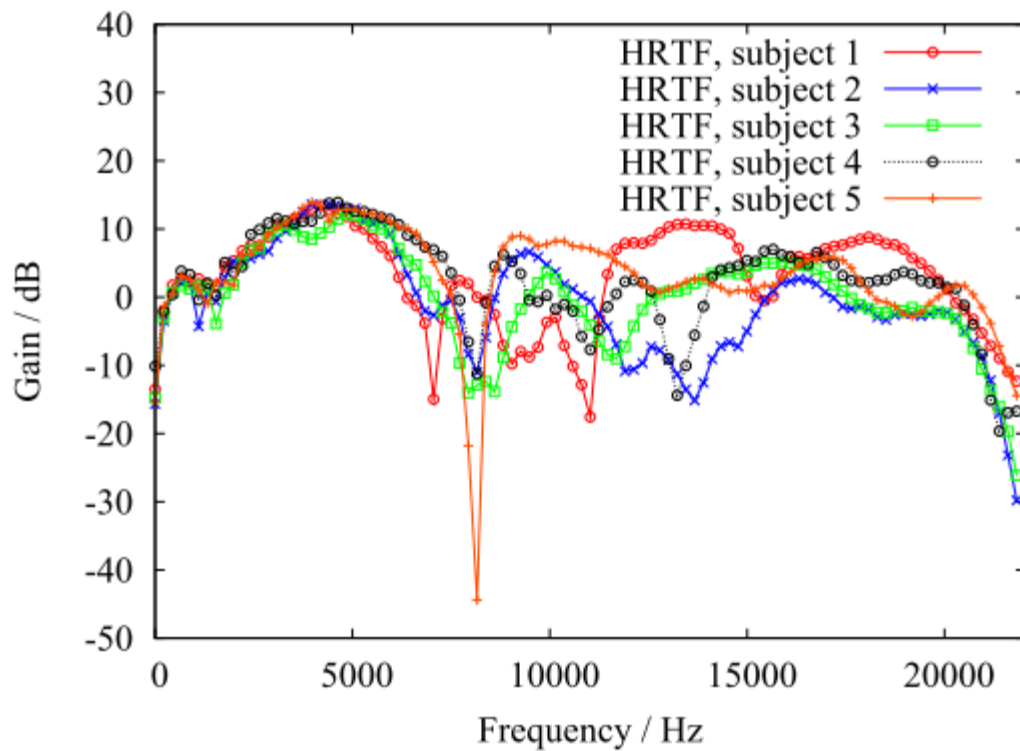


Figura 3.2. Los HRTF ejemplares de 5 personas. Se ve las diferencias claras arriba de 4 kHz (Toni Liitola, 2006).

Las divergencias son causadas por las diferencias fisiológicas, en gran medida por las formas diferentes de aurícula que son casi tan únicas como las líneas papilares. Debido a su forma complicada la aurícula funciona como resonador ampliando o atenuando diferentes bandas de frecuencia. Las consecuencias de las diferencias mencionadas en el contexto del sonido binaural son confusiones delante-detrás y localización dentro de la cabeza. Éstas tienen lugar durante escucha de las grabaciones en que se ha usado HRTF de otra persona o HRTF medio.

### 3.4 Otros mecanismos de localización.

Aparte de los factores mencionados anteriormente en el proceso de localización para las frecuencias entre 100-2000 Hz también participa la parte superior del cuerpo (Ruff, 2009) que causa una reflexión de la onda y una interferencia con la onda directa. La consecuencia es por ejemplo amplificación de 4 dB de nivel de sonido procedente del

frente en el margen de 250-500 Hz y la atenuación de 2 dB en la banda 800-1200Hz (Begault, 2000). Es conviene mencionar, que para las frecuencias bajas el sonido se recibe no sólo por el canal de oído, sino que también a través del cuerpo y los huesos.

Aparte de los factores físicos el proceso de localización está apoyado por la vista y las costumbres relacionadas a ella. El hecho de que la procedencia del sonido sea frontal nos permite ver su fuente sonora, esto puede explicar hasta cierto punto, la localización incorrecta por la falta del estímulo visual y que por ello la podamos percibir desde el otro punto.

Un impulso natural que hace más fácil la localización es el movimiento de la cabeza (más frecuentemente en el plano horizontal). El hombre automáticamente dirige la cabeza hacia el sonido oído y en base a las indicaciones que cambian (ILD, ITD, HRTF) localiza la fuente mucho mejor que en el caso de falta de movimiento (Cengarle, 2012). Los movimientos esenciales están por encima de 1 grado. En el caso de la configuración básica de escucha binaural, es decir sin seguir los movimientos de la cabeza, el oyente se queda sin la posibilidad de usar este mecanismo.

Así como en el caso de los movimientos de la cabeza, la localización del sonido será más fácil si la fuente de sonido es móvil. Los movimientos importantes son los que tienen al menos 3 grados (Begault, 2000).

El tipo de sonido, o más precisamente su envolvente, también influyen sobre las posibilidades de localización de la fuente. Los sonidos de carácter explosivo, es decir de envolvente creciente rápidamente, son localizados más fácilmente.

## 4 TÉCNICAS DE GRABACIONES BINAURALES

### 4.1 Micrófonos binaurales.

Una de las técnicas básicas de grabaciones binaurales es el uso de micrófonos colocados en los oídos y que a simple vista pueden parecer unos auriculares convencionales. Como podemos ver en la siguiente ilustración.



*Figura 4.1. Los micrófonos Soundman OKM usados durante la investigación (www.mixingroom.de).*

Debido a la colocación de los micrófonos en los oídos, en las grabaciones están presentes ITD, ILD y HRTF de la persona que está grabando.

### 4.2 Dummy head.

Otra técnica de grabación es el empleo de cabeza artificial con los micrófonos colocados dentro, junto a las salidas de los canales de oído. A veces se usa también un tronco artificial que tiene como objetivo la simulación de los rebotes del sonido en tórax y hombros. Abajo un ejemplo de la cabeza artificial usada para grabaciones binaurales.



Figura 4.2. Neumann KU100 ([www.neumann.com](http://www.neumann.com)).

## 5 IMPORTANCIA DE LOS AURICULARES EN LA ESCUCHA BINAURAL

En el proceso de localización, sobre todo en la eliminación de confusiones delante-detrás, la corrección individual de frecuencia de los auriculares juega un papel importante (Sang-Myeong, 2004). A pesar de eso en este estudio se ha utilizado la corrección individual por dos razones. La corrección mencionada requiere el empleo de las herramientas y las técnicas usadas en el laboratorio de acústica y este estudio se centra en las posibilidades de transformar el sonido con las herramientas típicas de estudio de postproducción. Además, supone que el producto audiovisual está dirigido al amplio círculo de espectadores, por las razones prácticas, no es posible aplicar la corrección individual a cada oyente.

Otro parámetro que influye sobre la percepción correcta de las grabaciones binaurales es la impedancia acústica de auriculares. Este valor puede ser expresado por el parámetro PDR (*Pressure Division Ratio*) que define la relación entre la presión acústica generada por los auriculares justo al principio del canal de oído con un oyente y en el campo libre (sin oyente). En el caso ideal este parámetro es igual a 1, que significa que los auriculares corresponden a la escucha en el campo libre (en este caso son definidos como abiertos). Esta condición es cumplida por los auriculares semi profesionales AKG K530 que han sido usados durante la investigación. Su margen lineal es según la especificación 17-26500 Hz. En el caso de grabaciones binaurales el uso de auriculares cerrados, aísla de los sonidos externos pero cambia la impedancia acústica junto al oído, que a su vez, influye negativamente en la externalización.

## **6 REALIZACIÓN DE GRABACIONES**

Como este trabajo tiene como objetivo verificar de manera práctica las posibilidades de utilización de las técnicas binaurales en postproducción cinematográfica, en el estudio de televisión se ha capturado unas secuencias de vídeo con diferentes movimientos de objetos, tanto como con diferentes tipos de sonido (voz, pasos, monedas, explosión).

Las grabaciones se ha realizado con dos técnicas, la primera usando los micrófonos binaurales Soundman OKM para grabar el sonido del plano. La señal de los micrófonos colocados en los oídos del operador de cámara estaba registrada con la grabadora Zoom H4n. Este método obliga a la persona que está grabando a conservar la conformidad de movimientos de la cámara con los de la cabeza que en los casos de las producciones profesionales no siempre es posible. Teniendo en cuenta lo indicado anteriormente merece la pena considerar el uso de *dummy head* fijado a la cámara.

La primera de las escenas grabadas, *Círculo*, muestra una persona hablando mientras da la vuelta alrededor de la cámara en la distancia de un metro.

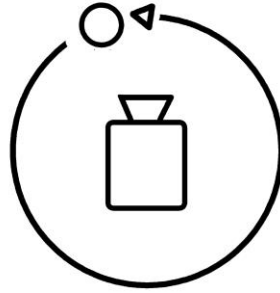


Figura 6.1. La escena Círculo.

En la escena siguiente, *Dinero*, una persona esta hablando, mientras se va acercando en el centro del plano y meciendo unas monedas con las manos, desde tres metros hasta medio metro de la cámara. Luego la persona se aleja a la localización inicial.



Figura 6.2. La escena Dinero.



El vídeo *Dos Personas* presenta una conversación con las personas colocadas como en la ilustración. Mientras las personas están hablando, la cámara, que esta alejada dos metros de las persona, se está moviendo de una persona a otra.

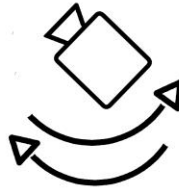


Figura 6.3. La escena *Dos Personas*.

*Globos* muestra una persona hablando, pinchando los globos y, a la vez, cruzando el plano de la derecha a la izquierda. La distancia de la cámara son dos metros.



..

Figura 6.4. La escena *Globos*.

En la escena *Soledad* se ve una persona hablando, cruzando el plano como se ha mostrado en la ilustración. La distancia es de medio metro cuando está más cerca de la cámara.

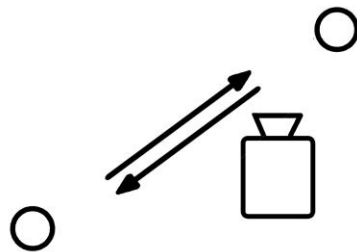


Figura 6.5. La escena *Soledad*.

El vídeo siguiente, *Balón*, presenta un movimiento similar a el de la escena *Globos*, pero la persona está botando el balón y la distancia son 4 metros.

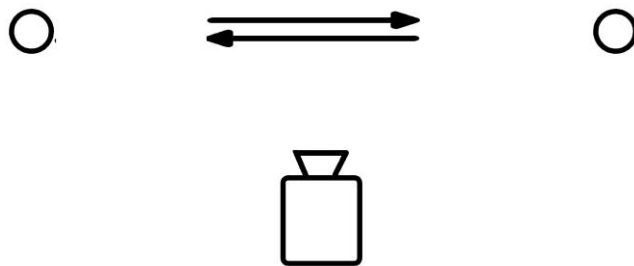


Figura 6.6. La escena *Balón*.

Se ha grabado también otras escenas, pero, teniendo en cuenta el nivel de complicación de la postproducción, se ha elegido para estudio sólo unas, con los movimientos y los tipos de sonido diferenciados.

Una vez capturado el material de vídeo, todos los sonidos e incluso las cuestiones habladas, han vuelto a ser grabadas en el estudio de postproducción con un micrófono mono. Para obtener la conformidad entre el sonido del plano y el del estudio, sobre todo en el caso de las cuestiones habladas, los actores grababan escuchando la versión original con auriculares. Además, para facilitar el proceso, podían ver el vídeo en la pantalla. Las grabaciones tenían lugar en un espacio de poca reverberación (para obtenerlo se ha colocado en casi todos lados los paneles absorbentes) para que se pudiera controlar este parámetro en las etapas siguientes. Posteriormente las grabaciones han sido procesadas para obtener un sonido binaural y estéreo; este procesado será comentado en las siguientes partes del trabajo.

## **7 USO DE HERRAMIENTAS SOFTWARE**

### **7.1 Especificación de las herramientas usadas.**

Para editar y procesar el material grabado se ha trabajado con el programa Logic Pro 9, es un estándar en la postproducción de audio. La mayoría de las herramientas empleadas (excepto el plug-in Panorama 5) están integradas dentro de este programa.

### **7.2 Preparación de trabajo.**

Una vez creado un nuevo proyecto en Logic Pro, han sido importadas las grabaciones de la cámara, las binaurales (captadas con la grabadora Zoom) y las mono realizadas en el estudio de postproducción. Todas han sido sincronizadas con el sonido del plano registrado por la cámara. Para sincronizar los archivos de audio con el sonido original, se han ajustado viendo la forma de onda y buscando los puntos característicos (como transitorios). Además las grabaciones del estudio de postproducción requerían

usar el modo Flex, ante todo para sincronizar la voz hablada con los movimientos de labios.

Las grabaciones binaurales se han dejado en una pista estéreo con la salida *Stereo Out* y las grabaciones hechas en el estudio de postproducción se han dirigido paralelamente a los buses independientes para procesar posteriormente ese mismo material de maneras diferentes (mezcla estéreo y procesamiento binaural con Binaural Panner y Panorama 5). Más adelante se ha mostrado la ventana del mixer de la grabación *Dinero*.

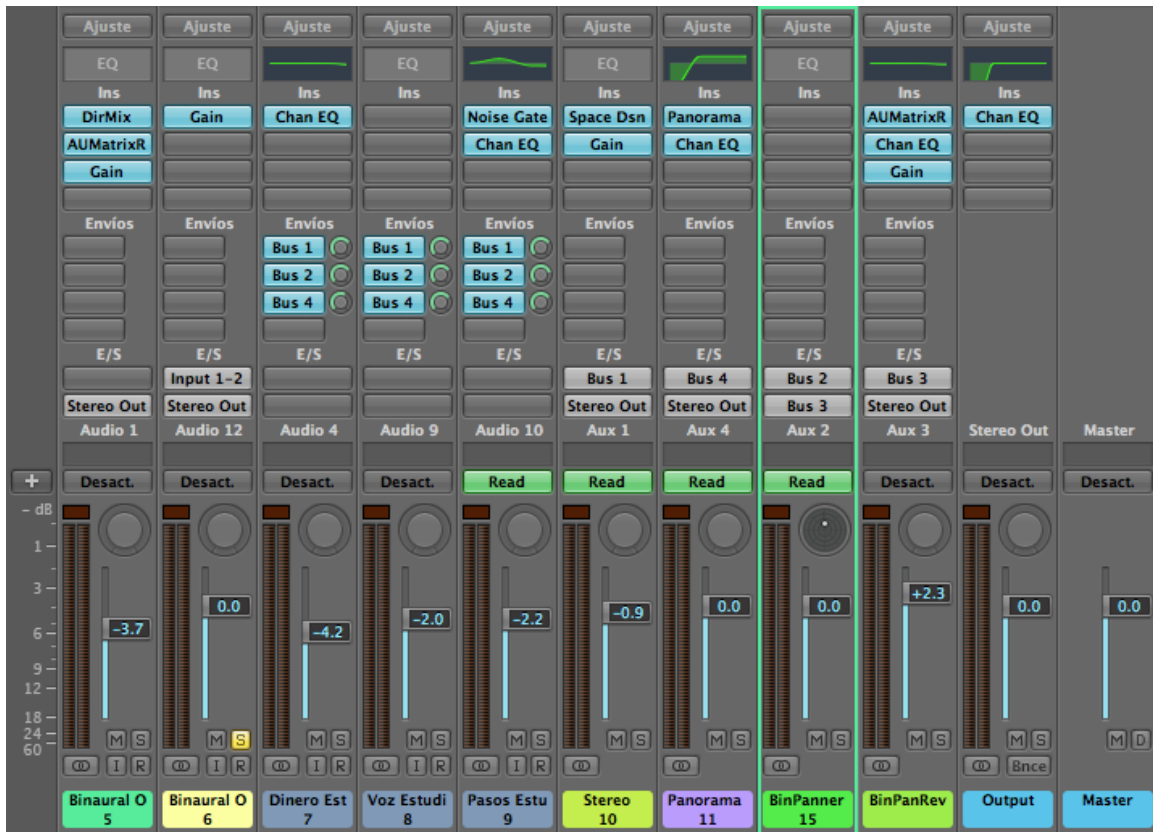


Figura 7.1. La vista del mixer en la sesión ejemplar de Logic Pro.

### 7.3 Mezcla estéreo.

Después de limpiar y sincronizar las grabaciones mono del estudio se ha realizado una mezcla estéreo básica con la automatización de los niveles de volumen, panorama y timbre con ecualizador. En algunos casos se ha usado expander para atenuar

suavemente el nivel de ruido entre los sonidos presentes en el plano. Como las grabaciones han sido realizadas en un espacio bastante insonorizado, en la mezcla se ha aplicado reverberación con la intención de introducir los sonidos en el contexto acústico adecuado. El uso de la reverberación permite evitar un error común que está presente en las películas dobladas y que consiste en un procesado de la voz que provoca percibir al actor como si estuviera justo en frente del espectador y no en la escena. Una compresión fuerte, falta de reverberación y acentuación de los tonos altos y graves mejoran la integridad de la habla, pero perjudicando de manera importante la naturalidad de recepción. En la ilustración que podemos observar más abajo, se ha presentado la automatización de parámetros usada durante el trabajo.

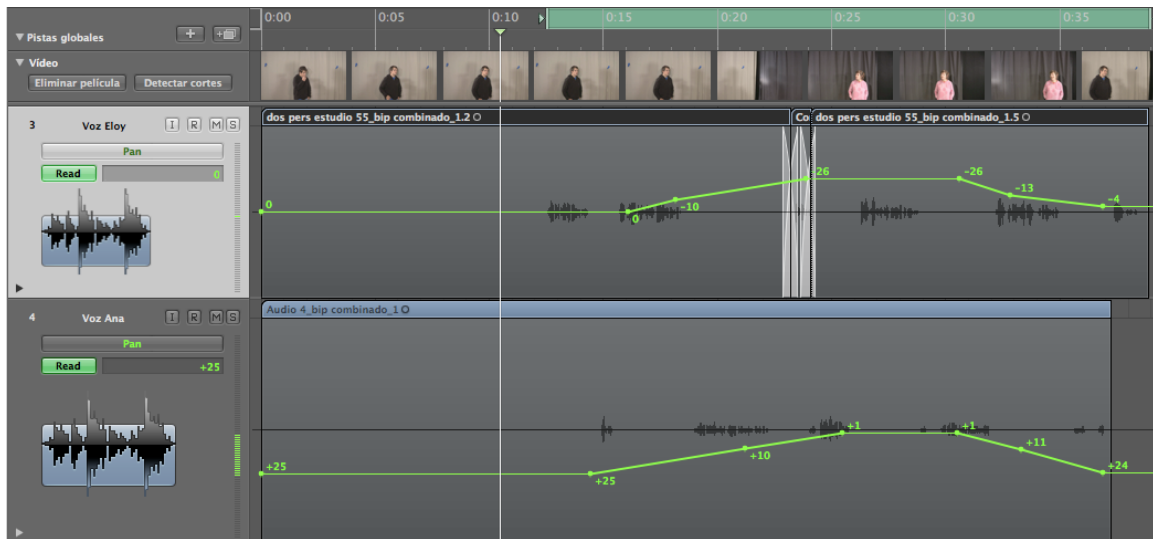


Figura 7.2. Automatización del panorama en grabación estéreo.

## 7.4 Tratamiento de las grabaciones binaurales.

### 7.4.1 Los obstáculos encontrados.

El material registrado con los micrófonos binaurales Soundman OKM comparado con la mezcla estéreo ofrece una percepción del espacio mucho más realista, pero se puede notar numerosas imperfecciones.

La primera de ellas es la falta de coincidencia entre el movimiento del objeto en el plano horizontal en el vídeo y el sonido correspondiente. Un movimiento pequeño del objeto fuera de la parte central del plano causa la localización del sonido justo al lado de un oído. Se puede decir que el efecto binaural está exagerado.

El siguiente problema es la localización parcial dentro de la cabeza. Este fenómeno ocurre sobre todo con los objetos que se encuentran durante la grabaciones justo en frente de la persona que está grabando. La impresión no es tan significativa como en el caso de la mezcla estéreo donde la localización dentro de la cabeza es una consecuencia natural del ajuste de sonido a la escucha con altavoces, pero causa que las grabaciones sean recibidas muy por debajo de las previsiones.

Las imperfecciones también pueden darse por la colocación imprecisa de la cabeza de la persona que está grabando. En una de las escenas la persona grabada que se encontraba en la parte central es percibida lateralmente durante la escucha.

Hay que acordarse que en las grabaciones binaurales no se pueden colocar los micrófonos libremente con el fin de registrar los sonidos separadamente y por eso puede resultar difícil obtener el equilibrio deseado entre los niveles de sonido de los objetos diferentes. Ese fenómeno es muy visible en la escena donde aparece paralelamente la voz hablada y mucho más alto sonido de los globos pinchados.

Un error esperado, que aparece frecuentemente con las grabaciones binaurales, son las confusiones delante-detrás que han resultados inevitables también en este caso.

Además, según las previsiones, se ha notado una difícil localización durante la escucha de las grabaciones realizadas en una sala con mucha reverberación.

En el material registrado a veces es posible oír la respiración de la persona que lleva los micrófonos especialmente en el caso de usar los micrófonos colocados en los oídos, debido a eso hay que intentar permanecer en silencio, ya que los sonidos como la respiración pueden resultar difícil de eliminar a la hora de edición.

### 7.4.2 Mejora de calidad de las grabaciones.

Se han propuesto soluciones que han permitido mejorar la localización y externalización del material registrado, mediante la utilización de herramientas software de audio, aplicando los conocimientos sobre el funcionamiento de sonido binaural, pero también experimentando con nuevos métodos.

El primero de los problemas presentes, la localización incorrecta en el plano horizontal, ha sido eliminado prácticamente por completo estrechando la base estéreo. Con el fin de eliminar dicho problema se ha usado la herramienta Direction Mixer que deja definir la proporción entre las señales *Mid* (M) y *Side* (S), donde *Mid* es la parte común del canal izquierdo (*L*) y el derecho (*R*) y *Side* la diferencia:

$$\begin{aligned} M &= L + R, \\ S &= L - R. \end{aligned}$$

El parámetro *Spread* que toma los valores de 0 (que significa sólo la parte *Mid*, es decir, la señal mono) a 2 (sólo la señal *Side*) se ha ajustado, según el caso, entre 0.2 y 0.4, por lo tanto el oyente no percibe el objeto cerca del oído o indudablemente al lado mientras éste está en frente, un poco desviado del centro de la escena. Después del usar la herramienta las grabaciones no son tan espectaculares como al principio, sin embargo la localización con el oído y con la vista son más congruentes.



Figura 7.3. La herramienta Directional Mixer.

La localización dentro de la cabeza no ha sido eliminada por completo, pero se ha conseguido reducirla hasta cierto punto. Basándose en que el ser humano localiza la fuente del sonido fuera de la cabeza si la onda acústica contiene la información sobre la acústica de un espacio dado, se ha aplicado una reverberación corta que simula colocación de la fuente en un cuarto. La relación entre el sonido directo y la reverberación llamada *Acoustic Ratio* (AR) es el factor más importante de la localización fuera de la cabeza (Wilde, 1993). Por supuesto las grabaciones contienen la respuesta acústica de la sala, sin embargo esas informaciones son insuficientes o equívocas. Como el mecanismo de localización es sensible sobre todo a los primeros no dispersados rebotes, el tiempo de reverberación es posiblemente corto. Se ha observado la mejora de la externalización junto con la reducción de ese parámetro, pero el límite inferior del tiempo de reverberación es el valor que causa aparición de coloración de sonido típico para las señales parecidas desplazadas en fase. Durante las investigaciones se ha usado un plug-in disponible en Logic Pro AUMatrixReverb (Figura 7.4).

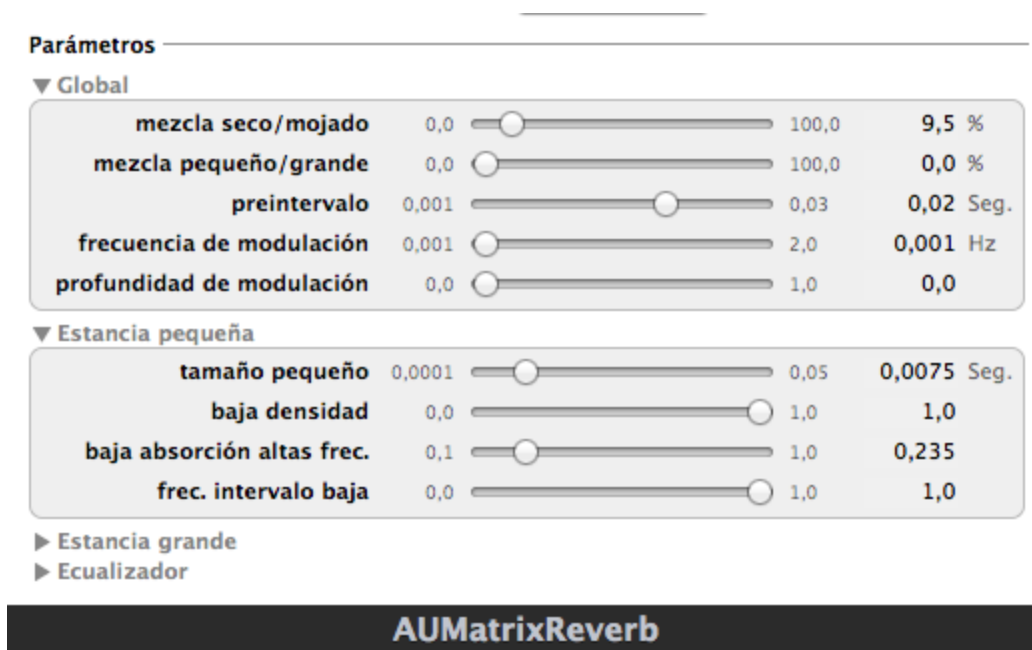


Figura 7.4. La vista del plug-in AUMatrixReverb usado en el proyecto.

El tiempo de reverberación ha sido ajustado de manera experimental, el fabricante no informa de cual es la relación entre los valores disponibles y el tiempo expresado en



segundos. Igualmente la situación parece en el caso del filtro paso bajo que se ha usado con fin de dejar sólo los rebotes para las frecuencias medias y bajas ya que solo esas tienen importancia en el proceso de localización (aquí utilizado es el mecanismo de ITD).

Por supuesto la aplicación de reverberación lleva tras sí no sólo localización de la fuente, sino también la acústica de cuarto y el timbre de sonido, por lo tanto su uso implica conseguir el equilibrio entre esos parámetros.

No se ha logrado encontrar una herramienta eficaz que permita eliminar las confusiones delante-detrás, sin embargo se ha notado una mejoría en este campo después del uso de la reverberación mencionada arriba.

El problema del pequeño desplazamiento del objeto en plano horizontal se ha solucionado con la automatización de panorama que se ha mostrado en la foto.

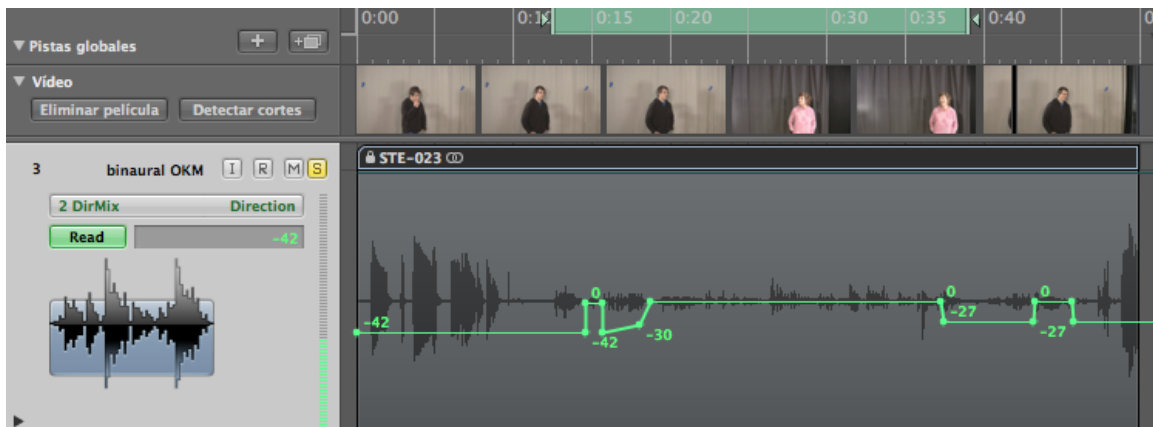


Figura 7.5. Corrección de posición de la fuente en grabación binaural.

No obstante hay que tener en cuenta que eso no soluciona el caso de que en el plano haya más fuentes de sonido localizadas incorrectamente. El mencionado arriba no ha sido estudiado.

Los inconvenientes causados por el enmascaramiento de la voz por las explosiones de los globos han estado reducidas al grado mínimo con el uso de compresor que reacciona al valor momentáneo (*Peak*). Debido a eso se ha obtenido una buena

inteligibilidad de habla y además se ha disminuido la molestia causada por las explosiones escuchadas con auriculares. La configuración del compresor se ha presentado en la foto más abajo.



*Figura 7.6. El compresor usado para mejorar la dinámica en la grabación „Globos”. Resulto necesario aplicar los ajustes causantes el Gain Reduction de unos 20 dB.*

La localización del sonido registrado en la sala de una reverberación grande ha sido mejorada gracias al uso de expander, que ha reducido subjetivamente la cantidad de la respuesta acústica del espacio que enmascaraba los más importantes mecanismos de localización.



Figura 7.7. Expander usado en la grabación Balón.

El efecto de funcionamiento del expander se ve gráficamente en la foto siguiente. La amplitud justo antes de cada impulso siguiente (el rebote del balón) es menor después de la aplicación del procesador que significa la respuesta acústica del espacio más pequeña.

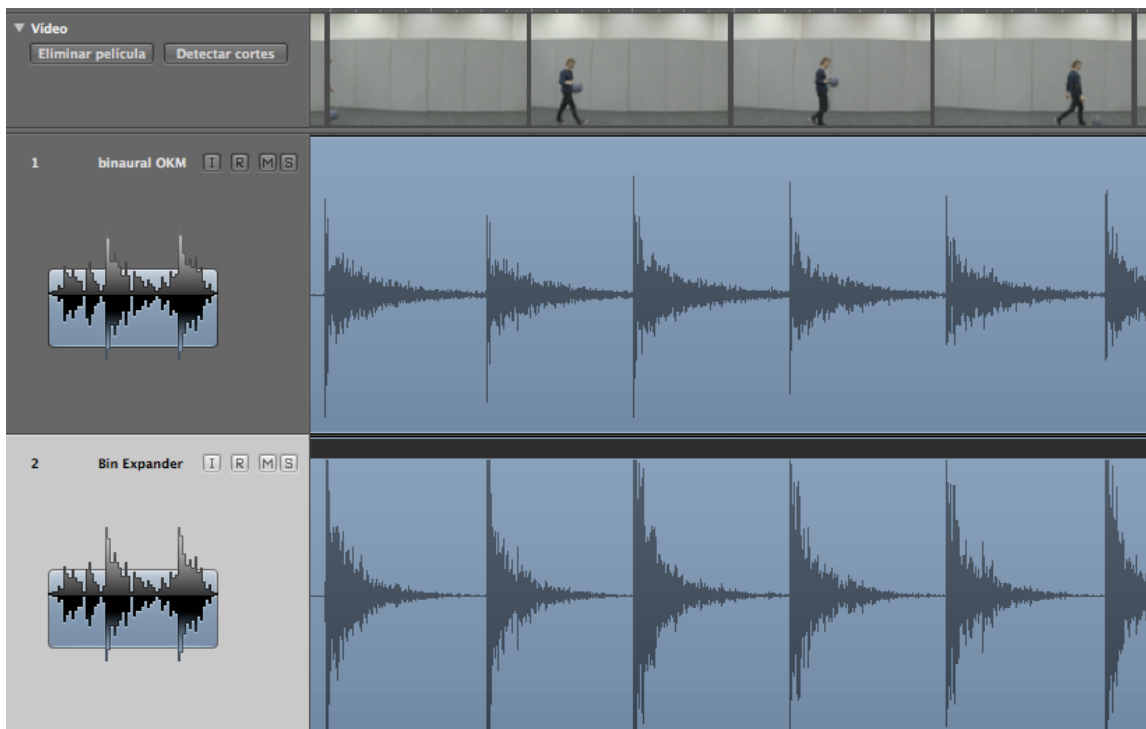


Figura 7.8. La forma de onda de la grabación binaural sin procesador de dinámica (arriba) y después de la aplicación del expander (abajo).

## 7.5 Mezcla con el uso de la herramienta Binaural Panner.

Accesible en Logic Pro Binaural Panner ha sido diseñado con la intención de transformar las grabaciones estándares al sonido binaural. La herramienta añade a la señal las informaciones como ITD, ILD, HRTF. El tablero de usuario permite definir unos parámetros virtuales, entre otros la localización del objeto a través de apreciar la distancia del oyente y el grado respecto al eje central (tanto en el plano horizontal como vertical).

Cada grabación mono realizada en el estudio de postproducción ha sido transformada por medio de la herramienta Binaural Panner, con el uso de automatización de los parámetros que definen la posición de la fuente de sonido.

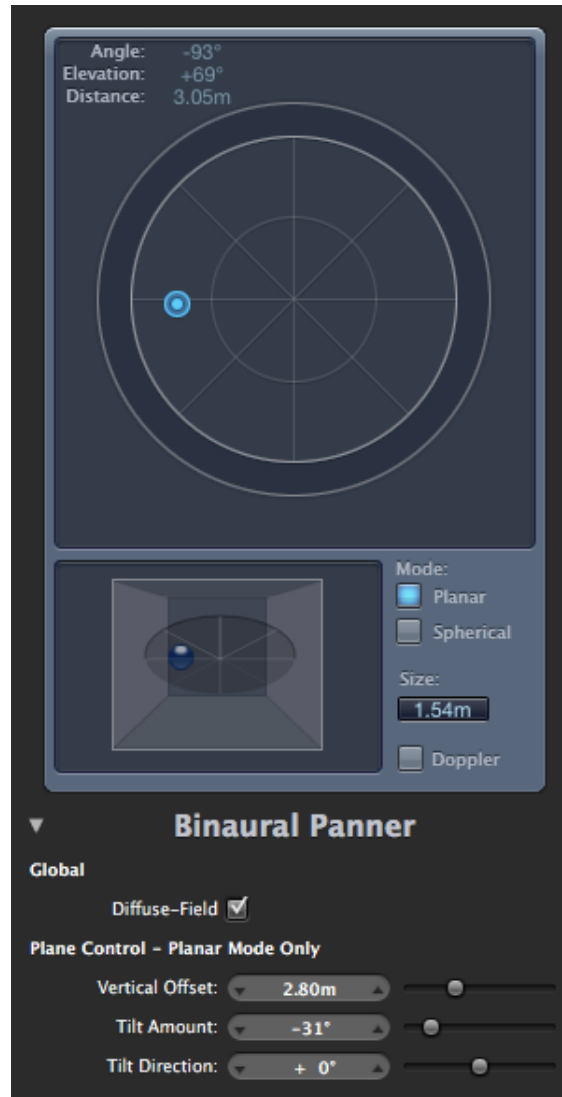


Figura 7.9. La vista de la herramienta Binaural Panner en el modo Planar.

Han sido testadas diferentes configuraciones del tamaño del espacio, la distancia del oyente y la inclinación; a pesar de eso, todos los resultados obtenidos han sido de baja calidad ya que resultaba difícil establecer su localización correcta (el autor ha notado que las localizaciones más frecuentes se encontraban dentro de la cabeza o detrás en lugar de delante).

Según la evaluación subjetiva por el autor, en comparación con otros tipos de procesamiento (estéreo, Panorama 5), Binaural Panner introduce a la señal una cantidad de ruido y empeora la calidad de sonido (que se oye particularmente en el margen de frecuencias altas).

En unos casos ha resultado favorable el cambio del ángulo de la inclinación del plano en que se mueve el objeto o usar el modo esférico que hace posible ajustar el ángulo de la inclinación de manera dinámica, sin embargo no se ha conseguido encontrar la relación entre las posibilidades de mejorar la calidad de la grabación por medio de dichos parámetros y el tipo de escena.

Hay que tener presente, que los parámetros como la distancia, con respecto a la del oyente no siempre se ajustan a los valores indicados por el programa. Por lo tanto es necesario ajustar los valores de manera experimental. Más abajo se muestra la automatización de parámetros de la grabación *Soledad*.

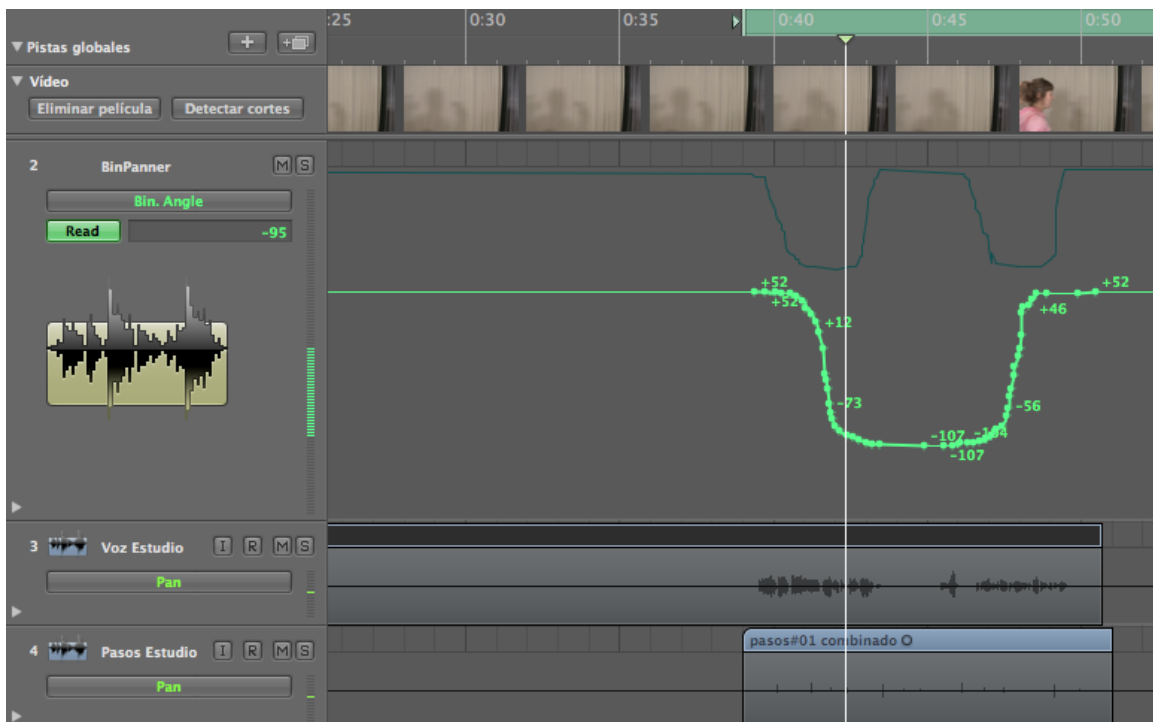


Figura 7.10. Automatización del ángulo y de la distancia en Binaural Panner.

En algunos casos, después de Binaural Panner se ha usado AUMatrixReverb con los ajustes similares a los de grabaciones binaurales del plano, que ha permitido mejorar la externalización.

## 7.6 El uso de la herramienta Panorama 5.

El siguiente paso consistía en la aplicación del plug-in Panorama 5 de Wave Arts. La herramienta procesa las grabaciones de mono a sonido binaural, pero con la posibilidad de mantener un control de los parámetros mucho más preciso, que en el caso de Binaural Panner. En esta imagen podemos observar la herramienta descrita, usada para procesar los archivos mono del estudio de postproducción.



Figura 7.11. La herramienta Panorama 5.

A diferencia de Binaural Panner, Panorama 5 hace posible aplicar modelos de HRTF diferentes, lo que ha resultado tener mucha influencia en el proceso de localización y externalización. La elección del modelo HRTF ha sido dictado por las impresiones subjetivas y por eso no garantiza la percepción correcta en el caso de otras personas. Hay que tener en cuenta que no es posible elegir un modelo que de resultados positivos con todos los oyentes, puesto que ha sido imprescindible una elección arbitraria. El problema de variedad de los modelos de función de transferencia derivada de la cabeza es muy complejo y excede el campo de este trabajo, sin embargo hay materiales

fácilmente accesibles que tratan ampliamente de ese tema. Cabe resaltar en este apartado del proyecto, que los mejores resultados que proporcionaban los modelos HRTF dependían de la escena en concreto, el modelo más adecuado depende de la situación acústica particular, que no era lo que se podía esperar. En la foto siguiente se ve el menú de los HRTF accesibles.



Figura 7.12. Los modelos de HRTF y la vista a la posición virtual del objeto de Panorama 5.

La herramienta estudiada permite generar los primeros rebotes y su control bastante preciso.



Figura 7.13. El tablero de los primeros rebotes de Panorama 5.

Para cada pared del espacio hexagonal virtual es posible definir su posición y el material con que está cubierta. Como los primeros rebotes sostienen el proceso de localización más en el margen de frecuencias bajas en la mayoría de los casos se ha elegido un material muy absorbente (*Heavy Carpet*) que además tiene la ventaja de que no causa mucha coloración del sonido. Según las previsiones, al añadir los primeros rebotes a la señal original se ha mejorado la localización y externalización.

Igual que en el caso de Binaural Panner, las distancias definidas en el plug-in no reflejan la realidad, por lo tanto hay que seleccionar dichas distancias basándonos en nuestra experiencia y el estudio de la herramienta.

## **8 EVALUACIÓN PERCEPTIVA**

### **8.1 Análisis de datos cuantitativos.**

Se ha realizado una encuesta sobre la calidad de las grabaciones hechas y sobre la congruencia entre la localización del sonido y la posición del objeto en la pantalla. Los resultados se presentan a continuación.

Se ha entregado a los sujetos cinco vídeos (*Círculo, Dinero, Dos Personas, Globos, Soledad*), cada uno en cinco versiones (la grabación bruta de OKM, esa misma grabación procesada en Logic Pro, el sonido grabado en el estudio de postproducción y procesado con Binaural Panner de Logic Pro, luego sonido procesado en Panorama 5 de Wave Arts y mezcla estéreo). Las grabaciones han sido cifradas para que los oyentes no conozcan el tipo de procesamiento.

Se ha preguntado a los sujetos sobre la calidad del sonido y la congruencia entre la localización del objeto en la pantalla y la del sonido (indicando que se debe tomar en cuenta tanto los movimientos horizontales como los delante-detrás). Esos dos criterios han sido considerados como más importantes y fáciles de entender para los sujetos que en la mayoría no se dedican al trabajo con sonido. Cada vídeo se evaluaba en la escala



entre uno y cinco (donde el uno es la nota más baja). Los sujetos escuchaban con su propio equipo las grabaciones subidas a la plataforma politube.upv.es y marcaban las respuestas en el documento Word.

Estadísticos de fiabilidad. Los sujetos (N=7) tienen una concordancia baja en las respuestas de congruencia (Alfa de Cronbach=,483), y más alta en las respuestas de calidad (Alfa de Cronbach=,815 tras eliminar el sujeto 7 que tiene repuestas opuestas al resto de participantes y no se introducirá en los cálculos).

Es importante remarcar que la concordancia de respuestas es alta para las escenas “dinero” y “dos personas” (alfa>,75) y es negativa para el resto.

Una correlación de Pearson entre las repuestas de calidad y congruencia muestra que los sujetos separan las dos valoraciones ya que la correlación es baja ( $r=,318$ ;  $p<.001$ ). Para algunos sujetos hay una correlación muy significativa entre calidad y congruencia (e.g. s3) mientras que para otros no hay ninguna relación (e.g. s5).

Se normalizan las respuestas por sujeto (puntuación Z) y se realiza un análisis de varianza multifactor, utilizando los factores escena y tipo de procesado, y como variables dependientes, las respuestas normalizadas de congruencia y calidad.

### **8.1.1 Congruencia.**

Análisis de varianza: hay una influencia significativa tanto del procesado ( $F=3,7$ ;  $p<.01$ ;  $\eta^2=,091$ ), como de la escena ( $F=4,4$ ;  $p<.01$ ;  $\eta^2=,106$ ). Si se analiza escena por escena, solamente en la escena “dinero” hay una influencia claramente significativa del procesado ( $F=4,8$ ;  $p<.01$ ;  $\eta^2=,391$ ), en el que el binaural panner destaca por una menor congruencia, y las grabaciones binaurales (OKM) y estéreo como más congruentes, como se puede ver en la figuras 8.1 y 8.2.

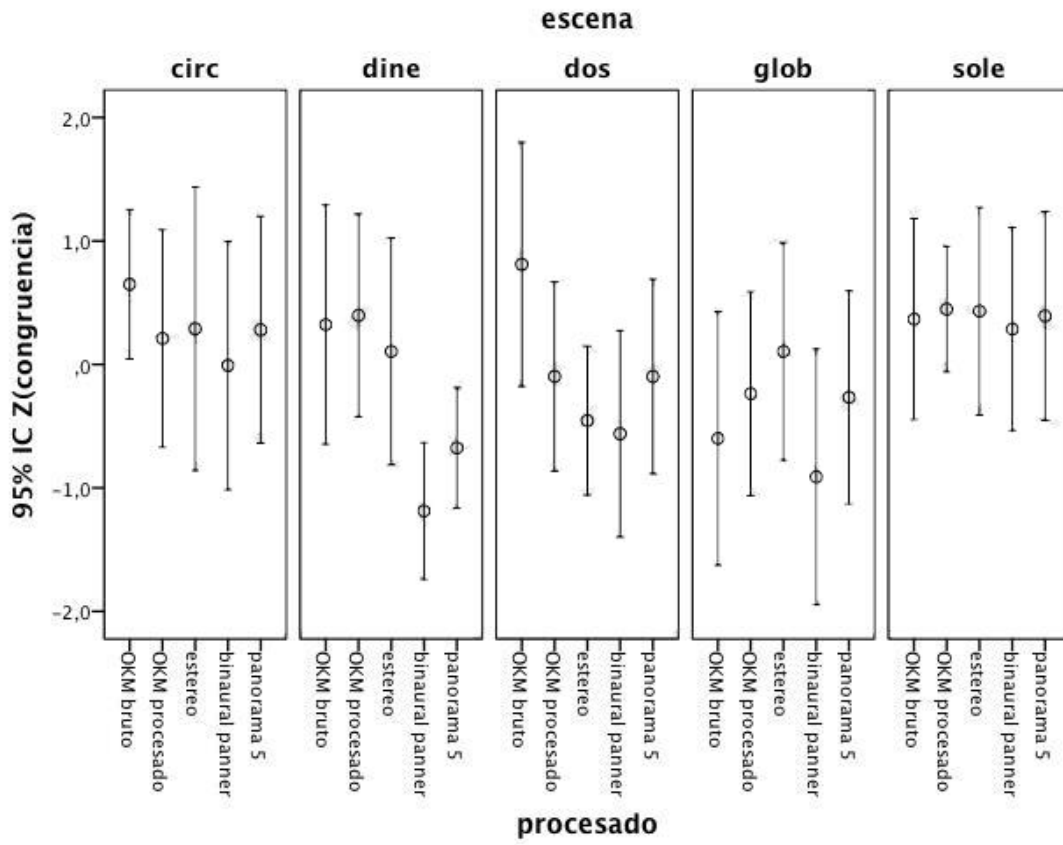


Figura 8.1. Medias con intervalo de confianza para la congruencia por procesado y escena (circ – Círculo, dine – Dinero, dos – Dos Personas, glob – Globos, sole – Soledad).

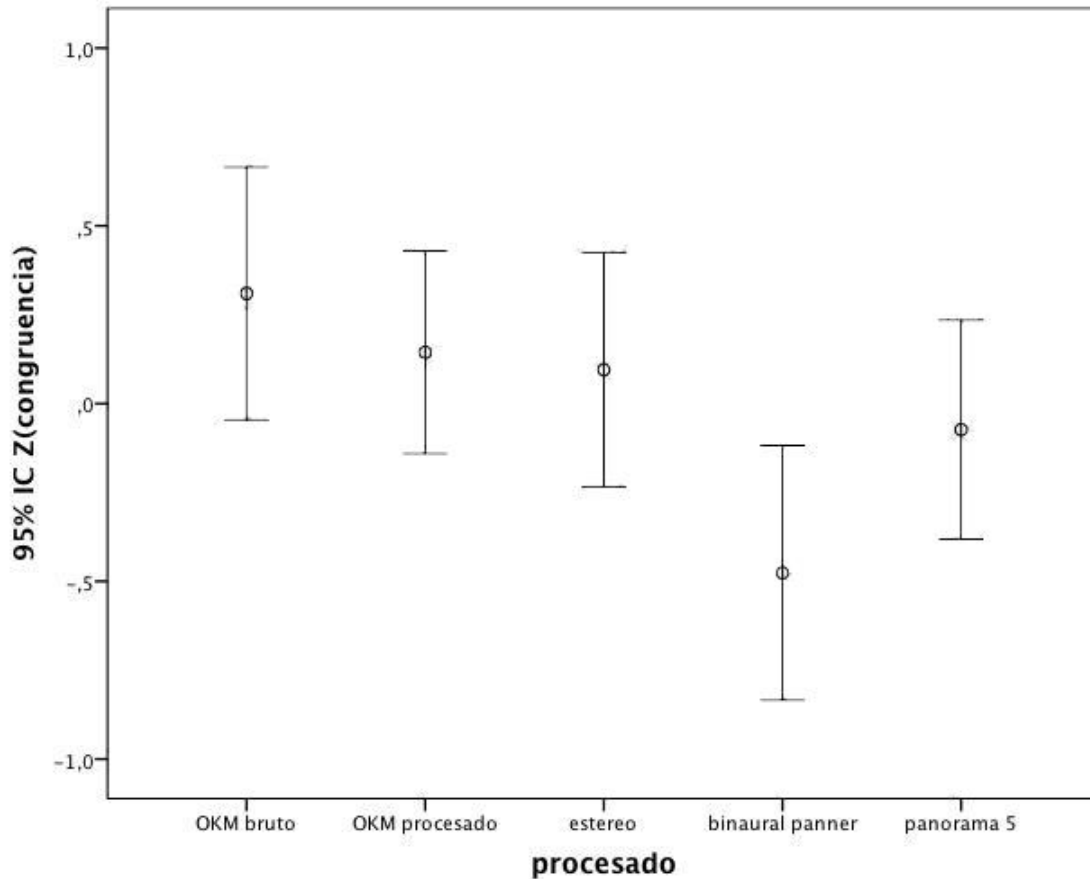


Figura 8.2. Medias con intervalo de confianza para la congruencia por procesado.

### 8.1.2 Calidad.

En todos los casos, menos el vídeo *Dos Personas*, los sujetos indican las grabaciones binaurales (OKM) como las de la peor calidad y a parte de eso, las demás con calidad similar. Eso significa que los oyentes valoran la calidad según el método de grabación y no el tipo de procesamiento. Los resultados se han mostrado en las ilustraciones a continuación.

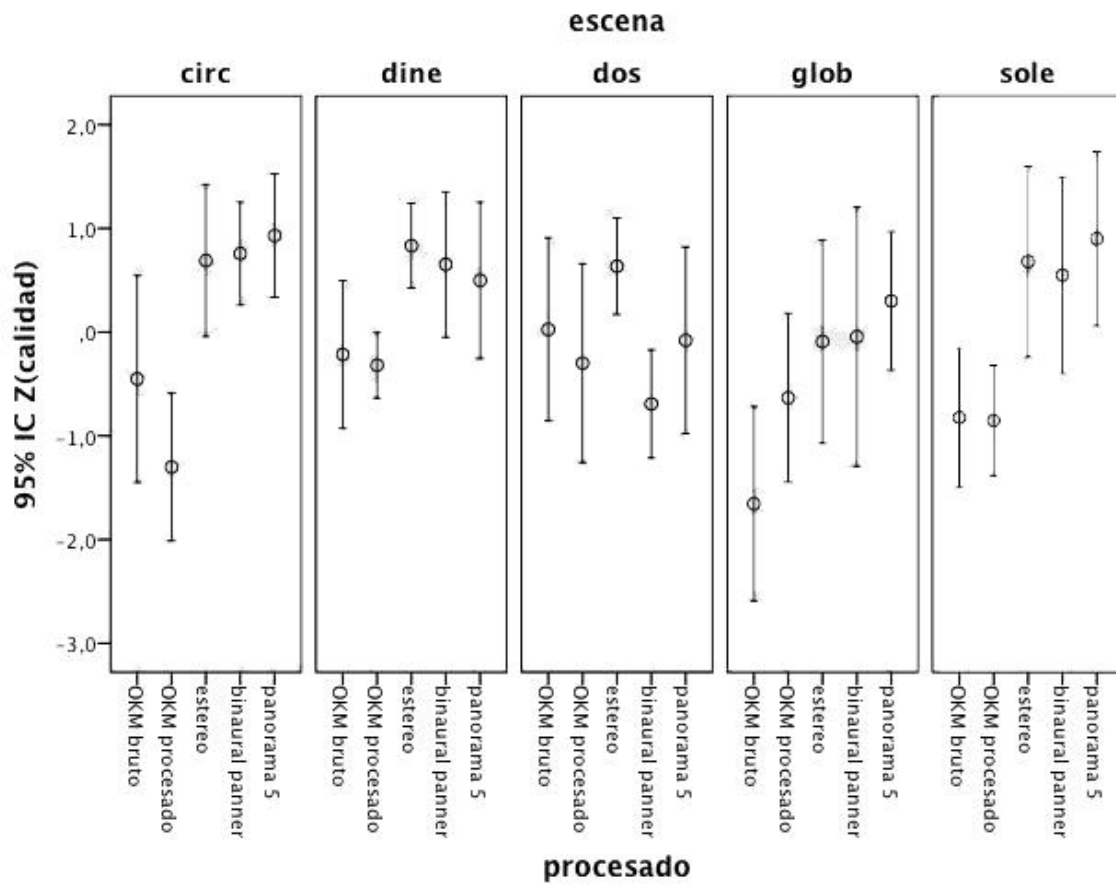


Figura 8.3. Medias con intervalo de confianza para la calidad por procesado y escena (circ – Círculo, dine – Dinero, dos – Dos Personas, glob – Globos, sole – Soledad).

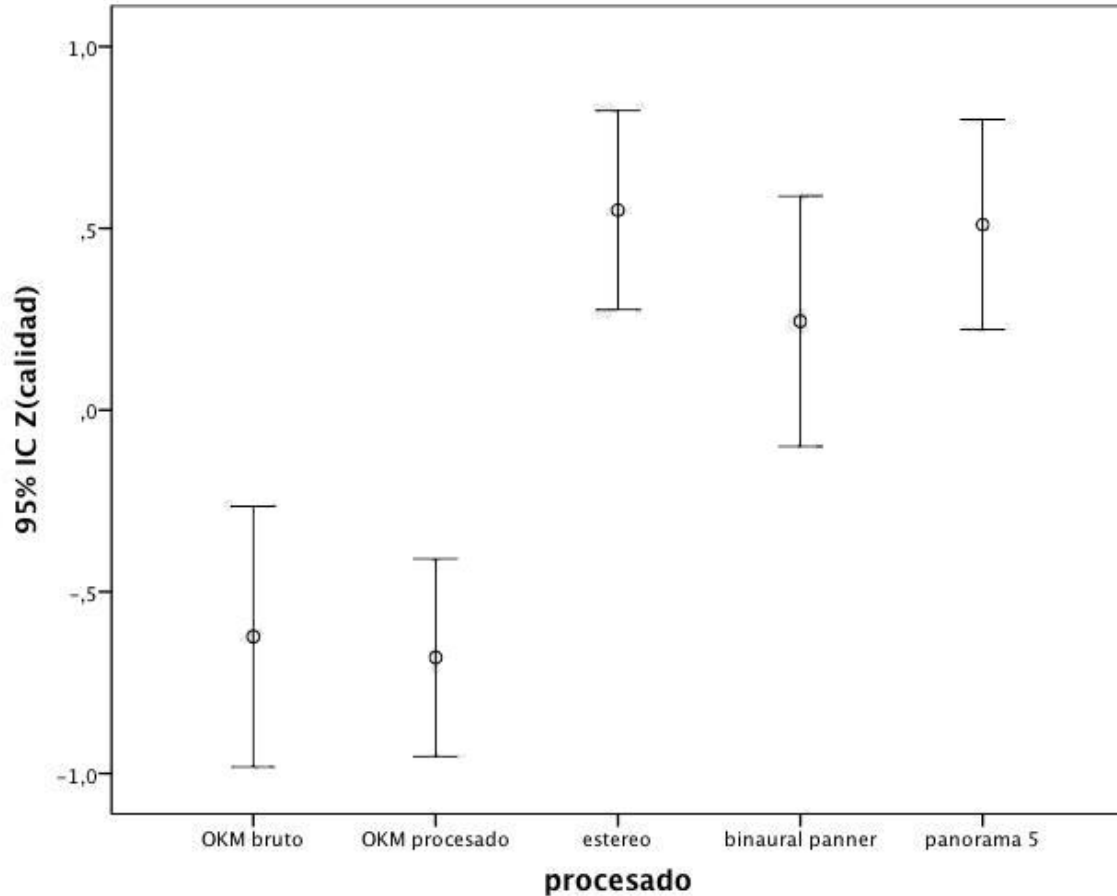


Figura 8.4. Medias con intervalo de confianza para la congruencia por procesado.

## 8.2 Análisis cualitativo.

El siguiente análisis ha sido hecho con los ingenieros de sonido en activo. Después de entregar esos mismos vídeos que en el caso del análisis cuantitativo, se ha pedido a los oyentes que expresen su juicio sobre la congruencia, calidad, naturalidad, comodidad y, ante todo, lo que puede molestar a los espectadores.

Lo que indican los oyentes como más molesto es el nivel de ruido bastante alto en el caso de las grabaciones con los micrófonos binaurales (tanto las procesadas como las brutas). Esto es en consecuencia a grabar los objetos desde cierta distancia (más alejada que en el caso de grabaciones en el estudio). Hay que tener en cuenta que para mantener la localización correcta, el objeto no puede acercarse a los micrófonos con el

fin de bajar la relación señal/ruido, ya que es muy recomendable usar los micrófonos y la grabadora o la interfaz de alta calidad.

De acuerdo con el autor, las grabaciones brutas de OKM Soundman son percibidas como si el objeto se encontrara más cerca de lo que se ve. Como se ha mencionado antes, el problema ha sido solucionado con una reverberación corta y a través de disminuir la base estéreo. Sin embargo, como indican los oyentes, los procesos cambian negativamente la acústica y la localización en unas escenas (no todas). Por lo tanto, hay que considerar hacer más experimentos con los ajustes de procesamiento de este tipo para obtener resultados óptimos.

En el caso de la gran mayoría de las escenas los oyentes opinan que Panorama 5 ofrece mejor calidad del sonido que Binaural Panner. Además indican que el sonido de Binaural Panner suena antinatural y evidentemente procesado de alguna manera.

En la escena *Globos* unos oyentes indican la aparición de flutter eco que causa la coloración del sonido. El motivo de ese fenómeno son los primeros rebotes que han sido añadidos durante el procesamiento. En el futuro, para evitar ese problema habrá que ajustar más precisamente la cantidad de los rebotes.

Además se ha notado en el caso de Binaural Paner en la escena *Dos Personas* que, a pesar del cambio sólo del ángulo, cuando el objeto está en la posición lateral, se le percibe como si estuviera más cerca que en la posición frontal. Es la única escena en el que todos los oyentes prefieren la versión estéreo. Los motivos de la elección así pueden ser diferentes, pero hay que acordarse que *Dos Personas* es el único vídeo con movimiento de cámara, por eso se puede suponer que la simulación de los movimientos de la cabeza causan unas molestias.

## 9 CONCLUSIONES

A pesar de las posibilidades teóricas que ofrece el sonido binaural, siempre quedan muchos elementos a desarrollar. Existen unas herramientas que permiten conseguir unos resultados bastante buenos, pero no eliminan las imperfecciones más importantes del sistema (las confusiones delante-detrás y la localización dentro de la cabeza). Antes de usar el sistema binaural como estándar en lugar del estéreo, hay que considerar el nivel de complicación del sonido en el caso particular. Como muestran los datos de la encuesta, mucha gente no fija la atención en el sonido 3D, es decir observa sólo los movimientos en el eje horizontal y no delante-detrás, y valora más la inteligibilidad. Por eso, a pesar de todas las ventajas que ofrece la solución descrita, antes de decidirse a uno de los sistemas, hay que preguntarse, si el sonido binaural realmente puede proporcionar los valores deseados, manteniendo la calidad del contenido.

Por supuesto el trabajo presente no agota el tema de la postproducción con el uso de la técnica binaural, pero presenta unos problemas que hay que tener en cuenta al decidirse emplear ese sistema en la película. Siempre quedan para investigar cuestiones como el uso de *dummy head*, uso de otros programas de espacialización, grabaciones al campo libre (afuera), el uso de más de un par de micrófonos binaurales, aplicación de *head-tracking* en los programas DAW, grabaciones postsincrónicas con los micrófonos binaurales. En la opinión del autor, la técnica descrita va a ganar en importancia, pero a la vez va a pasar mucho tiempo para que pueda competir con los tradicionales sistemas de altavoces.

## 10 BIBLIOGRAFÍA

Begault, D. R., *3-D Sound For Virtual Reality And Multimedia*, Academi Press, Boston, 2000.

Cengarle, G., *3D Audio Technologies: Applications To Sound Capture, Post-Production And Listener Perception*, Tesi Doctoral UPF, Department Of Information And Communication Technologies, 2012.

Liitola, T., *Headphone Sound Externalization*, Master's Thesis, Helsinki University Of Technology, Tampere, 2006.

Litovsky, R., *Binaural Hearing*.

Nussel, D., 14 de enero de 1978, "Artista Has 1<sup>st</sup> Stereo/Binaural Disk". *Billboard*.

Ruff, R., *Comparison Between Soundman OKM II Studio Classic And Neumann Dummy Head KU81i In Technical And Timbral Aspects*, SAE Institute In Association With University Of Middlesex, Berlin, 2009.

Sang-Myeong, K., Wonjae, C., *On The Externalization Of Virtual Sound Images In Headphone Reproduction: A Wiener Filter Approach*, 2004.

Sęk, A., Skrodzka, E., Marszałkiewicz, M., *Psychoakustyka W Pigułce*, Instytut Akustyki UAM, Poznań, 2000.

Stern, R. M., Wang, D. L., Brown, G., *Binaural Sound Localization*, Chapter In Computational Auditory Scene Analysis, Wiley/IEEE Press, New York, 2006.

Intrerior <interiormovie.com> [Consulta: 15 de junio de 2013]

Mind Theatre <mind-theatre.com> [Consulta: 15 de junio de 2013]