

Dado que el efecto de las distintas respuestas en frecuencia producen efectos perceptivamente claros, se diseñó una prueba según la recomendación ITU-R 1534-2 [6] que describe el test perceptivo conocido como MUSHRA. Este tipo de test se emplea como método para la evaluación subjetiva del nivel de calidad intermedia de sistemas de audio y también describe todos los requisitos necesarios para realizar la prueba con rigor. Utiliza una escala continua de 0 (mala) a 100 (excelente) para evaluar la calidad y otros parámetros subjetivos, empleando siempre un sonido de referencia. En este tipo de test todos los sistemas se comparan con una referencia de máxima calidad, además de todos los sistemas entre sí.

Los participantes de la prueba debían responder dos cuestiones distintas. Su primera tarea fue evaluar la calidad del sonido con respecto a la referencia. Después de un descanso, la segunda tarea consistió en evaluar la imagen espacial (localización, sensación de profundidad, inmersión y realismo del sonido) con respecto a la referencia.

Como fuentes sonoras se usaron cinco fragmentos de audio diferentes (de entre 12 y 14 segundos), todos ellos reproducidos mediante la simulación de los distintos auriculares empleados en el estudio. El primer fragmento de sonido era una pieza musical con voz femenina y guitarra (*Haugebonden* del CD de música *Women en Mi* de Bettina Flater). El segundo fragmento era una canción con voz masculina y guitarra (*Zambra Gitana* del CD *Canción Andaluza* de Paco de Lucía). Como tercer fragmento de audio se emplearon unos segundos de música orquestal con gran variación dinámica (*Night Boarders* de la BSO de *The Mummy* compuesta por Jerry Goldsmith). Los fragmentos cuarto y quinto fueron grabaciones binaurales, una pieza de jazz (*Basin Street Blues*, versión de The Chad Fisher Group) y un sonido ambiente de la audiencia de un concierto con sonidos de batería (*Smashing Pumpkins Live at Northrop Auditorium 2000*). Todos los fragmentos sonoros se eligieron por sus atributos tímbricos, espaciales y estereofónicos.

Para cada uno de los fragmentos de sonido descritos, se realizó una simulación de la escucha de los cinco auriculares bajo estudio referidos en el punto 2.1. Estas cinco simulaciones de los auriculares se presentaron de manera aleatoria, en series por cada fragmento de sonido. Además en la serie se incluía una señal de referencia (REF) y dos señales de ancla. La señal de referencia era el mismo fragmento sonoro sin ningún tratamiento ni simulación añadida. La primera señal de ancla (ANC1) era una versión del fragmento sonoro filtrado con un filtro paso bajo de 7kHz de frecuencia de corte (se corresponde con el ancla de calidad media de la recomendación ITU-R 1534-2). La segunda señal de ancla (ANC2) era una versión monoaural del sonido, escogida para servir de referencia a la pregunta sobre impresión espacial. Según lo anterior, el número de estímulos empleado en este test fue de (5 simulaciones de auriculares + 1 referencia oculta + 2 señales de ancla) x 5 fragmentos sonoros = 40 estímulos, presentados en series de 8 estímulos más su referencia.

Tal y como recomienda la norma ITU, se desarrolló una interfaz gráfica para que fueran los propios participantes los que dirigieran el test de percepción (método doble ciego). De esta manera podían escuchar los diferentes estímulos así como la referencia de manera completamente libre y tantas veces como quisieran. Los distintos sonidos se presentaban en series aleatorias de las distintas simulaciones, anclas y referencia oculta, para compararse con el sonido de referencia. Una vez que el participante había puntuado todos los fragmentos de la serie, se presentaba un nuevo fragmento sonoro con sus simulaciones para que se evaluaran igualmente. Este proceso se repitió dos veces para todos los fragmentos, la primera vez para la pregunta sobre calidad y, después de un descanso, una segunda vez para la cuestión sobre impresión espacial.

11 personas realizaron el test, 7 hombres y 4 mujeres (de entre 21 a 37 años, media de edad de 30). El tiempo medio de ejecución fue de 22 minutos para la primera parte y 16 para la segunda. Cada participante realizó una sesión de entrenamiento antes de realizar el test real, con la intención de que pudieran haber escuchado todos los estímulos y adquirieran familiaridad con la interfaz y las tareas asignadas.

2.3. Resultados

La figura 3 muestra la media de las respuestas a la pregunta sobre calidad percibida (normalizadas de 0 a 100), para la referencia oculta, los cinco auriculares simulados y las dos señales de ancla. Como se puede ver, la referencia ha sido correctamente identificada en la mayoría de los casos. Los tres auriculares de supuesta buena calidad tienen las puntuaciones más altas, mientras que los dos auriculares de supuesta peor calidad tienen las puntuaciones más bajas. Las dos señales de ancla se mantienen en una puntuación intermedia entre estos dos grupos.

Un análisis de varianza confirma que la respuesta en frecuencia de los auriculares tiene una influencia muy significativa ($F=58,33$ $gl=7$ $p<0,001$) sobre la calidad percibida.

En la figura 4 se puede ver la media de las respuestas a la impresión espacial (normalizadas de 0 a 100) para la referencia oculta, los cinco auriculares simulados y las dos señales de ancla. Los resultados tienen bastante similitud con las respuestas sobre calidad, con una correlación muy significativa de valor $r=0,648$. Sin embargo hay algunas diferencias en las puntuaciones que tienen interés. Los tres auriculares de supuesta buena calidad tienen las puntuaciones más altas, al igual que en la cuestión sobre calidad, aunque en este caso los intervalos de confianza no tienen tanta separación entre ellos. Con respecto a la pregunta anterior sobre calidad, hay una mayor diferencia entre los dos auriculares de supuesta peor calidad. Los auriculares de Bajo Coste multimedia (BCmul) se sitúan ahora en el mismo rango que las señales de ancla. Además, llama la atención que la señal de ancla 2 (ANC2) no tenga una puntuación mucho menor dado que es la versión monoaural de la señal.

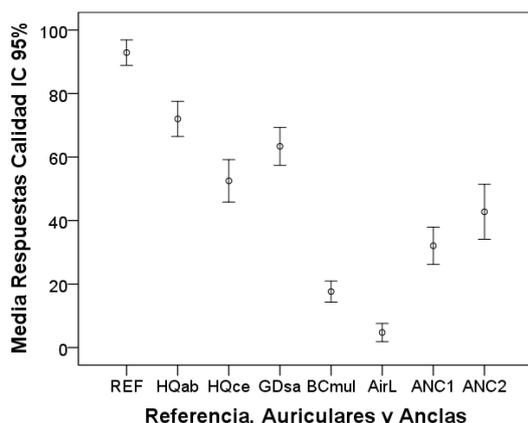


Figura 3. Media de las respuestas de calidad para la referencia oculta, los auriculares simulados y las señales de ancla

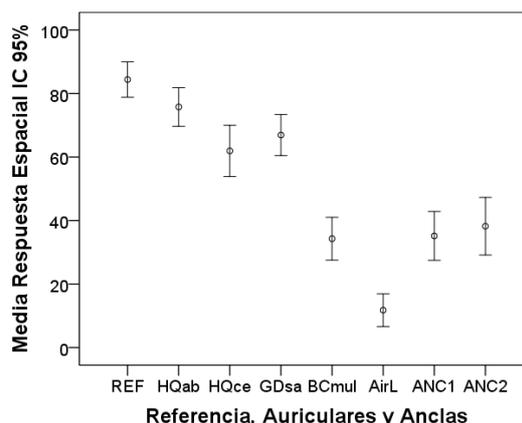


Figura 4. Medida de las respuestas de impresión espacial para la referencia oculta, los auriculares simulados y las señales ancla

En cualquier caso, un análisis ANOVA confirma que la respuesta en frecuencia de los auriculares tiene una influencia muy significativa ($F=58,33$ $gl=7$ $p<0,001$) en la percepción de la impresión espacial. No se ha encontrado ninguna influencia significativa por parte del tipo de sonido, incluso a pesar de que algunos de los fragmentos sonoros eran grabaciones binaurales.

3. TEST DISTORSIÓN NO LINEAL

3.1. Método de simulación

El objetivo de este test es medir los efectos de la distorsión armónica de los auriculares sobre la impresión espacial sonora. Es necesario establecer previamente un método de medida de la distorsión de los distintos auriculares y su simulación por medio de los auriculares de referencia.

Para medir la distorsión armónica se empleó un método de barrido sinusoidal [5] que nos dio la respuesta en frecuencia así como los armónicos segundo y tercero de la respuesta en frecuencia de manera simultánea.

En esta prueba se emplearon los mismos cinco auriculares del test anterior. En la figura 5 se muestra la respuesta en frecuencia y los órdenes de distorsión segundo y tercero de los auriculares de Aerolínea (AirL), que son los que presentaban mayor distorsión.

Para simular la distorsión no lineal escogimos el método descrito en [7] que emplea núcleos de Volterra y series de convoluciones lineales. Con este método la función de transferencia de un sistema queda descrita por medio de una expansión de series de Volterra. La señal de salida se puede representar como la suma de las convoluciones lineales de las respuestas al impulso medidas con la señal de entrada y sus correspondientes versiones desplazadas en frecuencia. Al aplicar la transformada de Fourier a estas series se obtiene un sistema lineal de ecuaciones. La solución de este sistema permite el cálculo de los núcleos diagonales de Volterra, obteniendo así los términos de respuesta al impulso de la respuesta principal y de los dos primeros órdenes de distorsión, (Ecuación 2).

$$\begin{cases} H_1 = H'_1 + H'_3 \\ H_2 = -2\hat{H}'_2 \\ H_3 = -4H'_3 \end{cases} \quad (2)$$

Donde H'_1, H'_2, H'_3 son las respuestas en frecuencia medidas y H_1, H_2, H_3 son los núcleos de Volterra ($\hat{}$ representa la transformada de Hilbert).

Usando estas ecuaciones se pueden simular los órdenes de distorsión segundo y tercero mediante la convolución. Se deben aplicar los núcleos en la siguiente ecuación (Ecuación 3), donde $x(n)$ es la señal de entrada y M es el número de muestras del núcleo:

$$y(n) = \sum_{i=0}^{M-1} h_1(i) \cdot x(n-i) + \sum_{i=0}^{M-1} h_2(i) \cdot x^2(n-i) + \sum_{i=0}^{M-1} h_3(i) \cdot x^3(n-i) \quad (3)$$

Pueden verse más detalles de esta técnica en la referencia [7]

3.2. Descripción del test

Para evaluar la percepción de la distorsión en este test se emplearon estímulos con y sin distorsión. El efecto de las distorsiones empleadas es bastante sutil, por ese motivo esta prueba se diseñó según la recomendación ITU-R 1116-2 [8] que describe un método para evaluar de manera subjetiva pequeñas degradaciones en los sistemas de audio, así como estrictos requisitos que deben cumplir la sala y el equipamiento con el que se realizará la prueba. Según esta recomendación, se debe emplear una escala del 1-muy molesto al 5-imperceptible para evaluar las degradaciones detectadas con respecto a la señal de referencia. El tipo de test propuesto se denomina ABC, donde dos estímulos (A y B) son presentados cada vez al participante para que sean comparados con una referencia dada (C). De entre esos dos estímulos presentados uno de ellos A o B, será siempre una referencia oculta (es decir, igual a C) y el otro la señal degradada.

Se emplearon los mismos cinco fragmentos de audio descritos en el test anterior, así como los mismos cinco auriculares descritos en el punto 2.1. En este caso, se simuló tanto la respuesta en frecuencia como las distorsiones no lineales. No se añadieron señales de ancla adicionales. Los auriculares simulados así como sus distorsiones fueron reproducidos mediante los auriculares de referencia (REF - Sennheiser HD800).

Durante el test se simularon dos versiones de cada auricular, una con y otra sin distorsión. La mecánica de la prueba consistía en evaluar de manera ciega la versión distorsionada y la no distorsionada del mismo estímulo sonoro y compararla con la versión no distorsionada que servía como referencia conocida.

En este caso a los participantes se les pidió responder a una sola cuestión: “¿Qué degradación de calidad y sensación espacial percibes con respecto a la referencia?”.

El número de estímulos empleado fue: 5 auriculares simulados x 2 versiones (con y sin distorsión) x 5 fragmentos sonoros = 50 estímulos, que fueron presentados en veinticinco series de dos estímulos con su referencia correspondiente. Todas estas parejas de estímulos se mostraban aleatorizadas para cada participante, así como el orden interno de muestras con o sin distorsión (A o B).

Para llevar a cabo la prueba se desarrolló una interfaz gráfica, según la describe la recomendación, de manera que permitía a los participantes escuchar cada sonido a evaluar tantas veces como quisieran.

Las mismas 11 personas que realizaron el test anterior hicieron también éste, 7 hombres y 4 mujeres (de entre 21 a 37 años, media de edad de 30). El tiempo medio de ejecución fue de 16 minutos. Cada participante realizó una sesión de entrenamiento antes de efectuar el test real, con la intención de que pudieran haber escuchado todos los estímulos y adquirieran familiaridad con la interfaz y las tareas asignadas.

3.3. Resultados

Para llevar a cabo la evaluación de los resultados obtenidos, la recomendación exige calcular la diferencia entre las puntuaciones obtenidas por la referencia oculta y la puntuación obtenida por la señal distorsionada. La figura 6 muestra la media de estas diferencias para cada uno de los auriculares simulados.

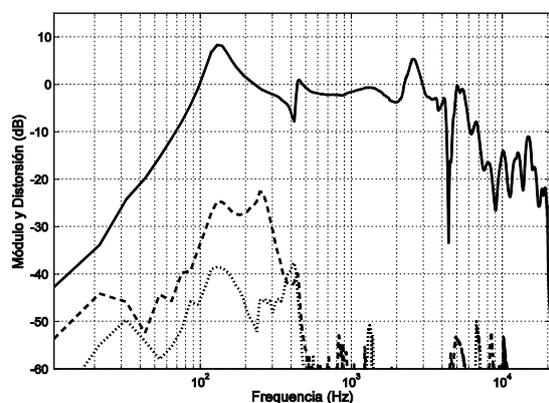


Figura 5. Respuesta en frecuencia, segundo y tercer órdenes de distorsión de los auriculares de Aerolínea (AirL).

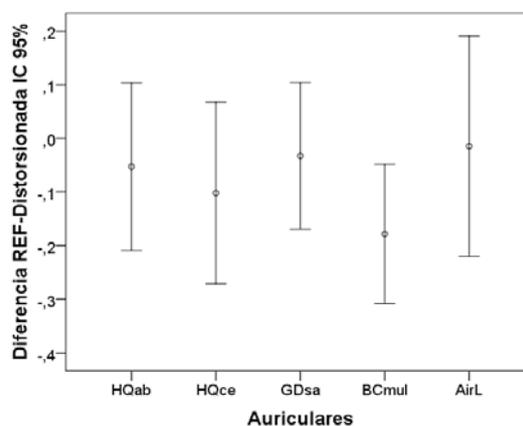


Figura 6. Diferencia entre la referencia oculta y la señal distorsionada, para cada uno de los auriculares.

Tal como se desprende de la gráfica anterior, un análisis de varianza confirmó que no hay ninguna relación de significancia en las respuestas de los participantes debido a los efectos que la distorsión introducida pueda generar.

4. CONCLUSIONES

En este trabajo se ha realizado un estudio sobre la influencia de diferentes parámetros de calidad de auriculares, en el contexto de la reproducción de sonido espacial. En un trabajo previo [4] se analizaron los efectos de la disparidad de nivel entre los transductores y en el

presente escrito se han estudiado los efectos de la respuesta en frecuencia y de la distorsión no lineal mediante pruebas de evaluación subjetiva.

Se ha comprobado que la respuesta en frecuencia tiene una influencia muy importante en la impresión espacial. También se ha visto que la impresión espacial tiene una correlación muy alta con respecto a la calidad percibida. Sin embargo, no se ha encontrado ninguna influencia significativa en el hecho de usar grabaciones binaurales con respecto a mezclas estereofónicas de dos canales, lo que puede deberse quizá a que todas las muestras sonoras fueron seleccionadas con una buena sensación estereofónica. Queda pendiente para un trabajo futuro el intentar cuantificar los posibles efectos de la desviación de la respuesta en frecuencia de la ideal.

Por otro lado, la distorsión no lineal que introducen los auriculares bajo estudio no parece afectar a la percepción espacial o a la imagen binaural de las grabaciones, al menos al nivel de volumen y distorsión empleados en el test.

Como trabajo futuro también sería interesante medir parámetros objetivos de los auriculares, así como de los sonidos empleados como estímulos, para estudiar más relaciones estadísticas y buscar posibles influencias de otros factores. Además nos gustaría incluir en el estudio un mayor número de auriculares, especialmente más cantidad de baja calidad, y evaluar un mayor número de grabaciones binaurales.

5. AGRADECIMIENTOS

Financiado por el Ministerio Español de Economía y Competitividad, proyecto TEC2012-37945-C02-01.

6. REFERENCIAS

- [1] J. Blauert, "Spatial Hearing: The psychophysics of human sound localization". Cambridge, MA: MIT Press, 1996.
- [2] D. R. Begault, "3D Sound for Virtual Reality and Multimedia Applications", Academic Press Professional Inc., San Diego (California), 1994.
- [3] S. Bech and N. Zacharov, "Perceptual Audio Evaluation-Theory, Method and Application", John Wiley & Sons Ltd., Sussex (UK), 2006.
- [4] J. López, P. Gutiérrez-Parera, A. Aguilera, "Influencia de la calidad de los auriculares en la inmersión acústica producida por grabaciones binaurales", 45º Congreso Español de Acústica y 8º Congreso Ibérico de Acústica, SEA-Tecniacústica 2014, Murcia, 29-31 Octubre 2014.
- [5] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique", 108th AES Convention, Paris 18-22 February 2000.
- [6] Rec. ITU-R BS. 1534-2 "Method for the subjective assessment of intermediate quality level of audio systems", International Union of Telecommunications, Geneva, 2014.
- [7] A. Farina, E. Armelloni, "Emulation of Non-Linear, Time-Variant devices by the convolution technique", Congresso AES Italy 2005, Como, 3-5 November 2005.
- [8] Rec. ITU-R BS. 1116-2 "Methods for the subjective assessment of small impairments in audio systems", International Union of Telecommunications, Geneva, 2014.