

## INFLUENCIA DE LA CALIDAD DE LOS AURICULARES EN LA INMERSION ACUSTICA PRODUCIDA POR GRABACIONES BINAURALES

REFERENCIA PACS: 43.66.Pn, 43.66.Qp, 43.60.Dh

Lopez, José Javier; Gutiérrez-Parera, Pablo; Aguilera, Emanuel

Universidad Politécnica de Valencia  
ITEAM, Instituto de Telecomunicaciones y Aplicaciones Multimedia  
Camino de Vera s/n, Edificio 8G, acceso D  
46022 Valencia (España)  
Tel: +34 963 877 007 (Ext. 73008) Fax: +34 963 879 583  
[jjlopez@ocom.upv.es](mailto:jjlopez@ocom.upv.es), [pabgupa@iteam.upv.es](mailto:pabgupa@iteam.upv.es), [emagmar@iteam.upv.es](mailto:emagmar@iteam.upv.es)

### ABSTRACT

The binaural sound recordings obtained using an acoustic manikin produce a very realistic sound immersion when played through high quality headphones. However, most people commonly use headphones of inferior quality and in many cases the ones provided accompanying devices such as smartphones or music players. Factors such as frequency response, distortion, and the disparity between the left and right transducer could be some of the degrading factors. This preliminary study focuses on the analysis of how the disparity in levels between the two transducers affects the final result by performing a subjective test with a jury of 20 people. The results confirm that from differences of 1 dB it appears degradation in azimuth location of sound sources and it increases significantly at higher levels.

### RESUMEN

El sonido binaural obtenido mediante grabaciones con un maniquí acústico produce una sensación de inmersión acústica muy realista cuando se reproduce a través de unos auriculares de alta calidad. Sin embargo, la mayor parte de usuarios utilizan auriculares de inferior calidad y en muchos casos los que acompañan de serie a dispositivos móviles como teléfonos o reproductores de música. Factores como la respuesta en frecuencia, la distorsión y la disparidad entre el transductor izquierdo y derecho podrían ser algunos de los factores degradantes. Este estudio preliminar se ha centrado en el análisis de cómo la disparidad de niveles entre los dos transductores afecta al resultado final realizando un estudio subjetivo con jurado de 20 personas. Los resultados confirman que a partir de diferencias de 1 dB se produce degradación en la localización en azimut de las fuentes sonoras y que aumenta sensiblemente a niveles superiores.

## 1. INTRODUCCIÓN

Con la llegada de la alta definición, el 3D y las tecnologías móviles, el audio espacial ha ganado gran popularidad en los últimos años. Las configuraciones de altavoces han evolucionado desde el clásico estéreo hacia los sistemas de muchos canales que no sólo consideran configuraciones 2D, sino también altavoces elevados. La variedad de formatos (desde auriculares a sistemas 22.2) y de técnicas de reproducción (VBAP, WFS, Ambisonics...) abre muchas posibilidades para la recreación de ambientes acústicos y, especialmente, la creación de nuevas experiencias musicales.

Los sistemas basados en altavoces son los más populares, pero los sistemas binaurales basados en auriculares están ganando cada vez mayor popularidad tanto por la privacidad de escucha que proporcionan en cualquier tipo de ambiente como por la proliferación de dispositivos móviles de hoy en día. La reproducción de sonido binaural mediante auriculares utiliza los principios del sistema de audición humano basado en dos oídos. Asumiendo el hecho de que si somos capaces de reproducir en los oídos del oyente mediante auriculares las mismas presiones sonoras que dicho oyente experimentaría en la realidad podemos simular una inmersión acústica realista.

Para que la sensación de inmersión espacial sea correcta, se deben de utilizar tanto micrófonos de alta calidad para la grabación con maniqués acústicos como auriculares de calidad para la reproducción. Sin embargo los auriculares de baja gama son utilizados en la mayoría de los casos por parte de usuarios, bien por cuestiones de coste o bien porque en muchos casos se utilizan por comodidad los que acompañan de serie a dispositivos móviles como teléfonos o reproductores de música, que en raras ocasiones son óptimos.

En general se sabe que un auricular de bajo coste suele proporcionar una sensación de inmersión más deficiente, pero no se ha estudiado suficientemente que factores degradantes son exactamente los que provocan dicha merma de calidad y a qué nivel afectan cada uno.

Podemos suponer que factores como la respuesta en frecuencia, la distorsión y la disparidad entre el transductor izquierdo y derecho podrían ser algunos de los factores degradantes comentados.

En este estudio preliminar pretendemos adentrarnos en este tema investigando uno de los factores más evidentes como es la disparidad de niveles entre los dos transductores (derecho e izquierdo) y como afecta al resultado final. Para ello se ha realizado un estudio subjetivo por medio de un jurado de 20 personas que han valorado cuatro señales diferentes situadas en diferentes posiciones azimutales y con diferentes niveles de error de niveles entre los dos transductores.

Los resultados de este trabajo confirman que a partir de diferencias de 1 dB entre transductores empieza a notarse una degradación en la localización en azimut de las fuentes sonoras. Además conforme aumentamos esta diferencia a niveles superiores el error de localización va aumentando progresivamente,

El resto del artículo está organizado de la siguiente manera: el apartado 2 realiza un repaso de los fundamentos del sonido binaural, la sección 3 describe el experimento subjetivo y sus objetivos. El apartado 4 expone los resultados así como su análisis y discusión. Por último la sección 5 presenta las conclusiones de este trabajo.

## 2. FUNDAMENTOS DEL SONIDO BINAURAL

Los sistemas de sonido binaural tratan el problema de la reproducción espacial de sonido teniendo como referencia al oyente. Aunque la escena sonora sea muy compleja el oyente únicamente va a analizar la misma atendiendo a lo que llega a sus dos oídos.

Según la teoría dúplex [2] existen dos factores que influyen en la localización de sonidos en el plano horizontal, también denominada localización en acimut, en relación con la percepción binaural: las Diferencias de Tiempo Interaural ITD (*Interaural Time Difference*) y las Diferencias de Intensidad Interaural IID (*Interaural Intensity Difference*).

Las ITD se explican teniendo en cuenta la diferencia de tiempo que tarda en llegar un mismo frente de onda a los dos oídos. Consideremos una onda de sonido proveniente de una fuente distante que golpea una cabeza esférica de radio  $r$  desde una dirección que viene dada por el ángulo de llegada  $\theta$  en el plano horizontal (acimut). Claramente el sonido llega antes a un oído que al otro, ya que tiene que atravesar una distancia extra para alcanzar el oído más lejano. Dividiendo esta distancia por la velocidad del sonido, obtendremos la siguiente ecuación para la diferencia de tiempo interaural:

$$ITD = \frac{r}{c}(\theta + \text{sen}\theta)$$

Por tanto, la ITD es cero cuando la fuente se encuentra justo enfrente del oyente, y es máxima cuando la fuente está situada justo en un lado. Esto representa una diferencia de tiempo de llegada de aproximadamente 0,7 ms para una cabeza humana típica y es fácilmente percibido.

La ecuación anterior puede invertirse para obtener el acimut partiendo del valor de las ITD. El sistema de audición humana realiza una operación equivalente a ésta para obtener el acimut de la señal percibida a partir de la información de ITD.

Las IID se producen por la diferencia de intensidad con que llega a cada oído un mismo sonido. Como cabría esperar, las IID son muy dependientes de la frecuencia. A bajas frecuencias, donde la longitud de onda del sonido es grande comparada con el diámetro de la cabeza, no existe apenas diferencia en la presión del sonido en los oídos. Sin embargo, a altas frecuencias, donde la longitud de onda es pequeña, puede haber fácilmente una diferencia de 20 dB o más. A esta característica se la denomina *efecto de sombra de la cabeza* y el oído más alejado es el que se encontraría en la zona de sombra.

La teoría dúplex afirma que las IID y las ITD son complementarias. A bajas frecuencias (por debajo de 1,5 kHz), existe muy poca información de IID, sin embargo las ITD son detectadas fácilmente por los cambios de fase en fracciones inferiores a un ciclo. A mayores frecuencias (por encima de 1,5 kHz), existe una ambigüedad en las ITD, dado que pueden existir varios ciclos de desfase, pero las IID resuelven esta ambigüedad en la dirección. La combinación de los dos mecanismos, ITD e IID, proporciona información de localización en todo el rango de frecuencias audible. Este umbral de frecuencia de 1,5 kHz se debe a que su longitud de onda se corresponde con el diámetro promedio de una cabeza humana. Este valor, por tanto, no es muy preciso, pues determinados estudios sugieren que más que un valor debería contemplarse un rango de frecuencias comprendido entre los 1,5 kHz y los 3kHz [3].

Así, las ITD son capaces de dominar a las IID siempre que el contenido en bajas frecuencias sea suficiente, como por ejemplo en señales de banda ancha [4]. Por otro lado las IID son significativas en ausencia de bajas frecuencias, especialmente para señales de banda estrecha [5]. Esta complementariedad funcional se suele denominar *time-intensity trading* (negociación tiempo-intensidad).

Mediante una grabación en la que se coloquen dos micrófonos en los oídos de un oyente mientras se reproduce la escena, o sustituyendo a la persona por un maniquí acústico, se

consiguen registrar las ITD e IID y realizar por tanto una grabación binaural. Posteriormente dicha grabación podrá ser reproducida y escuchada directamente mediante auriculares, consiguiendo reproducir las sensaciones de localización espacial.

### 3. DESCRIPCIÓN DEL EXPERIMENTO

Tal como se comentó en la introducción, se llevó a cabo un test de percepción subjetiva [1] para evaluar cómo influyen en la localización horizontal pequeñas variaciones de nivel aplicadas sobre una serie de sonidos binaurales.

Durante el test los participantes escuchaban mediante auriculares unos sonidos binaurales grabados en unos ángulos específicos con respecto al oyente. A los sonidos presentados a los participantes se les aplicaba una ligera modificación del nivel original entre el canal izquierdo y el derecho. Los participantes debían entonces indicar el ángulo de procedencia del que les parecía que provenía el sonido.

Las variaciones de nivel aplicadas eran de 0dB (sin modificación), 1dB, 2dB o 4dB más en el canal izquierdo que en el derecho.

Se escogieron cuatro ángulos de procedencia de los sonidos,  $-30^\circ$ ,  $0^\circ$ ,  $65^\circ$  y  $90^\circ$  de acimut en el plano horizontal.

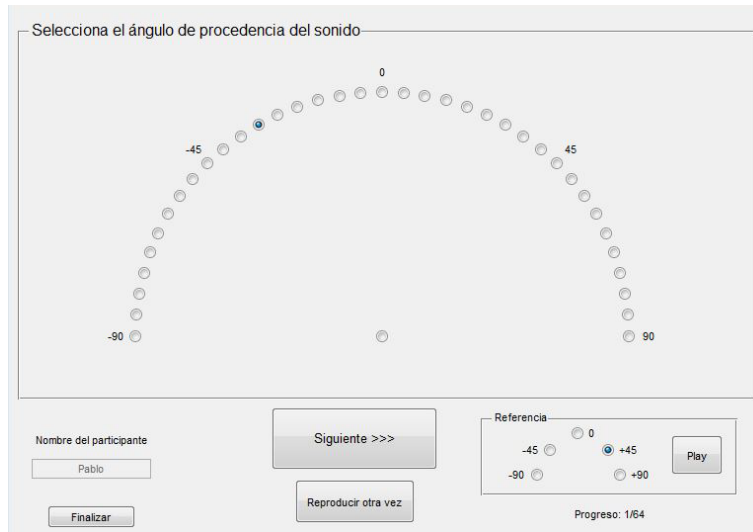
Además, se estudió la influencia de distintos tipos de sonido. Se realizó una grabación binaural de varios sonidos en los ángulos específicos bajo estudio, empleando finalmente cuatro tipos: un golpe de timbal, voz, un silbido y ruido rosa. El carácter impulsivo del sonido del timbal es interesante para su uso en localización sonora, así como su contenido en bajas frecuencias. Tanto la voz como el silbido son sonidos comunes fácilmente reconocibles, lo que los hace particularmente útiles en el test. También es interesante el reducido contenido espectral del silbido. Para evaluar una señal de amplio espectro se utilizó como estímulo ruido rosa. Todos los sonidos fueron grabados en los ángulos correspondientes a un metro de distancia, empleando un maniquí B&K Type 4100 y posteriormente reproducidos mediante unos auriculares Sennheiser HD 545.

Teniendo en cuenta las distintas características de los sonidos descritas antes, el número total de estímulos a presentar a cada participante en el test fue de: 4 ángulos x 4 tipos sonido x 4 variaciones de nivel = 64 estímulos. Dichos estímulos se presentaron aleatoriamente en el test, con la posibilidad para el participante de escuchar el estímulo a evaluar todas las veces que deseara.

Paralelamente durante la realización del test el participante también tenía la posibilidad de escuchar un estímulo de referencia en cualquier momento, a elegir entre los ángulos  $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $45^\circ$  y  $90^\circ$  de acimut.

Para realizar esta prueba se diseñó una interfaz gráfica que controlaba el propio participante (figura 1) en la que podía evaluar el ángulo de procedencia aparente de cada sonido en un arco de  $-90^\circ$  a  $90^\circ$  de acimut, con una resolución de  $5^\circ$ . También podía controlar y escuchar libremente el estímulo de referencia.

El test fue realizado por 20 personas, 10 hombres y 10 mujeres de entre 21 y 45 años, con una edad media de 32. El tiempo medio de ejecución del test fue de 9 minutos. Todos los participantes realizaron una pequeña prueba de entrenamiento antes de iniciar el test en la que podían escuchar los distintos tipos de sonidos que se les presentarían, así como familiarizarse con la tarea asignada y la interfaz.

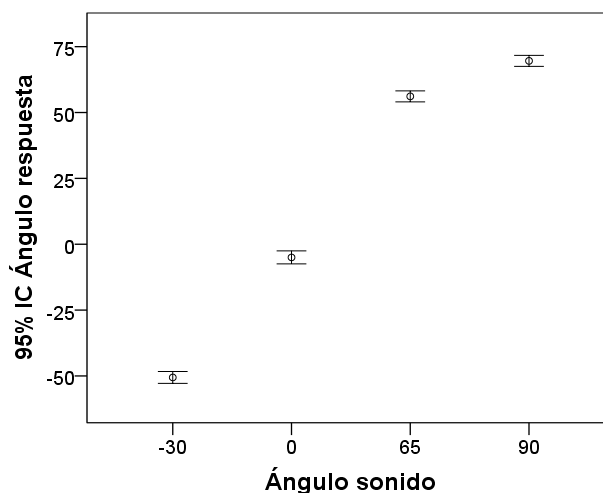


**Figura 1:** Interfaz gráfica para la realización del test subjetivo.

## 4. RESULTADOS Y DISCUSIÓN

### 4.1. Resultados del primer experimento

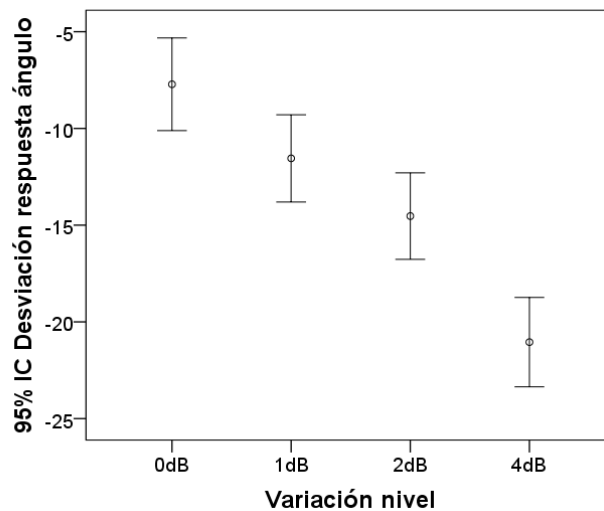
En la figura 2 se representan las medias de las respuestas de los ángulos (para todos los casos de variación de nivel) en función de los ángulos reproducidos. Como se puede ver la media de las respuestas tiene una desviación hacia la izquierda. Esto es lo esperado dado que las variaciones (0dB, 1dB, 2dB, 4dB) eran siempre de más nivel en el canal izquierdo que en el derecho.



**Figura 2:** Media de los ángulos respuesta en función de los ángulos reproducidos (en grados).

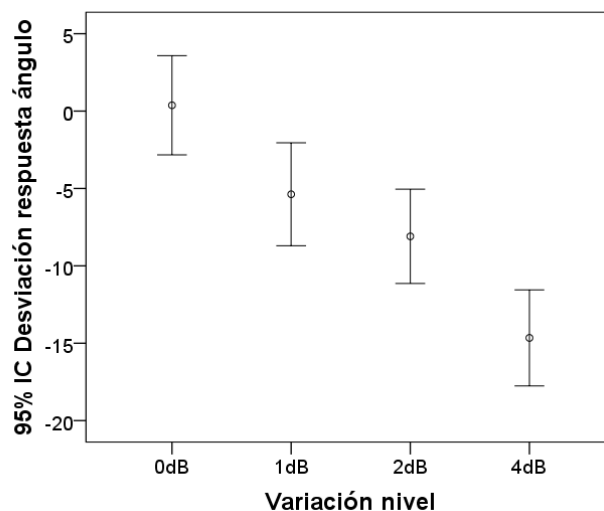
La tendencia de esta desviación de grados hacia la izquierda se puede observar en la figura 3, teniendo en cuenta la variación de nivel aplicada (0dB, 1dB, 2dB, 4dB).

Un análisis de varianza (ANOVA) indica que la variación de nivel tiene una influencia muy significativa ( $F=27.338$ ,  $gl=3$ ,  $p<0.001$ ) sobre la desviación en las respuestas.



**Figura 3:** Media de la desviación de los ángulos respuesta (en grados) según la variación de nivel (en dB).

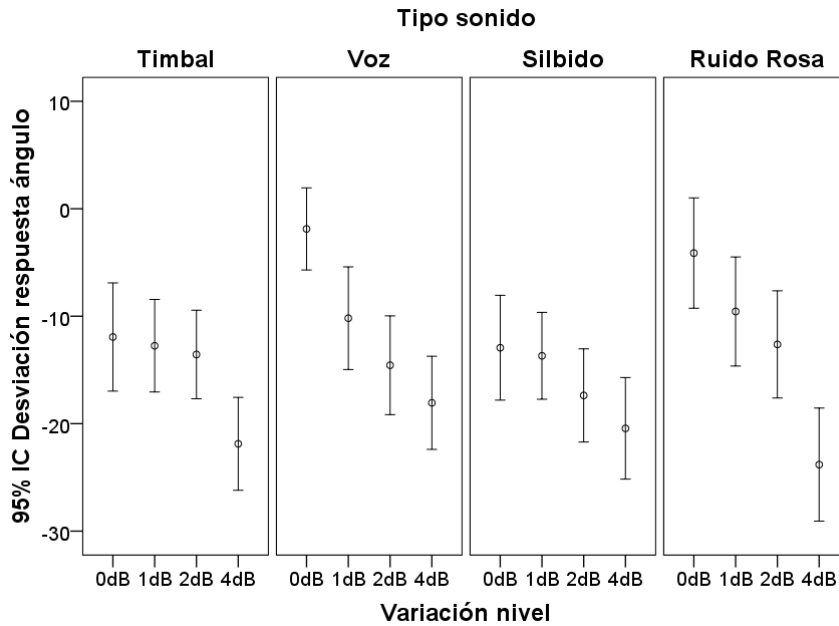
Teniendo en cuenta sólo los ángulos centrales usados en el experimento ( $0^\circ$  y  $65^\circ$ ), se puede apreciar una media de desviación más reducida (figura 4). Esto nos lleva a pensar que los oyentes tendían a desviar más la localización de los sonidos percibidos en los laterales, es decir, las variaciones de nivel introducidas les hacían dispersar más las posiciones de los ángulos laterales que la de los ángulos centrales.



**Figura 4:** Media de la desviación de los ángulos respuesta (en grados) según la variación de nivel (en dB), teniendo en cuenta sólo los ángulos  $0^\circ$  y  $65^\circ$ .

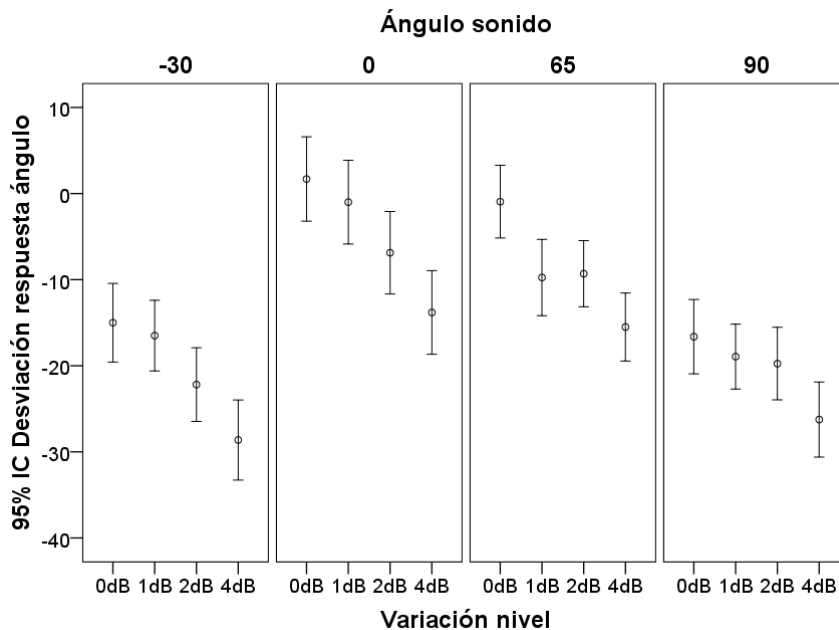
Por otro lado, la influencia del tipo de sonido (timbal, voz, silbido o ruido rosa) sobre la desviación en las respuestas se puede observar en la figura 5. Tanto la voz como el ruido rosa manifiestan una desviación menor que los sonidos del timbal y el silbido, especialmente en los casos de 0 y 1dB de variación. Además los estímulos de voz y ruido rosa manifiestan una desviación algo más separada y clara según la variación de nivel aplicada.

La influencia del tipo de sonido sobre la desviación en las respuestas es significativa ( $F=4.409$ ,  $gl=3$ ,  $p=0.004$ ) según un análisis de varianza.



**Figura 5:** Media de la desviación de los ángulos respuesta (en grados) en función de la variación de nivel (en dB), teniendo en cuenta el tipo de sonido.

El ángulo de reproducción del sonido tiene una influencia muy significativa ( $F=54.932$ ,  $gl=3$ ,  $p<0.001$ ) sobre la desviación de las respuestas. En la figura 6 están representadas la desviación de las respuestas para cada ángulo de reproducción de los sonidos. Los ángulos  $0^\circ$  y  $65^\circ$  son los que presentan una desviación menor hacia la izquierda. La mayor desviación de las respuestas para el ángulo  $-30^\circ$  puede ser debida a que era el único ángulo presente en el sector izquierdo de escucha (entre  $-90^\circ$  y  $0^\circ$ ) y esto pudo producir una dispersión hacia la izquierda más pronunciada que en los otros casos.



**Figura 6:** Media de la desviación de los ángulos respuesta (en grados) en función de la variación de nivel (en dB), teniendo en cuenta el ángulo de reproducción.



## 5. CONCLUSIÓN Y TRABAJO FUTURO

En este trabajo se ha realizado un estudio de la influencia de la disparidad de niveles entre los dos transductores de un auricular en la precisión de localización espacial de fuentes. Para ello se ha realizado un estudio subjetivo por medio de un jurado de 20 personas que han valorado cuatro señales diferentes situadas en diferentes posiciones azimutales y con diferentes niveles de error entre los dos transductores.

Los resultados de este trabajo confirman que a partir de diferencias de 1 dB entre transductores empieza a notarse una degradación en la localización en azimut de las fuentes sonoras. Además conforme aumentamos esta diferencia a niveles superiores el error de localización va aumentando progresivamente.

También se ha podido comprobar que el tipo de sonido tiene una fuerte influencia en la localización, siendo las muestras de ruido rosa y voz las que presentan una menor desviación en su localización. La elección de los ángulos para la realización del test también es un factor determinante, presentando una mayor desviación los ángulos más lateralizados. Un aumento del número de posiciones angulares bajo estudio puede ser de interés en posteriores trabajos.

Como trabajo futuro queda pendiente el análisis de otros factores degradantes que se presupone que también pueden afectar a la calidad de la localización y al efecto de inmersión final que percibe el oyente. Entre estos factores estarían la respuesta en frecuencia general del transductor y la distorsión que produce.

## AGRADECIMIENTOS

Financiado por el Ministerio Español de Ciencia e Innovación, proyecto TEC2012-37945-C01.

## REFERENCIAS

- [1] S. Bech and N. Zacharov, "Perceptual Audio Evaluation-Theory, Method and Application", John Wiley & Sons Ltd., Sussex (UK), 2006
- [2] J. W. S. Rayleigh. "Our perception of sound direction". *Philosophical Magazine*, vol13, pp. 214-232, 1907.
- [3] J. Blauert, "Spatial Hearing: The psychophysics of human sound localization". Cambridge, MA: MIT Press, 1996.
- [4] F. Wightman & D. Kistler, "The dominant role of low-frequency interaural time differences in sound localization". *Journal of the Acoustical Society of America (JASA)*, 19 (3), 1648-1661, 1992
- [5] J. Middlebrooks & D. Green, "Directional dependence of interaural envelope delays". *Journal of the Acoustical Society of America (JASA)*, 87, 2149-2162, 1989.