

FIA 2020/22

XII CONGRESSO/CONGRESO IBEROAMERICANO DE ACÚSTICA
XXIX ENCONTRO DA SOCIEDADE BRASILEIRA DE ACÚSTICA - SOBRAC

Florianópolis, SC, Brasil

Comparison of perceived auditory distance between real and virtual sound sources

Lombera, E. N.¹; Manuel Viskovic, M.²; Cerviño, J.¹ and Vergara, R.¹

¹ Laboratorio de Acústica y Percepción Sonora (LAPSo), Universidad Nacional de Quilmes, CONICET, R. S. Peña 352, Bernal, Buenos Aires, B1876BXD, Argentina. ramiro.vergara@unq.edu.ar

² Universidad Nacional de Tres de Febrero (UNTREF). Departamento de Ciencia y Tecnología. CONICET. Sáenz Peña, Argentina.

Abstract

Virtual acoustic technology has been commonly used in psychoacoustic experiments as a means of realistically simulating the acoustic environment. Of course, beyond the advantages of virtual technology, a researcher will consider it essential that virtual acoustic technology manages to generate realistic environments. However, comparative studies on the perception of auditory distance between real and virtual environments are scarce. In this study we compared the perceived distance of real sources and three different types of virtual sources: individualized HRTF, non-individualized HRTF, and non-individualized simulated. In all cases, the stimulus was a musical fragment played from sources located in front of the listener at the following distances: 1, 2, 3, 4, 5 and 6 meters. The stimuli were played through open headphones in the same room where the HRTFs were recorded. Our results showed that the responses obtained with the simulated room were significantly more biased than those obtained with real sources. In contrast, the responses obtained with both individualized and non-individualized HRTFs were indistinguishable from those obtained in the real setting. Our main findings were: 1- When played through open-back headphones and in the same room where they were recorded, the HRTF recording technique robustly simulates the distance from the sound source. 2- Under these conditions, the use of non-individualized HRTFs was sufficient to robustly simulate the distance from the sound source.

Keywords: auditory distance perception, virtual reality, auditory context.

PACS: look at the instructions inside this template.

Comparación de la distancia auditiva percibida entre fuentes de sonido reales y virtuales

Resumen

La tecnología acústica virtual se ha utilizado habitualmente en experimentos psicoacústicos como un medio para simular de forma realista el entorno acústico. Por supuesto, más allá de las ventajas de la tecnología virtual, un investigador considerará fundamental que la tecnología acústica virtual consiga generar entornos realistas. Sin embargo, los estudios comparativos sobre la percepción de la distancia auditiva entre entornos reales y virtuales son escasos. En este estudio comparamos la distancia percibida de fuentes reales y tres tipos diferentes de fuentes virtuales: HRTF individualizadas, HRTF no individualizadas y simuladas no individualizadas. En todos los casos, el estímulo fue un fragmento musical reproducido desde fuentes ubicadas frente al oyente a las siguientes distancias: 1, 2, 3, 4, 5 y 6 metros. Los estímulos se reprodujeron a través de auriculares abiertos en la misma habitación donde se grabaron los HRTFs. Nuestros resultados mostraron que las respuestas obtenidas con la sala simulada estaban significativamente más sesgadas que las obtenidas con fuentes reales. Por el contrario, las respuestas obtenidas con HRTFs tanto individualizadas como no individualizadas fueron indistinguibles de las obtenidas en el entorno real. Nuestros principales hallazgos fueron: 1- Cuando se reproducen a través de auriculares abiertos y en la misma habitación donde se grabaron, la técnica de grabación HRTF simula de forma robusta la distancia desde la fuente de sonido. 2- En estas condiciones, el uso de HRTF no individualizados fue suficiente para simular de forma robusta la distancia desde la fuente de sonido.

Palabras claves: percepción auditiva de distancia, realidad virtual, contexto auditivo.



1. INTRODUCTION

El estudio de la percepción auditiva a distancia (PAD) se ha desarrollado a través de diferentes metodologías experimentales, en diferentes entornos acústicos (tanto reales como virtuales) y utilizando estímulos de diversas características. La PAD es un fenómeno complejo que involucra una gran cantidad de señales diferentes. La intensidad se ha considerado como la clave principal de la PAD y, como consecuencia, se ha investigado con mayor frecuencia. Otro indicio importante es la relación entre la energía sonora directa y la reverberante. Por otro lado, hay algunas señales no acústicas, como la visión y la familiaridad con el sonido, que también influyen en la PAD (Para una revisión reciente, ver [1]). Al planificar un experimento de PAD en entornos reales, un investigador debe enfrentarse a varios desafíos metodológicos como, por ejemplo, la necesidad de espacios amplios, la necesidad de configuraciones para mover las fuentes de sonido de forma silenciosa, la dificultad de aislar los múltiples indicios manteniendo la naturalidad de los estímulos sonoros, entre otros. Una forma de evitar estos inconvenientes es mediante el uso de tecnología acústica virtual. Para generar un entorno acústico virtual realista, es necesario simular correctamente una amplia variedad de señales auditivas tanto de fuentes sonoras como de su relación con el entorno circundante. La mayoría de las señales auditivas espaciales se producen por cambios en el sonido original provocados por el cuerpo del oyente antes de que llegue al tímpano [2]. De esta forma, la percepción del entorno auditivo depende de la fisonomía del oyente. Por ejemplo, tanto el tiempo interaural como las diferencias de nivel (las principales señales de azimut) dependen del tamaño de la cabeza del oyente. En este orden, la interacción del sonido incidente con la cabeza, los oídos y el torso de un oyente genera cambios en el espectro sonoro (señales espectrales) que utiliza el sistema auditivo para determinar la elevación de la fuente como también si se encuentra delante o detrás de la cabeza. El filtrado que induce el cuerpo del oyente sobre la onda incidente se describe mediante funciones de transferencia relacionadas con la

cabeza o HRTF [3]. Los HRTF dependen de la morfología del oyente, por lo que son individualizados [4].

La tecnología acústica virtual se ha utilizado de forma rutinaria en experimentos psicoacústicos como un medio para simular de forma realista el entorno acústico [ver por ejemplo 5]. Por supuesto, más allá de las ventajas técnicas de la tecnología virtual, un investigador considerará fundamental que la tecnología acústica virtual consiga generar estímulos virtuales indistinguibles de los reales. Es por esto que es importante abordar estudios en los que se comparen estímulos acústicos virtuales con sus contrapartes reales. Sin embargo, los estudios comparativos entre la distancia auditiva simulada y su equivalente real son escasos, si no inexistentes. El objetivo de este estudio fue comparar la distancia percibida entre fuentes reales y fuentes virtuales obtenidas con tres técnicas diferentes (HRTF individualizadas, HRTF no individualizadas y fuentes simuladas no individualizadas).

2. METHODS

2.1 Participants

En total participaron 24 personas (17 hombres y 7 mujeres), con una edad media en años de 29.1 ± 2.51 . Todos/as los/as participantes informaron tener una audición normal. Los experimentos se realizaron con el entendimiento y consentimiento por escrito de cada participante, siguiendo el Código de Ética de la Asociación Médica Mundial (Declaración de Helsinki), y fueron aprobados por el Comité de Ética de la Universidad Nacional de Quilmes y la Universidad Nacional de Tres de Febrero. Los/as participantes no tenían conocimiento previo de la sala experimental ni de los objetivos del estudio.

2.2 Descripción de la sala experimental

Todos los experimentos se realizaron en una sala semi-reverberante de forma paralelepípeda con un tamaño de $8 \times 4,2 \times 4$ m (largo $\times$ ancho $\times$ alto). El tiempo de reverberación promedio de la sala fue de $T20 = 0,6$ s. El tiempo de reverberación se obtuvo utilizando un barrido exponencial medido en la posición del oyente.

El ruido de fondo de la sala era de 43,9 dBA (medido con un sonómetro SVAN 959 en la posición del oyente). La sala estaba compuesta techo de hormigón y piso de madera. Las paredes laterales y la del fondo estaban pintadas y contenían material fonoabsorbente (cubriendo la superficie en un 35% aproximadamente). En la pared frontal se estaba cubierta por un telón de teatro.

2.3 Aparatos y estímulos

El estímulo utilizado fue un fragmento musical anecoico de 8 segundos correspondiente a la canción “Hell of a guy”, registrada por Benjamin Bernschutz [6]. El procedimiento experimental fue controlado por una computadora portátil con Matlab R2016b con Psychtoolbox (PTB-3) y a través de una placa de audio RME Fireface USB, con una frecuencia de muestreo de 44,1 kHz. Todos los participantes se sentaron en una silla regulable en altura situada en una esquina de la sala (delante del telón de teatro) y fijaron la cabeza (con 0° elevación y azimut) durante todo el experimento a través de un apoya menton. La posición de la silla se ajustó de tal manera que las orejas de cada participante se encuentre a 1,3 m del suelo. Las distancias utilizadas para el experimento fueron: $D = 1, 2, 3, 4, 5$ y 6 metros.

La configuración experimental para la reproducción de estímulos reales, consistió en un parlante activo (Genelec 8010a) ubicado en un soporte de metal con el eje central colocado a la misma altura que las orejas del sujeto. El mismo se ubico en diagonal a la sala de forma tal de formar una línea recta hasta el vértice opuesto a la ubicación del participante. Para evitar la oclusión acústica entre los altavoces de sonido, un experimentador colocó manualmente la fuente sonora frente al participante en alguna de las 6 distancias establecidas según la secuencia del experimento. Para ello, los experimentadores podían ver la distancia correspondiente a cada ensayo en una pantalla del ordenador situada detrás del participante. Entre ensayos, se presentó un sonido de banda ancha a través de dos parlantes (Edifier R1000TCN 25W) ubicados a ambos lados del participante para

enmascarar el posible ruido relacionado con el procedimiento de movimiento del altavoz.

Los estímulos virtuales se obtuvieron convolucionando el fragmento musical con la respuesta al impulso binaurales (BIRs) (individualizada, no individualizada y simulada) para cada distancia. Las BIRs para obtener las HRTF individualizadas y no individualizada fueron medidas utilizando la técnica del Log Sine Sweep [7]. Para ello se utilizó un micrófonos binaurales DPA4560 en cada participante con el canal auditivo bloqueado [8] y la misma fuente sonora y distancias que en la configuración experimental real. La figura 1 presenta un ejemplo de esta medida. Las BIRs simuladas se obtuvieron mediante el programa EASE utilizando el método de rayos, su herramienta de auralización y su base de datos de HRTF estándar. Se simuló la misma sala ubicando el receptor y las fuentes emisoras en las mismas posiciones que en la condición real. Previo a ello, se realizó una validación de los parámetros acústicos otorgados por el programa de la sala simulada con las medidas in situ de la sala real.

La configuración experimental para la reproducción de estímulos virtuales, consistió en el uso de unos auricular abierto (Sennheiser HD800) por donde se reprodujo el estímulo en alguna de las 6 distancias establecidas según la secuencia del experimento. Al igual que en la configuración experimental para reproducir estímulos reales, y con el fin de generar diferencias entre cada condición, entre ensayos, se presentó un sonido de banda ancha a través de dos altavoces ubicados a ambos lados del participante.

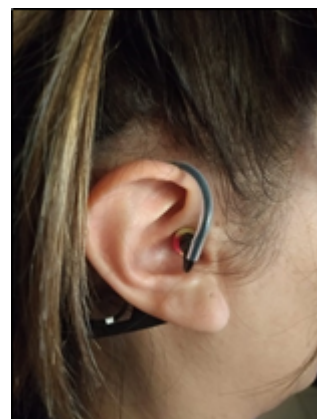




Figura 1: Fotografía del micrófono DPA4560 y el bloque de canal auditivo utilizado.

2.4 Procedimiento

Cada participante realizó dos sesiones en días separados por una semana. En ambas sesiones, los participante fueron recibidos en una habitación contigua a la sala experimental, donde se les informó y detalló la tarea a realizar antes de firmar el formulario de consentimiento. Durante la primera sesión, solo se realizó el proceso de registro de las señales con HRTF individualizadas. Para ello, se le explicó a cada participante como colocarse los tapones, luego se le vendaron los ojos y se lo acompañó (con ayuda del experimentador) hasta la sala experimental. Una vez que el participante se sentó y la silla se ajustó a la posición adecuada, se fijó la cabeza utilizando un apoya menton y se le colocó los micrófonos en cada oído. Allí permanecieron sentados durante toda la grabación. Al finalizar, se acompañó al participante hasta la sala contigua y se le quitó la venda de los ojos. Esta sesión tuvo una duración promedio de 1 hora. En la segunda sesión se realizó el experimento principal. Nuevamente, luego de leer las instrucciones, a cada participante se le vendaron los ojos y se lo acompañó (con ayuda del experimentador) a la sala experimental hasta ser ubicados correctamente (sentados en la silla con la cabeza fija). Allí permanecieron sentados durante toda la tarea. Ninguno de los participantes tuvo conocimiento previo de la sala.

El experimento consistió en 4 bloques. El orden de los bloques fue aleatorio de forma contrabalanceada. En cada bloque se presentaron de forma aleatoria 4 repeticiones por distancia. Los bloques fueron:

- Real: los estímulos se reproducen a través de parlantes ubicados a las distancias definidas dentro de la sala experimental.
- Estímulos virtuales no individualizados (No ind virtual): el oyente escucha, a través de los auriculares, estímulos registrados de manera binaural a las

distancias definidas dentro de la sala experimental con las HRTF de un sujeto de referencia.

- Estímulos virtuales individualizados (Ind virtual): el oyente escucha, a través de auriculares, estímulos registrados con micrófonos binaurales ubicados en la entrada del canal auditivo del mismo participante. Las HRTFs individualizadas se obtienen a las distancias definidas dentro de la sala experimental.
- Estímulos virtuales simulados (Sim virtual): el oyente escucha, a través de auriculares, sonidos simulados mediante el programa EASE.

En total cada participante realizó 96 trials. Para cada trial el/la participante informó estimaciones de distancia verbalmente utilizando una escala de metro (o centímetro) con un decimal. Al finalizar, se acompañó al participante hasta la sala contigua y se le quitó la venda de los ojos. Esta sesión tuvo una duración promedio de 1 hora.

2.5 Modelos estadísticos

Ajustamos modelos lineales (LM, por sus siglas en inglés) y modelos lineales de efectos mixtos (LME, por sus siglas en inglés) para estimar las asociaciones de las magnitudes (distancia percibida y sesgos relativos) con el bloque y la distancia. Definimos el sesgo absoluto con signo como la diferencia entre la distancia física y la percibida para una prueba determinada. El sesgo relativo con signo es el sesgo absoluto dividido por la distancia física. Ambos sesgos con signo indican la sobreestimación (si es positiva) o la subestimación (si es negativa) de la percepción de los sujetos y, por lo tanto, son útiles para medir la precisión de las estimaciones. El sesgo relativo con signo también es útil para comparar estimaciones de fuentes que abarcan un amplio rango, como es el caso de nuestro estudio. Calculamos el sesgo relativo para cada sujeto, posición de la fuente sonora y bloque, y lo promediamos entre posiciones de fuente sonora, para obtener el sesgo colapsado individual para cada bloque. Esta magnitud nos

permitió medir la precisión global de las estimaciones de los sujetos.

3. RESULTADOS

Percepción de distancia

Se ajustó un LMEM con la respuesta de cada participante y las distancias físicas en logaritmo con el fin de ajustar una función de potencia [9]. La figura 2 presenta los ajustes del modelo para cada bloque. Tanto las estimaciones de los parámetros, como los valores p ($p > 0.05$), indican que las pendientes no son responsables de las diferencias observadas entre cada bloque ($\alpha = 0,71$; 95% IC = [0,59; 0,82] para virtual individualizado; $\alpha = 0,70$; 95% IC = [0,58; 0,82] para real; $\alpha = 0,72$, 95% IC = [0,6; 0,84] para virtual simulado, $\alpha = 0,74$; 95% IC = [0,62; 0,86] para virtual no individualizado). Esto quiere decir, que los sujetos sostienen una misma comprensión en sus respuestas de PAD a medida que aumenta la distancia independientemente del bloque. Sin embargo, la diferencia en la intersección (es decir, el desplazamiento vertical de las curvas de respuesta para cada bloque) es de 37,2 cm (IC del 95% = [30,8; 43,4]) para la curva virtual simulada con un efecto asociado estadísticamente significativo ($p < 2e-16$) en relación a las otras condiciones. Las demás intersecciones no presentan diferencias. Al realizar un análisis de varianza (Anova) del modelo lineal de efectos mixtos con distancia y bloque como efectos fijos e intersecciones aleatorias para cada sujeto, confirmamos que los objetivos son subestimados para la condición virtual simulada ($\beta = 37,2\%$, fixed-effect test: $p < 2,2e-16$). Esto sugiere que cuando los sujetos utilizan estímulos sonoros simulados para realizar tareas de PAD, subestiman en casi un 40% las distancias. En contraposición, cuando se utilizan estímulos sonoros virtuales ya sea individualizados o no, los sujetos responden del mismo modo que en una condición real. Al estudiar el factor distancia, encontramos diferencias significativas ($p < 1,437e-13$) que pueden sugerir una forma de representar la capacidad de externalización de los sujetos en general. Es decir, si no se encontrarán diferencias en este factor se podría entender

que a medida que varía la distancia los sujetos responden lo mismo, lo que representaría curvas constantes en cada bloque. Finalmente, no se encontraron diferencias en la interacción entre factores ($p = 0,88$).

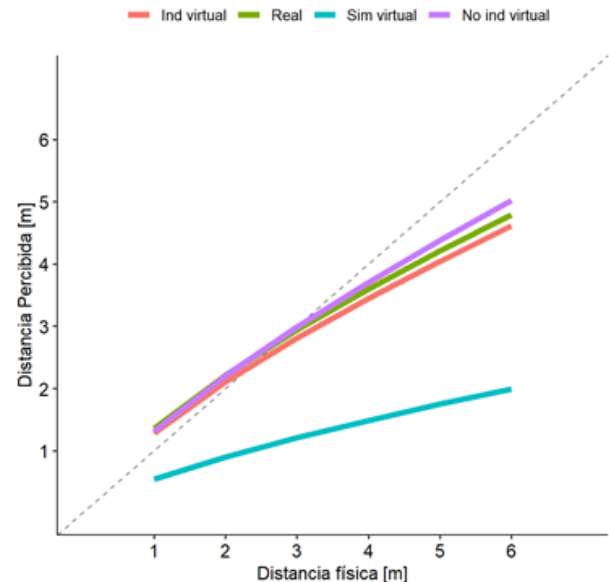


Figura 2: Distancias percibidas con los ajustes del modelo LMEM vs. distancias reales. La línea punteada representa la respuesta ideal.

Sesgo relativo.

La Figura 3 presenta los resultados del sesgo colapsado promedio para cada bloque. Los puntos representan los datos individuales de cada sujeto. Podemos observar una subestimación para el bloque virtual simulado a diferencia del resto ($M = -0,49$; 95% IC = [-0,92; -0,05]). Para los casos real, virtual individualizado y virtual no individualizado se puede ver una muy leve sobreestimación respecto a la línea de referencia ($M = 0,09$, 95% IC = [-0,34; 0,52] para real; $M = 0,02$; 95% IC = [-0,16; 0,2] para virtual individualizado; $M = 0,15$; 95% IC = [-0,27; 0,58] para virtual no individualizado). Al realizar un análisis de varianza (Anova) del modelo lineal con bloque como efectos fijos e intersecciones aleatorias para cada sujeto, confirmamos que los objetivos son subestimados para la condición virtual simulada ($\beta = 49,3\%$, fixed-effect test: $p < 6,675e-6$). Estos resultados son consistentes con los obtenidos en las curvas de distancia, sugiriendo que cuando los sujetos utilizan estímulos sonoros simulados para realizar



tareas de PAD, subestiman las distancias en relación con los otros bloques. De acuerdo a los intervalos de confianza, es posible afirmar con buena certeza que el desempeño para el bloque virtual simulado empeora al menos un 29% más que para los demás bloques. Además, se realizaron ad-hoc pruebas de hipótesis (t-test) pareada de dos colas comparando los resultados individuales del sesgo relativo entre cada bloque. Estos resultados confirman las diferencias significativas con el bloque virtual simulado.

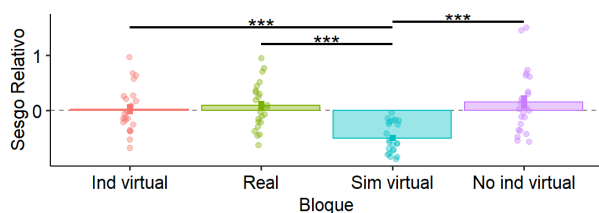


Figura 3: Sesgo colapsado para cada bloque. Las barras representan los estimadores del modelo lineal (LM) en conjunto con el error estándar medio. Los puntos representan los sesgos colapsados de cada sujeto. Los asteriscos representan los resultados de p-valor ($p < 0,005$) de los t-test entre bloques. La línea punteada negra representa la respuesta ideal

4. DISCUSIÓN Y CONCLUSIÓN

Los resultados obtenidos son interesantes ya que indican que basta la utilización de HRTFs no individualizadas reproducidas a través de auriculares abiertos en la misma sala donde fueron medidas para obtener respuestas de percepción de distancia auditiva indistinguibles de las obtenidas con fuentes reales. Las respuestas obtenidas con el programa EASE fueron claramente más sesgadas que las reales. Es decir, las fuentes no fueron realistas a pesar de haber simulado cuidadosamente la reverberación de la sala. Este resultado indica que no basta que los estímulos tengan reverberación para simular correctamente la distancia auditiva sino que es necesario que la reverberación sea similar a la que los oyentes escuchan a través de los auriculares abiertos. Resultados similares fueron reportados previamente por Plenge [10]. Este trabajo mostró que aunque el efecto de la reverberación en la externalización es sólido, para una externalización completa las señales recibidas deben ser plausibles, naturales y

consistentes con las expectativas del oyente. Plenge sugirió que la externalización puede romperse si las señales recibidas no coinciden con la "información almacenada" del oyente sobre una habitación. Sería interesante entonces probar estos mismos estímulos en una habitación con características acústicas diferentes con el fin de comprobar esta hipótesis.

En resumen nuestros resultados sugieren que la percepción auditiva de distancia no utiliza información individual provocada por las orejas del participante sobre la onda incidente. Indican además que si se utilizan grabaciones binaurales no individualizadas reproducidas en la misma sala donde fueron grabadas es posible simular correctamente la distancia a la fuente sonora.

5. AGRADECIMIENTOS

El trabajo fue financiado por la Universidad Nacional de Quilmes, Argentina (PUNQ 1394/15); Agencia Nacional de Promoción Científica y Tecnológica, Argentina (PICT 2016-0738); y Agencia Nacional de Promoción Científica y Tecnológica, Argentina (PICT 2018-4586).

REFERENCIAS

- [1] Kolarik, A. J., Moore, B. C., Zahorik, P., Cirstea, S., & Pardhan, S. (2016). Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, & Psychophysics*, 78(2), 373-395. <https://doi.org/10.3758/s13414-015-1015-1>
- [2] Middlebrooks, J. C. (2015). Sound localization. *Handbook of clinical neurology*, 129, 99-116.
- [3] Moller, H. Fundamentals of Binaural Technology, *Applied Acoustics*. 36 (3), 171-218 (1992).
- [4] Jenny, C., & Reuter, C. (2020). Usability of individualized head-related transfer functions in virtual reality: Empirical study with perceptual attributes in sagittal plane sound localization. *JMIR Serious Games*, 8(3), e17576. DOI: 10.2196/17576
- [5] Kolarik, A. J., Pardhan, S., Cirstea, S., & Moore, B. C. (2017). Auditory spatial representations of the world are compressed in blind humans. *Experimental Brain Research*, 235(2), 597-606.
- [6] Bernschutz B., Fragmento de "Hell of a guy" de Anechoic Recordings, Cologne University of Applied Sciences, Institute of Communication Systems,

Betzdorfer Str. 2, 50679 Cologne, Germany (2013)
<http://www.audiogroup.web.fh-koeln.de>

[7] Farina, A. (2000, February). Simultaneous measurement of impulse response and distortion with a swept-sine technique. In Audio Engineering Society Convention 108. Audio Engineering Society.

[8] Hiipakka, M., Kinnari, T., & Pulkki, V. (2012). Estimating head-related transfer functions of human subjects from pressure-velocity measurements. *The Journal of the Acoustical Society of America*, 131(5), 4051-4061. <https://doi.org/10.1121/1.3699230>

[9] Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *ACTA Acustica united with Acustica*, 91(3), 409-420.

[10] Plenge, G. (1974). On the differences between localization and lateralization. *The Journal of the Acoustical Society of America*, 56(3), 944-951. <https://doi.org/10.1121/1.1903353>