TECNIACUSTICA ELCHE · 2022

53º CONGRESO ESPAÑOL DE ACÚSTICA
XII CONGRESO IBÉRICO DE ACÚSTICA

UTILIZACIÓN DE MODELOS ACÚSTICOS PARA LA VALIDACIÓN DE RESULTADOS DE LOCALIZACIÓN DEL ALGORITMO SRP-PHAT

PACS: 43.60.Acoustic signal processing.

García-Barrios, Guillermo; Gutiérrez-Arriola, Juana Mª; Gómez-Alfageme, Juan José; Sáenz-Lechón, Nicolás; Fraile, Rubén.
CITSEM de la Universidad Politécnica de Madrid (UPM).
Edificio La Arboleda. Campus Sur UPM. Calle Alan Turing 3. 28031 Madrid.
Madrid
España
+34914524900 ext 20787
guillermo.garcia.barrios@upm.es

Palabras Clave: Localización de fuentes sonoras, Simulación acústica, Arrays de micrófonos, Steered-response power

ABSTRACT

This communication presents a comparative study of a sound source localization algorithm performance between simulated and real environments. The purpose of the study is to explore the possibility of using validated acoustic models to evaluate localization algorithms. The experiments are performed in an anechoic chamber and a big room using two arrays of different sizes. The analysis of the results shows a similar position estimation between real recordings and simulations for the small array. However, the large array results are very different due to the inaccuracies present in the validated acoustic models.

RESUMEN

Este artículo compara el funcionamiento de un algoritmo de localización de fuentes sonoras en entornos simulados y reales. La finalidad es explorar la posibilidad de utilizar modelos acústicos validados para evaluar algoritmos de localización. Los experimentos se llevan a cabo en una cámara anecoica y una sala de grandes dimensiones utilizando dos arrays de micrófonos de tamaños diferentes. El análisis de los resultados muestra estimaciones de posición similares entre grabaciones reales y simulaciones para el array pequeño. Sin embargo, los resultados para el array grande son muy distintos debido a las inexactitudes de los modelos acústicos validados.

1. INTRODUCCIÓN

La localización de fuentes sonoras (SSL, del inglés, Sound Source Localization) es un problema importante cuya relevancia ha crecido en los últimos años, lo que ha llevado al desarrollo de diferentes algoritmos de localización basados en la utilización de arrays de micrófonos [1], [2]. En la actualidad, este tipo de sistemas se han utilizado en distintas aplicaciones como sistemas de vigilancia, dispositivos de *Smart Home*, teleconferencia o salud [3].

Entre los algoritmos existentes, el *steered-response power* con filtrado *phase transform* (SRP-PHAT) es el que ha demostrado ser más robusto en condiciones de reverberación y ruido [4], [5]. La mayoría de los trabajos de investigación en este tema han validado los resultados utilizando simulaciones [6]-[9]. Esto se debe a la complejidad de realizar medidas reales, lo cual requiere equipo de audio específico, una gran precisión en la colocación de los micrófonos y las fuentes sonoras, y mucho tiempo. Hasta el momento, algunos trabajos han validado los resultados de investigación utilizando audios reales, que podían pertenecer a audios de bases de datos [3], [10], o a grabaciones realizadas en el propio estudio [6], [11], [12]. Investigaciones más recientes basadas en redes neuronales profundas probaron sus modelos utilizando simulaciones con el método de la fuente-imagen para generar grandes cantidades de datos de entrenamiento [13], [14]. Velasco et al. afirmaron que el entrenamiento únicamente basado en simulaciones podía limitar su aplicabilidad en escenarios reales, por lo que validaron un modelo analítico de SRP-PHAT utilizado grabaciones reales de audios [3]. Otro artículo más reciente comprobó la precisión de la técnica propuesta utilizando 3 bases de datos de grabaciones reales [10].

La opción de utilizar audios reales para el proceso de entrenamiento y evaluación de algoritmos choca con la poca cantidad de bases de datos públicas disponibles. Durante nuestro proceso de búsqueda, solamente se han encontrado tres accesibles vía web: la AV16.3, que incluye grabaciones de audio de 2 arrays de micrófonos circulares dentro de la sala rectangular *Smart Meeting Room* del instituto de investigación IDIAP [15]; la base de datos audiovisual CAVD3D, que utilizó un array circular de 8 micrófonos [16]; y la base de datos CHIL-CLEAR, donde se realizaron medidas en 5 salas diferentes [17]. Todas las bases de datos contienen grabaciones de habla, lo cual resulta un problema cuando se quieren localizar otro tipo de eventos acústicos. Por otra parte, las bases de datos de respuestas al impulso de salas (RIRs) resultan de gran utilidad, ya que permiten generar audios reales a partir de grabaciones anecoicas calculando su convolución. Se han encontrado dos bases de datos de RIRs que se pueden utilizar en tareas de SSL: la base de datos SMARD [18] y la base de datos RIR de la Universidad de Londres [19]. Hay más colecciones, pero no son de utilidad ya que no especifican las coordenadas de las fuentes y los micrófonos [20], [21].

Como ya se ha comentado, el escaso número de bases de datos disponibles de RIRs y grabaciones reales resulta un problema para la validación de algoritmos de localización de fuentes sonoras. Además, las grabaciones de las bases de datos están limitadas a determinadas posiciones de micrófonos y fuentes sonoras, características acústicas de la sala o tipo de evento sonoro. De esta forma, resulta de interés encontrar un método de simulación acústica que permita obtener resultados comparables a los de las medidas reales. Por esta razón, en esta comunicación se estudia la posibilidad de utilizar modelos acústicos de salas que han sido validados con procedimientos estándar, para poder comparar los resultados de localización con los obtenidos de audios grabados en salas reales. El resto del artículo se estructura de la siguiente forma. En la Sección 2, se describen las tres variantes del algoritmo SRP-PHAT que han sido evaluadas. Las bases teóricas de la simulación de acústica de salas se presentan en la Sección 3. En la Sección 4, se explican los modelos acústicos y el proceso de validación, así como la base de datos de audios anecoicos y las grabaciones realizadas. Los resultados de los experimentos se analizan en la Sección 5. Finalmente, las conclusiones se resumen en la Sección 6.

2. SRP-PHAT CON GCC DE ANCHO DE BANDA VARIABLE

El SRP-PHAT es un algoritmo de *beamforming* donde la posición $\vec{r}_s$ de la fuente acústica que emite una señal de audio capturada por K micrófonos se calcula de la siguiente forma [4]:

$$\vec{r}_s \approx \arg\max P(\vec{r}), \tag{1}$$

donde $P(\vec{r})$ es el valor del mapa SRP en la posición $\vec{r}$. Este se puede calcular como:

$$P(\vec{r}) = 2\pi \sum_{k=1}^{K} \sum_{l=1}^{K} R_{kl}\left(\tau_{kl}(\vec{r})\right), \quad (2)$$

donde $\tau_{kl}(\vec{r}) = \tau_l(\vec{r}) - \tau_k(\vec{r})$, siendo $\tau_k(\vec{r})$ el retardo de propagación entre la posición $\vec{r}$ y la posición del micrófono k, y $R_{kl}\left(\tau_{kl}(\vec{r})\right)$ es la función de correlación cruzada generalizada (GCC) entre las señales acústicas capturadas por los micrófonos k y l, respectivamente $s_k(t)$ y $s_l(t)$, evaluadas al tiempo de retardo $\tau_{kl}(\vec{r})$. Cuando se aplica la función de peso PHAT, la GCC se calcula de la siguiente forma:

$$R_{kl}(\tau) = \int_{-\infty}^{\infty} \frac{S_k(\omega) S_l^*(\omega) \cdot e^{j\omega\tau}}{2\pi |S_k(\omega) S_l^*(\omega)|} d\omega, \quad (3)$$

donde $S_k(\omega)$ es la transformada de Fourier de $s_k(t)$, y j es la unidad imaginaria.

En [22], presentamos una versión de la GCC-PHAT limitada en banda demostrando su robustez cuando las fuentes sonoras están cerca del array de micrófonos. De esta forma, limitar la GCC definida en (3) a un cierto intervalo de frecuencias $\omega_{\min} \leq \omega \leq \omega_{\max}$ es equivalente a limitar el intervalo de integración:

$$R_{kl}(\tau) = \int_{\omega_{\min} \leq |\omega| \leq \omega_{\max}} \frac{S_k(\omega) S_l^*(\omega) \cdot e^{j\omega\tau}}{2\pi |S_k(\omega) S_l^*(\omega)|} d\omega. \quad (4)$$

Aunque esta modificación permite reducir las oscilaciones de frecuencia de la GCC e incrementar el ancho del pico principal de la GCC, también reduce la amplitud de este pico. Para compensar dicho efecto se definió una versión normalizada de (4):

$$\widetilde{R_{kl}}(\tau) = \frac{\omega_{\max} - \omega_{\min}}{\widehat{\omega}_{max} - \omega_{\min}} \cdot R_{kl}(\tau), \quad (5)$$

donde $\widehat{\omega}_{max}$ es la máxima frecuencia que depende de la resolución de la rejilla del mapa SRP y la norma Euclidea del gradiente de la diferencia de tiempos de llegada (TDOA) [22].

Para facilitar la comparación de los resultados las 3 variaciones del SRP-PHAT se nombran de la siguiente forma: S-SRP se refiera a (3), B-SRP representa (4), y BN-SRP se corresponde con (5).

3. MODELADO ACÚSTICO DE SALAS

El modelado acústico geométrico consiste en construir un modelo geométrico de una sala, normalmente mediante una herramienta CAD (Computer-Aided Design), añadiendo información sobre las propiedades acústicas de los materiales de construcción y las estructuras más relevantes de la sala. Estos permiten la generación de respuestas al impulso similares a las medidas en las salas reales, para así simular sonidos virtuales que se perciban lo más parecido posible a los sonidos producidos en dicha sala [23].

La validación de modelos acústicos consiste en la comparación de parámetros utilizados en acústica de salas, como el tiempo de reverberación o la claridad, extraídos de las RIRs obtenidas de la simulación y las medidas *in situ*. Los parámetros acústicos evaluados están definidos en la norma ISO 3382-1 [24]. La comparación entre los parámetros medidos y los obtenidos del modelo acústico se realiza de manera separada en bandas de octava [25]. Cuando las diferencias son muy grandes, el modelo se ajusta para volver a realizar la comparación. Normalmente el indicador utilizado para medir dichas referencias es el Just Noticeable Difference (JND), que se

define como la mínima variación de un parámetro acústico que se puede percibir por un oyente estándar [25].

La mayor ventaja de usar modelos acústicos es su flexibilidad, que permite posicionar los micrófonos y las fuentes de sonido en cualquier punto de la sala, y simular los canales acústicos de manera sencilla. De esta forma, las RIRs se obtienen fácilmente para convolucionarlas con los eventos anecoicos. Hay diferentes herramientas software para la generación de modelos acústico como EASE® y ODEON®. En este estudio, el modelo acústico se ha desarrollado utilizando EASE®. Este software obtiene los resultados de un modelo estadístico utilizando las ecuaciones de Sabine y Eyring, calculando el trazado de rayos con el método fuente-imagen. Además, introduce un método llamado AURA, que es un motor híbrido que permite incluir la dispersión de cada material en las simulaciones [26].

4. EXPERIMENTOS

Como se ha señalado anteriormente, el objetivo de esta comunicación es estudiar la similitud de los resultados de localización entre salas reales y su simulación. Para llevarlo a cabo, se han seleccionado una cámara anecoica y una oficina de grandes dimensiones.

4.1. Escenarios Acústicos

Como se ha dicho, uno de los escenarios es una cámara anecoica localizada en la Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación (ETSIST) de la Universidad Politécnica de Madrid (Fig. 1, izq.). Las dimensiones de la cámara son 5.7 m x 4.3 m x 2.8 m, siendo completamente anecoica con absorbentes acústicos debajo de la rejilla de metal. Por ello, el modelo acústico se reduce únicamente a aplicar el retardo correspondiente a las señales acústicas.

La oficina vacía es una sala de grandes dimensiones que se encuentra en el Centro de Investigación de Tecnologías Software y Sistemas Multimedia para la Sostenibilidad (CITSEM) de la Universidad Politécnica de Madrid. La forma de la sala es irregular, con una superficie total de 435.8 m^2 y un volumen de 1220 m^3 (Fig.1, der.). El tiempo de reverberación medido es de 1.22 s. El modelo acústico fue generado y validado por Benutti en su Trabajo Fin de Máster utilizando el software EASE® [27]. La validación se hizo utilizando el método estadístico y el geométrico en relación con el tiempo de reverberación promedio (T20) evaluado en bandas de octava y tercios de octava calculando el JND.

Figura 1 – Esquema de medida para la cámara anecoica (izquierda) y la sala grande (derecha).

4.2. Base de Datos de Audios

Las señales utilizadas para las simulaciones fueron eventos anecoicos de la tarea *Sound event detection in synthetic audio task* del DCASE2016 Challenge [28]. Hay 11 tipos de eventos sonoros que se pueden clasificar en cuatro categorías según el análisis espectral realizado en [29]. Para utilizar eventos de las cuatro categorías, se ha seleccionado uno por grupo: llamar a la puerta, hablar, caída de llaves y tono de llamada de un teléfono. Los audios tienen una frecuencia de muestreo de 44.1 kHz, una resolución de 16 bits y una duración comprendida entre 0.13 s y 3.34 s.

4.3. Medidas

Las grabaciones se realizaron utilizando dos tipos de arrays de micrófonos. Ambos formados por 4 micrófonos colocados en las esquinas de un tetraedro regular. La diferencia entre ellos era la distancia entre micrófonos: un array con un lado de 0.5 m (array pequeño), y el otro con 2.5 m (array grande). El centro de los arrays se colocó en el centro aproximado de las regiones evaluadas de las salas. Atendiendo al equipo utilizado, las señales se grabaron utilizando micrófonos omnidireccionales de condensador Superlux ECM99 y la tarjeta de sonido Behringer UMC1820. Los eventos generados en las salas fueron voz de hombre, llamada de teléfono, llamar a la puerta y caída de llaves, cada uno correspondiente a una de las 4 categorías.

La región para evaluar en cada una de las salas fue de 5 m x 4 m x 2.5 m. Se seleccionaron 10 posiciones de fuentes sonoras distribuidas uniformemente en el plano horizontal, evitando pares simétricos. Para cada una de estas posiciones, 2 alturas diferentes fueron evaluadas dependiendo del tipo de evento. Al final, se tuvieron un total de 20 posiciones por tipo de evento. Las mismas posiciones de micrófonos y fuentes acústicas se seleccionaron para la extracción de las RIRs de los modelos acústicos validados.

5. RESULTADOS

El número total de audios evaluados fue de 80 (20 posiciones por tipo de evento) por cada array de micrófonos. La distribución de error de localización, medida en metros, obtenida de aplicar el algoritmo SRP-PHAT y sus dos variantes descritas en la Sección 2, se presentan en la Figura 2 mediante diagramas de cajas y bigotes. La resolución de la rejilla del mapa SRP fue de 0.5 m y el ancho de banda entre 100 Hz y 6000 Hz. Es remarcable que la mayoría de los errores de localización están por debajo de 1 m.

Una vez verificado que la precisión de la localización era aceptable, se calculó la diferencia de localización entre las simulaciones y las grabaciones reales. La Tabla 1 muestra el percentil 75 (3er cuartil) de la distancia euclídea entre la posición estimada en las simulaciones y medidas reales. En el caso del array pequeño, el error del SRP con la GCC limitada en banda es menor que la longitud de la diagonal de una unidad cúbica de rejilla ($0.5 \cdot \sqrt{3} \approx 0.87$ m). Esto quiere decir que el 75 % de las estimaciones se encuentran en la misma celda de la rejilla o en una adyacente. Para los otros dos métodos SRP el error es mayor, lo cual corresponde a una distancia de dos o más celdas. Sin embargo, en el caso del array grande, no se puede asegurar que la estimación SRP sea equivalente entre el escenario simulado y el real para ningún método.

Cuando se analizan las distribuciones de los errores de la Figura 2, se puede ver como las distribuciones de los escenarios simulados y reales son diferentes, pero se pueden llegar a conclusiones similares cuando se comparan. De hecho, los algoritmos que obtienen los mejores resultados en relación con la mediana del error son los mismos para ambos escenarios. Además, los algoritmos con la mayor y menor dispersión de error coinciden.

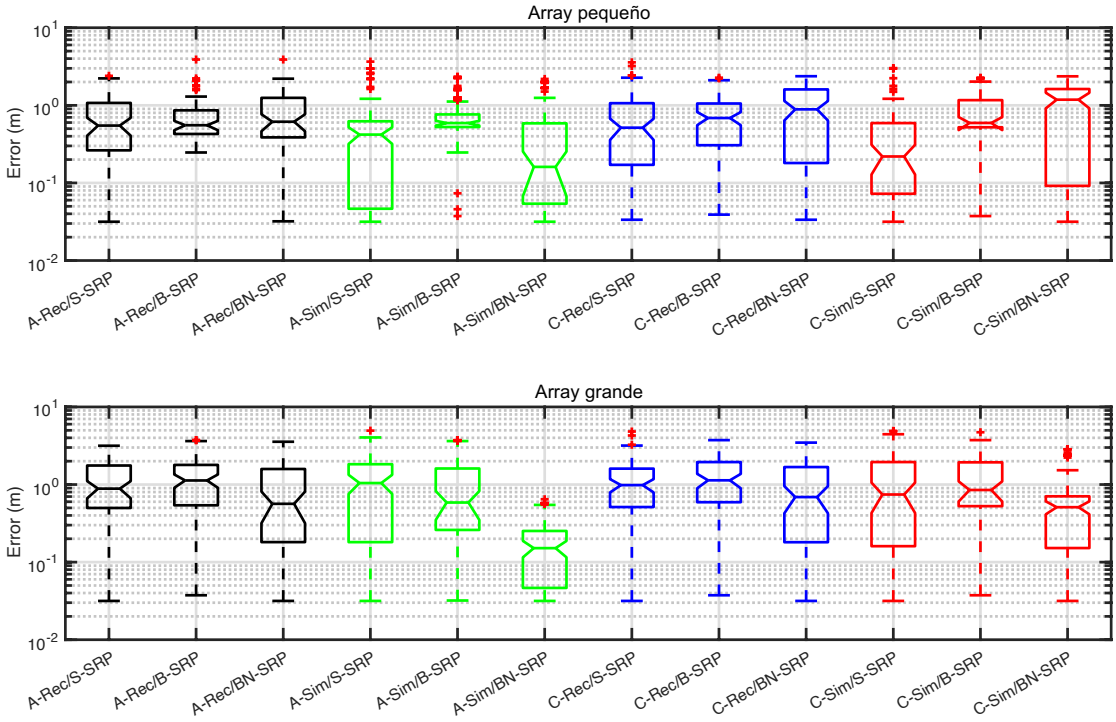

Figura 2 – Distribución de los errores de localización para el array pequeño (arriba) y el array grande (abajo). Para tener nombre compacto, los tipos de audios se han etiquetado como A-Rec (grabaciones anecoicas), A-Sim (simulaciones acústicas anecoicas), C-Rec (grabaciones CITSEM), y C-Rec (simulaciones acústicas CITSEM).

Tabla 1 – Percentil 75 de la distribución del error de la distancia euclídea en metros entre la posición de fuente estimada utilizando las simulaciones y las grabaciones reales.

	Cámara anecoica			**Sala CITSEM**		
	S-SRP	**B-SRP**	**BN-SRP**	**S-SRP**	**B-SRP**	**BN-SRP**
Array pequeño	1.22 m	0.50 m	1.12 m	1.22 m	0.60 m	1.12 m
Array grande	2.64 m	2.18 m	1.58 m	2.64 m	2.29 m	1.87 m

Una comparación complementaria aparece en los resultados mostrados en la Tabla 1. Estos muestran que, para el array pequeño, el algoritmo B-SRP produce estimaciones de localización similares entre las simulaciones y los audios reales. Sin embargo, en el array grande esto no es así. Esta diferencia se puede deber a los procesos de simulación acústica y validación del modelo. En el caso de la simulación acústica, el software no puede simular todas las reflexiones que tienen lugar en las superficies de la sala. Por ello, puede que las señales simuladas y reales sean perceptualmente iguales, pero las RIRs no. Estas diferencias en la simulación de la reverberación generan diferentes picos secundarios en la GCC afectando así a los mapas SRP. En el caso del array de micrófonos pequeño, como los micrófonos están más juntos, estas variaciones presentan mayor correlación y afectan en al cálculo de la GCC en menor medida. Por otro lado, La validación del modelo se hace calculando valores de parámetros acústicos en diferentes puntos de la sala, y comparando los valores promedio en diferentes bandas de frecuencia. Esto quiere decir, que el modelo se ajusta en función de esas posiciones, y el resto de las posiciones se extrapolan. Esto podría explicar las diferencias observadas entre la sala real y la simulada.

6. CONCLUSIONES

La evaluación de algoritmos de SSL mediante grabaciones implica un alto coste. Por otro lado, las evaluaciones basadas en simulaciones se consideran limitadas cuando no van acompañadas de experimentos reales. En esta comunicación se ha presentado un análisis comparativo de los resultados obtenidos a partir de las grabaciones en las salas y de los audios simulados generados a partir de los modelos acústicos. Como estos presentan simulaciones realistas, se esperaban resultados similares a los de las salas reales.

Los resultados de las Figura 2 muestran que, aunque la distribución de los errores de localización entre audios simulados y reales son diferentes, si que tienen un comportamiento similar cuando la comparación se realiza entre algoritmos. Esto quiere decir que la utilización de las simulaciones para realizar comparaciones entre algoritmos podría ser válida.

Por otro lado, los resultados que aparecen en la Tabla 1 indican que las estimaciones para el array pequeño entre escenarios reales y simulados son muy similares. Esto significa que los modelos acústicos se pueden utilizar para validar resultados de localización cuando la distancia entre micrófonos del array es pequeña. Lo cual resulta de gran utilidad para el entrenamiento de redes neuronales profundas. Sin embargo, en el caso del array grande las diferencias son mucho mayores. Una posible explicación es la falta de correlación entre las RIRs ya que se obtienen por simulación con un modelo que ha sido validado midiendo en puntos concretos del recinto. Esto ocasiona la generación de GCCs diferentes, y como consecuencia diferentes mapas SRP.

Para finalizar, los resultados de este estudio están limitados a dos salas, por lo que para aumentar su relevancia se deben ampliar a más entornos acústicos. Además, la relación entre el tamaño del array y la validez de los resultados de simulación necesita un análisis más profundo para determinar las causas de las diferencias de estimación entre audios simulados y reales, así como determinar qué distancia limite entre la fuente y el array asegura que el modelo acústico se puede utilizar para validar resultados de localización.

AGRADECIMIENTOS

Este trabajo ha sido apoyado por la Universidad Politécnica de Madrid a través del Programa Propio de I+D+I, en concreto la convocatoria predoctoral, y los recursos computacionales del supercomputador Magerit.

REFERENCIAS

[1] Brandstein, M. S.; Silverman, H. F. A practical methodology for speech source localization with microphone arrays. *Computer Speech & Language*, 11 (2), 1997, 91-126.

[2] Cobos, M.; Antonacci, F. A Survey of Sound Source Localization Methods in Wireless Acoustic Sensor Networks. *Wireless Communications and Mobile Computing*, 2017, 2017.

[3] Velasco, J; Martín-Arguedas, C. J.; Macias-Guarasa, J.; Pizarro, D.; Mazo, M. Proposal and validation of an analytical generative model of SRP-PHAT power maps in reverberant scenarios. *Signal Processing*, 119, 2016, 209-228.

[4] DiBiase, J. H.; Silverman, H. F.; Brandstein M. S. Robust localization in reverberant rooms, en *Microphone Arrays*, Springer, Berlín (Alemania), 2001.

[5] Crocco, M.; Cristani, M.; Trucco, A.; Murino, V. Audio surveillance: A systematic review. *ACM Computing Surveys (CSUR)*, 48 (4), 2016, 52:1-52:46.

[6] Gustafsson, T.; Rao, B. D.; Trivedi, M. Source localization in reverberant environments: Modeling and statistical analysis. *IEEE Transactions on Speech and Audio Processing*, 11 (6), 2003, 791-803.

[7] Zhang, C.; Florencio, D.; Zhang, Z. Why does PHAT work well in low-noise, reverberative environments?. *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2565-2568.

[8] Cobos, M.; Lopez, J. J.; Spors, S. Analysis of room reverberation effects in source localization using small microphone arrays. *2010 4th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, 2010, 1-4.

[9] Fang, Y.; Xu, Z. Multiple sound source localization and counting using one pair of microphones in noisy and reverberant environments. *Mathematical Problems in Engineering*, 2020, 2020, 1-12.

[10] Vera-Diaz, J. M.; Pizarro D.; Macias-Guarasa J. Acoustic source localization with deep generalized cross correlations. *Signal Processing*, 187, 108169.

[11] Arabaci, M.; Strickland, R. N. Direction of arrival estimation in reverberant rooms using a resource-constrained wireless sensor network. *IEEE International Conference on Pervasive Services*, 2007, 29-38.

[12] Perez-Lorenzo, J.; Viciana-Abad, R.; Reche-Lopez, P.; Rivas, F.; Escolano J. Evaluation of generalized cross-correlation methods for direction of arrival estimation using two microphones in real environments. *Applied Acoustics*, 73 (8), 2012, 698-712.

[13] Krause, D.; Politis A.; Kowalczyk, K. Comparison of convolution types in CNN-based feature extraction for sound source localization. *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, 820-824.

[14] Diaz-Guerra D.; Miguel A.; Beltran J. R. Robust sound source tracking using SRP-PHAT and 3D convolutional neural networks. I*EEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 2021, 300-311.

[15] Lathoud G.; Odobez J. M.; Gatica-Perez D. AV16.3: An audio- visual corpus for speaker localization and tracking, en *Machine Learning for Multimodal Interaction*, Springer, Berlín (Alemania), 2005.

[16] Qian, X.; Brutti, A.; Lanz, O.; Omologo, M.; Cavallero, A. Multispeaker tracking from an audio–visual sensing device. *IEEE Transactions on Multimedia*, 21 (10), 2019, 2576-2588.

[17] Waibel, A.; Stiefelhagen, R.; Carlson R.; Casas, J.; Kleindienst, J.; Lamel, L.; Lanz, O.; Mostefa, D.; Omologo, M.; Pianesi, F.; et al. Computers in the human interaction loop, en *Handbook of Ambient Intelligence and Smart Environments*, Springer, Boston (USA), 2010.

[18] Nielsen, J. K.; Jensen, J. R.; Jensen, S. H.; Christensen, M. G. The single- and multichannel audio recordings database (SMARD). *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2014, 40-44.

[19] Stewart, R.; Sandler, M. Database of omnidirectional and B-format room impulse responses. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, 165-168.

[20] Wen, J. Y. C.; Gaubitch, N. D.; Habets, E. A. P.; Myatt, T.; Naylor, P. A. Evaluation of speech dereverberation algorithms using the MARDY database. *International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2006.

[21] Jeub, M.; Schafer, M.; Vary, P. A binaural room impulse response database for the evaluation of dereverberation algorithms. *2009 16th International Conference on Digital Signal Processing*, 2009, 1-5.

[22] García-Barrios, G.; Gutiérrez-Arriola, J. M.; Sáenz- Lechón, N.; Osma-Ruiz, V. J.; Fraile, R. Analytical model for the relation between signal bandwidth and spatial resolution in steered-response power phase transform (SRP-PHAT) maps. *IEEE Access*, 9, 2021, 121549-121560.

[23] Savioja, L.; Svensson, U. P. Overview of geometrical room acoustic modeling techniques. *Journal of the Acoustical Society of America*, 138 (2), 2015, 708-730.

[24] *Measurement of room acoustic parameters — Part 1: Performance spaces*, ISO Std., 3382-1:2009.

[25] Bork, I. A comparison of room simulation software - The 2nd round robin on room acoustical computer simulation. *Acta Acustica united with Acustica*, 86 (6), 2000, 943-956.

[26] Inc, R. H. *EASE. User's Guide & Tutorial*, Acoustic Design Ahnert.

[27] Benutti, M. Analysis of acoustic parameters for validation of room acoustic simulation models. Master's thesis, Universidad Politécnica de Madrid, España, 2019.

[28] Mesaros, A.; Heittola, T.; Benetos, E.; Foster, P.; Lagrange, M.; Virtanen, T.; Plumbley, M. D. Detection and classification of acoustic scenes and events: Outcome of the DCASE 2016 challenge. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26 (2), 2018, 379-393.

[29] Gutiérrez-Arriola, J.; Fraile, R.; Camacho, A.; Durand, T.; Jarrín, J. L. Synthetic sound event detection based on MFCC. *Proc. of DCASE 2016 Workshop*, 2016, 6929-6936.