



44º CONGRESO ESPAÑOL DE ACÚSTICA ENCUENTRO IBÉRICO DE ACÚSTICA SIMPOSIO EUROPEO DE ACÚSTICA AMBIENTAL Y MAPAS DE RUIDO

CARACTERIZACIÓN Y DETECCIÓN AUTOMÁTICA DE REFLEXIONES ACÚSTICAS EN RECINTOS UTILIZANDO ARRAYS CIRCULARES Y PROCESADO DE IMAGEN

REFERENCIA PACS: 43.60.FG

Ana M. Torres¹; Maximo Cobos²; Francesc J. Ferri²; y Jaume Segura³

1 Departamento IEEAC, Univ. de Castilla-La Mancha, ana.torres@uclm.es

2 Dpt Informàtica, ETSE, Universitat de València, Poligon de la Coma s/n, macose2@uv.es

3 IRTIC, Universitat de València, Poligon de la Coma s/n, jsegura@uv.es

ABSTRACT

Microphone arrays are a very useful tool for analyzing the spatial properties of acoustic fields in rooms. To this end, impulse responses are acquired simultaneously at all the sensors in the array and several processing techniques can be applied to study the distribution of the room reflections both in time and spatial direction. Due to the circular shape of the array, the multi-trace impulse response of the array can be seen as an image showing different cosine-like footprints corresponding to plane-wave reflections occurring within the room.

This paper studies the performance of two image processing techniques applied to the detection of plane-wave reflections, namely, the generalized Hough transform and cross-correlation. Experiments using a set of real microphone recordings in different rooms are described, showing the potential of image analysis techniques in acoustic applications.

RESUMEN

Los arrays de micrófonos son una herramienta muy útil para analizar las características espaciales del campo acústico dentro de un recinto. Para ello, se han de capturar las respuestas al impulso entre una fuente de prueba y todos los micrófonos del array simultáneamente, aplicando distintos tipos de técnicas para estudiar la distribución tanto temporal como espacial de las distintas reflexiones en la sala. Debido a la forma circular del array, la respuesta al impulso multi-micrófono puede ser vista como una imagen donde se pueden observar distintas huellas cosenoidales que se corresponden con las reflexiones de onda plana capturadas en la sala.

Este artículo estudia la validez de dos tipos de técnicas de procesamiento de imagen para detectar estas reflexiones, concretamente, la Transformada de Hough generalizada y la correlación-cruzada. Se describen varios experimentos realizados en varias salas reales, mostrando el potencial de las técnicas de análisis de imagen en aplicaciones acústicas.

1 INTRODUCCIÓN

Los arrays de micrófonos son ampliamente utilizados en múltiples tareas de procesado con señales acústicas, tales como para la mejora del habla, la localización de fuentes acústicas o la cancelación de eco. Una de las más interesantes aplicaciones de los arrays es el análisis de características espaciales del campo sonoro, permitiendo a los diseñadores acústicos investigar de manera sistemática los beneficios y limitaciones de los fenómenos de reflexión que aparecen en el interior de las salas. De hecho, medir las respuestas al impulso de manera simultánea en diferentes posiciones espaciales, permite a los ingenieros acústicos un análisis completo del campo sonoro pudiendo estimar de la geometría de la sala en cuestión [1], la detección de las principales reflexiones que aparecen en ella [2] y la identificación de otros fenómenos en la propagación del sonido por la sala [3].

En los últimos años se han desarrollado múltiples métodos de análisis para investigar el impacto de las reflexiones acústicas dentro de una sala [4-6]. Algunos de estos métodos se basan en la medición de múltiples respuestas al impulso con arrays circulares de micrófonos, los cuáles proporcionan una cobertura completa en azimut [7]. Generalmente el procesado de las respuestas se realiza con el objetivo de descomponer el campo sonoro en ondas planas utilizando armónicos cilíndricos [6]. Aunque estos métodos proporcionan un correcto marco teórico para el análisis del campo sonoro, tienen el hándicap de necesitar un alto número de micrófonos para satisfacer las necesidades de un muestreo espacial, lo cuál hace muy difícil su implementación, tanto desde un punto de vista experimental como computacional.

En este trabajo se presentan dos métodos para la identificación de reflexiones en salas mediante la captura de respuestas al impulso por medio de un array circular de micrófonos. El conjunto de datos medidos correspondientes a esas respuestas al impulso, conforman una matriz que puede interpretarse como una imagen donde las reflexiones de la sala aparecen en formas de coseno. De esta forma, los métodos propuestos en este trabajo están basados en la Transformada de Hough para detectar formas semejantes a un coseno (también llamados ecogramas). La identificación de los máximos de esas formas coseno en los ecogramas permitirán identificar la dirección de llegada (DOA) y el tiempo de llegada de las ondas planas más importantes y significativas capturadas por el array.

El artículo se estructura de la siguiente forma. La Sección 2 describe brevemente los fundamentos de los arrays circulares de micrófonos y de las respuestas al impulso. La Sección 3 explica los métodos de detección de ondas planas propuestos en este trabajo. Las Secciones 4 y 5 describen respectivamente los experimentos realizados en salas reales y sus correspondientes resultados. Finalmente, la Sección 6 resume las principales conclusiones de este trabajo.

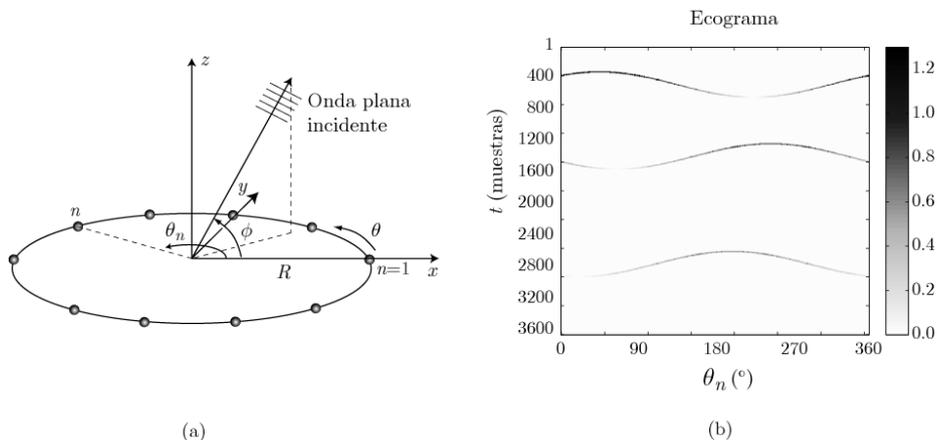


Figura 1. (a) Geometría de un array circular. (b) Representación de las respuestas al impulso con 3 ondas planas llegando con los siguientes ángulos $\theta_i = \{45^\circ, 245^\circ, 200^\circ\}$ llegando en los siguientes instantes de tiempo en muestras $\tau_i = \{400, 1400, 2800\}$.

2 ONDAS PLANAS Y ARRAYS CIRCULARES

Un array de micrófonos es un conjunto de micrófonos distribuidos espacialmente utilizados para grabar y procesar señales sonoras. El procesado simultáneo de señales capturadas por el array, puede utilizarse para mejorar una señal de interés, localizar una fuente o para reunir información sobre un entorno acústico. Aunque los arrays de micrófonos pueden tener cualquier forma, este artículo se centra en arrays circulares de micrófonos.

Consideremos un array circular de micrófonos con N elementos y de radio R posicionado en el plano horizontal (x, y) tal y como muestra la Figura 1 (a). Las ondas planas inciden en el array con un ángulo azimut θ_i , un ángulo de elevación ϕ_i y una velocidad de propagación c ($c \approx 343m/s$). La posición de un micrófono dentro del array, se especifica con un ángulo

azimut con $\theta_n = (n-1)\frac{2\pi}{N}$, con $n = 1, \dots, N$. El instante de tiempo en el cuál la onda plana

llega al centro del array es conocido como *intercept time* y es denotado como τ_i . Los tiempos de llegada de una onda plana incidiendo en el array circular están descritas en función del ángulo del micrófono y vienen representadas curvas semejantes a formas de onda coseno. La cresta o máximo de la curva coseno se produce para el micrófono que capta primero la onda y que es precisamente el que tiene el ángulo más cercano a θ_i . En la Figura 1 (b), se muestra un ejemplo simulado para diferentes ondas planas que llegan en las direcciones 45° , 245° y 200° e *intercept times* de 400, 1400 y 2800 muestras para un array con $R = 1.5m$ y $N = 288$ micrófonos.

Para una determinada frecuencia de muestreo f_s , estas curvas pueden ser matemáticamente expresadas como:

$$t_n(\tau_i, \theta_i) = \left\lceil f_s \tau_i - f_s \frac{R}{c} \cos(\theta_n - \theta_i) \right\rceil,$$

donde $t_n(\tau_i, \theta_i)$ representa el instante de tiempo (en muestras) en el que una onda plana con DOA θ_i e *intercept time* τ_i llega al n-ésimo micrófono.

2.1 Representación de respuestas al impulso

La representación que conforman las ondas sonoras en el espacio, puede ser descrita de forma completa mediante la captura de las respuestas al impulso, las cuáles proporcionan una descripción de los cambios que sufre una señal acústica cuando viaja de un punto a otro dentro

de una sala [8, 9]. Cuando se miden de manera simultánea N respuestas al impulso, se puede construir una matriz espacio-tiempo mediante el almacenamiento de todas las respuestas al impulso en azimut capturadas por el array:

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N],$$

donde $\mathbf{h}_n = [h_n(1), \dots, h_n(L)]^T$ es la respuesta al impulso de longitud L capturada por el micrófono n . Los elementos de \mathbf{H} se indican como $H(t, n)$, donde t y n corresponden a un determinado tiempo y micrófono, respectivamente. Esta matriz permite observar de forma completa el comportamiento de una sala mediante la representación conjunta de todas las respuestas al impulso medidas.

3 MÉTODOS PROPUESTOS PARA LA DETECCIÓN DE ONDAS

3.1 Transformada de Hough- modelo parábola

Aunque la clásica Transformada de Hough fue diseñada inicialmente para la identificación de líneas en una imagen [10], más tarde se ha utilizado y extendido para la identificación de formas arbitrarias. Hoy en día es un método muy popular dentro de la literatura en procesamiento de imagen para reconocer diferentes tipos de formas en múltiples aplicaciones [11]. Los fundamentos de la Transformada de Hough se basan en el diseño de un espacio que acumula los valores de un conjunto de píxeles seleccionados en una zona a analizar dentro de una imagen. Estos píxeles seleccionados se supone que pertenecen a las instancias imperfectas de la forma para ser detectados, por lo que su valor sólo se acumula para aquellos puntos en el espacio de parámetros que definen una forma a la que el píxel pertenecería. Este procedimiento se puede interpretar como un procedimiento de votación de la que se obtienen los candidatos de objeto como máximos locales en el espacio de parámetros. En resumen, el objetivo de la transformada de Hough es realizar agrupaciones de los puntos que pertenecen a los bordes de posibles figuras a analizar a través de un procedimiento de votación sobre un conjunto de figuras parametrizadas.

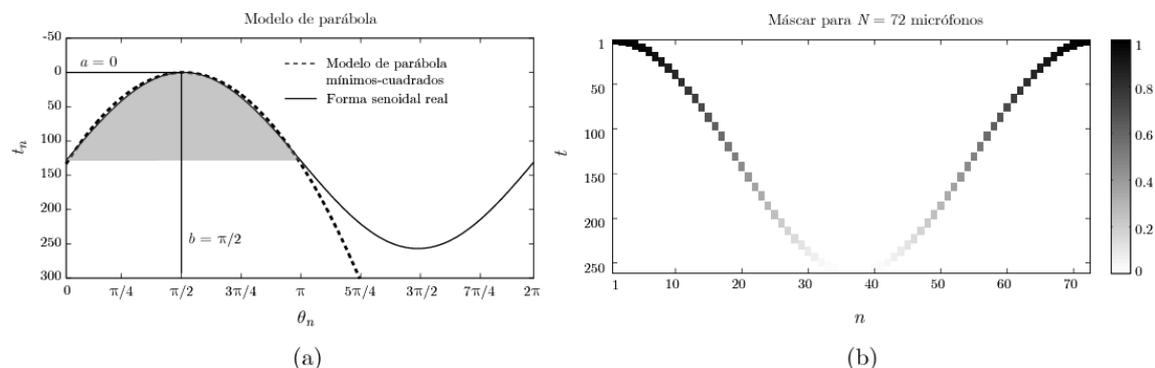


Figura 2. (a) Modelo Parábola con los parámetros a y b . El área sombreada indica el rango de ajuste de mínimos cuadrados para modelar la cresta sinusoidal. (b) Máscara cosenoidal para la Correlación cruzada.

El primer modelo utilizado para la detección de ondas planas en los datos medidos por el micrófono es una parábola que tiene el correspondiente desfase tanto en el eje de tiempo y el eje del micrófono. Ya que sólo estamos buscando los máximos correspondientes a las formas coseno en la imagen, se define un modelo de parábola de mínimos cuadrados ajustada a un rango de ángulos longitud π :

$$t_n = a + 0.4185 \left(\frac{Rf_s}{c} \right) (\theta_n - b)^2$$

donde a es el eje de desplazamiento de tiempo y b es el ángulo del micrófono tomado de referencia. Por lo tanto, a representa el momento en que la onda plana llega al micrófono con un ángulo $\theta_n = b = \theta_i$, que es el ángulo del micrófono correspondiente al DOA. En la Figura 2

(a) se muestra el modelo de parábola para $a = 0$ y $b = \frac{\pi}{2}$. La transformada se calcula mediante el uso de un conjunto de puntos resultantes a partir de un proceso de umbralización de amplitud. A partir de aquí, el espacio acumulado de Hough se digitaliza de nuevo mediante el uso de otro umbral con el objetivo de encontrar los máximos locales correspondientes a los casos más similares de forma.

3.2 Correlación cruzada

Un método alternativo de detección se basa en la correlación cruzada de una máscara en forma de coseno circularmente-desplazada con una imagen binaria a partir de la imagen original. La máscara de correlación debe coincidir con las propiedades de las curvas que se encuentran en las representaciones de las respuestas al impulso, y que están caracterizadas por la Eq. (1). Para este fin, se utilizan los parámetros específicos del array de micrófonos para construir una máscara en forma de coseno adaptada a los parámetros de procesado. Por otra parte, puesto que se utilizan micrófonos cardioides, la amplitud registrada por cada micrófono se puede modelar como

$$A_n(\theta_i) = 0.5 + 0.5 \cos(\theta_n - \theta_i)$$

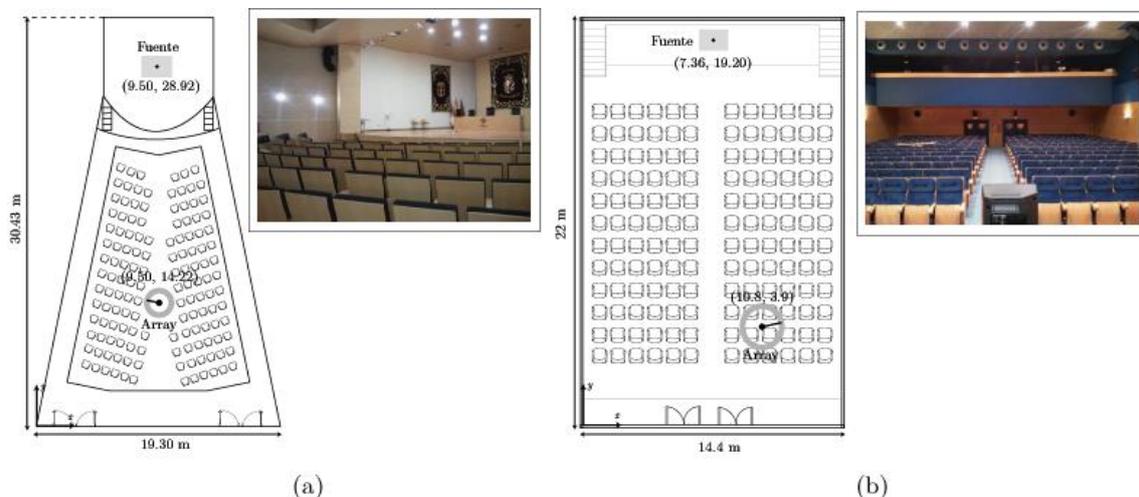


Figura 3. Fotografías y planos de las salas medidas. (a) Sala I- Auditorio. (b) Sala II- Sala de Conferencias.

De esta forma, la máscara de correlación \mathbf{M} y de dimensiones $\left[\left[f_s \frac{2R}{c} \right], N \right]$ se completará de la siguiente forma:

$$M(t, n) = \begin{cases} A_n(0) & \text{if } t = \left[f_s \frac{R}{c} (1 - \cos(\theta_n)) \right] + 1 \\ 0 & \text{otros} \end{cases} \quad \forall t, n.$$

En la Figura 2 (b) se muestra un ejemplo de esta máscara de correlación para un array de 72 micrófonos. De manera análoga a la detección de picos con el método Hough, los máximos locales en el método de correlación cruzada deben de ser obtenidos siguiendo a continuación otro paso de umbralización.

4 EXPERIMENTOS

En esta sección se presenta la aplicación de los métodos de detección de ondas mencionados anteriormente, para las respuestas al impulso obtenidas en dos salas reales localizadas en la Universidad de Castilla-La Mancha. En primer lugar, se describen las salas y el set-up experimental. A continuación, se analizan las respuestas al impulso medidas para detectar las principales reflexiones de las salas.

4.1 Salas analizadas

Sala I-Auditorio. En la Figura 3 (a), se muestran tanto el plano como las dimensiones de la sala nombrada como Auditorio. El tiempo de reverberación en la banda de 1000 Hz, fue de aproximadamente 1,41 s. Por otra parte, su forma es significativamente diferente a la de la Sala II, ya que tiene una planta en forma de trapecio.

Sala II-Sala de Conferencias. Esta sala, tiene forma rectangular con la excepción de una galería que está construida a mitad de altura de la misma, destinada al control de los equipos audiovisuales. Sus dimensiones y plano se muestran en la Figura 3 (b). En este caso, el tiempo de reverberación medido fue de 1,23 s y los materiales de los que constan sus paredes son la madera y el mármol.

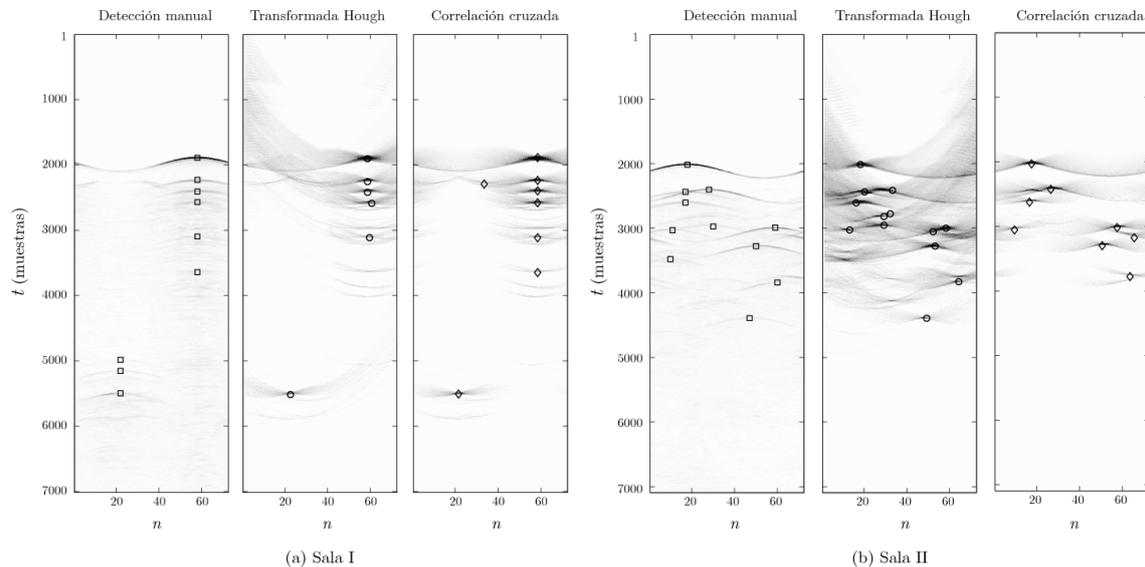


Figura 4. Ecogramas detectados mediante inspección visual y de forma automática utilizando la Transformada de Hough y la correlación cruzada. (a) Sala I. (b) Sala II. I. (b) la Sala II.

4.2 Método de medida

Las medidas se llevaron a cabo con un mismo altavoz colocado en el centro de las salas analizadas, y el array de micrófonos posicionado en el área de audiencia. Todas las salas estaban vacías y no se modificaron a la hora de realizar las mediciones. El array estaba compuesto por dos micrófonos cardioides sujetos a los extremos de una varilla larga de 2 m de longitud que a su vez estaba anclada en una placa giratoria con desplazamiento circular. Las medidas con el array se realizaron tomando muestras automáticas para todas las posiciones uniformemente distribuidas sobre el círculo de 2 m de diámetro dentro del área de escucha. Como fuente de excitación, se utilizaron secuencias de máxima longitud (Maximum Length Sequences, MLS) [12] con una frecuencia de muestreo de $f_s = 44100$ Hz. La señal fuente era controlada por medio de un ordenador portátil con el interface de audio Fast Track Pro de M-Audio y procesado con Matlab. Dado que las condiciones de la sala (temperatura, humedad, etc) no cambiaron significativamente durante la serie de mediciones, pudo suponerse que las medidas de todas las respuestas al impulso se tomaron en las mismas condiciones que si hubieran sido medidas de forma simultánea por un array de $N = 72$ micrófonos completo.

4.3 Detección de onda plana

Las diferentes respuestas al impulso capturadas, sus correspondientes transformadas de Hough y sus máscaras después de haber aplicado sobre ellas una correlación cruzada, se muestran en la Figura 4. La detección de las principales reflexiones en las medidas ha sido obtenida aplicando los siguientes métodos de detección de onda:

- *Detección manual*: Las ondas planas son seleccionadas manualmente por medio de un análisis visual. Para hacer esta selección, se preguntó a cinco personas diferentes familiarizadas con el procesamiento de imágenes para que identificaran de forma visual los picos que consideraran más significativos. Los cuadrados que aparecen en la parte izquierda de las Figuras 4 (a) y 4 (b) denotan estas reflexiones finales seleccionadas, que fueron al menos comunes por parte de cuatro de las cinco personas preguntadas. A pesar de no ser una

técnica muy precisa, estos valores manualmente seleccionados serán utilizados para comparar el funcionamiento de los otros métodos automáticos de detección.

- *Detección por Transformada de Hough*: Los eventos de onda son automáticamente detectados utilizando la Transformada de Hough descrita en la Sección 3.1. Estas reflexiones se muestran como círculos en los resultados centrales de las Figuras 4 (a) y 4 (b).

- *Detección basada en una correlación cruzada*: En este caso, las ondas planas son detectadas de forma automática siguiendo el procedimiento descrito en la Sección 3.2. Las reflexiones obtenidas se muestran como rombos en la parte derecha de las Figuras 4 (a) y 4 (b).

5 RESULTADOS

Los resultados obtenidos por los diferentes métodos de detección de ondas, están resumidos en la Tabla 1, donde se muestran el número de ondas planas detectadas correctamente (Correctas), el número de falsos positivos (FP), el número de falsos negativos (FN), y el error medio absoluto (Mean Absolute Error) tanto en tiempo (MAE_t) como en ángulo (MAE_θ). Estos errores sólo se han calculado para los puntos detectados correctamente. Por otra parte, en general, ambos métodos muestran una precisión relativamente buena en términos de MAE, donde el método basado en una correlación cruzada, parece proporcionar menor desviación en la magnitud angular, lo cuál es muy importante. Sin embargo, el número de ondas detectadas correctamente, falsos positivos y falsos negativos es muy dependiente de la sala analizada, por lo que es bastante difícil de evaluar en este estudio preliminar cuál de los dos métodos tiene un mejor rendimiento. Por lo tanto, es necesario seguir trabajando para analizar los efectos de los diferentes parámetros que intervienen en el proceso, como por ejemplo la robustez de diferentes tipos de salas, el rendimiento medio a partir de los datos obtenidos por las respuestas al impulso y el uso de modelos idóneos para la caracterización de ondas sonoras provenientes de diferentes alturas...

Tabla 1. Resultados de los métodos automáticos de detección

	Sala I					Sala II				
	Correctas	FP	FN	MAE_t	MAE_θ	Correctas	FP	FN	MAE_t	MAE_θ
Hough	6	0	3	4.0	4.2	10	3	1	6.0	10.0
Correlación	7	1	2	12.4	0.7	7	1	4	23.7	5.7

6 CONCLUSIÓN

En este trabajo, se han analizado dos técnicas de procesamiento de imágenes en el análisis de campos sonoros mediante arrays circulares de micrófonos: uno de ellos basado en la Transformada de Hough y otro basado en una correlación cruzada. A pesar de que la detección precisa de las reflexiones de onda plana no es trivial y en ocasiones es complicada, se ha podido demostrar que estos métodos son capaces de identificar muchas reflexiones de forma automática, cometiendo errores relativamente pequeños en tiempo y ángulo. No obstante y aunque el uso de estas técnicas es prometedor para el análisis de los campos sonoros, es

necesario todavía trabajar en esa línea para poder explorar y entender mejor sus beneficios y limitaciones.

Agradecimientos

Este trabajo, está apoyado bajo el proyecto TEC2012-37945-C02-02 por el Ministerio de Economía y Competitividad de España.

Referencias

1. Tervo, S., Tossavainen, T.: 3D room geometry estimation from measured impulse responses. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan (March 2012) 513–516
2. Tervo, S., Korhonen, T., Lokki, T.: Estimation of reflections from impulse responses. In: Proceedings of the International Symposium on Room Acoustics (ISRA). (2010)
3. Lokki, T.: Diffuseness and intensity analysis of spatial impulse responses. In: Proceedings of Acoustics'08, Paris, France (June 2008) 6495–6500
4. Vries, D., Baan, J.: Auralization of sound fields by wave field synthesis. In: Proceedings of the AES 106th Convention, Munich, Germany (2009) Paper No. 4927.
5. Vries, D., Hulsebos, E.: Auralization of room acoustics by wave field synthesis based on array measurement of impulsive responses. In: Proceedings of the XII European Signal Processing Conference (EUSIPCO), Vienna, Austria (September 2004) 1377–1380
6. Hulsebos, E.: Auralization using Wave Field Synthesis. PhD thesis, Delft University of Technology, The Netherlands (October 2004)
7. Hulsebos, E., Vries, D., Bourdillat, E.: Improved microphone array configurations for auralization of sound fields by wave field synthesis. *J. Acoust. Soc. Am.* 50 (October 2002) 779–790
8. Neely, S., Allen, J.: Invertibility of a room impulse response. *J. Acoust. Soc. Am.* 66 (1979) 165–169
9. Kuttruff, H.: Room acoustics. Taylor & Francis, Abingdon, Oxford, UK (October 2000) 368 pages.
10. Hough, P.V.C.: Machine analysis of bubble chamber pictures. In: Proceedings of the International Conference on High Energy Accelerators and Instrumentation. (1959)
11. Duda, R.O., Hart, P.E.: Use of the hough transformation to detect lines and curves in pictures. *Comm. ACM* 15(1) (1972) 11–15
12. Kite, T.D.: Techniques in audio and acoustic measurement. *J. Acoust. Soc. Am.* 114 (2003) 2381–2381