



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

ANÁLISIS DE LA ACÚSTICA DE SALAS  
MEDIANTE LA UTILIZACIÓN DE ARRAYS  
DE MICRÓFONOS CIRCULARES

TESIS DOCTORAL

DEPARTAMENTO DE COMUNICACIONES

AUTORA: ANA MARÍA TORRES ARANDA

DIRECTORES: DR. JOSÉ JAVIER LÓPEZ MONFORT

DR. BASILIO PUEO ORTEGA

17 DICIEMBRE 2012

# Capítulo 1

## Introducción

---

<b>1.1. Introducción . . . . .</b>	<b>2</b>
<b>1.2. Motivación . . . . .</b>	<b>4</b>
<b>1.3. Objetivos de la tesis . . . . .</b>	<b>5</b>
<b>1.4. Organización de la tesis . . . . .</b>	<b>7</b>

---

*En este primer capítulo se establece el contexto en el que se desarrolla el presente trabajo, así como los motivos que dieron lugar al mismo. A continuación, tras la presentación de la hipótesis principal, se enumeran los objetivos perseguidos y se realiza un breve resumen del contenido de cada uno de los capítulos, que trata de dar una visión global del trabajo realizado.*

## 1.1. Introducción

Desde la antigüedad, la acústica de los recintos dedicados a la oratoria, la música o el espectáculo en general era especialmente cuidada por los ingenieros y arquitectos que participaban en su construcción, conscientes de la importancia que tenía en el resultado final de su obra. Griegos y romanos construyeron teatros y circos con unas cualidades acústicas tan buenas que hasta hoy nos siguen sorprendiendo. En aquella época no existía una física formalizada del sonido (eso vendría en el siglo XVI), por lo que los diseños se hacían atendiendo al conocimiento acumulado de experiencias anteriores y a la prueba y error.

Los escritos más antiguos que se conocen sobre acústica arquitectónica datan del siglo I a. C., y se deben a Marco Vitrubio Polio, ingeniero militar de Julio César. En estos escritos describen varios diseños para la acústica de los antiguos teatros romanos. Por ejemplo, se utilizaban vasijas de bronce afinadas que actuaban como resonadores, bajos o agudos. Por su parte, los griegos tomaron en cuenta la geometría de sus construcciones para obtener una mejor acústica, donde, por ejemplo las gradas de sus teatros servían a su vez como grandes reflectores. Así, el sonido directo y reflejado se sumaba y por tanto aumentaba la sonoridad final del recinto por medios naturales. Actualmente, el legado dejado por las civilizaciones griega y romana sigue siendo válido y puede experimentarse en multitud de recintos abiertos. En la actualidad muchas construcciones siguen utilizando el modelo clásico que dejaron los griegos y romanos, sólo que en espacios cerrados cuyo factor fundamental de análisis es el tiempo que el sonido queda confinado en ese recinto.

Pero no sólo en Europa se tenían conocimientos de acústica arquitectónica, la civilización Maya, 1000 años antes de la formalización de la física moderna construyó en la ciudad de Chichén Itzá la pirámide de Kukulcán, donde se utiliza el fenómeno de difracción de una manera asombrosa a través de una de sus escalinatas para sintetizar el sonido del ave Quetzal. Otros efectos acústicos de amplificación de sonido para proyectarlo a grandes distancias se dan también en el juego de pelota de la misma ciudad o en otros templos de la ciudad

de Palenque en Chiapas.

Hasta el siglo XIX, el diseño acústico era puramente práctico y consistía principalmente, en imitar diseños anteriores que habían resultado satisfactorios. En la Edad Media incluso se consolidaron, no sólo la imitación, sino prácticas casi supersticiosas. La acústica arquitectónica moderna nació a finales del siglo XIX gracias al físico americano Wallace Clement Sabine. Entre otras aportaciones, Sabine estableció un modelo estadístico para el tiempo de reverberación de una sala que lo relaciona con su volumen  $V$  y su superficie absorbente efectiva  $S$  [1].

$$t_R = \frac{0,162V}{S} = \frac{0,162V}{\sum \alpha_i S_i}, \quad (1.1)$$

La fórmula de Sabine se suele utilizar sólo a modo de estimación, pero sigue siendo un buen punto de partida para valorar el tiempo de reverberación de una sala, que modifica de forma importante sus cualidades acústicas. Para que la sonoridad sea la adecuada, el tiempo no debe ser alto ni bajo, sino ajustarse al uso que tendrá la sala. Así, salas con tiempos bajos o «secas» pueden ser aptas para teatro o palabra hablada, pero poco adecuadas para la audición de música.

Sin embargo, conforme se ha ido avanzando en la física del sonido y la ingeniería, este principio básico de diseño atendiendo sólo al tiempo de reverberación ha resultado ser insuficiente para describir completamente el comportamiento acústico de una sala. Así, nos encontramos salas con tiempos de reverberación similares, donde en unas la música clásica suena adecuadamente y en otras donde suena muy pobre. Por tanto se ha visto necesario recurrir a parámetros que describan mejor las propiedades de una sala. La norma ISO 3382 [2] introduce estos parámetros adicionales como la claridad, ( $C80$ ), definición ( $D50$ ) o el tiempo de caída temprana ( $EDT$ ), que permiten describir mejor su comportamiento acústico. Todos ellos se pueden derivar de lo que se conoce como respuesta impulsional de la sala ( $RIR$ ), que no es más que el modelo de señal que describe la sala cuando se considera un sistema de transmisión lineal e invariante en el tiempo ( $LTI$ ) entre dos puntos de la misma.

Las  $RIRs$  pueden describir acústicamente a una sala proporcionando un correcto detalle de los cambios que sufre una señal sonora cuan-

do viaja de un punto a otro de la misma, [3], [4], [5]. Desde la década de los 80 las técnicas de medida precisa de la RIR como el método de las secuencias de longitud máxima (MLS), [6], o la integración de pulsos o *Sweep* Logarítmico [7] han popularizado el uso de la RIR y la extracción de parámetros a través de la misma.

Durante mucho tiempo las salas se han descrito completamente mediante sus respuestas impulsionales en lugares representativos como el escenario o ciertos puntos del patio de butacas. El análisis del comportamiento en estos lugares se dió como representativo del total del recinto de tal forma que los cambios o ajustes se realizaban para satisfacer los criterios en esos puntos.

Sin embargo, el campo sonoro de un recinto es lo suficientemente complejo como para que una RIR puntual de uno o varios lugares dispersos de la sala no sea suficiente para definir algunos parámetros ligados a la componente espacial del campo acústico. Hay que tener en cuenta que el sistema de audición humano es más complejo que un receptor puntual como lo es un micrófono. El ser humano posee un sistema de audición binaural que le capacita para discriminar direccionalmente y hasta cierto punto las señales que llegan desde distintas direcciones. El conjunto de las reflexiones en una sala puede llegar a ser extremadamente complejo, pero además la forma en la que nuestro cerebro interpreta dichas reflexiones para obtener una sensación de la espacialidad de la sala es todavía más compleja. La función de transferencia relacionada con la cabeza (HRTF) [8], [9], modela hasta cierto punto las interacciones del sistema auditivo humano externo con el entorno. El estudio de la HRTF es una disciplina compleja e interesante, y por tanto muy tratada en la literatura científica, aunque no es el tema principal de esta tesis.

## 1.2. Motivación

De la discusión anterior se extrae que el comportamiento acústico de una sala teniendo en cuenta todos los factores espaciales no es trivial. Existen diferentes técnicas de capturar este campo acústico siendo la captura del campo mediante un array circular de micrófonos y su posterior descomposición de ondas planas una de las que mayor

aceptación ha tenido en los últimos años [10], [11], [12].

La metodología y el aparato matemático para el empleo de estas técnicas han sido propuestos y validados en la literatura científica, como atestiguan las citas anteriores, así como algunos ejemplos de capturas reales en salas. Sin embargo los resultados de estas medidas, no han sido explotados convenientemente para obtener conclusiones relevantes sobre la acústica final de dichas salas y tampoco se han relacionado con los aspectos geométricos de las mismas.

En general, se puede afirmar que actualmente no existe un estudio exhaustivo experimental sobre el análisis de la acústica de salas utilizando técnicas como las mencionadas anteriormente basadas en descomposición de ondas planas. Por lo tanto, se decidió iniciar la presente tesis doctoral para explorar las prometedoras aplicaciones en un campo que suscita tanta importancia como el análisis acústico de salas.

El proceso experimental se basa en la toma de una muestra sonora mediante el uso de arrays circulares, con la que realizar a posteriori la preceptiva descomposición de ondas planas. Con la información capturada por tal array se podrían aplicar otras técnicas avanzadas, como la de *beamforming*, que, junto con la descomposición de ondas planas, permitiría obtener, analizar y comparar la acústica de las salas (reflexiones, absorciones, etc) permitiendo extrapolar una serie de conclusiones sobre sus características, comportamiento y calidad.

### 1.3. Objetivos de la tesis

Teniendo en cuenta el contexto anterior, el principal objetivo de la tesis se describe de la siguiente forma:

*“Contribuir al estudio de la acústica de salas mediante medidas con arrays circulares de micrófonos, realizando un análisis exhaustivo de los resultados de la aplicación de las técnicas de descomposición en ondas planas del campo sonoro y de las técnicas de beamforming basadas en procesado modal, y analizar los ecogramas y descomposiciones de ondas a partir del desarrollo de un algoritmo de localización de reflexiones y fuentes sonoras basado en morfología de imagen”.*

Del objetivo principal se desprenden una serie de objetivos particulares concretos que se pueden enumerar en:

- Construir un dispositivo de captura de sonido mediante un array circular, verificar su funcionamiento y ponerlo en práctica midiendo una serie de salas reales que sean representativas.
- Implementar todas las etapas del algoritmo de descomposición de ondas planas aplicándolo a medidas de salas reales obtenidas con el array circular.
- Implementar un algoritmo alternativo de análisis de las capturas del array basado en técnicas de *beamforming* modal para arrays circulares.
- Analizar las medidas obtenidas por el método de descomposición en ondas planas, intentando asociar los resultados obtenidos con la geometría de las diferentes salas analizadas para corroborar el buen funcionamiento de estos métodos.
- Identificar en los ecogramas y descomposiciones de ondas las reflexiones significativas de la sala desarrollando métodos específicos de búsqueda de máximos locales basados en técnicas de morfología de imagen.
- Comparar los resultados obtenidos por los métodos implementados, destacando las ventajas e inconvenientes de cada uno de ellos así como indicar qué método es más conveniente para cada aplicación o parámetro a analizar.
- Como objetivo adicional al margen del análisis de la acústica de salas, aplicar el análisis modal para la localización de fuentes acústicas bajo condiciones adversas de las salas, es decir, con alto coeficiente de reverberación, y baja relación señal a ruido.

## 1.4. Organización de la tesis

El resto de esta tesis describe la investigación que se ha llevado a cabo para desarrollar los objetivos mencionados anteriormente. Los capítulos están organizados y presentados con el siguiente contenido:

- El Capítulo 2 introduce el estado del arte y la teoría necesaria para establecer el marco de trabajo de la descomposición de onda plana mediante armónicos cilíndricos. En él se hace un repaso de los trabajos ya realizados sobre esta materia y se presenta el método matemático implementado para llevar a cabo las medidas que serán presentadas en los siguientes capítulos.
- El Capítulo 3 recoge los fundamentos teóricos del *beamforming* de forma general y del *beamforming* modal de manera específica. De nuevo se establece el marco teórico y estado del arte de la técnica implementada así como la base matemática y metodología que se ha realizado para el resto de este trabajo.
- El Capítulo 4 presenta un algoritmo de detección de ondas basado en morfología de imagen que permitirá identificar y separar eventos significativos a partir de las respuestas al impulso capturadas por un array circular de micrófonos. Este algoritmo será necesario en el análisis de ecogramas y descomposición de ondas planas obtenidos en las mediciones de las salas reales presentadas en el siguiente capítulo.
- El Capítulo 5 presenta un exhaustivo estudio experimental basado en descomposición de onda plana mediante la medida y el análisis de diferentes salas reales. En este estudio se comprobará la operatividad de los métodos de descomposición de onda plana y de detección de ondas para identificar y separar reflexiones, estableciendo un marco de trabajo práctico en estudios de acústica. Estos resultados también podrían aplicarse en el futuro a acústica virtual o reproducción de sonido 3D.



- En el Capítulo 6 se realiza una aplicación particular del *beam-forming* modal para localización de fuentes sonoras en condiciones acústicas adversas. Los resultados obtenidos en diferentes pruebas experimentales, tanto simuladas como reales, así como varias comparaciones con otros métodos afines en la literatura avalan el método implementado en esta tesis.
- Por último, el Capítulo 7 recoge las conclusiones obtenidas en toda esta tesis y propone las líneas de trabajo futuras pertinentes.

## Capítulo 2

# Descomposición en onda plana con arrays de micrófonos circulares

---

<b>2.1. Fundamentos teóricos del campo de onda . . . . .</b>	<b>11</b>
2.1.1. Teoría ondulatoria . . . . .	11
2.1.2. Integral de Kirchhoff-Helmholtz . . . . .	13
2.1.3. Integral de Rayleigh . . . . .	15
2.1.4. Integrales 2D de Kirchhoff-Helmholtz y Rayleigh	18
<b>2.2. Extrapolación del campo de onda . . . . .</b>	<b>21</b>
<b>2.3. Descomposición en armónicos cilíndricos . . . . .</b>	<b>23</b>
<b>2.4. Descomposición en onda plana . . . . .</b>	<b>29</b>
<b>2.5. Auralización . . . . .</b>	<b>30</b>
<b>2.6. Modelo de array circular de micrófonos . . . . .</b>	<b>32</b>
<b>2.7. Aliasing espacial y efectos de baja frecuencia . . . . .</b>	<b>35</b>
<b>2.8. Representaciones gráficas de la descomposición . . . . .</b>	<b>36</b>
2.8.1. Representación espacio-tiempo . . . . .	36
2.8.2. Descomposición en armónicos cilíndricos . . . . .	37
2.8.3. Descomposición frecuencia-espacio . . . . .	37

2.8.4. Descomposición tiempo-espacio . . . . .	39
<b>2.9. Grabación mediante un array circular de micrófonos</b>	<b>39</b>
2.9.1. Simulaciones . . . . .	41
2.9.2. Comprobación experimental . . . . .	56
2.9.3. Discusión de los resultados . . . . .	63

---

*En este capítulo se analizará la teoría general en la que se basa la descomposición de ondas planas mediante armónicos cilíndricos. Primeramente se dará información sobre la ecuación de onda, origen de toda la base matemática que se presentará en el capítulo. Se abordará brevemente las diferentes configuraciones en array para llevar a cabo la descomposición, implementada bajo una estructura en array circular. Por último, se tratará el método de descomposición con armónicos cilíndricos, mostrando diferentes simulaciones así como varias pruebas reales para corroborar la correcta funcionalidad del método que se llevará a cabo en el resto de este trabajo.*

## 2.1. Fundamentos teóricos del campo de onda

La descomposición del campo de onda es una técnica que divide el campo de onda en soluciones propias ortogonales de la ecuación de onda acústica en un sistema de coordenadas adaptable a la apertura o array bajo estudio. Aquí, las aperturas son dispositivos distribuidos espacialmente para analizar las características espacio-temporales de los campos de onda. Aunque las aperturas pueden ser procesadores continuos en el espacio tridimensional, en este caso, serán aperturas continuas muestreadas en puntos discretos en el espacio, conocidas como arrays. De manera general existen aperturas lineales, planares, circulares y esféricas, pero para el método de descomposición de ondas planas (PWD), en esta tesis se utilizarán arrays circulares, los cuáles son particularmente idóneos para la captura y descripción del campo de onda en dos dimensiones debido a sus propiedades espaciales en  $360^\circ$  de acimut. Además, los arrays circulares, se pueden implementar de forma óptima en situaciones donde el campo de onda es invariante con respecto al eje  $z$ , como por ejemplo en salas donde las reflexiones del techo y suelo, están suficientemente atenuadas [13].

En este capítulo, a partir de la teoría ondulatoria, se detalla la descomposición de onda plana en armónicos cilíndricos bajo una apertura o array circular. Así mismo, se plantean una serie de medidas simuladas y experimentales comprobando la operatividad del método que se utilizará en posteriores capítulos.

### 2.1.1. Teoría ondulatoria

De forma general, se define sonido a cualquier fenómeno que involucre la propagación en forma de ondas elásticas (sean audibles o no), generalmente a través de un fluido (u otro medio elástico) que esté generando el movimiento vibratorio de un cuerpo. El sonido audible consiste en ondas mecánicas que producen oscilaciones de la presión del aire, que son convertidas en movimiento en el oído humano y percibidas por el cerebro. Al propagarse, el sonido transporta energía pero no materia. Las vibraciones se generan en la misma dirección en la que se propaga el sonido: se trata, por lo

tanto, de ondas longitudinales, [14]. Esta propagación de vibraciones, puede ser expresada por medio de ecuaciones, que conducen a la ecuación de onda acústica. La presión y la velocidad de partículas serán dos parámetros fundamentales y básicos, describiéndose su comportamiento por la presión,  $p$  en una posición determinada  $\mathbf{r}$  y medida en  $[\text{N/m}^2]$  o pascales [Pa] y la velocidad de partícula  $\mathbf{v}$  cuya magnitud se mide en  $[\text{m/s}]$ .

Las ecuaciones linealizadas de la ecuación de presión y velocidad, vienen dadas por:

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = 0, \quad (2.1)$$

$$\nabla^2 \cdot \mathbf{v}(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 \mathbf{v}(\mathbf{r}, t)}{\partial t^2} = 0, \quad (2.2)$$

donde  $\cdot$  representa el producto escalar,  $c$  es la velocidad del sonido, determinada por la densidad  $\rho_0$  y el módulo  $K$  del medio, [15]

$$c = \sqrt{\frac{K}{\rho_0}}, \quad (2.3)$$

y el operador Nabla que viene definido por [16]

$$\Delta = \nabla \cdot \nabla = \nabla^2. \quad (2.4)$$

De tal forma, la ecuación de onda para la presión, llega a ser

$$\nabla^2 p(\mathbf{r}, t) + \frac{\rho_0}{K} \frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = 0. \quad (2.5)$$

Muchos cálculos en teoría de arrays se realizan, como se verá más adelante en esta tesis, en el dominio de la frecuencia. Una de las razones es que la convolución en el dominio del tiempo se simplifica a una multiplicación de señales en el dominio de la frecuencia.

A partir de la segunda ley de Newton, y la ley de Hooke para los medios de compresión, es mucho más fácil obtener la ecuación de onda tal y como se muestra en la siguiente derivación:

$$-\nabla P(\mathbf{r}, \omega) = j\omega\rho V(\mathbf{r}, \omega), \quad (2.6)$$

$$-\nabla \cdot V(\mathbf{r}, \omega) = \frac{j\omega}{K} P(\mathbf{r}, \omega). \quad (2.7)$$

donde  $P(\mathbf{r}, \omega)$  y  $V(\mathbf{r}, \omega)$  representan la presión y el vector velocidad de partícula en el dominio frecuencial. Combinando estas dos ecuaciones en este dominio se obtiene la ecuación de onda de la presión:

$$\nabla^2 P(\mathbf{r}, \omega) + \frac{\rho_0 \omega^2}{K} P(\mathbf{r}, \omega) = 0, \quad (2.8)$$

y siendo  $k$  el número de onda, y  $\omega$  la velocidad angular

$$k = \frac{\omega}{c}, \quad (2.9)$$

se llega finalmente a esta ecuación de onda para la presión sonora en el dominio de la frecuencia

$$\nabla^2 P(\mathbf{r}, \omega) + k^2 P(\mathbf{r}, \omega) = 0. \quad (2.10)$$

### 2.1.2. Integral de Kirchhoff-Helmholtz

La integral de Kirchhoff-Helmholtz tomará un papel importante dentro de la descomposición de onda que se verá más adelante. Para derivarla es necesario utilizar el segundo teorema de Green:

$$\int_V f \nabla^2 g - g \nabla^2 f dV = \oint_S \left( g \frac{\partial f}{\partial n} - f \frac{\partial g}{\partial n} \right) dS, \quad (2.11)$$

donde  $f$  y  $g$  son dos funciones que se definirán a continuación.

Supóngase ahora la geometría mostrada en la Figura 2.1 donde  $V$  es un volumen de un medio que no contiene fuentes sonoras en su interior ya que todas están fuera de dicho volumen y  $\mathbf{n}$  es un vector normal unitario apuntando hacia su interior.

Sustituyendo  $f$  por la presión del campo sonoro  $P(\mathbf{r}, \omega)$  dentro del volumen, y  $g$ , por una función de Green  $G(\mathbf{r}|\mathbf{r}_R, \omega)$  definida como:

$$G(\mathbf{r}|\mathbf{r}_R, \omega) = \frac{e^{-jk|\mathbf{r}-\mathbf{r}_R|}}{|\mathbf{r}-\mathbf{r}_R|}, \quad (2.12)$$

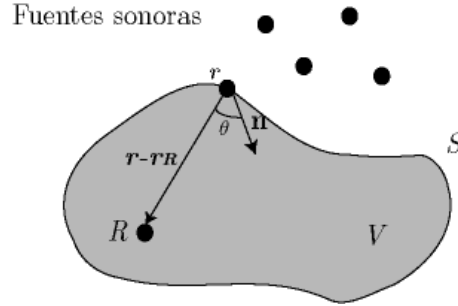


Figura 2.1: Geometría utilizada para el segundo teorema de Green: un volumen  $V$  rodeado por una superficie  $S$ .

la ecuación de onda que se satisface en  $V$ :

$$\nabla^2 P(\mathbf{r}, \omega) + k^2 P(\mathbf{r}, \omega) = 0, \quad (2.13)$$

$$\nabla^2 G(\mathbf{r}|\mathbf{r}_R, \omega) + k^2 G(\mathbf{r}|\mathbf{r}_R, \omega) = -4\pi\delta(\mathbf{r} - \mathbf{r}_R), \quad (2.14)$$

hace que el segundo teorema de Green, quede como:

$$-4\pi \int_V P(\mathbf{r}, \omega) \delta(\mathbf{r} - \mathbf{r}_R) dV = \quad (2.15)$$

$$\oint_S \left( P(\mathbf{r}, \omega) \frac{\partial G(\mathbf{r}|\mathbf{r}_R, \omega)}{\partial n} - G(\mathbf{r}|\mathbf{r}_R, \omega) \frac{\partial P(\mathbf{r}, \omega)}{\partial n} \right) dS,$$

y utilizándose la Ecuación (2.6), se puede simplificar a:

$$P(\mathbf{r}_R, \omega) = \frac{1}{4\pi} \oint_S \left( P(\mathbf{r}, \omega) \frac{\partial G(\mathbf{r}|\mathbf{r}_R, \omega)}{\partial n} + jk\rho_0 c V_n(\mathbf{r}, \omega) G(\mathbf{r}|\mathbf{r}_R, \omega) \right) dS. \quad (2.16)$$

Esta ecuación significa que  $P(\mathbf{r}_R, \omega)$  puede ser calculada dentro del volumen  $V$ , si la presión del campo y la velocidad normal de partícula en los límites de la superficie  $S$  son conocidas. La función de Green no es única, de hecho, cualquier solución de una ecuación de onda homogénea en  $V$ , puede añadirse a la función de Green y

proporcionar una nueva función de Green que todavía satisface a la Ecuación (2.13). La integral (2.16), puede tomar además diferentes formas, aunque todas ellas obtendrán idéntica solución permitiendo derivar las consiguientes funciones fuente de un monopolo y dipolo en la superficie, cuyas ecuaciones respectivas son [15]

$$S_{\text{monopolo}}(\mathbf{r}, \omega) = jk\rho_0cV_n(\mathbf{r}, \omega), \quad (2.17)$$

$$S_{\text{dipolo}}(\mathbf{r}, \omega) = P(\mathbf{r}, \omega), \quad (2.18)$$

La Ecuación (2.16), forma la base para la extrapolación del campo de onda que será utilizada en el capítulo siguiente.

### 2.1.3. Integral de Rayleigh

En principio, dado que las integrales de Kirchoff-Helmholtz son bastante generales, éstas se pueden simplificar a integrales Rayleigh si, en lugar de medir presión y velocidad en una curva cerrada o infinita cualquiera, las medimos en una línea recta infinita. De esta forma se presentan dos simplificaciones: medir únicamente la velocidad normal en la línea, resultando la integral Rayleigh I y medir únicamente presión en la línea, dando lugar a la integral de Rayleigh II. A continuación se van a desarrollar ambas integrales.

Para poder simplificar la integral de Kirchoff se utiliza una función de Green diferente para poder satisfacer de manera más conveniente la condición de contorno. Si  $\partial G/\partial |\mathbf{n}| = 0$  se elige como condición de contorno en  $S$ , el primer término de la parte derecha de la Ecuación (2.16) desaparecería, y si se toma  $G=0$ , lo haría también el segundo término. La primera condición implica que  $S$  actúe como una perfecta superficie rígida reflectante. Bajo la segunda condición,  $S$  se comporta como una perfecta superficie reflectante.

Considerando la geometría de la Figura 2.2, el volumen  $V$  está encerrado por una superficie plana  $S_0$  y una esfera  $S_1$  con radio  $R$ . Las fuentes del campo sonoro están situadas en la mitad superior externa a  $V(z > z_s)$ .

Se considera ahora el caso  $\partial G/\partial |\mathbf{n}| = 0$  en  $S_0$ . Puesto que ahora la superficie  $S_0$  actúa como un perfecto reflector rígido, una función adecuada de Green sería la suma de dos monopolos iguales situados



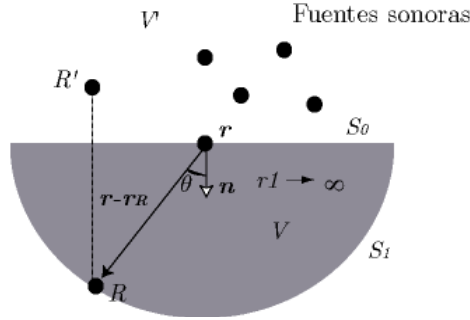


Figura 2.2: Geometría utilizada para las integrales de Rayleigh y la función de Green.  $R'$  es el punto espejo de  $R$  en la superficie  $S_0$ .

simétricamente con respecto a  $S_0$ , tal y como puede verse en la Figura 2.2.

$$G(\mathbf{r}|\mathbf{r}_R, \omega) = \frac{e^{-jk|\mathbf{r}-\mathbf{r}_R|}}{|\mathbf{r}-\mathbf{r}_R|} + \frac{e^{-jk|\mathbf{r}-\mathbf{r}'_R|}}{|\mathbf{r}-\mathbf{r}'_R|}. \quad (2.19)$$

Utilizando la función de Green y aproximando el radio de la esfera  $S_1$  a infinito, se puede mostrar que la integral sobre  $S_1$  desaparece y el teorema de Green se reduce a:

$$P(\mathbf{r}_R, \omega) = \frac{1}{2\pi} \int_{S_0} \left( jk\rho_0 c V_n(\mathbf{r}, \omega) \frac{e^{-jk|\mathbf{r}-\mathbf{r}_R|}}{|\mathbf{r}-\mathbf{r}_R|} \right) dS_0, \quad (2.20)$$

llamada integral de Rayleigh I. Esta ecuación indica que cualquier campo debido a fuentes presentes en la mitad espacial  $V'(z < z_S)$  puede ser reconstruido en la otra mitad espacial  $V(z > z_S)$  por medio de una distribución continua de monopolos en la superficie  $S_0$  con una fuente  $S_{\text{monopolo}}(\mathbf{r}, \omega)$ , tal y como muestra la Figura 2.3.

Hay que tener en cuenta que esta distribución de monopolos radiará un campo de onda "espejo" en la mitad espacial  $V'(z < z_s)$ , ya que no hay dipolos para cancelar este campo de onda. De esta forma, la solución de Rayleigh I podrá aplicarse para la síntesis, siempre que el campo de onda en el espacio  $V(z > z_s)$  no sea de interés, [17].

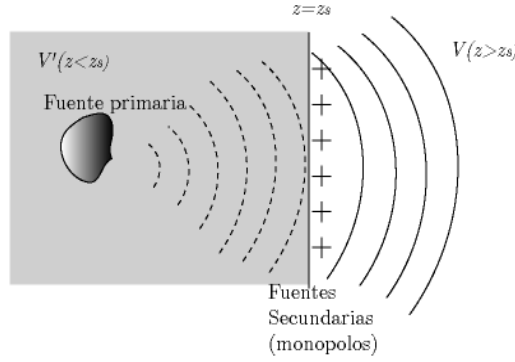


Figura 2.3: Síntesis del campo de onda por una distribución de monopolos en  $z=z_s$  donde la velocidad en el plano es conocida.

Este resultado es fundamental por ejemplo para las técnicas de holografía acústica, como la síntesis del campo de onda (WFS), las cuáles se basan en reconstruir el campo sonoro y la velocidad de vibración de un objeto o fuente sonora a partir de medidas con micrófonos, en un plano paralelo y cercano a la fuente. Las medidas se hacen normalmente en una superficie plana (plano del holograma), y estos datos se utilizan para reconstruir el campo tridimensional. Esta información se puede obtener mediante la utilización de una función de Green conocida. Además, el campo medido obedece a la ecuación de onda.

A continuación se considera el caso de  $S = 0$  en  $S_0$ . Puesto que la superficie  $S_0$  actúa como un perfecto reflector, la función de Green será la diferencia de dos monopolos iguales situados simétricamente con respecto a  $S_0$ , tal y como puede verse en la Figura 2.2.

$$G(\mathbf{r}|\mathbf{r}_R, \omega) = \frac{e^{-jk|\mathbf{r}-\mathbf{r}_R|}}{|\mathbf{r}-\mathbf{r}_R|} - \frac{e^{-jk|\mathbf{r}-\mathbf{r}'_R|}}{|\mathbf{r}-\mathbf{r}'_R|}. \quad (2.21)$$

Utilizando la función de Green y aproximando de nuevo el radio de la esfera  $S_1$  a infinito, se puede mostrar que el teorema de Green se reduce a:

$$P(\mathbf{r}_R, \omega) = \frac{1}{2\pi} \int_{S_0} \left( P(\mathbf{r}, \omega) \frac{1 + jk|\mathbf{r}-\mathbf{r}_R|}{|\mathbf{r}-\mathbf{r}_R|} \cos \theta \frac{e^{-jk|\mathbf{r}-\mathbf{r}_R|}}{|\mathbf{r}-\mathbf{r}_R|} \right) dS_0, \quad (2.22)$$

expresión llamada de Rayleigh II y que indica cualquier campo sonoro creado por fuentes posicionadas en la mitad espacial  $V'(z < z_s)$  puede ser reconstruido en la otra mitad espacial  $V(z > z_s)$  por medio de una distribución continua de dipolos en la superficie  $S_0$  con una fuente  $S_{\text{dipolo}}(\mathbf{r}, \omega)$ , como muestra la Figura 2.4. De nuevo hay que destacar que tal distribución de dipolos emitirá un campo de onda 'espejo' en el espacio  $V'(z < z_s)$ , cuya presión estará en oposición de fase con la presión en  $V(z > z_s)$ , [17].

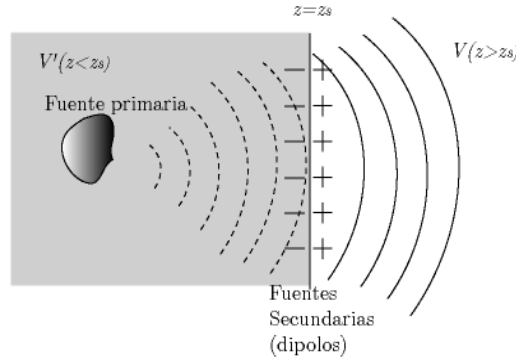


Figura 2.4: Síntesis del campo de onda por una distribución de dipolos en  $z=z_s$  donde la velocidad en el plano es conocida.

Por último, las integrales inversas de Rayleigh pueden ser derivadas utilizando las funciones de Green con exponenciales complejas conjugadas. El resultado de las integrales inversas de Rayleigh son:

$$P(\mathbf{r}_R, \omega) = \frac{1}{2\pi} \int_{S_0} \left( jk\rho_0 c V_n(\mathbf{r}_R, \omega) \frac{e^{-jk|\mathbf{r}-\mathbf{r}_R|}}{|\mathbf{r}-\mathbf{r}_R|} \right) dS_0, \quad (2.23)$$

$$P(\mathbf{r}_R, \omega) = \frac{1}{2\pi} \int_{S_0} \left( P(\mathbf{r}, \omega) \frac{1 - jk|\mathbf{r}-\mathbf{r}_R|}{|\mathbf{r}-\mathbf{r}_R|} \cos\theta \frac{e^{jk|\mathbf{r}-\mathbf{r}_R|}}{|\mathbf{r}-\mathbf{r}_R|} \right) dS_0. \quad (2.24)$$

#### 2.1.4. Integrales 2D de Kirchhoff-Helmholtz y Rayleigh

En dos dimensiones (2D) y considerando de nuevo la Figura 2.2, el segundo teorema de Green sigue siendo válido, salvo que

las funciones de Green a utilizar son diferentes a las anteriormente expuestas. Para el caso de dos dimensiones, las integrales Kirchhoff-Helmholz directa e inversa, utilizan las siguientes funciones de Green [18]:

$$G(\mathbf{r}|\mathbf{r}_R, \omega) = -j\pi H_0^{(2)}(k|\mathbf{r} - \mathbf{r}_R|), \quad (2.25)$$

$$G(\mathbf{r}|\mathbf{r}_R, \omega) = -j\pi H_0^{(1)}(k|\mathbf{r} - \mathbf{r}_R|), \quad (2.26)$$

donde  $H_n^{(i)}$  es la función de Hankel de orden  $n$  y especie  $i$ . Estas ecuaciones anteriores incluidas en la Ecuación (2.16), dan lugar a una integral 2D Kirchhoff-Helmholz directa,

$$P(\mathbf{r}_R, \omega) = \frac{-jk}{4} \oint_S (P(\mathbf{r}, \omega) \cos \theta H_1^{(2)}(k|\mathbf{r} - \mathbf{r}_R|) + j\rho c V_n(\mathbf{r}, \omega) H_0^{(2)}(k|\mathbf{r} - \mathbf{r}_R|)) dL, \quad (2.27)$$

e inversa también en dos dimensiones:

$$P(\mathbf{r}_R, \omega) = \frac{-jk}{4} \oint_S (P(\mathbf{r}, \omega) \cos \theta H_1^{(1)}(k|\mathbf{r} - \mathbf{r}_R|) + j\rho c V_n(\mathbf{r}, \omega) H_0^{(1)}(k|\mathbf{r} - \mathbf{r}_R|)) dL. \quad (2.28)$$

Por otra parte y en campo lejano  $kr \gg 1$ , las funciones de Hankel pueden ser simplificadas a:

$$H_n^{(1)}(k\mathbf{r}) \approx (-j)^n \frac{1 - j e^{-jk\mathbf{r}}}{\sqrt{\pi} \sqrt{k\mathbf{r}}}, \quad (2.29)$$

$$H_n^{(2)}(k\mathbf{r}) \approx j^n \frac{1 + j e^{-jk\mathbf{r}}}{\sqrt{\pi} \sqrt{k\mathbf{r}}}. \quad (2.30)$$

Incluyendo estas aproximaciones en las Ecuaciones (2.27) y (2.28), se obtiene unas integrales 2D directa e inversa de Kirchhoff-Helmholz para campo lejano:

$$\begin{aligned}
P(\mathbf{r}_R, \omega) &= \frac{1+j}{2} \sqrt{\frac{k}{\pi}} \oint_S \{P(\mathbf{r}, \omega) \cos \theta + \rho c V_n(\mathbf{r}, \omega)\} \frac{e^{-jk|\mathbf{r}-\mathbf{r}_R|}}{\sqrt{|\mathbf{r}-\mathbf{r}_R|}} dL, \\
P(\mathbf{r}_R, \omega) &= \frac{1-j}{2} \sqrt{\frac{k}{\pi}} \oint_S \{P(\mathbf{r}, \omega) \cos \theta - \rho c V_n(\mathbf{r}, \omega)\} \frac{e^{jk|\mathbf{r}-\mathbf{r}_R|}}{\sqrt{|\mathbf{r}-\mathbf{r}_R|}} dL.
\end{aligned} \tag{2.31}$$

Analizando las anteriores integrales, se puede comprobar que las integrales directas de Rayleigh I y II vienen dadas por las expresiones siguientes:

$$P(\mathbf{r}_R, \omega) = \frac{-jk}{2} \oint_S j \rho c V_n(\mathbf{r}, \omega) H_0^{(2)}(k|\mathbf{r}-\mathbf{r}_R|) dL, \tag{2.32}$$

$$P(\mathbf{r}_R, \omega) = \frac{-jk}{2} \oint_S P(\mathbf{r}, \omega) \cos \theta H_1^{(2)}(k|\mathbf{r}-\mathbf{r}_R|) dL, \tag{2.33}$$

e inversas:

$$P(\mathbf{r}_R, \omega) = \frac{-jk}{2} \oint_S j \rho c V_n(\mathbf{r}, \omega) H_0^{(1)}(k|\mathbf{r}-\mathbf{r}_R|) dL, \tag{2.34}$$

$$P(\mathbf{r}_R, \omega) = \frac{-jk}{2} \oint_S P(\mathbf{r}, \omega) \cos \theta H_1^{(1)}(k|\mathbf{r}-\mathbf{r}_R|) dL, \tag{2.35}$$

y si se consideran las aproximaciones de campo lejano, se tienen las siguientes ecuaciones directas:

$$P(\mathbf{r}_R, \omega) = \frac{1+j}{2} \sqrt{\frac{k}{\pi}} \oint_S \rho c V_n(\mathbf{r}, \omega) \frac{e^{-jk|\mathbf{r}-\mathbf{r}_R|}}{\sqrt{|\mathbf{r}-\mathbf{r}_R|}} dL, \tag{2.36}$$

$$P(\mathbf{r}_R, \omega) = \frac{1+j}{2} \sqrt{\frac{k}{\pi}} \oint_S P(\mathbf{r}, \omega) \cos \theta \frac{e^{-jk|\mathbf{r}-\mathbf{r}_R|}}{\sqrt{|\mathbf{r}-\mathbf{r}_R|}} dL, \tag{2.37}$$

e inversas:

$$P(\mathbf{r}_R, \omega) = \frac{1-j}{2} \sqrt{\frac{k}{\pi}} \oint_S \rho c V_n(\mathbf{r}, \omega) \frac{e^{jk|\mathbf{r}-\mathbf{r}_R|}}{\sqrt{|\mathbf{r}-\mathbf{r}_R|}} dL, \quad (2.38)$$

$$P(\mathbf{r}_R, \omega) = \frac{1-j}{2} \sqrt{\frac{k}{\pi}} \oint_S P(\mathbf{r}, \omega) \cos \theta \frac{e^{jk|\mathbf{r}-\mathbf{r}_R|}}{\sqrt{|\mathbf{r}-\mathbf{r}_R|}} dL. \quad (2.39)$$

## 2.2. Extrapolación del campo de onda

Los arrays de micrófonos permiten medir el campo sonoro de una sala en diferentes localizaciones espaciales. El sonido captado en cada posición del array, puede ser procesado para analizar las propiedades del campo acústico utilizando el concepto de extrapolación del campo de onda. Cuando se utiliza un array circular de radio  $R$ , las integrales 2D directas e inversas de Kirchhoff-Helmoltz se convierten en las herramientas matemáticas necesarias para la reconstrucción de ese campo sonoro dentro del array circular [19]. De tal forma, si  $P(\mathbf{r}, \omega)$  y  $V(\mathbf{r}, \omega)$  representan la presión y la velocidad de las partículas en el array circular en el dominio de la frecuencia, donde  $\theta$  es el ángulo acimut, y  $\phi$  es el ángulo de elevación, el campo total en la posición  $\mathbf{r}$  vendrá dado por la suma de los campos de onda extrapolados directo e inverso:

$$P(\mathbf{r}, \omega) = P^{(1)}(\mathbf{r}, \omega) + P^{(2)}(\mathbf{r}, \omega), \quad (2.40)$$

$$P^{(1)}(\mathbf{r}, \omega) = \frac{-jk}{4} \int_0^{2\pi} P(\mathbf{r}, \omega) \cos \theta H_1^{(1)}(k\Delta\mathbf{r}) R d\theta + \frac{-jk}{4} \int_0^{2\pi} j\rho c V_n(\mathbf{r}, \omega) H_0^{(1)}(k\Delta\mathbf{r}) R d\theta, \quad (2.41)$$

$$P^{(2)}(\mathbf{r}, \omega) = \frac{-jk}{4} \int_0^{2\pi} P(\mathbf{r}, \omega) \cos \theta H_1^{(2)}(k\Delta\mathbf{r}) R d\theta + \frac{-jk}{4} \int_0^{2\pi} j\rho c V_n(\mathbf{r}, \omega) H_0^{(2)}(k\Delta\mathbf{r}) R d\theta. \quad (2.42)$$

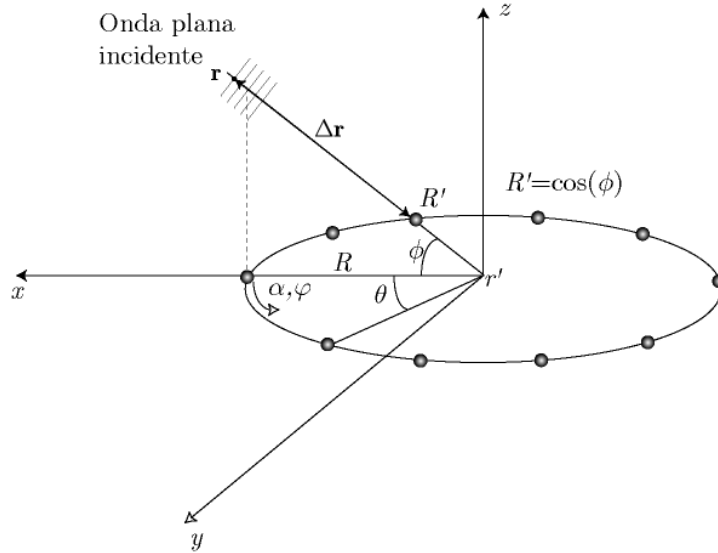


Figura 2.5: Geometría de un array circular.

En la Figura 2.5 se muestran los anteriores parámetros dentro de la geometría considerada.

Es resaltable comentar, que las integrales Kirchhoff-Helmholtz sólo son capaces de reconstruir el campo sonoro dentro de la circunferencia. Fuera de ésta, el campo es cero. La Figura 2.6 muestra la reconstrucción de una grabación circular para  $P$  y  $\mathbf{V}_n$  de una onda plana en  $t = 0$ .

De forma general, la descomposición de onda plana del campo sonoro puede ser obtenido por aproximación de la presión extrapolada directa en su origen por el campo sonoro extrapolado inverso  $P^{(1)}(\mathbf{r}, \theta, \omega)$  sobre una circunferencia con radio infinito [10]. Así pues, la descomposición de ondas planas entrantes  $s^{(1)}$  y salientes  $s^{(2)}$  para cuando  $\mathbf{r}$  tiende a  $\infty$  sería

$$s^{(1)}(\theta, \omega) = \frac{1+j}{\sqrt{4\pi}} \lim_{r \rightarrow \infty} \sqrt{kr} P^{(1)}(\mathbf{r}, \theta, \omega) e^{-jkr}, \quad (2.43)$$

$$s^{(2)}(\theta, \omega) = \frac{1-j}{\sqrt{4\pi}} \lim_{r \rightarrow \infty} \sqrt{kr} P^{(2)}(\mathbf{r}, \theta, \omega) e^{jkr}. \quad (2.44)$$

Esta descomposición en ondas planas da una completa descripción

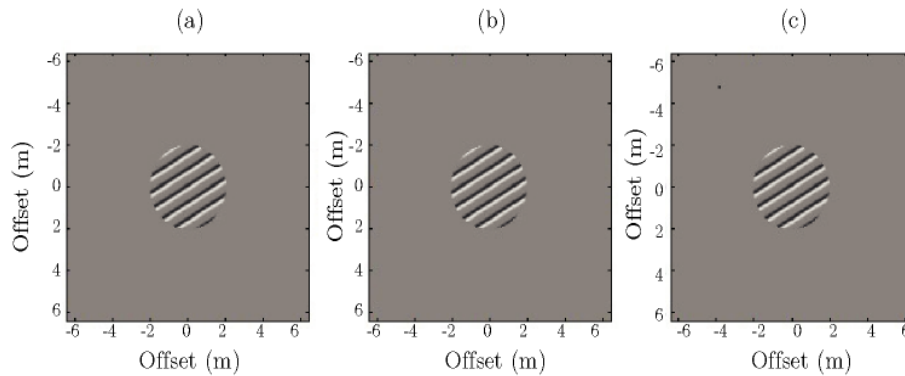


Figura 2.6: Reconstrucción de una grabación circular utilizando las integrales de Kirchhoff-Helmholtz directa (a), inversa (b) y suma (c).

del campo sonoro no sólo en el origen, sino también en cualquier posición si dichas ondas planas son extrapoladas. Sin embargo es importante recordar que las integrales de Kirchhoff-Helmholtz no son capaces de reconstruir el campo sonoro fuera de la circunferencia del array, y la descomposición en ondas planas no puede ser directamente calculada. De tal forma es necesario un paso intermedio a seguir, que consiste en la descomposición del campo sonoro en armónicos cilíndricos. Utilizando estos armónicos, el campo acústico puede ser extrapolado fuera del círculo del array, eliminando la anterior restricción mencionada.

## 2.3. Descomposición en armónicos cilíndricos

Uno de los requerimientos de Kirchhoff-Helmholtz es que la extrapolación se realice dentro de un volumen cerrado 3D o de superficie 2D. Con arrays lineales, la extrapolación puede alcanzarse para un área infinita, donde las aproximaciones de campo lejano son válidas y por lo tanto, la descomposición en ondas planas es posible. Para el caso de arrays circulares, Hulsebos [10] propuso una solución basada en descomposición de armónicos cilíndricos para



poder realizar una extrapolación en el exterior de este tipo de arrays.

Físicamente, al procedimiento de aplicar una transformada espacial de Fourier con respecto a coordenadas polares en una apertura circular o array circular, puede interpretarse como una descomposición en onda plana con armónicos cilíndricos en un círculo de radio  $r = R$  [13]. Los campos sonoros de fuentes como el monopolo o el dipolo, corresponden a armónicos esféricos o cilíndricos de orden cero y primer orden respectivamente. En dos dimensiones, los campos sonoros de éstos o de armónicos de orden superior vienen dados por:

$$P_{k_\theta}^{(1)}(\mathbf{r}, \theta, \omega) = H_{k_\theta}^{(1)}(k\mathbf{r})e^{jk_\theta\theta}, \quad (2.45)$$

$$P_{k_\theta}^{(2)}(\mathbf{r}, \theta, \omega) = H_{k_\theta}^{(2)}(k\mathbf{r})e^{jk_\theta\theta}, \quad (2.46)$$

los cuales representan los campos de presión de los armónicos cilíndricos de entrada y salida respectivamente.  $k_\theta$  es el orden del armónico y puede ser un número entero positivo o negativo. El campo sonoro de un monopolo corresponde al caso de  $k_\theta = 0$  y el campo producido por un dipolo puede ser obtenido tomando la apropiada combinación lineal entre  $P_{-1}^{(2)}$  y  $P_1^{(2)}$ .

Para calcular la descomposición de la presión y velocidad grabadas en términos de armónicos cilíndricos, primero se ha de calcular la componente radial de velocidad de estos armónicos cilíndricos utilizando la segunda ley de Newton,

$$-\nabla P = j\omega\rho\mathbf{V}_n. \quad (2.47)$$

Si esta ecuación se transforma en coordenadas cilíndricas y sólo se considera la componente radial interior de la velocidad entonces:

$$\frac{\partial P}{\partial \mathbf{r}} = j\omega\rho\mathbf{V}_n. \quad (2.48)$$

Utilizando este resultado junto con las derivadas de las funciones de Hankel,

$$H_n^{(1)} = \frac{1}{2}H_{n-1}^{(1)} - \frac{1}{2}H_{n+1}^{(1)}, \quad (2.49)$$

$$H_n^{(2)} = \frac{1}{2}H_{n-1}^{(2)} - \frac{1}{2}H_{n+1}^{(2)}, \quad (2.50)$$

las componentes radiales de velocidad de los armónicos cilíndricos pueden ser calculadas de la siguiente forma:

$$j\rho c V_{k_\theta}^{(1)}(\mathbf{r}, \theta, \omega) = H_{k_\theta}^{\prime(1)}(k\mathbf{r})e^{jk_\theta\theta} = \frac{1}{2} \left[ H_{k_\theta-1}^{(1)}(k\mathbf{r}) - H_{k_\theta+1}^{(1)}(k\mathbf{r}) \right] e^{jk_\theta\theta}, \quad (2.51)$$

$$j\rho c V_{k_\theta}^{(2)}(\mathbf{r}, \theta, \omega) = H_{k_\theta}^{\prime(2)}(k\mathbf{r})e^{jk_\theta\theta} = \frac{1}{2} \left[ H_{k_\theta-1}^{(2)}(k\mathbf{r}) - H_{k_\theta+1}^{(2)}(k\mathbf{r}) \right] e^{jk_\theta\theta}. \quad (2.52)$$

La presión y velocidad de partícula en un array circular de micrófonos, pueden descomponerse en armónicos cilíndricos de la siguiente forma:

$$P(\theta, \omega) = \sum_{k_\theta} \mathcal{M}^{(1)}(k_\theta, \omega) P_{k_\theta}^{(1)}(R, \theta, \omega) + \sum_{k_\theta} \mathcal{M}^{(2)}(k_\theta, \omega) P_{k_\theta}^{(2)}(R, \theta, \omega), \quad (2.53)$$

$$V_n(\theta, \omega) = \sum_{k_\theta} \mathcal{M}^{(1)}(k_\theta, \omega) V_{k_\theta}^{(1)}(R, \theta, \omega) + \sum_{k_\theta} \mathcal{M}^{(2)}(k_\theta, \omega) V_{k_\theta}^{(2)}(R, \theta, \omega), \quad (2.54)$$

donde  $\mathcal{M}^{(1)}(k_\theta, \omega)$  y  $\mathcal{M}^{(2)}(k_\theta, \omega)$  representan los coeficientes de expansión de los armónicos cilíndricos de entrada y salida

$$P_{k_\theta}^{(1,2)}(\mathbf{r}, \theta, \omega) = H_{k_\theta}^{(1,2)}(k\mathbf{r})e^{jk_\theta\theta}, \quad (2.55)$$

de orden  $k_\theta$ .

Para simplificar estas ecuaciones, es conveniente tomar transformadas de Fourier de  $p$  y  $v$  con respecto a  $\theta$ ,

$$P(k_\theta, \omega) = \frac{1}{2\pi} \int_0^{2\pi} P(\theta, \omega) e^{-jk_\theta\theta} d\theta, \quad (2.56)$$

$$V_n(k_\theta, \omega) = \frac{1}{2\pi} \int_0^{2\pi} \mathbf{V}_n(\theta, \omega) e^{-jk_\theta\theta} d\theta. \quad (2.57)$$

Las series de Fourier pueden ser escritas entonces como se muestra a continuación:

$$P(\theta, \omega) = \sum_{k_\theta} P(k_\theta, \omega) e^{jk_\theta \theta}, \quad (2.58)$$

$$V_n(\theta, \omega) = \sum_{k_\theta} V_n(k_\theta, \omega) e^{jk_\theta \theta}. \quad (2.59)$$

La aplicación de estas series de Fourier dan como resultado las siguientes relaciones:

$$P(k_\theta, \omega) = \mathcal{M}^{(1)}(k_\theta, \omega) H_{k_\theta}^{(1)}(kR) + \mathcal{M}^{(2)}(k_\theta, \omega) H_{k_\theta}^{(2)}(kR), \quad (2.60)$$

$$j\rho c V_n(k_\theta, \omega) = \mathcal{M}^{(1)}(k_\theta, \omega) H_{k_\theta}^{\prime(1)}(kR) + \mathcal{M}^{(2)}(k_\theta, \omega) H_{k_\theta}^{\prime(2)}(kR). \quad (2.61)$$

Para descomponer el campo sonoro en armónicos cilíndricos, los coeficientes de expansión  $\mathcal{M}^{(1)}$  y  $\mathcal{M}^{(2)}$  pueden utilizarse para reconstruirlo en cualquier punto en el espacio. Aplicando de nuevo la segunda ley de Newton  $\frac{\partial P}{\partial \mathbf{r}} = j\omega\rho\mathbf{V}_n$  y teniendo en cuenta las propiedades de las funciones de Hankel, se obtienen los coeficientes expandidos del campo sonoro en términos de armónicos cilíndricos en función de  $P(k_\theta, \omega)$  y  $V(k_\theta, \omega)$ ,

$$\mathcal{M}^{(1)}(k_\theta, \omega) = \frac{H_{k_\theta}^{\prime(2)}(kR)P(k_\theta, \omega) - H_{k_\theta}^{(2)}(kR)j\rho c V_n(k_\theta, \omega)}{H_{k_\theta}^{(1)}(kR)H_{k_\theta}^{\prime(2)}(kR) - H_{k_\theta}^{(2)}(kR)H_{k_\theta}^{\prime(1)}(kR)}, \quad (2.62)$$

$$\mathcal{M}^{(2)}(k_\theta, \omega) = \frac{H_{k_\theta}^{\prime(1)}(kR)P(k_\theta, \omega) - H_{k_\theta}^{(1)}(kR)j\rho c V_n(k_\theta, \omega)}{H_{k_\theta}^{(2)}(kR)H_{k_\theta}^{\prime(1)}(kR) - H_{k_\theta}^{(1)}(kR)H_{k_\theta}^{\prime(2)}(kR)}. \quad (2.63)$$

Además, y para el caso de que las fuentes sólo se localicen fuera del array circular  $\mathcal{M}^{(1)} = \mathcal{M}^{(2)}$ , es posible considerar un conjunto de coeficientes del tipo:

$$\mathcal{M} = \frac{1}{2}(\mathcal{M}^{(1)} + \mathcal{M}^{(2)}). \quad (2.64)$$

Combinando esta ecuación en las Ecuaciones (2.60) y (2.61) se obtiene

$$\frac{1}{2} \{P + \rho_0 c V(k_\theta, \omega)\} = \frac{1}{2} \mathcal{M}^{(1)} + \quad (2.65)$$

$$\frac{1}{2} \mathcal{M}^{(2)} \left[ (H_{k_\theta}^{(1)}(kR) + H_{k_\theta}^{(2)}(kR)) \right] - \frac{1}{2} \mathcal{M}^{(2)} \left[ (jH_{k_\theta}'^{(1)}(kR) + jH_{k_\theta}'^{(2)}(kR)) \right]$$

dando lugar finalmente a la siguiente ecuación de descomposición

$$\mathcal{M}(k_\theta, \omega) = \frac{\frac{1}{2} \{P + \rho_0 c V(k_\theta, \omega)\} (k_\theta, \omega)}{H_{k_\theta}^{(1)}(kR) + H_{k_\theta}^{(2)}(kR) - jH_{k_\theta}'^{(1)}(kR) - jH_{k_\theta}'^{(2)}(kR)}. \quad (2.66)$$

Aunque en general es necesario tener tanto la presión como la velocidad normal para hacer una apropiada extrapolación del campo sonoro, esta expresión sugiere que un array circular utilizando únicamente micrófonos cardioides orientados en sentido saliente al array pueden utilizarse para analizar el campo con fuentes externas al array. Esto es debido a que el término de velocidad  $V(k_\theta, \omega)$  para un micrófono de figura en ocho combinado con el término de presión  $P$  resulta el patrón de un micrófono cardioide, el cuál es muy conveniente desde una perspectiva práctica [20]. Sin embargo, en el caso de que se utilizaran sólo micrófonos omnidireccionales o micrófonos de figura en ocho, no podrían ser obtenidos todos los armónicos cilíndricos en tal grabación.

En la Figura 2.7, se muestran las amplitudes para las transformaciones de Fourier 2D de una onda plana grabada en un array circular, utilizando micrófonos omnidireccionales, de figura en ocho y cardioides, junto con sus correspondientes descomposiciones armónicas cilíndricas.

Las gráficas correspondientes a las grabaciones realizadas por micrófonos omnidireccionales y figura en ocho, Figura 2.7 (b) y (c) contienen componentes cero en sus correspondientes transformadas de Fourier, no pudiendo ser usadas para el cálculo correcto de las componentes de los armónicos cilíndricos de la onda plana de la Figura 2.7 (a). Sin embargo, puesto que las componentes cero para los arrays de micrófonos omnidireccionales y figura en ocho, aparecen en componentes frecuenciales distintas, pueden utilizarse de

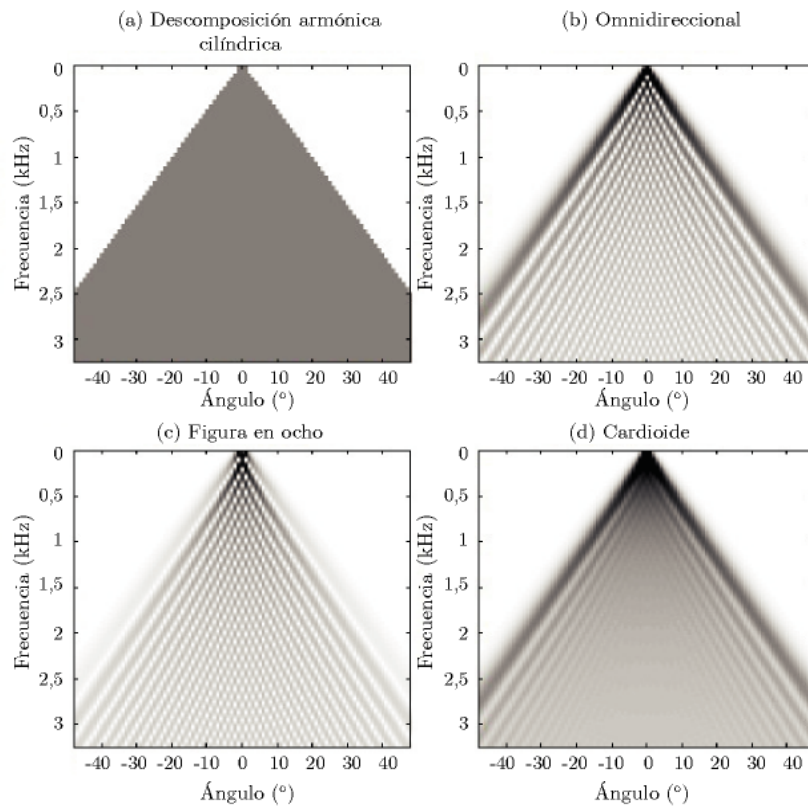


Figura 2.7: Transformada 2 D de Fourier de una onda plana simulada mediante un array circular con (b) micrófonos omnidireccionales, (c) de figura en ocho, (d) cardioides. Estas respuestas son comparadas con la descomposición en armónicos cilíndricos presentada en (a).

manera conjunta para poder descomponer apropiadamente mediante armónicos cilíndricos el campo sonoro utilizando las Ecuaciones (2.62) y (2.63). Para el caso de un array con micrófonos cardioides (caso de la Figura 2.7 (d) ) la transformada de Fourier 2D está libre de ceros y puede ser utilizada fácilmente para el mismo propósito utilizando en este caso la Ecuación (2.66).

## 2.4. Descomposición en onda plana

Anteriormente se ha comprobado que si tanto la presión como la velocidad normal de partícula eran grabadas en un array circular, el campo sonoro podía ser reconstruido de forma completa dentro de éste utilizando la integral de Kirchhoff-Helmholtz. Además, gracias a la descomposición en armónicos cilíndricos del campo, la reconstrucción no sólo se limitaría al interior del array sino que podría también reconstruirse en el exterior del mismo. Teniendo en cuenta todo el análisis anterior, ahora es necesario adecuarlo dentro del marco experimental que se llevará a cabo en esta tesis, concretamente en el Capítulo 5. Cabe destacar dos aspectos fundamentales que particularizarán las ecuaciones antes analizadas. Por un lado, la fuente sonora se ubicará fuera del array de micrófonos, lo cuál implicará no tener que discriminar entre el campo de entrada y de salida de este array. Otra consideración a tener en cuenta es la utilización de micrófonos cardioides en todas las medidas. El uso de este tipo de micrófonos permite no tener que grabar la presión y la velocidad de partícula como se ha visto en la Sección 2.3. Con este tipo de micrófonos es suficiente grabar el campo sonoro del array utilizando la posición de los mismos apuntando hacia afuera del array y haciendo uso de la ecuación de directividad de este tipo de micrófonos.

De tal forma, sustituyendo únicamente la parte de entrada del campo de onda al array en la Ecuación (2.43)

$$P^{(1)}(\mathbf{r}, \theta, \omega) = \sum_{k_\theta} \mathcal{M}^{(1)}(k_\theta, \omega) P_{k_\theta}^{(1)}(\mathbf{r}, \theta, \omega), \quad (2.67)$$

y asumiendo las aproximaciones en campo lejano dadas por las Ecuaciones (2.29) y (2.30), la descomposición de onda plana del campo

sonoro en términos de armónicos cilíndricos sería finalmente:

$$s_{\infty}(\theta, \omega) = \frac{1}{2\pi} \sum_{k_{\theta}} e^{-jk_{\theta}\pi/2} \mathcal{M}(k_{\theta}, \omega) e^{jk_{\theta}\theta}. \quad (2.68)$$

La flexibilidad de este resultado pone de manifiesto que el campo sonoro al poder ser calculado en cualquier posición, cumple una de las características fundamentales y necesarias en propósitos de auralización y análisis de acústica de salas [21], [22], [23]. A continuación se va a comentar de forma muy breve en qué consiste la auralización y cómo el anterior desarrollo matemático cobra sentido en la misma.

## 2.5. Auralización

Generalmente, la captura del campo sonoro en una sala y concretamente la PWD, se realiza con el propósito de realizar una posterior auralización. La auralización es la técnica que permite hacer audible en una sala de reproducción, la acústica de otra sala diferente. Para ello, es necesario un proceso de grabación de respuestas impulsionales mediante un array de micrófonos o mediante un único micrófono colocado en diferentes posiciones, para reproducir posteriormente el campo de onda desde la sala en la que se han registrado dichas RIRs en otra sala final de reproducción [11], [24].

Un sistema de auralización está compuesto por tres partes, una parte de grabación, otra de procesado y por última una de reproducción, Figura 2.8.

La grabación corresponde a la fase de medición de las respuestas al impulso y se pueden realizar tres configuraciones de arrays: array lineal, array cruzado y array circular.

En la parte de procesado, se calculan los filtros de convolución para las señales que emitirán los altavoces dentro del proceso de síntesis, como por ejemplo un método WFS, [14], [25], [22] y es necesaria siempre que las posiciones de los altavoces en la sala de reproducción no se correspondan a las posiciones de los micrófonos de la parte de grabación. En este caso, el campo sonoro es necesario que se extrapole

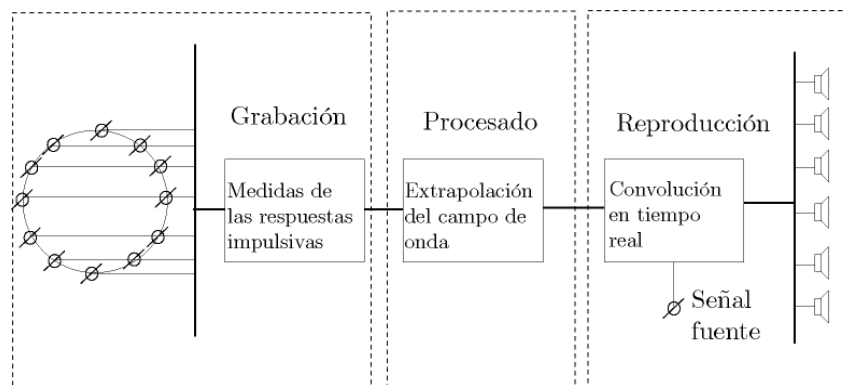


Figura 2.8: Diagrama del sistema de auralización, consistente en una parte de grabado, otra de procesado y una parte de reproducción.

desde las posiciones de los micrófonos (array de micrófonos), a las posiciones de los altavoces en el sistema de síntesis.

La fase de reproducción es donde se graba una fuente sonora independiente, para ser convolucionada posteriormente con los filtros de la parte de procesado y reproducirse finalmente en los altavoces del sistema de síntesis.

Por otra parte, hay tres posibilidades de llevar a cabo la auralización: mediante 'holofonía', mediante 'extrapolación', y mediante 'descomposición de onda' dándose en este trabajo especial mención a los dos últimos tipos.

En una auralización mediante holofonía, la disposición del array de micrófonos coincide exactamente con las posiciones de los altavoces en el sistema de síntesis de ondas planas dentro de la sala de reproducción. Sin embargo, cuando las posiciones del array de micrófonos no se corresponden con la disposición del sistema de síntesis, para auralizar correctamente las respuestas impulsivas o reproducir el sonido grabado, es necesario extrapolar las RIRs del array de grabación a las posiciones de los altavoces del sistema WFS por ejemplo, lo cual permite mayor flexibilidad pero implica mayor procesado. Esto se puede llevar a cabo con las integrales de Kirchhoff-Helmholtz o de Rayleigh. Aunque la extrapolación del



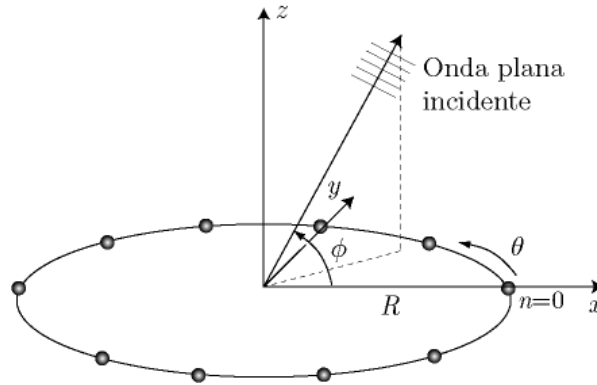


Figura 2.9: Geometría de un array circular donde incide una onda plana.

campo de onda da una gran flexibilidad en términos de configuración de altavoces en la parte de síntesis, la cantidad de datos y de procesamiento para alcanzar la correcta extrapolación en un área suficientemente grande, es muy elevada. Es entonces cuando se aplica una auralización mediante descomposición de ondas planas donde en el caso de un array circular, es más conveniente emplear la descomposición en armónicos cilíndricos derivados de la matemática presentada en anteriores secciones.

## 2.6. Modelo de array circular de micrófonos

Como ya se ha comentado, los arrays circulares de micrófonos son conocidos por su idoneidad en un completo rango azimutal y por poseer una geometría adecuada para el procesamiento modal de arrays, lo cual permite mejoras significativas en el análisis del campo de onda, [10, 14, 26, 27]. La geometría de array utilizada a partir de ahora es la mostrada en la Figura 2.9. En ella, se considera un array circular de  $N$  micrófonos y radio  $R$ , posicionado en el plano horizontal  $(x, y)$ . A partir de ahora, se considerarán que las ondas planas incidirán sobre el array con un ángulo acimut  $\theta$  y un ángulo de elevación  $\phi$ . La posición de un micrófono determinado en el array vendrá especificado por un ángulo  $\theta_n$ , donde  $n = 1, \dots, N$ .

Cuando una onda plana viaja a través del plano horizontal, es

decir, con un ángulo de elevación  $\phi = 0^\circ$ , y un ángulo de acimut  $\theta = 0^\circ$ , ésta alcanzará en primera posición al micrófono localizado en  $\theta_n = 0^\circ$ . Los tiempos de llegada de la onda a los micrófonos del array circular, estarán en función de la posición del micrófono  $\theta_n$  y describirán una onda en forma de coseno, donde el máximo se obtiene para el micrófono más cercano a  $0^\circ$  el mínimo para el más cercano a  $\pm 180^\circ$  y el valor cero para aquellos micrófonos localizados en  $\pm 90^\circ$ , siempre y cuando se tome como referencia temporal ( $t = 0$ ), el centro del array. En lo siguiente, el tiempo medio de llegada al centro del array se denotará como  $\tau_i$  (también conocido como *intercept-time*). La amplitud geométrica de la curva de tiempo y forma cosenoidal se calcula como  $R/c$  donde  $c$  es la velocidad del sonido.

Cuando una onda plana alcanza al array bajo un ángulo de elevación  $\phi_i$  diferente a cero y un ángulo de acimut  $\theta_i = 0^\circ$ , llegará en primer lugar al micrófono ubicado en  $\theta_n = 0^\circ$  y por último al localizado en  $\theta_n \approx \pm 180^\circ$ . De nuevo, la curva de tiempo mantendrá una forma de onda cosenoidal, pero en este caso, la amplitud geométrica con respecto al tiempo de llegada  $\tau_i$  decrecerá según  $(R/c) \cos \phi$ . Por otro lado, la variación en el ángulo de acimut conduce a un desplazamiento espacial de fase de la curva de tiempos de llegada de forma que el micrófono cuya localización sea  $\theta_n \neq 0^\circ$  será el primero en recibir dicha onda. De hecho, cuando una onda plana llega al array con incidencia vertical ( $\phi = 90^\circ$ ), es cuando alcanza a todos los micrófonos en el mismo instante de tiempo, y la curva de tiempo de llegada en este caso sería una línea recta. Esto permite sacar la conclusión de que las curvas de tiempo en las respuestas al impulso de una onda plana medidas a lo largo de un array circular permiten obtener información sobre los ángulos de elevación de ésta. Sin embargo y debido a la simetría en la disposición del array con respecto al plano horizontal, no es posible distinguir entre ondas planas que lleguen por encima o por debajo del array. En la Figura 2.10 se muestra un ejemplo simulado a partir de un array con un radio de  $R=1,5$  m y  $N = 288$  micrófonos al que le llegan tres ondas planas con direcciones  $45^\circ$ ,  $245^\circ$  y  $200^\circ$  en diferentes *intercept-times*.

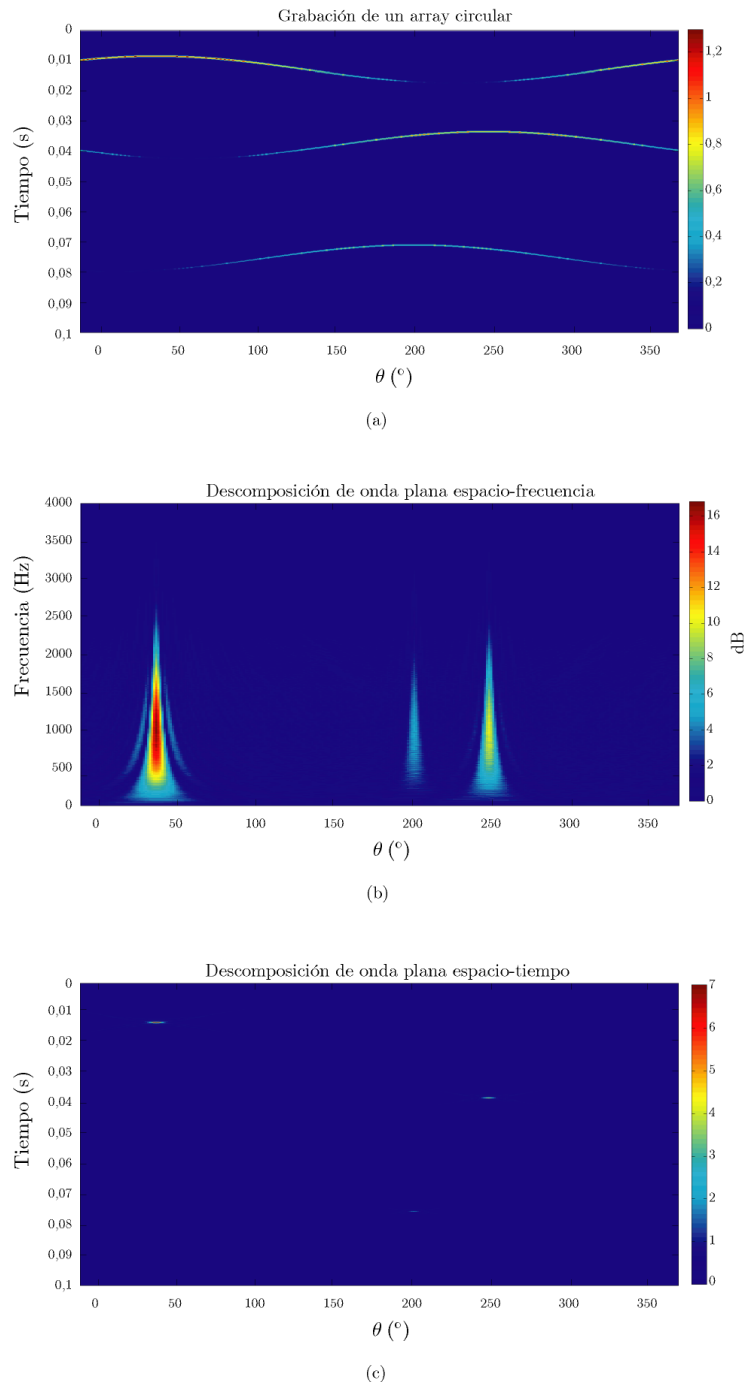


Figura 2.10: Ejemplo simulado para tres ondas planas incidiendo en un array circular con direcciones  $45^\circ$ ,  $245^\circ$  y  $200^\circ$ . (a) Grabación del array circular. (b) Descomposición de onda plana frecuencia-espacio. (c) Descomposición de onda plana tiempo-espacio.

## 2.7. Aliasing espacial y efectos de baja frecuencia

Generalmente en la práctica y en particular en esta tesis, las medidas se llevan a cabo con arrays discretos. Esto implica tener que muestrear por medio de un número finito de sensores las aperturas continuas que se han analizado teóricamente en las anteriores secciones. De forma análoga a la frecuencia de Nyquist en un muestreo temporal, esta discretización conllevará una restricción a una determinada frecuencia a partir de la cuál el campo sonoro se verá seriamente distorsionado. Esta distorsión se conoce como *aliasing* espacial y la frecuencia a partir de la cuál ocurre este efecto se llama frecuencia de *aliasing* espacial. Considerando el array circular uniforme, mostrado en la Figura 2.9, la distancia entre micrófonos puede ser calculada como

$$d = 2R \operatorname{sen} \left( \frac{\pi}{N} \right), \quad (2.69)$$

La anterior distancia determina la frecuencia de *aliasing* espacial  $f_{al}$  que viene dada por

$$f_{al} = \frac{c}{2d}. \quad (2.70)$$

Cuanto mayor sea el espaciado entre micrófonos menor será esa frecuencia de *aliasing*. Para aplicaciones que requieran una gran calidad, será necesario un espaciado muy pequeño para evitar el *aliasing* espacial en la extrapolación inversa del campo de onda. Esto se conseguirá con un gran número de micrófonos para un radio determinado. No obstante, el *aliasing* espacial no es el único inconveniente para llevar a cabo las medidas reales mediante un array de este tipo. El coste computacional se eleva de forma considerable con el aumento de micrófonos. De tal forma, es necesario llegar a un compromiso entre resultados, material y memoria computacional asociada. En los ejemplos simulados en la Sección 2.9.1, se mostrará gráficamente cómo se representan los artefactos ocasionados por el *aliasing* espacial y cómo éstos dificultan la descomposición en onda plana.

Por otra parte, es necesario considerar también que el rango frecuencial cubierto por un determinado array no está limitado únicamente a los efectos de *aliasing* espacial. La premisa de campo lejano hace que el modelo de propagación de onda plana no llegue a cumplirse en las bajas frecuencias donde la separación entre los micrófonos es relativamente pequeña comparada con la longitud de onda [28]. Además, el valor de algunos órdenes de la función de Hankel pueden ser muy pequeños cuando se calculan los coeficientes de expansión en la Ecuación (2.66). Esta limitación en el rango de bajas frecuencias se hará evidente por la aparición de pérdida de resolución frecuencial en representaciones PWD frecuencia-espacio. Estos efectos serán mostrados tanto en la Sección 2.9.1 como en el Capítulo 5 .

## 2.8. Representaciones gráficas de la descomposición

Cuando las RIRs son obtenidas a partir de un array de micrófonos en vez de un único sensor, aparecen ciertos patrones al dibujar de manera conjunta esas respuestas. A continuación se va a introducir brevemente las principales representaciones que se obtendrán y analizarán después de aplicar el método de descomposición de ondas planas en el resto de esta tesis.

### 2.8.1. Representación espacio-tiempo

Como ya se ha explicado, la captura de las señales en un array circular tiene forma de coseno en el tiempo. La Figura 2.10 (a) muestra la grabación de tres ondas planas con diferentes atenuaciones, tiempos de llegada y direcciones de incidencia ( $45^\circ$ ,  $245^\circ$  y  $200^\circ$ ). El array utilizado para este ejemplo simulado tenía un radio de  $R = 1,5$  m y  $N = 288$  micrófonos. Cuando se trabaja con medidas capturadas en salas reales, se obtiene una representación similar a partir de las respuestas al impulso capturadas por los micrófonos. Esta matriz espacio-tiempo, conocida como ecograma, está formada por todas las

respuestas al impulso capturadas por el array:

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N], \quad (2.71)$$

donde  $\mathbf{h}_n = [h_n(1), \dots, h_n(L)]^T$  es la respuesta al impulso de longitud  $L$  obtenida por el micrófono  $n$  a partir de una fuente de medida. De tal forma, los elementos de  $\mathbf{H}$  son denotados como  $H(t, n)$ , donde  $t$  y  $n$  son el tiempo dado de la muestra y el micrófono, respectivamente. Esta matriz permite observar el comportamiento completo de la salas mediante la recopilación y representación gráfica de todas las RIRs a la vez y para cualquier posición del array [14]. Esta representación, anterior al método de descomposición plana, revela mucho más que la estructura espacial y temporal del campo de onda. Como se verá más adelante en esta tesis, en ecogramas reales es posible discriminar diferentes frentes de onda reflejados en distintas direcciones. Además el análisis de esta representación puede servir para obtener información adicional en cuanto a las propiedades acústicas como la difusividad relacionada con posiciones específicas dentro de la sala en cuestión. No obstante y puesto que la presión sonora será medida en este trabajo con dos micrófonos cardioides, este tipo de representación no permitirá una discriminación en el plano de elevación al array.

### 2.8.2. Descomposición en armónicos cilíndricos

A partir de la Ecuación (2.66) y una vez transformadas las RIRs capturadas por un array circular al dominio  $(k_\theta, \omega)$ , es posible obtener la representación mostrada en la Figura 2.11 donde se puede observar cómo se distribuyen las frecuencias (eje de ordenadas) según el orden de armónico en el espacio  $k_\theta$  (eje de abcisas). Para el ejemplo de la figura y para un array de 144 micrófonos, se obtendrán los unos valores de  $k_\theta$  entre  $-\frac{144}{2}$  y  $\frac{144}{2}$  [ $\text{m}^{-1}$ ].

### 2.8.3. Descomposición frecuencia-espacio

Después de la aplicación del método de descomposición de onda plana utilizando la Ecuación (2.68), el campo sonoro puede ser

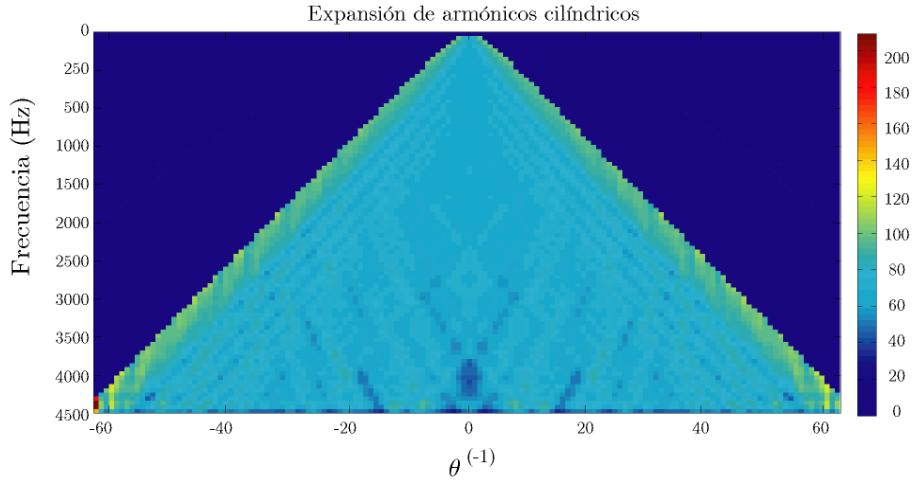


Figura 2.11: Descomposición en armónicos cilíndricos, donde el eje  $y$  representa la frecuencia y el eje  $x$  representa el número de armónicos.

fácilmente interpretado como la suma de ondas planas en diferentes direcciones de llegada. Hay que tener en cuenta que la entrada para la descomposición del campo sonoro en ondas planas son las respuestas al impulso reales grabadas por los micrófonos cardioides del array. De esta forma, el término  $(P(\mathbf{r}, \omega) + \rho_0 c V(\mathbf{r}, \omega))$  corresponde a la transformada de Fourier espacial de  $H_n(\omega) = \text{DFT}\{h_n(t)\}$ , es decir, la transformada de Fourier discreta de las respuestas al impulso en el dominio del tiempo como demostró [14]. La descomposición final es la siguiente:

$$\mathbf{s}^\omega = [\varrho_1^\omega, \varrho_2^\omega, \dots, \varrho_N^\omega], \quad (2.72)$$

donde  $\varrho_n^\omega = [s(\theta_n, \omega_1), \dots, s(\theta_n, \omega_L)]^T$  es la PWD correspondiente al micrófono de ángulo  $\theta_n$  en el dominio de la frecuencia. El contenido espectral para cada dirección de onda plana en el ejemplo simulado puede observarse en la Figura 2.10 (b). Esta representación muestra la distribución de frecuencia y de la dispersión de energía en función del ángulo acimut. Como se puede comprobar, la energía depende del radio del array y encuentra su máximo en el ángulo de incidencia de la onda plana. Sin embargo y debido a que el tamaño del array es finito comparado con la longitud de onda a bajas frecuencias, se mostrará en la Sección 2.9.1, cómo la resolución frecuencial variará según

lo comentado en la Sección 2.7. Aunque en este ejemplo simulado se presenta una señal simple consistente en tres ondas planas, en medidas reales, esta representación podría utilizarse por ejemplo para analizar posibles defectos debido a obstáculos, geometría o materiales de la sala...

#### 2.8.4. Descomposición tiempo-espacio

Por último, otra importante representación a tener en cuenta es la PWD en tiempo, derivada de la Transformada Inversa Discreta de Fourier de la Ecuación (2.68).

$$\mathbf{s}^t = [\varrho_1^t, \varrho_2^t, \dots, \varrho_N^t], \quad (2.73)$$

donde  $\varrho_n^t = [s(\theta_n, 1), \dots, s(\theta_n, L)]^T$  es la PWD correspondiente al micrófono de ángulo  $\theta_n$  en el dominio del tiempo, donde  $s(\theta_n, t) = \text{IDFT} \{s(\theta_n, \omega)\}$ . La correspondiente representación para el ejemplo simulado, se muestra en la Figura 2.10 (c). Es importante observar que las ondas planas pueden ser identificadas como puntas de flecha correspondientes a sus tiempos de llegada y direcciones de acimut. Aunque esta descomposición es un método útil para separar los diferentes eventos de las ondas planas incidentes en una sala, se requiere cálculos adicionales, lo cual es altamente dependiente del número de micrófonos que forman el array.

## 2.9. Grabación mediante un array circular de micrófonos

Con respecto a la grabación de las RIRs por un array circular, comentar que existen dos formas de medir estas respuestas al impulso en una sala: utilizando un array completo de micrófonos o un array discreto. En esta tesis hemos optado por la segunda opción debido a las ventajas que se comentarán a continuación.

Medir las respuestas al impulso mediante un array discreto puede realizarse mediante la colocación de un micrófono sujeto a una barra anclada en un motor giratorio. De esta forma se pueden realizar



diferentes medidas repetitivas con un giro acorde al número de micrófonos que se quieran utilizar. Con este método de medida, evitamos el sobrecoste económico que implicaría tener un array completo de micrófonos. Además los resultados deberían ser los mismos que en un array completo siempre que las condiciones de temperatura, humedad, etc... no cambien significativamente durante la serie de medidas en la sala. Sin embargo, el utilizar únicamente un sensor trae consigo la desventaja de la gran cantidad de tiempo que se ha de invertir en las medidas.

Por otra parte, hay dos formas de realizar las medidas con un array circular discreto. Por un lado, utilizando pasos discretos en un eje giratorio y por otro realizando una rotación continua.

Realizar pasos discretos para la captura de las RIRs, implica un aumento en el tiempo total de medida. En cada paso o salto, el sistema debe esperar a que el micrófono se estabilice y no oscile hasta la siguiente rotación. En el caso de realizar una medida de forma continua, el estímulo debe de ser repetido y grabado ininterrumpidamente mientras el eje giratorio va rotando de forma constante. No obstante, la principal desventaja de utilizar un eje rotativo continuo es que el sistema no es invariante en el tiempo y esto podría distorsionar los resultados de medida por el hecho de que durante un estímulo, el micrófono podría moverse una pequeña trayectoria.

Si se utilizara un array de micrófonos completo para medir las respuestas al impulso en una sala, éstas podrían medirse simultáneamente ahorrando mucho tiempo en la toma de medidas. No obstante, utilizar este tipo de micrófonos trae consigo desventajas. Además del aspecto económico comentado anteriormente, la primera de todas es que en el caso de utilizar muchos sensores en un array, la presencia de uno de ellos podría actuar como borde reflectante afectando al campo sonoro medido en las demás posiciones. Por otro lado, si se utilizan muchos micrófonos, y sabiendo que en la práctica es imposible encontrar dos exactamente iguales, la medición del array circular aparecería deteriorada por estas pequeñas diferencias.

### 2.9.1. Simulaciones

Con el fin de comprobar el método matemático, se han llevado a cabo varias simulaciones. En ellas se han tenido en cuenta diferentes escenarios para obtener mediante su comparativa, resultados que permitieran confirmar la correcta operatividad de todo el proceso de descomposición de ondas planas.

Los escenarios han sido los siguientes:

- variación en la posición de la fuente respecto al array con un radio fijo, y un número de micrófonos también fijo,
- variación en la frecuencia de corte del ancho de banda de la respuesta impulsiva,
- variación del número de micrófonos implicados en el array para una posición fija de fuente, y radio fijo de array,
- variación del radio del array para un mismo número de micrófonos.

En todos los escenarios simulados, se han obtenido las cuatro gráficas reseñadas en la anterior sección, donde la fuente emitía una señal MLS con una frecuencia de muestreo de 44,1 kHz.

#### Primer escenario: Cambio de orientación de la fuente

En el primer caso, el radio del array establecido fue de 1 m y de 288 micrófonos para las siguientes orientaciones de fuente respecto al eje del array: 45°, 180° y 300°.

Las gráficas obtenidas de presión captada por los micrófonos así como las correspondientes a las descomposiciones referentes a armónicos cilíndricos y de ondas planas se ven mostradas en las Figuras 2.12 a 2.15.

Como se puede observar, la grabación circular de la onda plana tiene forma de coseno en el tiempo, donde la fase de esta función coseno está determinada por el ángulo de incidencia.

El cambio de ubicación de la fuente sonora con respecto al array de micrófonos, viene reflejado en principio por las gráficas de presión del

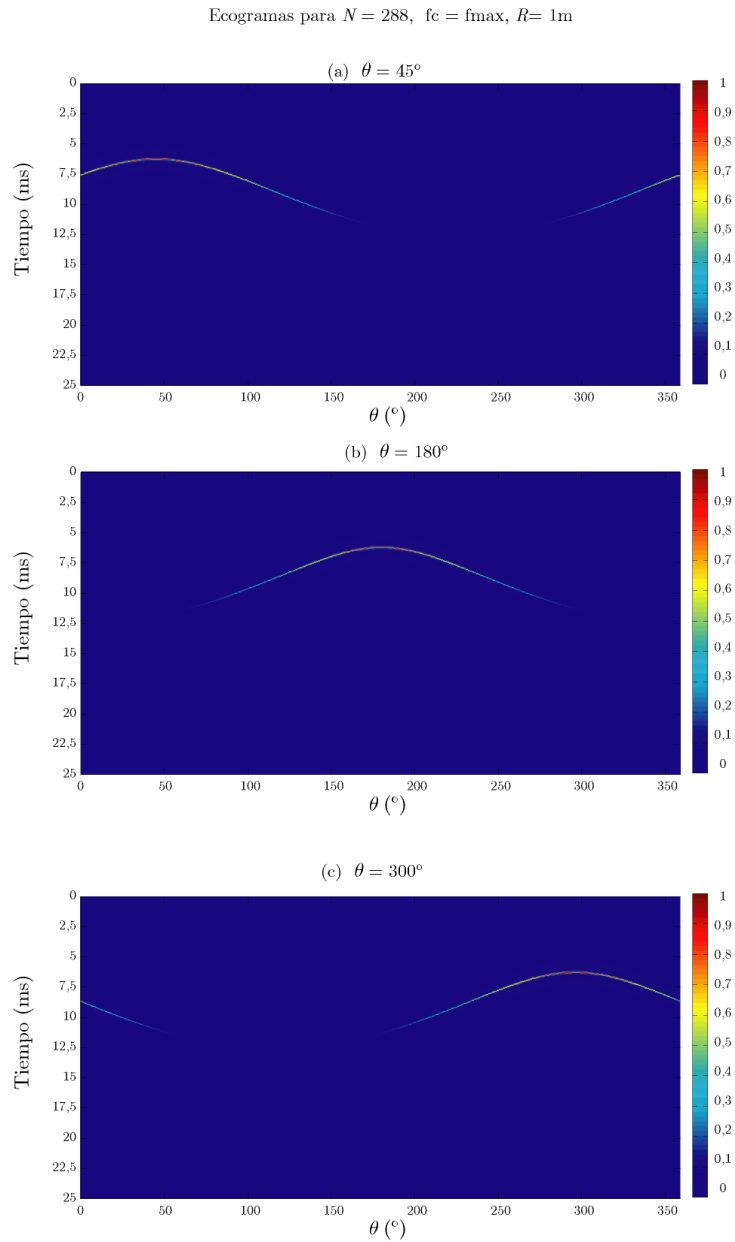


Figura 2.12: Grabación de la presión sonora captada por el array. (a) Posición de la fuente en  $45^\circ$ , (b) en  $180^\circ$ , y (c)  $300^\circ$ .

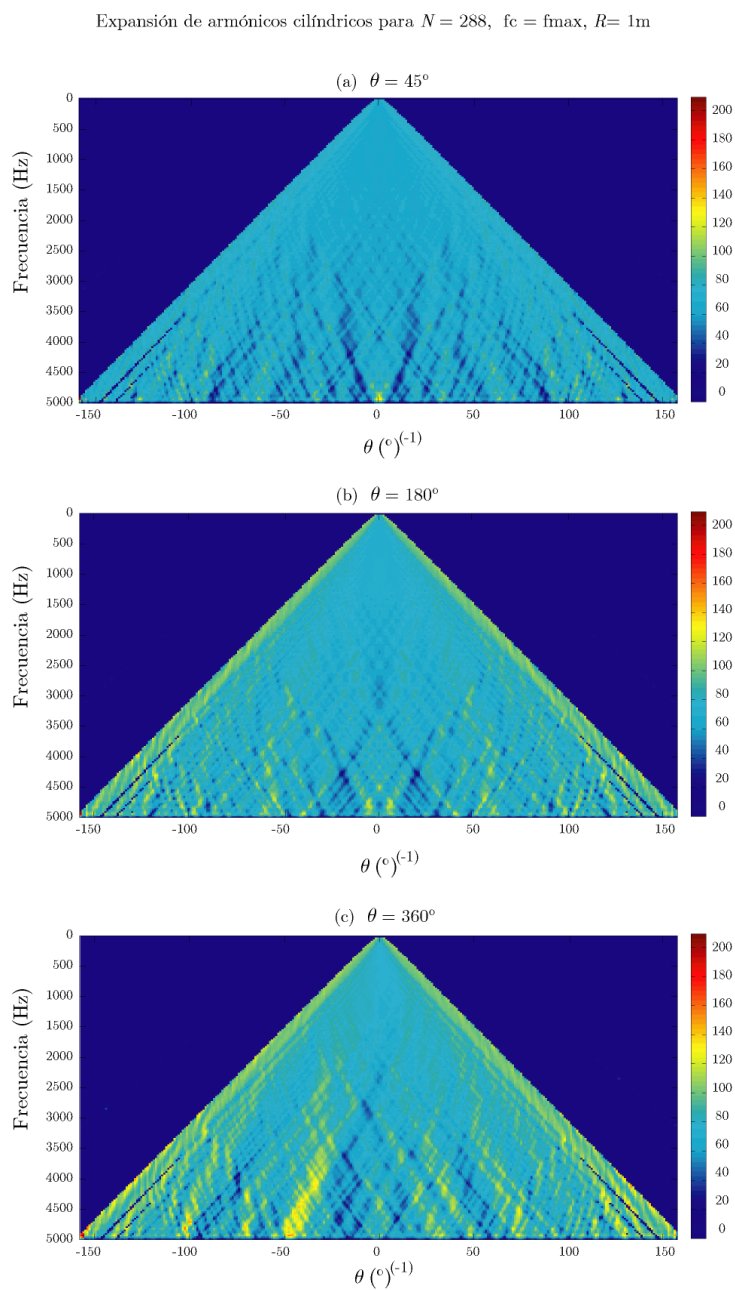


Figura 2.13: Expansión de armónicos cilíndricos para diferentes ángulos de acimut.

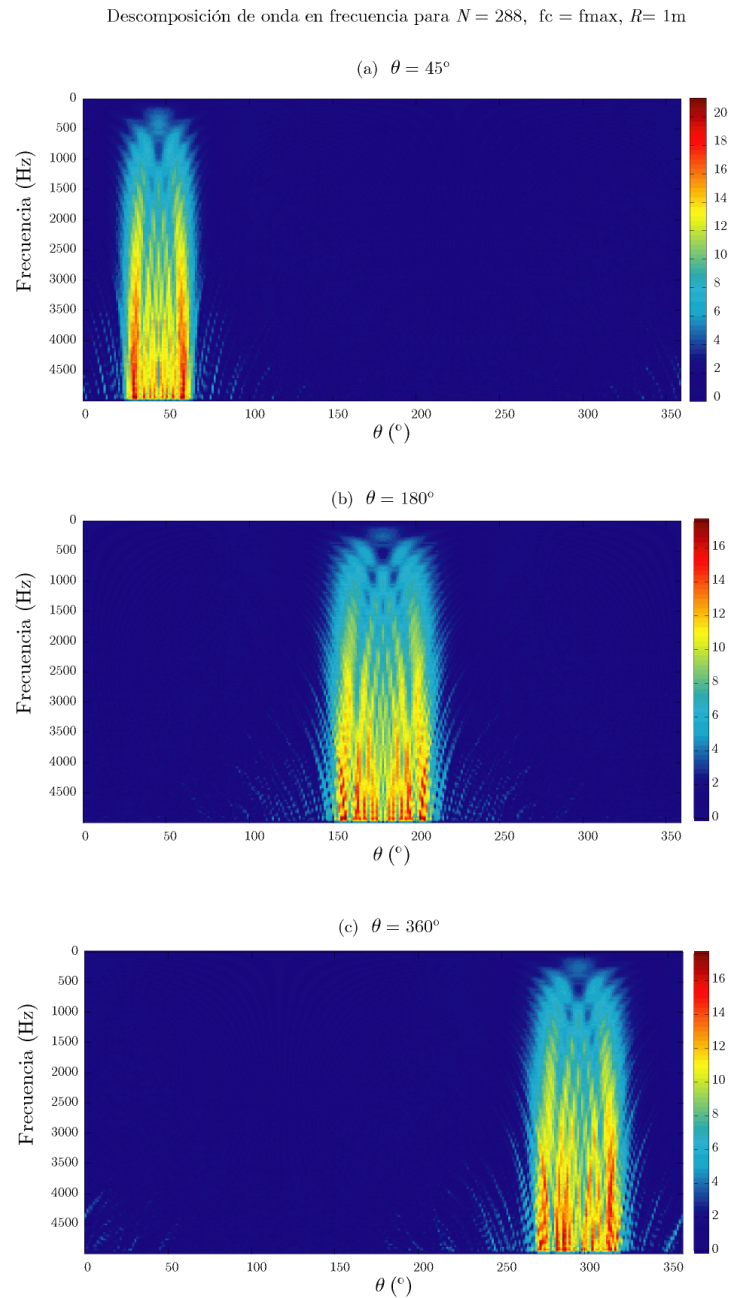


Figura 2.14: Descomposición de onda plana en frecuencia para diferentes ángulos de acimut.

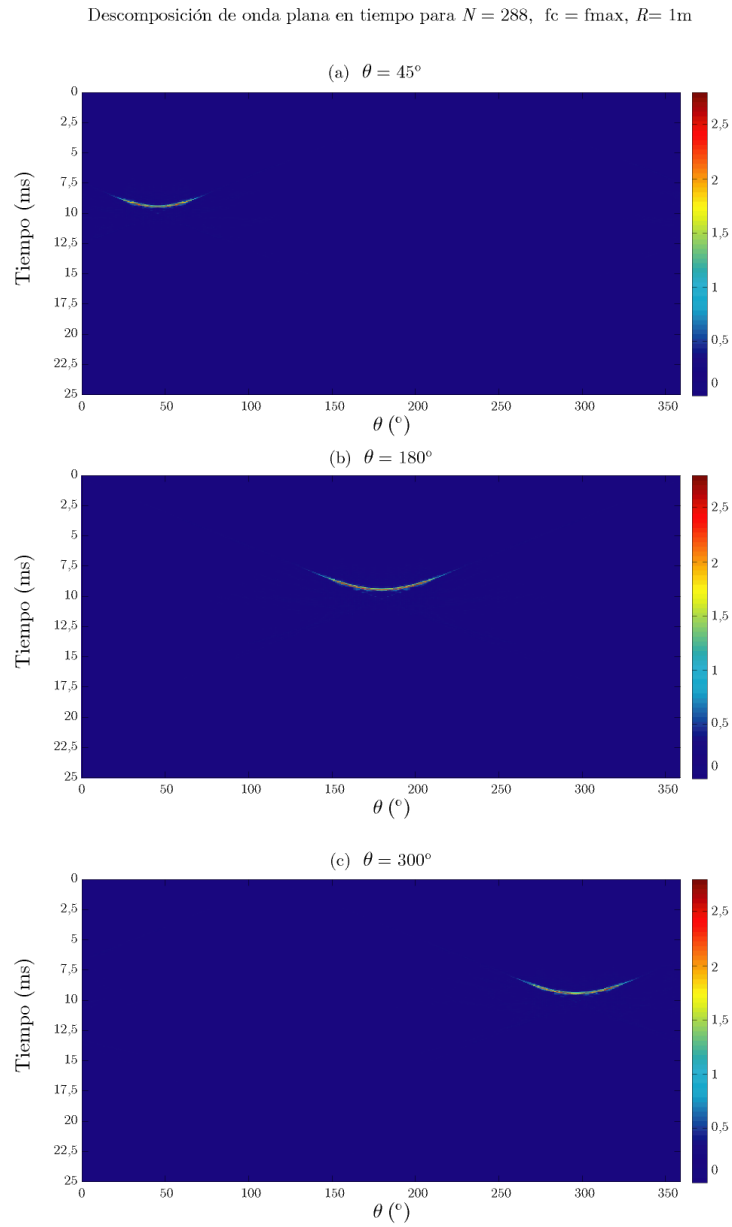


Figura 2.15: Descomposición de onda plana en tiempo para diferentes ángulos de acimut.

campo sonoro donde, como puede observarse se modifica el máximo de esta forma de onda cosenoidal, en función del ángulo que forma la fuente con el eje del array. Así pues, en todos los casos, la forma coseno es la misma (hay que recordar que sólo se varía el ángulo de la fuente con respecto al origen del eje y no el número de micrófonos ni el tamaño del array).

Para conseguir la descomposición en ondas planas según el desarrollo matemático analizado anteriormente, es necesario realizar una FFT2D y obtener las componentes armónicas que vienen mostradas en la Figura 2.13. En dichas representaciones, se ve como para el array simulado de 288 micrófonos, se obtienen unos coeficientes  $k_\theta$  entre  $-\frac{288}{2}$  y  $\frac{288}{2}$  [ $\sigma^{-1}$ ], mostrando las frecuencias que operan para cada coeficiente. En cuanto a las gráficas de descomposición de onda plana en frecuencia y tiempo (Figuras 2.14 y 2.15), se comprueba que el espectro en cada una de ellas no varía en cuanto a forma o tamaño, sólo su orientación se ve alterada donde el máximo aparece en todas ellas precisamente para las posiciones de la fuente.

### **Segundo escenario: Cambio de la frecuencia de corte del ancho de banda de la respuesta impulsiva.**

Es muy importante limitar el ancho de banda de las respuestas al impulso para poder obtener una representación correcta de la descomposición de onda. Como a continuación se ha demostrado, si se varía la frecuencia de corte del ancho de banda de la respuesta impulsiva ( $f_c$ ) en función de la frecuencia de *aliasing* del array ( $f_{al}$ ), se puede comprobar cómo la representación de la descomposición de onda varía de forma que estos resultados son correctos siempre que se limite en banda el contenido espectral de las señales del array por debajo de la frecuencia de *aliasing*.

En este caso, se han analizado tres valores de  $f_c$  en función de la  $f_{al}$ . El primero ha sido suponer una  $f_c=0, 2f_{al}$ . El segundo suponer una  $f_c=f_{al}$ , y por último una  $f_c= 3f_{al}$ . Las gráficas obtenidas para estos casos en este segundo escenario se ven mostradas en las Figuras 2.16, 2.17, 2.18 y 2.19.

Aunque en los ecogramas de la Figura 2.16, no viene reflejado de forma clara el caso en el que se produce *aliasing* espacial, sí se

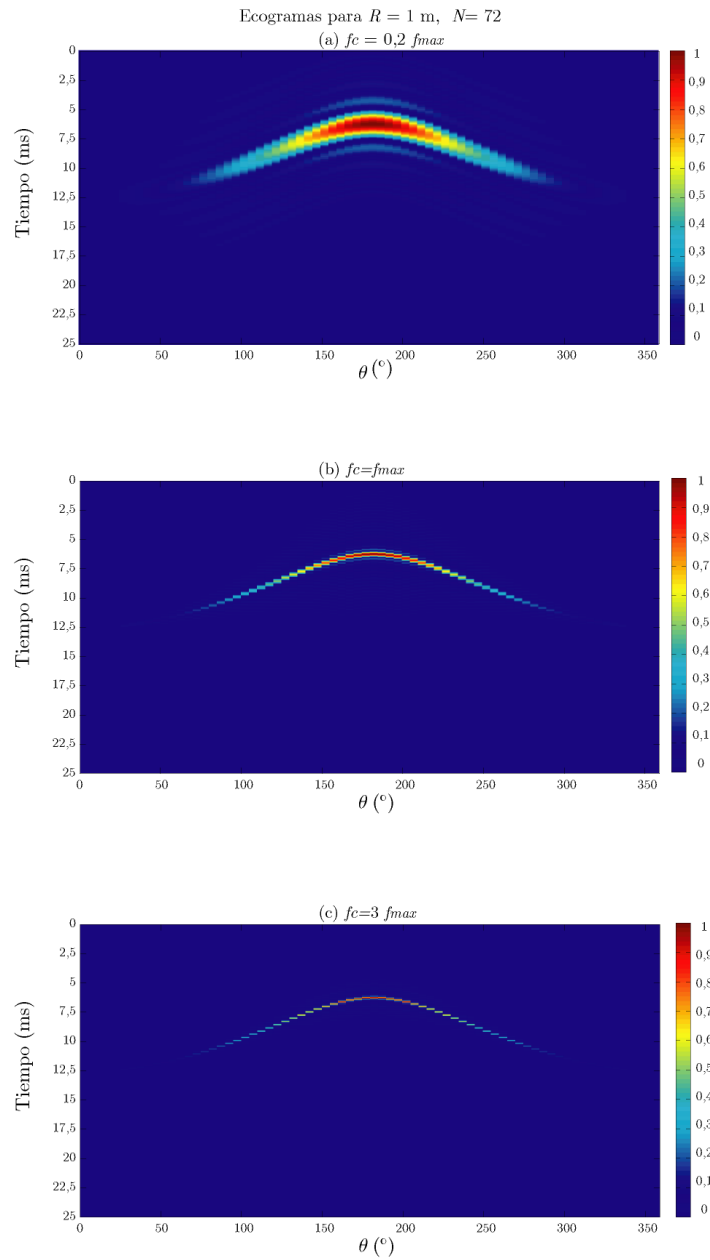


Figura 2.16: Ecogramas para diferentes frecuencias de corte del ancho de banda de la respuesta impulsiva.



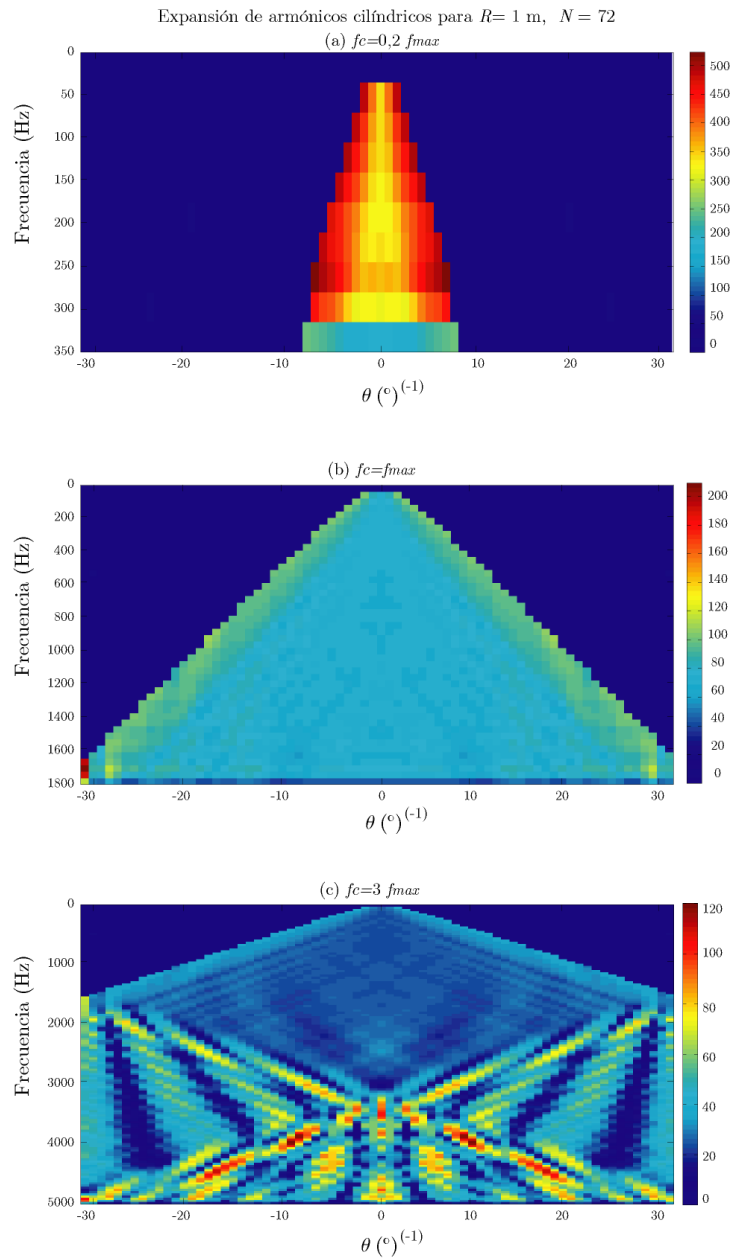


Figura 2.17: Expansión de armónicos cilíndricos para diferentes frecuencias de corte del ancho de banda de la respuesta impulsiva.

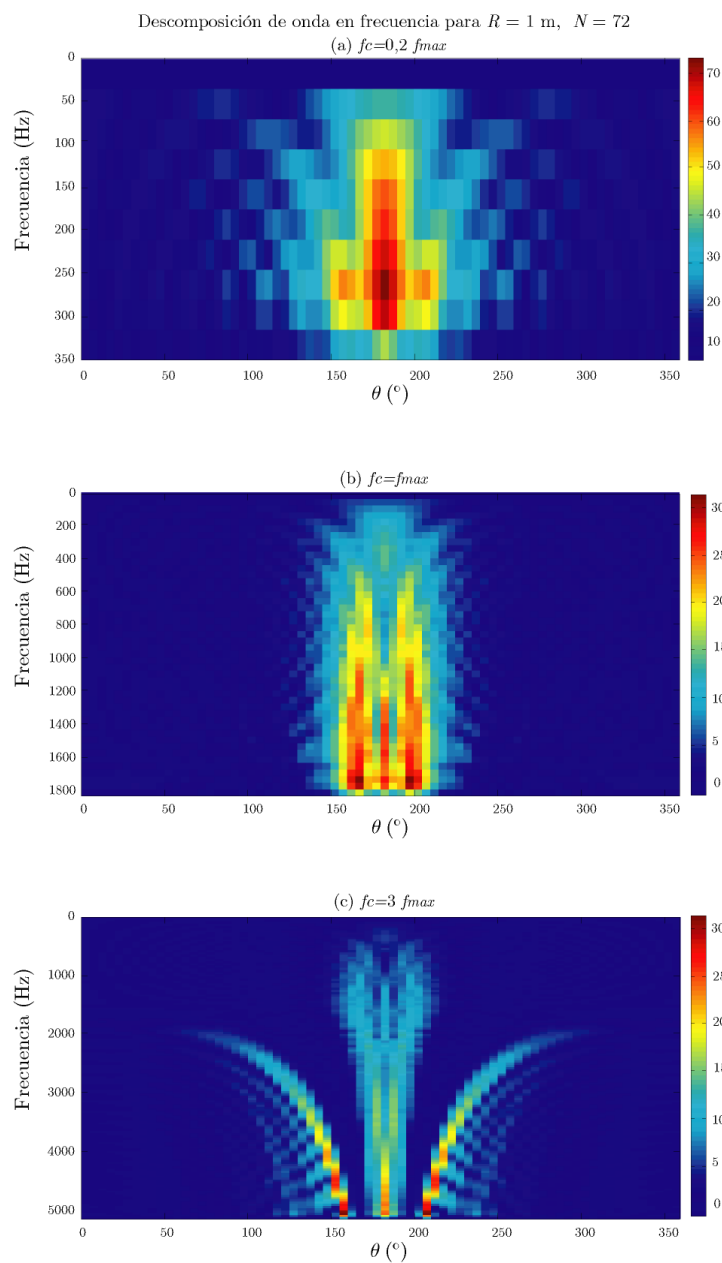


Figura 2.18: Descomposición de onda plana en frecuencia para diferentes frecuencias de corte del ancho de banda de la respuesta impulsiva.

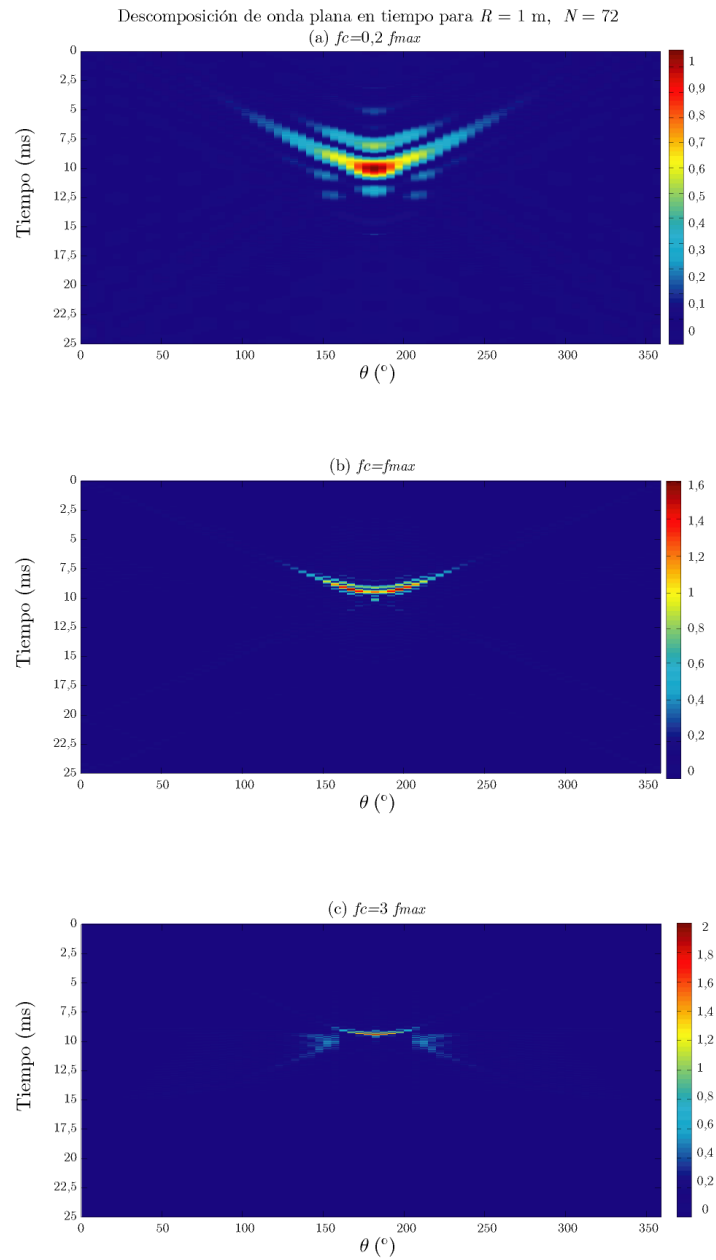


Figura 2.19: Descomposición de onda plana en tiempo para diferentes frecuencias de corte del ancho de banda de la respuesta impulsiva.

puede observar en la Figura 2.17. En este caso y al trabajar muy por debajo del ancho de banda posible de las RIRs ( $f_c=0, 2f_{al}$ ) muchas componentes frecuenciales son desaprovechadas. Este efecto también se observa en las correspondientes representaciones gráficas de las Figuras 2.18 y 2.19. Para el caso límite  $f_c=f_{al}$ , tanto la expansión en armónicos cilíndricos así como sus descomposiciones en frecuencia y en tiempo se muestran de forma correcta y sin *aliasing* espacial. Para el caso donde  $f_c$  supera considerablemente a la  $f_{al}$ ,  $f_c=3f_{al}$ , se observa cómo las distorsiones en los diferentes espectros (sobre todo en los de descomposición de onda plana) son evidentes producidas por la aparición del *aliasing*.

### **Tercer escenario: Variación del número de micrófonos presentes en el array**

El número de micrófonos en el array circular que capta las RIRs es otro de los parámetros fundamentales a la hora de obtener una correcta descomposición de onda plana sin tener *aliasing* espacial. En este escenario se propuso mantener la posición de la fuente con orientación de  $180^\circ$ , respecto al eje del array de 1 m de radio, y se simularon arrays de 36, 72 y 144 micrófonos obteniendo los resultados mostrados en las Figuras 2.20, 2.21, 2.22 y 2.23.

En este caso sí se produce variación en las formas de onda de las presiones recogidas por los micrófonos del array. Como se puede observar, en el caso de que haya menos micrófonos, las presiones de la Figura 2.20 vienen más distorsionadas en su forma coseno, que en el caso de grabar la onda con mayor número de sensores. No obstante existe un compromiso entre el tiempo de procesado y la forma de onda válida para el posterior análisis de descomposición. Hay que resaltar que el aumento por dos del número de micrófonos en cada simulación, elevaba a más del doble el tiempo de procesado, lo cuál puede ser un inconveniente en mediciones más complejas, ya que se requeriría también mayor número de recursos y sistemas potentes y robustos. Además, a mayor número de micrófonos, más fiel y completo es el campo sonoro emitido por la fuente. Esto viene reflejado en las Figuras 2.22 y 2.23 que muestran cómo la descomposición se lleva a cabo de forma más limpia y correcta conforme aumenta el número de sensores

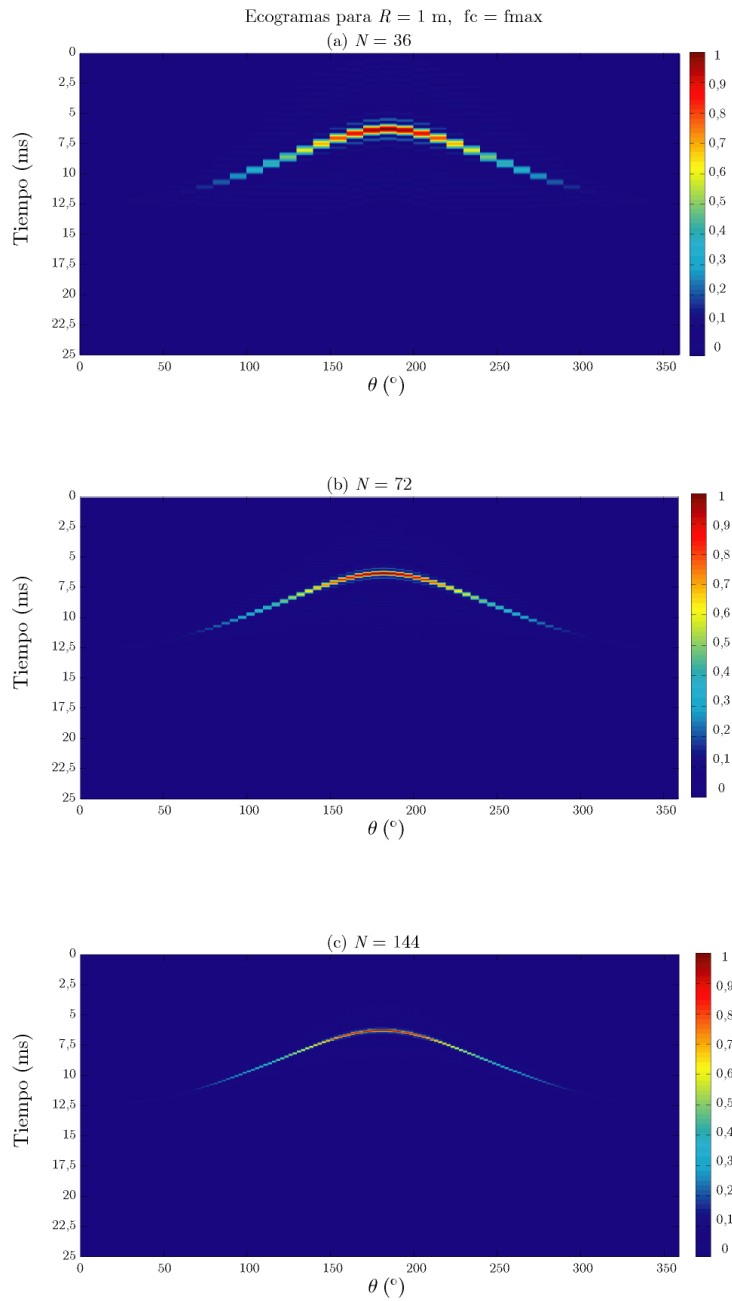


Figura 2.20: Presiones captadas por el array para diferente número de micrófonos.

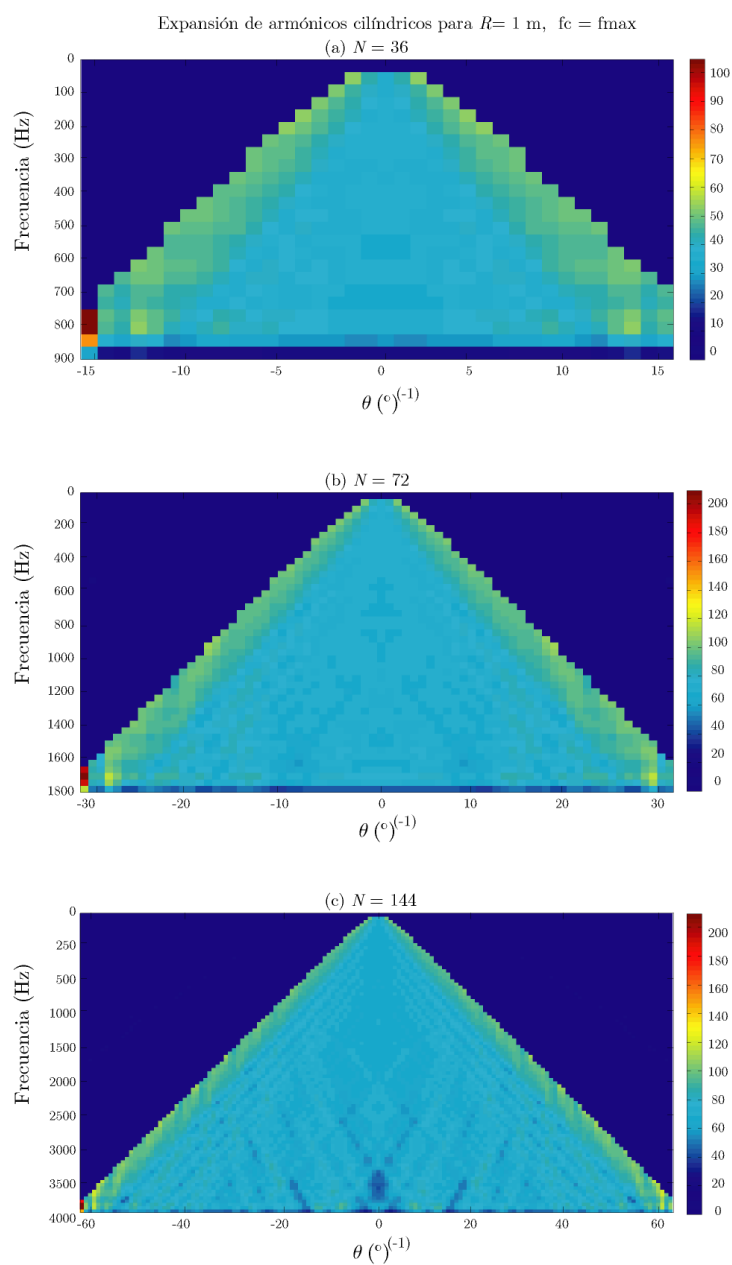


Figura 2.21: Expansión en armónicos cilíndricos para diferente número de micrófonos.

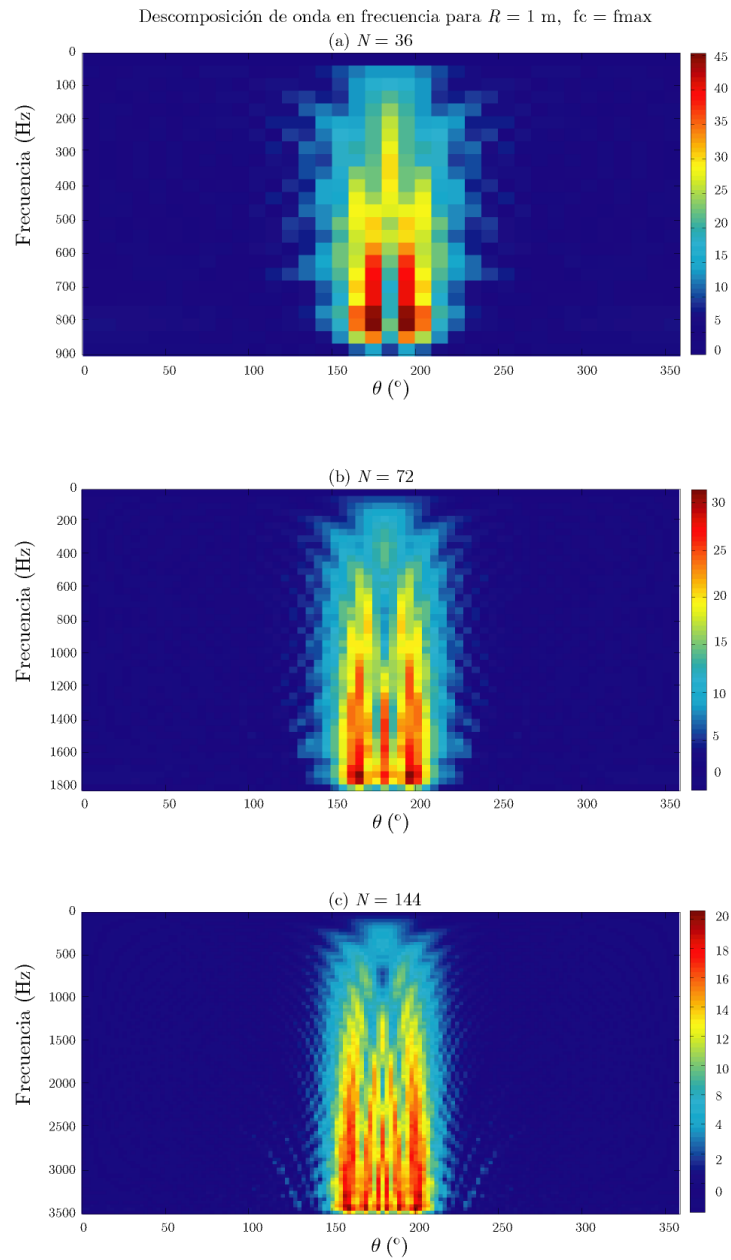


Figura 2.22: Descomposición de onda plana en frecuencia.

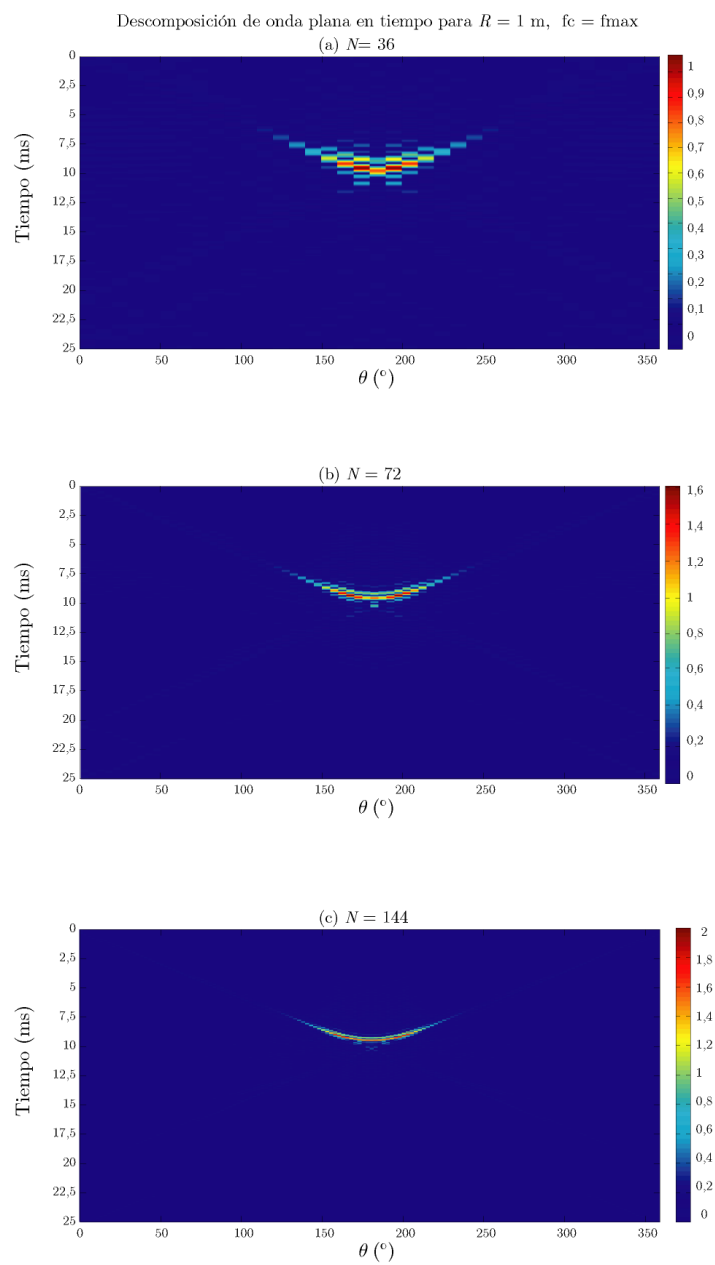


Figura 2.23: Descomposición de onda plana en tiempo.



que tiene el array.

#### **Cuarto escenario: Variación en el tamaño del array**

Por último, se quiso tener en cuenta cómo afectaba el tamaño del array a la descomposición de onda, simulándose varias medidas para un radio de 0,5 m, 1 m y 1,5 m, con un número de micrófonos fijo, Figuras 2.24, 2.25, 2.26 y 2.27.

En este caso se observa cómo la forma de onda coseno de la presión grabada por los micrófonos, se hace más directiva y menos plana conforme aumenta el radio del array como se muestra en la Figura 2.24. Esto se refleja también en las gráficas de la descomposición de onda, donde se comprueba que a mayor radio para una posición de fuente dada, el espectro comienza a presentar fenómenos de *aliasing* espacial como se ve en la Figura 2.26 (c) para un radio de 1,5 m. Esto se debe a la relación entre número de micrófonos, distancia entre ellos y radio del array, factores que están íntimamente relacionados de forma que el aumento de uno de ellos afectará al resto y en consecuencia a la frecuencia y aparición de *aliasing* espacial.

### **2.9.2. Comprobación experimental**

Una vez se simularon las ondas planas en un array circular para los anteriores casos, como paso último y previo a la captura de medidas en auditorios y salas de mayores dimensiones, se llevó a cabo la toma experimental de respuestas impulsivas en una pequeña sala del Instituto de Telecomunicaciones y Aplicaciones Multimedia (ITEAM) de la Universidad Politécnica de Valencia, concretamente en el Grupo de Tratamiento de Audio y Comunicaciones (GTAC). En ella aparecía un panel muy cercano al array y que resultó interesante a la hora de ver cómo repercutía la reflexión de éste en la descomposición de onda plana.

Se utilizó como estímulo un sweep logarítmico de 2 segundos de duración, y se posicionó una fuente (altavoz) formando  $90^\circ$  con el eje del array de micrófonos según la disposición mostrada en la Figura 2.28. Ésta orientación se escogió con la finalidad de observar

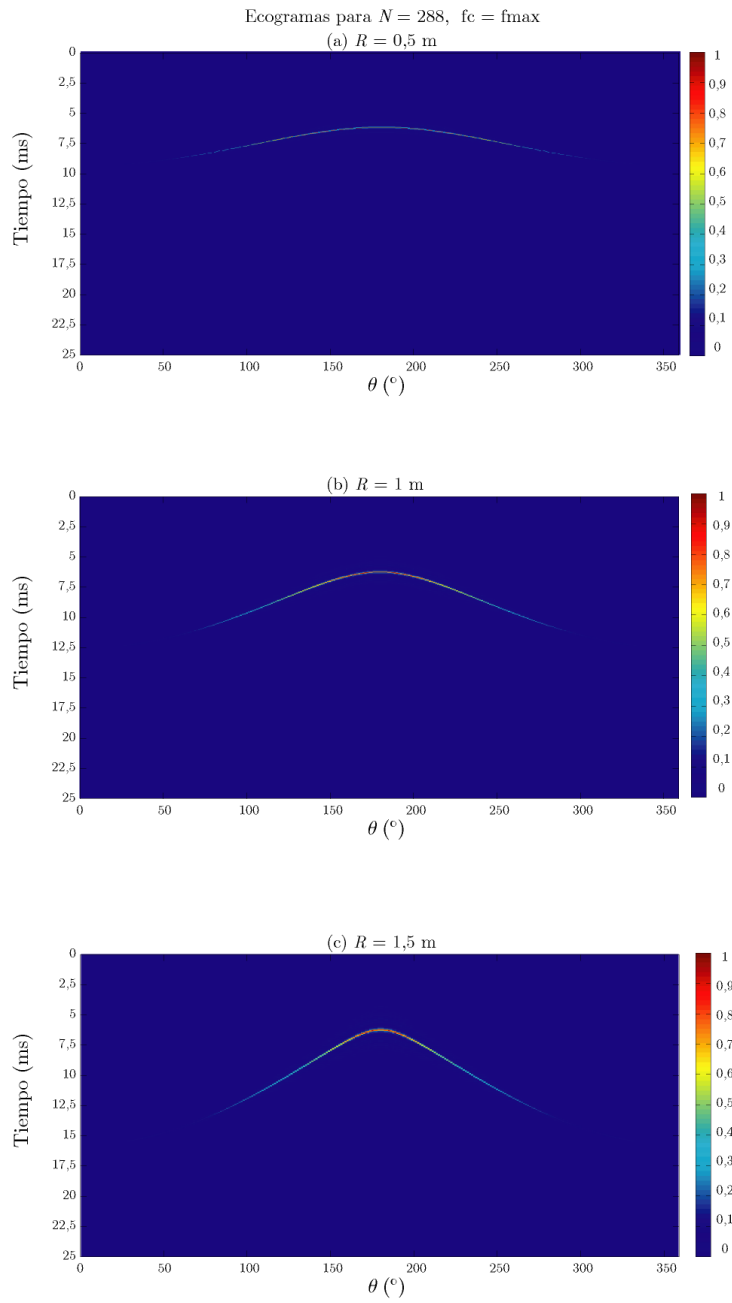


Figura 2.24: Ecogramas obtenidos para diferentes tamaños de arrays.

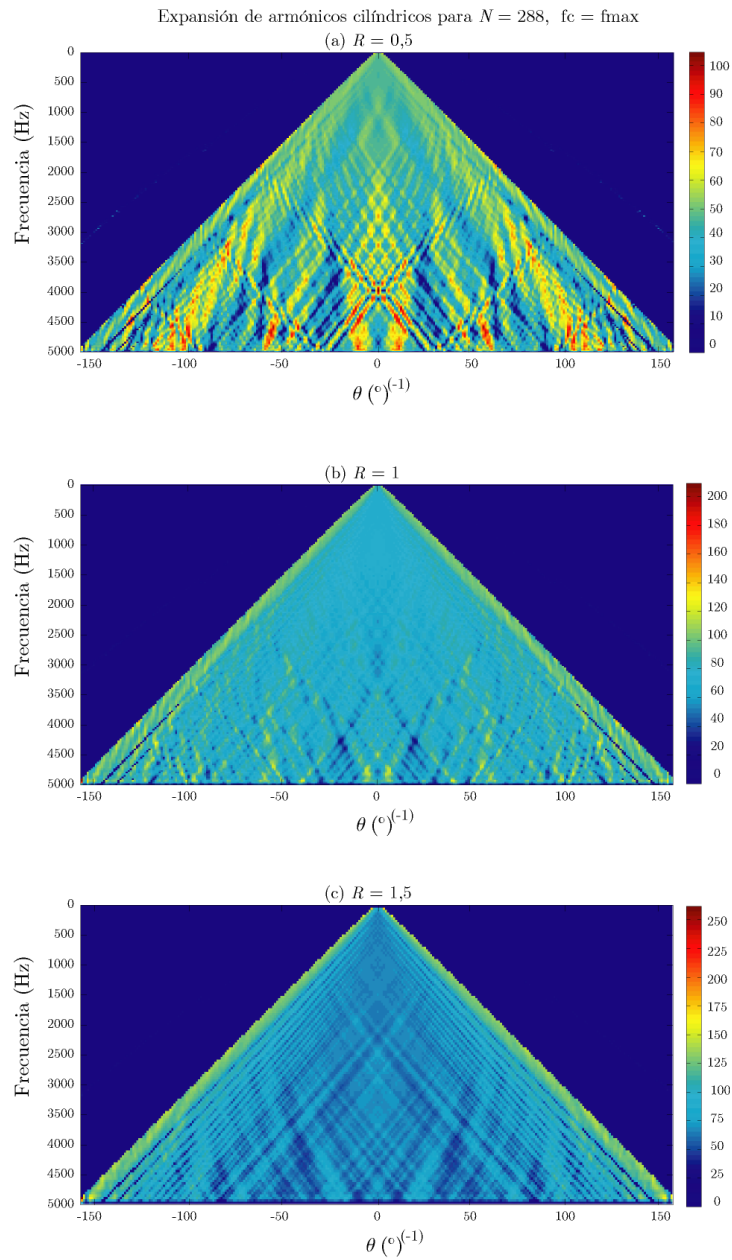


Figura 2.25: Expansión en armónicos cilíndricos obtenidos para diferentes tamaños de arrays.

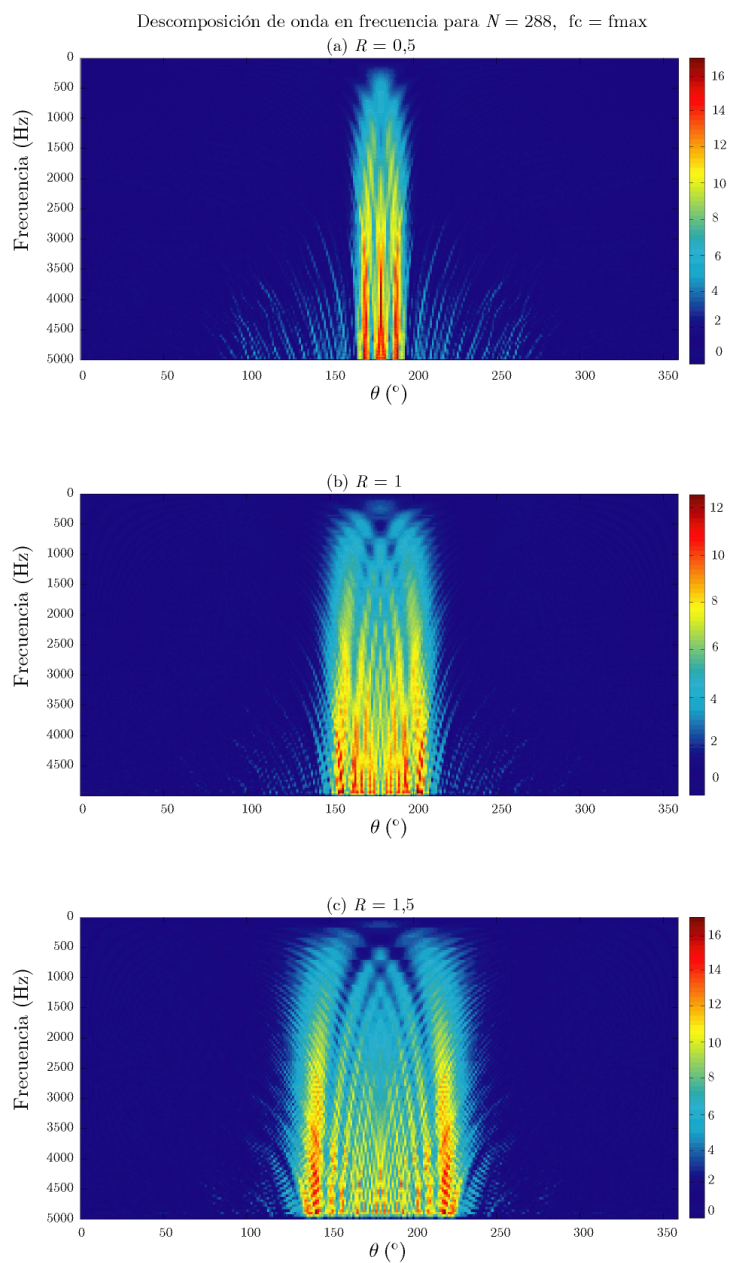


Figura 2.26: Descomposición de onda plana en frecuencia.

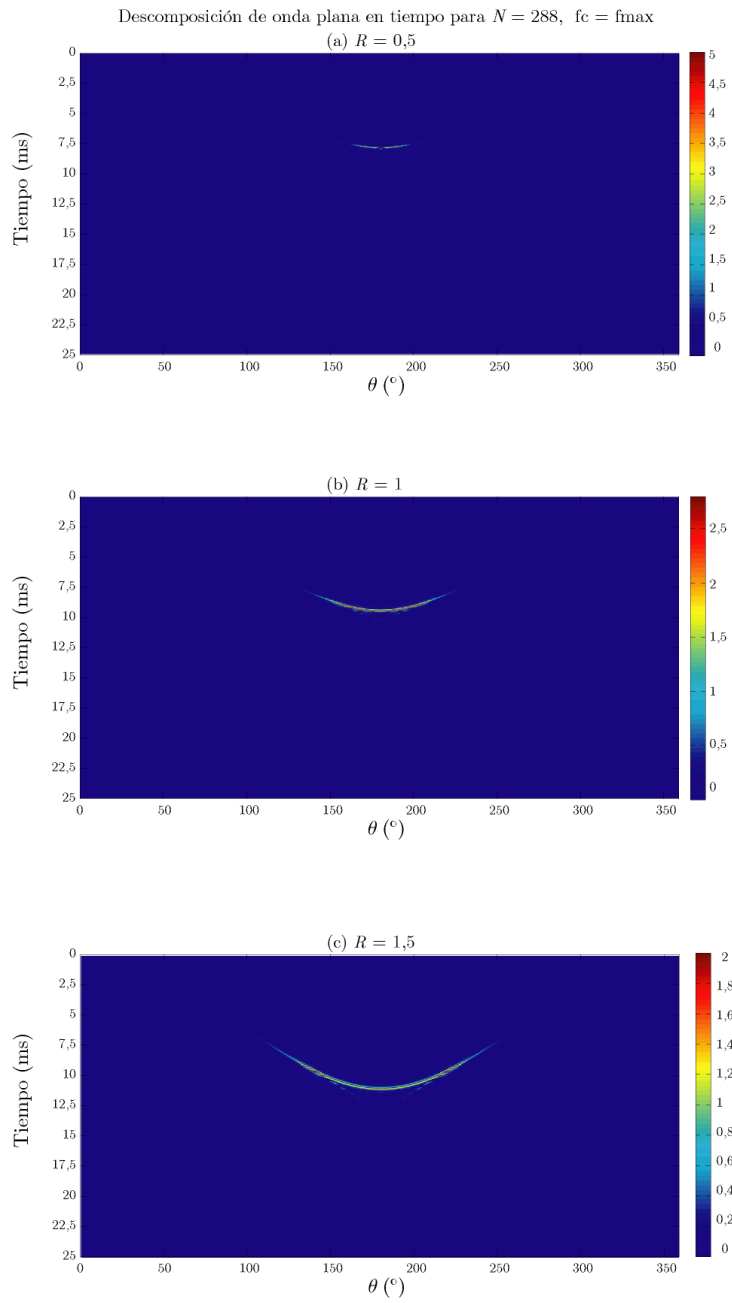


Figura 2.27: Descomposición de onda plana en tiempo.

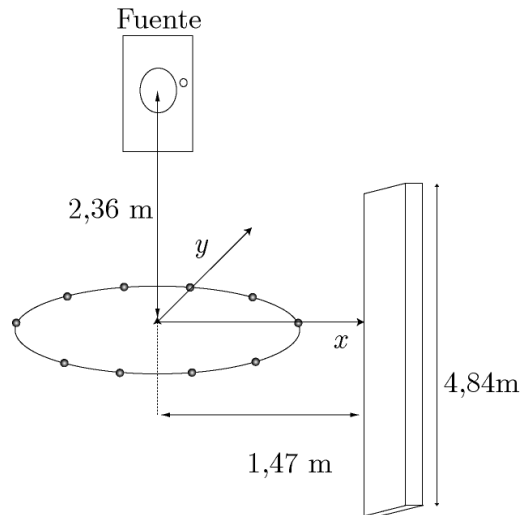


Figura 2.28: Montaje experimental con las posiciones del altavoz con respecto al eje del array.

la incidencia del panel que se muestra en la Figura 2.29 y poder comprobar que su realización se había hecho correctamente.

Las dimensiones del array circular y el número de micrófonos utilizados vinieron determinados por dos criterios de diseño: por un lado la resolución angular requerida para bajas frecuencias y por otro lado la frecuencia de Nyquist requerida para evitar el *aliasing* espacial. La resolución angular viene determinada por la apertura del array y es proporcional a la relación del radio del array la longitud de onda del sonido. Cuanto mayor es el array, mayor detalle angular se puede alcanzar a una frecuencia dada. La frecuencia Nyquist determina el mínimo espaciado necesario entre los micrófonos. Junto con el tamaño del array, determina el número de micrófonos que se necesitan finalmente. El array utilizado en las medidas de este trabajo tuvo un radio de  $R=0,82$  m y se utilizaron 2 micrófonos. El array rotó de manera discreta en pasos de cinco grados alrededor de la circunferencia, simulando una captura de 72 micrófonos ( $\frac{360^\circ}{5^\circ}=72$ ). Esto hizo que la frecuencia de Nyquist fuera de:

$$f_{al} = \frac{c}{2d} = 2375 \text{ Hz}, \quad (2.74)$$



Figura 2.29: Detalle del montaje llevado a cabo, donde se aprecia el parabán que tendrá especial relevancia en la descomposición de ondas planas.

para  $N=72$  micrófonos.

Como ya se ha mencionado, el array de micrófonos se realizó mediante la sujeción de dos micrófonos cardioides modelo *Earthworks SR77* apuntando hacia el exterior del array en una barra sujeta a un motor rotable, Figura 2.30. Por medio del software *SoundTenax* creado por el Director de esta tesis fue posible escoger tanto el tipo de estímulo a emplear como almacenar las diferentes RIR's mediante la rotación comentada simulando un array de 72 micrófonos. El tener dos micrófonos a  $180^\circ$  en esa barra que rotaría un determinado ángulo, permitió disminuir a la mitad el número de medidas necesarias.

Se comprobó por medio de la simulación previa que utilizando ese número de micrófonos en el análisis, la descomposición era bastante precisa, sin contener mucha distorsión y con un tiempo de procesado no excesivamente elevado. En la Figura 2.31 se muestran los resultados obtenidos concretamente para la posición (c) indicada en la Figura 2.28.

En la Figura 2.31 (a) se muestra el ecograma captado por el array con todas las reflexiones capturadas. En él se observa cómo en una



Figura 2.30: Vista de la barra sujeta al motor giratorio.

posición de aproximadamente  $90^\circ$  se produce la llegada del sonido de la fuente (sonido directo), y cómo en una posición de  $270^\circ$  se observa una reflexión muy notable proveniente del panel de la sala. La Figura 2.31 (b) muestra hasta qué frecuencia de trabajo aparecen componentes frecuenciales, y el número de armónicos cilíndricos obtenidos para ese número de micrófonos. Los resultados de la descomposición en onda plana tanto en frecuencia como en tiempo, vienen reflejados en las Figuras 2.31 (c) y (d) respectivamente. En ellas viene representado de forma precisa el espectro y la zona de energía donde se producen las principales reflexiones.

Como se puede comprobar, las medidas experimentales se ajustan a las obtenidas mediante simulación, pero reflejan mejor las diferentes reflexiones que tienen lugar en la sala de prueba. Esto ayudará en el resto del trabajo para poder estudiarlas de forma adecuada y para analizar la repercusión que tienen en la acústica de salas.

### 2.9.3. Discusión de los resultados

Los fundamentos teóricos, así como los resultados previos que se han visto en este capítulo, sirvieron para implementar un algoritmo



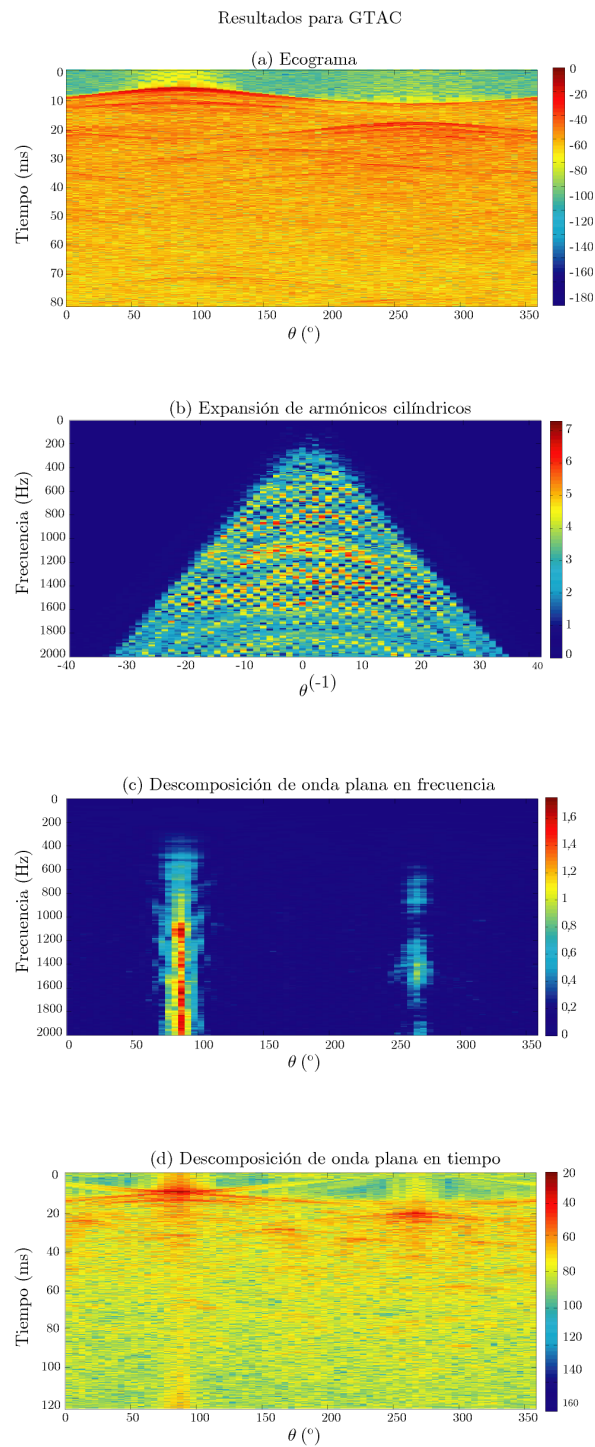


Figura 2.31: Descomposición de onda plana en la sala GTAC para medidas reales.

basado en la descomposición de ondas planas en un array circular gracias a los armónicos cilíndricos. La obtención de medidas tomadas en la sala, junto con los resultados que se observan en sus descomposiciones, conforman el punto de partida para llevar a cabo el resto de puntos de esta tesis.

En las simulaciones realizadas, no aparecían efectos reales de las salas, como son paredes, techos u otro tipo de obstáculos que afectan a las respuestas impulsionales que se pueden grabar y por tanto a la descomposición posterior de onda plana. Debido a ello, se hizo la toma experimental donde se reflejó lo que en Matlab se había analizado dando más información añadida sobre reflexiones y cómo éstas afectaban a las diferentes gráficas.

En la sala analizada se ha demostrado cómo aparece una reflexión de forma bastante acusada en la representación de descomposición de onda en frecuencia y que tiene que ver con un panel que se encontraba en la sala. Así mismo, la descomposición de onda viene afectada por todas esas reflexiones (que de nuevo, la más significativa es la del panel con ubicación a  $270^\circ$  respecto al eje del array), y que sumada a las del techo, suelo y otros obstáculos de menor tamaño, forma completamente el patrón espectral captado.

Se comprobó también que la disposición del array, junto con el número de micrófonos existentes en el mismo, puede condicionar enormemente la frecuencia de Nyquist y de *aliasing* y que las reflexiones de una sala real, quedan reflejadas en los espectros de onda ayudando a un posterior análisis y procesado. Asimismo, puesto que el array bajo estudio que se va a considerar en la tesis es circular y no esférico, los ángulos de elevación no han podido ser grabados. En aplicaciones prácticas no obstante, esto no se considera un gran problema incluso aunque se den fuertes reflexiones en el techo [14].

Por último, en este trabajo y considerando el compromiso entre la resolución y el tiempo de procesado, se ha tomado la determinación de elegir arrays de 72 micrófonos para el resto de la tesis, ya que se ha comprobado que con este número de sensores, los resultados han sido correctos y el tiempo computacional en el procesado de resultados no ha sido excesivamente elevado.



# Capítulo 3

## Beamforming

---

<b>3.1. Introducción . . . . .</b>	<b>69</b>
<b>3.2. Conceptos básicos . . . . .</b>	<b>69</b>
3.2.1. Patrón de directividad . . . . .	71
3.2.2. Aliasing espacial . . . . .	73
3.2.3. Conformador de haz o beamformer . . . . .	74
<b>3.3. Geometría de array . . . . .</b>	<b>78</b>
<b>3.4. Aperturas circulares . . . . .</b>	<b>79</b>
3.4.1. Apertura circular continua . . . . .	79
3.4.2. Apertura circular muestreada . . . . .	80
<b>3.5. Beamforming con armónicos circulares . . . . .</b>	<b>81</b>
3.5.1. Selección de modos y regularización . . . . .	82
<b>3.6. Beamforming Delay and Sum . . . . .</b>	<b>84</b>
<b>3.7. Comparativa entre CHB y DAS . . . . .</b>	<b>86</b>
<b>3.8. Aplicación del método CHB en la obtención de ecogramas . . . . .</b>	<b>89</b>
3.8.1. Ecograma simulado . . . . .	93
3.8.2. Ecograma real . . . . .	95

---

*En este capítulo se va a realizar una descripción de las dos técnicas de beamforming más utilizadas para localización en dos dimensiones a través de un array circular. Además se llevará a cabo una comparación entre ambas, que pondrá de manifiesto la buena operatividad del método de beamforming implementado en el Capítulo 6 de esta tesis.*

### 3.1. Introducción

Se conoce como *beamforming* a la técnica dentro del procesado de señal que permite que un receptor pueda discriminar entre diferentes señales incidentes, dependiendo de cuál sea la localización espacial de cada una de ellas. El *beamforming* es por tanto la manera más sencilla de mejorar una señal mediante un array [29] y puede asemejarse a un filtro espacio-temporal, ya que manipula las señales en el dominio del tiempo y del espacio [30]. Un *beamformer* extrae la señal deseada de una dirección específica dentro de un ambiente reverberante influenciado por fuentes que interfieren en ésta. Si se apunta el eje de máxima captación de un array hacia la fuente deseada se estará atenuando la señal procedente de otras fuentes no situadas en dicho eje, que se consideran como fuentes de ruido. Además, se atenuará el sonido procedente de las reflexiones en las paredes, el techo y el suelo de la sala reduciendo de tal forma el nivel de reverberación [31].

En este capítulo se derivan dos técnicas de *beamforming* para obtener, por un lado el ecograma de una sala y por otro lado para localizar fuentes en dos dimensiones a partir de arrays circulares. La primera técnica se llama *delay and sum* (DAS), y se considera como la técnica clásica en *beamforming* [32], [33]. La segunda es una técnica novel, que se conoce también como *beamforming* modal con armónicos circulares (CHB). Esta técnica puede incluirse dentro de las técnicas conocidas como de *eigenbeamforming*, que están basadas en la descomposición del campo sonoro en series de armónicos [34], [35].

A continuación, y antes de entrar en cada una de estas técnicas de forma más detallada, es conveniente introducir un marco teórico con los conceptos básicos del *beamforming*, para después poder abordarlo dentro del contexto de arrays circulares.

### 3.2. Conceptos básicos

Un array de micrófonos puede considerarse como una versión muestreada de una superficie continua (también llamada apertura como se explicó en la Sección 2.1), donde sólo es excitada en un número finito de puntos y cuya respuesta global se obtiene por la

superposición de todos y cada uno de los sensores que lo conforman. De esta forma se define por un lado la longitud efectiva de un array uniforme como la longitud de la superficie continua que es muestreada y por otro la longitud física, como la distancia entre el primer y el último micrófono.

La recepción de señal mediante la utilización de arrays de micrófonos ofrece importantes ventajas frente a las limitaciones de tener un único sensor. Básicamente, estas ventajas se refieren a la directividad y a la capacidad de rechazo de interferencias.

En primer lugar, la directividad de un array de sensores depende del radio de la circunferencia en el caso de arrays circulares, y de la longitud total en el caso de array lineales, pero no del número de elementos que lo componen. En consecuencia, a partir de un número determinado de sensores puede incrementarse la directividad del conjunto aumentando su separación. Sin embargo, si la localización de los elementos del array es equiespaciada, la distancia entre ellos debe de cumplir la Ecuación (2.70) comentada en la Sección 2.7. En caso contrario, el diagrama de radiación del array presentará lóbulos de difracción que derivarán en un aumento de la potencia de ruido recibida así como en la imposibilidad de cancelación de posibles señales incidentes desde las direcciones de dichos lóbulos, es decir, el array experimentará un decremento de la relación señal a ruido más interferencias (SNR) a su salida. Por otra parte, una distribución no uniforme o aleatoria de los elementos del array produce un incremento del nivel de los lóbulos laterales que redundará, por consiguiente, en una disminución de la SNR. No obstante, el valor esperado del diagrama de radiación resultante de una adecuada aleatorización de las posiciones de los sensores será proporcional al diagrama de radiación de la correspondiente apertura continua, el cual no presentará lóbulos de difracción ni niveles elevados de lóbulos secundarios.

En segundo lugar, la capacidad de rechazo de interferencias de un array adaptativo se basa en la diversidad espacial que utiliza, la cual permite el diseño de diagramas de radiación enfocados a la dirección deseada con nulos en las direcciones de incidencia de las señales interferentes. A fin de conseguir el diagrama de radiación

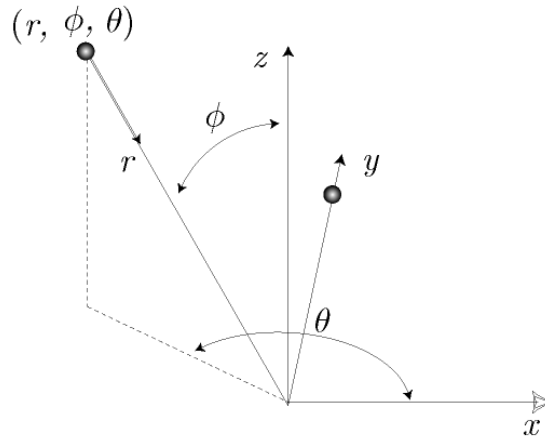


Figura 3.1: Convenio de representación en coordenadas esféricas.

óptimo para cada escenario, el vector de pesos del array o vector de enfoque se estima adaptativamente mediante las distintas técnicas de conformación de haz o *beamforming*, de forma que la SNR en recepción sea máxima.

### 3.2.1. Patrón de directividad

Como ya se ha comentado, la respuesta global de un array de micrófonos se obtiene por superposición de las respuestas individuales de cada uno. A esta respuesta se le denomina patrón de directividad y se puede expresar en su forma más simplificada (arrays lineales) como [32]

$$D(f, \theta, \phi) = \sum_{n=\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) e^{j \frac{2\pi}{\lambda} \text{sen } \phi \cos \theta n d}, \quad (3.1)$$

donde  $N$  es el número de elementos del array,  $w_n(f)$  es el peso complejo conjugado de cada elemento  $n$ , y  $d$  la distancia entre micrófonos. En la Figura 3.1 se muestra el convenio generalmente utilizado para la representación en coordenadas esféricas.

Considerando únicamente la dirección horizontal ( $\phi = \frac{\pi}{2}$ ) se obtiene



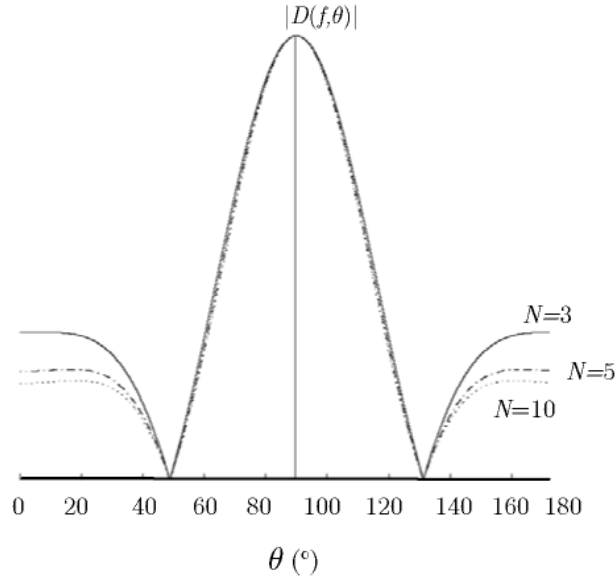


Figura 3.2: Patrón de directividad para un array lineal con  $f=1$  kHz, longitud=0,5 y número de elementos variable.

$$D(f, \theta) = \sum_{n=\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) e^{j \frac{2\pi}{\lambda} \cos \theta n d} = \sum_{n=\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) e^{j \frac{2\pi f}{c} \cos \theta n d}. \quad (3.2)$$

La fórmula anterior arroja interesantes conclusiones, ya que permite observar el comportamiento del patrón de directividad del array con la variación de diferentes parámetros como son el número de elementos  $N$  que conformarían el array, el tamaño de éste y la frecuencia.

La Figura 3.2 muestra el descenso de los lóbulos laterales con el consiguiente aumento de la directividad global conforme aumenta  $N$  y para un tamaño de array y frecuencia invariables. Por otro lado, la Figura 3.3 muestra la disminución de la anchura del lóbulo central conforme aumenta la longitud, y dejando el número de elementos  $N$  y la frecuencia fijos. Esto mismo ocurriría en el caso de aumentar la frecuencia manteniendo fijos tanto el número de micrófonos como el tamaño del array.

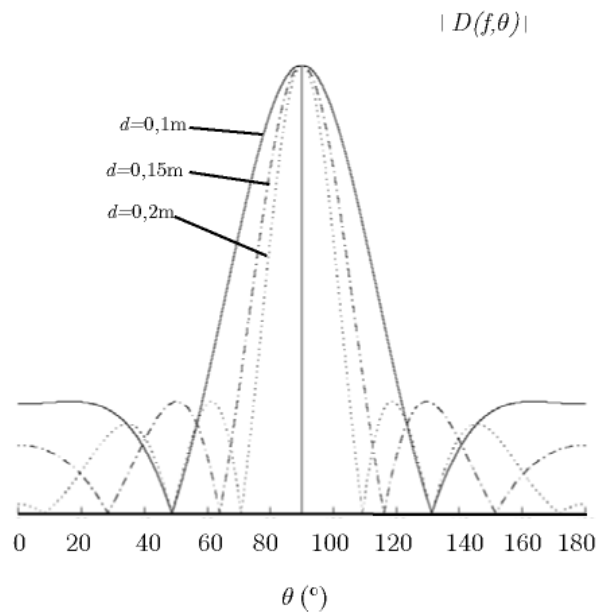


Figura 3.3: Patrón de directividad para un array lineal con  $f=1\text{kHz}$ ,  $N=5$  y longitud variable.

En consecuencia, existe una relación de compromiso que se debe alcanzar en sistemas que emplean arrays de micrófonos mediante una anchura de lóbulo principal lo suficientemente estrecha para conseguir un filtrado espacial adecuado que elimine la mayor cantidad posible de señales indeseadas, sin que esta anchura sea excesivamente estrecha y provoque que el algoritmo de localización (el cual no conoce a priori la posición de las fuentes) filtre también la señal deseada y por lo tanto se produzcan errores en la estimación de la posición.

### 3.2.2. Aliasing espacial

De acuerdo con el principio de Nyquist, la frecuencia de muestreo debe ser mayor o igual que el doble de la máxima componente frecuencial de la señal para evitar *aliasing* (o solapamiento) en frecuencia

$$f_s = \frac{1}{T_s} \geq 2f_{\text{máx}}. \quad (3.3)$$

Análogamente se puede aplicar este principio a los arrays de micrófonos, que constituyen una versión muestreada de una apertura continua [32]

$$f_{x_s} = \frac{1}{d} \geq 2f_{x_{\text{máx}}}, \quad (3.4)$$

donde  $f_{x_s}$  es la frecuencia de muestreo espacial en muestras por metro,  $f_{x_{\text{máx}}}$  es la mayor componente de frecuencia espacial y el array es lineal. La frecuencia a lo largo del eje  $x$  viene dada por:

$$f_{x_s} = \frac{\text{sen } \phi \cos \theta}{\lambda}, \quad (3.5)$$

alcanzando su máximo para

$$f_{x_{\text{máx}}} = \frac{1}{\lambda_{\text{mín}}}, \quad (3.6)$$

con lo que la condición final que se obtiene es:

$$d \leq \frac{\lambda_{\text{mín}}}{2}, \quad (3.7)$$

donde  $\lambda_{\text{mín}}$  representa la mínima longitud de onda de la señal de interés.

La Ecuación (3.7), conocida como teorema de muestreo espacial, se debe cumplir para evitar el *aliasing* espacial. El incumplimiento de esta condición provocaría la aparición de grandes lóbulos en direcciones indeseadas en el patrón de directividad, con la consiguiente integración de información procedente de dichas direcciones y los evidentes efectos negativos en cuanto a distorsión de la señal deseada y localización de los targets de la escena. La Figura 3.4 representa gráficamente dicha problemática [36].

### 3.2.3. Conformador de haz o beamformer

Como se verá más adelante, el patrón de la salida de un conformador de haz o *beamformer* varía de una técnica a otra y varía también simplemente cambiando la frecuencia. Por estas razones, es necesario evaluar el rendimiento de un *beamformer* por medio de las siguientes medidas: resolución y nivel máximo del lóbulo lateral (NMLL).

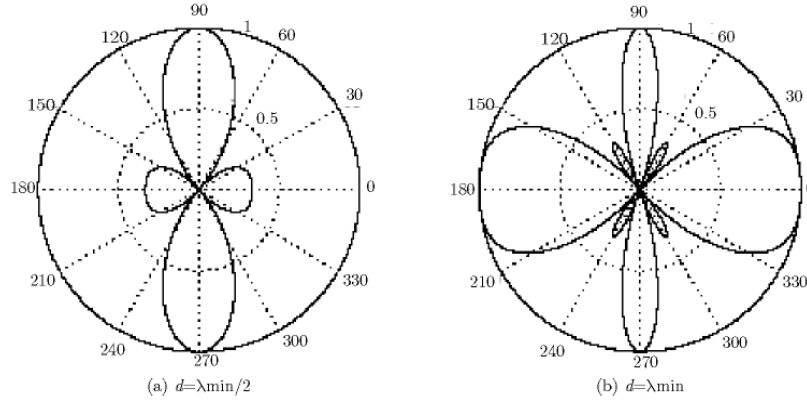


Figura 3.4: Patrón de directividad sin aliasing espacial (a) y con aliasing (b).

#### ■ Resolución

La resolución de un *beamformer* se define como la anchura a  $-3$  dB del lóbulo principal de su diagrama de radiación o *beam pattern* [37]. En la Figura 3.5, la resolución es calculada a partir de la salida de un *beamformer* que captura el sonido de una fuente localizada en un ángulo  $\theta = 180^\circ$ . Como se esperaba, el *beam pattern* o diagrama de radiación, presenta un lóbulo principal con un máximo en la dirección de la fuente. En el lóbulo principal, hay dos ángulos  $\theta_1$  y  $\theta_2$  cuya magnitud con respecto a la posición del máximo ha decrecido  $-3$  dB. Estos dos ángulos, determinan la resolución ( $\zeta$ ) de la siguiente forma:

$$\zeta = \theta_2 - \theta_1. \quad (3.8)$$

Este parámetro es de interés ya que da una aproximación de la mínima diferencia angular necesaria para que dos fuentes incoherentes puedan ser distinguidas. Es decir, si la diferencia angular entre fuentes es igual o excede a la resolución, el *beam pattern* presentará dos máximos que corresponderán a las posiciones de las fuentes y de esta forma, podrán ser distinguidas. La Figura 3.6, muestra el *beam pattern* medido con dos fuentes incoherentes de iguales características. Una de las fuentes está localizada en un ángulo  $\theta_2 = \theta_1 + \zeta$ . En esta situación, puede observarse que la salida del *beamformer*, revela dos

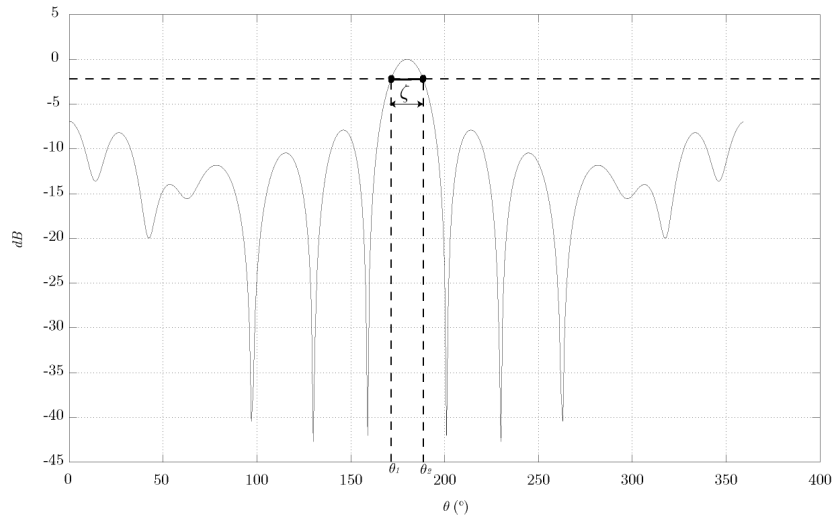


Figura 3.5: Cálculo de la resolución para un beampattern determinado.

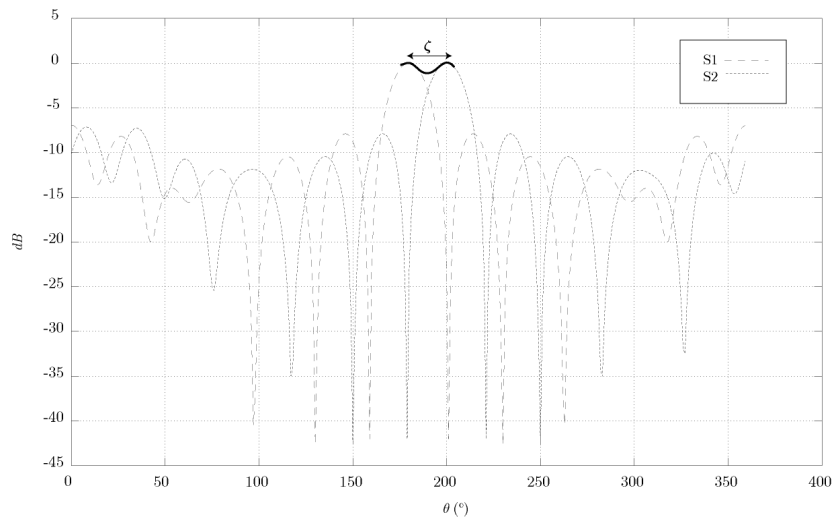


Figura 3.6: Salida del beamformer con dos fuentes activas, cuya distancia angular es igual a la resolución, pudiéndose distinguir cada fuente correctamente.

máximos independientes que coinciden con los de las fuentes. Por el contrario, cuando la diferencia angular entre las fuentes es más pequeña que la resolución, las fuentes no podrán ser identificadas. Este caso se muestra en la Figura 3.7. Como puede observarse, la salida del

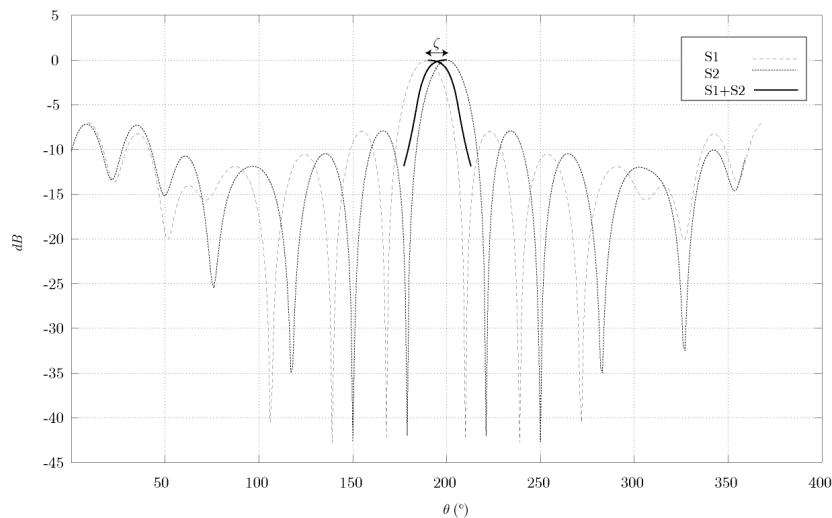


Figura 3.7: Salida del beamformer con dos fuentes activas, cuya distancia angular es menor a la resolución, no pudiéndose distinguir en este caso, las fuentes correctamente.

*beamformer* da un único máximo localizado entre los ángulos de las fuentes, y por tanto, sólo se identifica una fuente “fantasma” en vez de las dos fuentes reales.

- Nivel máximo de lóbulo lateral

Como se mencionó anteriormente, la salida del *beamformer* presenta generalmente lóbulos laterales alrededor del lóbulo principal. Este es un efecto no deseado ya que el *beam pattern* demuestra ser sensible no solamente en la dirección de enfoque, es decir, la dirección de la fuente “real”, sino también en direcciones de fuentes “fantasmas”. Por esta razón, es útil analizar la respuesta del *beamformer* por medio del parámetro NMLL, que viene dado por la diferencia entre el pico del lóbulo principal y el pico del lóbulo lateral más elevado [38]. El cálculo del NMLL se ilustra en la Figura 3.8.

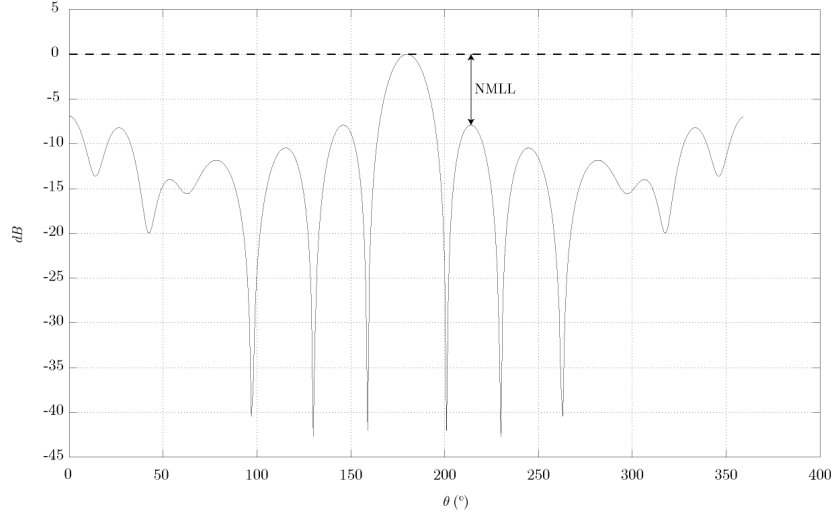


Figura 3.8: Cálculo del NMLL para una salida de beamformer determinada.

### 3.3. Geometría de array

La geometría de array en la que se basarán las técnicas de *beamforming* explicadas en las Secciones 3.5 y 3.6 viene descrita a continuación.

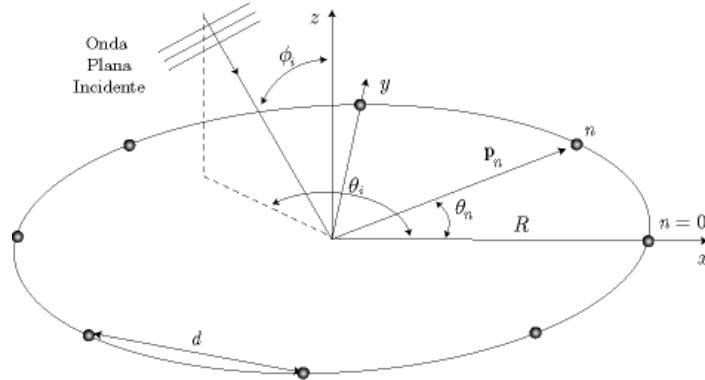
Se considera un array circular uniforme (UCA) que tiene  $N$  elementos equidistantes localizados en un array de radio  $R$  sobre el plano  $xy$ , tal y como muestra la Figura 3.9.

El centro del array está localizado en el origen de coordenadas y su elemento  $n$ -ésimo en coordenadas esféricas es  $(R, \theta_n, \pi/2)$ . El vector de localización de cada uno de los elementos en coordenadas cartesianas viene dado por

$$\mathbf{l}_n = [R \cos \theta_n, R \sin \theta_n, 0]^T, \quad n = 0, 1, \dots, N - 1, \quad (3.9)$$

donde  $(\cdot)^T$  indica transpuesta y el ángulo acimut de cada elemento es  $\theta_n = n \frac{2\pi}{N}$ . La distancia entre elementos puede ser calculada como

$$d = 2R \sin \frac{\pi}{N}. \quad (3.10)$$

Figura 3.9: Geometría de un UCA con  $N$  elementos.

La anterior distancia determina la frecuencia espacial de *aliasing*, la cual viene dada por

$$f_{al} = \frac{c}{2d}. \quad (3.11)$$

Asumiendo que las señales vienen del plano medio, el vector de apuntamiento o steering vector para un UCA depende del ángulo acimut de la siguiente forma:

$$\mathbf{a}(kR, \theta_i) = [e^{jkR \cos(\theta_i - \theta_0)}, \dots, e^{jkR \cos(\theta_i - \theta_{N-1})}]^T, \quad (3.12)$$

donde  $k = \frac{2\pi f}{c}$ ,  $c$  es la velocidad del sonido ( $c \approx 342$  m/s) y  $f$  es la frecuencia.

## 3.4. Aperturas circulares

### 3.4.1. Apertura circular continua

El anterior UCA puede ser considerado como una apertura circular muestreada espacialmente utilizando  $N$  sensores. Asumiendo que una onda plana incide desde la dirección  $(\theta_i, \phi_i = \pi/2)$ , la presión sonora en cualquier punto de una apertura circular continua puede ser escrita en coordenadas polares como

$$\bar{P}(kR, \theta) = P_0 e^{j\bar{k}_\theta R \cos(\theta - \theta_i)}, \quad (3.13)$$



donde  $P_0$  es la amplitud de la onda incidente y  $\bar{k}_\theta$  corresponde al número de onda en coordenadas esféricas. Hay que destacar que el término temporal  $e^{-j\omega t}$ , que aparece siempre al asumir una onda plana como es el caso, ha sido eliminado por simplificación. Aplicando un desarrollo matemático dado en [39], la presión incidente puede ser expresada como

$$\bar{P}(kR, \theta) = P_0 \sum_{\bar{k}_\theta=-\infty}^{\infty} j^{\bar{k}_\theta} J_{\bar{k}_\theta}(kR) e^{j\bar{k}_\theta(\theta-\theta_i)}, \quad (3.14)$$

donde  $J_{\bar{k}_\theta}$  es la función de Bessel de primera especie de orden  $\bar{k}_\theta$ . Destacar que la presión en la apertura puede ser considerada como series de Fourier, y de esta forma pueden ser representadas como

$$\bar{P}(kR, \theta) = \sum_{\bar{k}_\theta=-\infty}^{\infty} C_{\bar{k}_\theta} e^{j\bar{k}_\theta\theta}, \quad (3.15)$$

con los coeficientes de Fourier (o armónicos circulares) dados por

$$C_{\bar{k}_\theta}(kR, \theta_i) = P_0 j^{\bar{k}_\theta} J_{\bar{k}_\theta}(kR) e^{-j\bar{k}_\theta\theta_i}. \quad (3.16)$$

En la práctica, las aperturas continuas deben ser muestreadas por medio de un número finito de sensores. La siguiente subsección describe las consecuencias de este procedimiento muestreado.

### 3.4.2. Apertura circular muestreada

La discretización de aperturas circulares continuas por medio de un UCA con  $N$  micrófonos omnidireccionales dan lugar a los siguientes coeficientes de Fourier:

$$\tilde{C}_{\bar{k}_\theta}(kR) = \frac{1}{N} \sum_{n=0}^{N-1} \tilde{P}_n(kR) e^{-j\bar{k}_\theta\theta_n}, \quad (3.17)$$

donde  $\tilde{P}_n$  es la presión sonora medida en el micrófono  $n$  (localizado en el ángulo  $\theta_n$ ). Este procedimiento de muestreo implica un error en los coeficientes de Fourier [40]

$$\tilde{C}_{\bar{k}_\theta}(kR, \theta_i) = C_{\bar{k}_\theta}(kR, \theta_i) + \tilde{e}_{\bar{k}_\theta}(kR, \theta_i), \quad (3.18)$$

$$\tilde{e}_{\bar{k}_\theta}(kR, \theta_i) = P_0 \sum_{q=1}^{\infty} (j^g J_g(kR) e^{jg\theta_i} + j^h J_h(kR) e^{-jh\theta_i}), \quad (3.19)$$

donde  $g = Nq - \bar{k}_\theta$  y  $h = Nq + \bar{k}_\theta$ . Destacar que acorde a la Ecuación (3.15), serían necesarios para representar la presión sonora un número infinito de términos de Fourier. En la práctica el campo de onda incidente debe estar descompuesto en un orden máximo  $L$  de armónicos circulares, de tal forma que,  $N \geq 2L + 1$ . Como regla general se elige,  $L \approx kR$ , ya que el valor de la función particular de Bessel es pequeña cuando el orden  $\bar{k}_\theta > 0$  supera el argumento. La selección del número apropiado de términos de Fourier se discutirá de forma más amplia en la Sección 3.5.1.

### 3.5. Beamforming con armónicos circulares

Una vez descrito el modelo de señal y la geometría de array utilizados en todo el desarrollo matemático, a continuación se presenta la teoría del *beamforming* modal o *beamforming* con armónicos circulares junto con el procedimiento de selección de modos que se utilizará a lo largo de esta tesis para realizar tanto medidas en salas como para localizar fuentes sonoras.

El *beamforming* modal trata de combinar las diferentes componentes de armónicos circulares (o modos de fase) para formar un diagrama con propiedades espaciales apropiadas al caso en cuestión. De manera ideal, la respuesta del *beamformer* debería tener un máximo cuando éste apuntara en la dirección de la fuente  $\theta_i$  y debería ser cero en las demás direcciones (como se ha dicho en la Sección 3.1). Esta respuesta ideal puede ser representada como una función delta de la siguiente forma:

$$G_{\text{ideal}}(kR, \theta) = P_0 \delta(\theta - \theta_i). \quad (3.20)$$

En [41] se puede comprobar que esta respuesta ideal se alcanza añadiendo infinito número de modos obteniéndose un *beamformer* ideal cuya expresión matemática sería

$$G_{\text{ideal}}(kR, \theta) = \sum_{\bar{k}_\theta=-\infty}^{\infty} \frac{C_{\bar{k}_\theta}(kR, \theta_i)}{j^{\bar{k}_\theta} J_{\bar{k}_\theta}(kR)} e^{j\bar{k}_\theta\theta}. \quad (3.21)$$

Como se ha comentado en la anterior sección, cuando se utiliza un UCA implementado de forma real, el número de modos debe ser truncado a un orden máximo  $L$ . Además, los coeficientes modales corresponden a los de una apertura circular muestreada, dando como resultado la siguiente respuesta:

$$G_{\text{CHB}}(kR, \theta) = \sum_{\bar{k}_\theta=-L}^L \frac{\tilde{C}_{\bar{k}_\theta}(kR, \theta_i)}{j^{\bar{k}_\theta} J_{\bar{k}_\theta}(kR)} e^{j\bar{k}_\theta\theta}. \quad (3.22)$$

La salida del *beamformer* para una dirección de apuntamiento  $\theta_s$  puede expresarse como:

$$Y(kR, \theta_s) = \frac{1}{2L+1} \sum_{\bar{k}_\theta=-L}^L \tilde{C}_{\bar{k}_\theta}(kR) B_{\bar{k}_\theta}(kR) H_{\bar{k}_\theta}(\theta_s), \quad (3.23)$$

donde  $B_{\bar{k}_\theta}$  es un factor de ecualización dado por:

$$B_{\bar{k}_\theta}(kR) = j^{-\bar{k}_\theta} J_{\bar{k}_\theta}^{-1}(kR), \quad (3.24)$$

y  $H_{\bar{k}_\theta}$  es el factor de alineamiento de fase independiente de la frecuencia

$$H_{\bar{k}_\theta}(\theta_s) = e^{j\bar{k}_\theta\theta_s}. \quad (3.25)$$

El término de normalización  $\frac{1}{2L+1}$ , es igual al número de armónicos circulares en la suma, cuya función es la de mantener inalterable la amplitud de la onda plana incidente.

### 3.5.1. Selección de modos y regularización

Tal y como se ha analizado anteriormente, los filtros  $B_{\bar{k}_\theta}(kR)$  son los encargados de ecualizar las respuestas de los valores propios individuales, los cuáles dependen de la función de Bessel  $J_{\bar{k}_\theta}(kR)$ . La Figura 3.10 muestra la magnitud de los cuatro órdenes más bajos ( $\bar{k}_\theta = 0, \dots, 3$ ) de las funciones de primer tipo de Bessel para diferentes valores del argumento  $kR$ . Nótese que para un valor dado de  $kR$ , sólo hay algunos órdenes o modos con contribución significativa. Como ya se ha discutido, la regla general para seleccionar el máximo orden  $L$  es

$$L = \lceil kR \rceil, \quad (3.26)$$

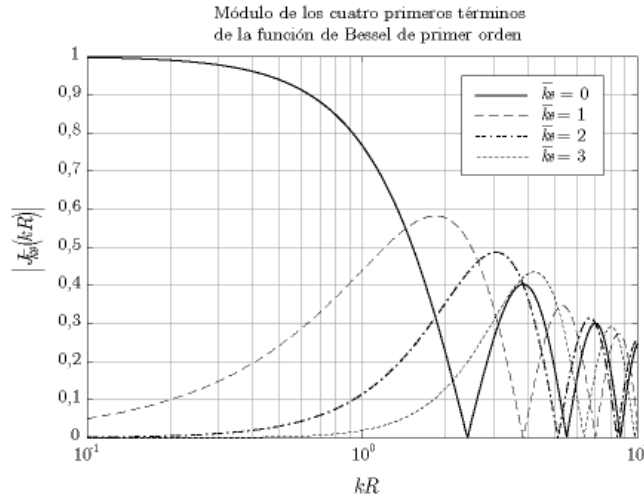


Figura 3.10: Magnitud de los cuatro órdenes más bajos de las funciones de Bessel de primer orden.

donde  $\lceil \cdot \rceil$  corresponde a la función límite ('ceiling'). Además de tener órdenes con magnitud baja, los diferentes modos contienen ceros periódicos y como resultado las señales que llevan componentes próximas a cero no pueden ser correctamente obtenidas. Para evitar este problema, la apertura circular puede ser montada en un cilindro como en [42], [43]. Sin embargo, se debe tener en cuenta que en este caso el array debe diseñarse para tener una relación altura-radio mayor a 2,8 y así aproximarse a un cilindro con longitud ideal e infinita [43]. Este requerimiento físico puede ser un problema en algunas aplicaciones prácticas (por ejemplo en el Array 3 de la Sección 6.3 se requeriría un cilindro mayor a 40,3 cm).

Para evitar la amplificación del ruido debido a grandes valores de ecualización, Parthy et al. [42] propuso el uso de los filtros regularizados Tikhonov, los cuales vienen dados por

$$B'_{\bar{k}_\theta}(kR) = \frac{w_{\bar{k}_\theta}^*(kR)}{\|w_{\bar{k}_\theta}(kR)\|^2 + \beta}, \quad (3.27)$$

donde  $w_{\bar{k}_\theta}(kR) = B_{\bar{k}_\theta}^{-1}(kR)$  y  $\beta$  es el coeficiente de regularización. El uso de la regularización, además de mejorar la ganancia de ruido

blanco, produce un *beam pattern* más robusto y suavizado. De hecho, la directividad y robustez, están relacionadas con el valor de  $\beta$  de forma que incrementando  $\beta$  se mejora la robustez y disminuye la directividad.

### 3.6. Beamforming Delay and Sum

Finalmente se va a abordar matemáticamente y de forma breve la técnica clásica de *beamforming* DAS, para poder comparar el comportamiento de ambos métodos (DAS y CHB) en la última sección de este capítulo.

El *beamforming* clásico, conocido como *Delay and Sum beamforming*, proporciona una máxima salida cuando el *beamformer* apunta en la dirección de la fuente y es cero en cualquier otra dirección, como ocurría en el caso de un *beamforming* CHB. Esta técnica toma las señales capturadas por los micrófonos del array circular, las retrasa y suma respectivamente. El valor de los retrasos viene determinado por la dirección de apuntamiento del array [29], [44]. La salida del *beamformer* DAS es máxima, cuando la dirección focalizada coincide con la posición de la fuente. Por simplificación, este *beamforming* suele tomarse en el dominio de la frecuencia [29], [37]. De esta forma, este método calcula los retrasos de fase entre los micrófonos y el apuntamiento del array en una determinada dirección. A continuación, los desplazamientos de fase opuestos son añadidos a las señales de los micrófonos para finalmente ser sumadas. Suponiendo que el *beamformer* está apuntando en la dirección de fuente, la salida del *beamformer* es

$$G_{DAS}(kR, \theta_S) = A \sum_{n=1}^N w_n(kR) \underbrace{\bar{P}(kR, \theta_n)}_{\text{medida}} \underbrace{\bar{P}^*(kR, \theta_n)}_{\text{teórica}} \quad (3.28)$$

donde  $A$  es un factor de escala,  $w_n$  es el coeficiente o peso del micrófono  $n$ ,  $\bar{P}(kR, \theta_n)$  es la presión capturada por cada micrófono debido a la onda plana creada por una fuente en  $\theta_S$  y  $P^*(kR, \theta_n)$  es igual en magnitud que  $\bar{P}(kR, \theta_n)$  pero con desplazamiento de fase. Debido a que el array está apuntando hacia la dirección de fuente,

$\bar{P}(kR, \theta_n)$  y  $\bar{P}^*(kR, \theta_n)$  son iguales en magnitud pero con oposición de fase, y por lo tanto, cuando todas las componentes son sumadas, la respuesta del *beamformer* es máxima. El punto clave es que el primer término de la Ecuación (3.28) es la auténtica señal que captura los micrófonos del array, mientras que el segundo término es la presión teórica que sería capturada teniendo la fuente ubicada en un ángulo  $\theta_S$ . De tal forma y utilizando el ángulo de la fuente en vez del ángulo de la onda incidente,  $\theta_S = \theta_i - \pi$ ,

$$\bar{P}(kR, \theta_S) = P_0 \sum_{\bar{k}_\theta = -L}^L -j^{\bar{k}_\theta} J_{\bar{k}_\theta}(kR) e^{j\bar{k}_\theta(\theta_n - \theta_S)} \quad (3.29)$$

donde en este caso el número de órdenes utilizados para el *beamformer* DAS son

$$L_{DAS} = \lceil kR \rceil + 1. \quad (3.30)$$

Introduciendo la Ecuación (3.29) en la Ecuación (3.28), la salida del *beamformer* queda como

$$G_{DAS}(kR, \theta_S) = AP_0 \sum_{n=1}^N \bar{P}(kR, \theta) \sum_{\bar{k}_\theta = -L}^L B_{\bar{k}_\theta}^*(kR) e^{-j\bar{k}_\theta(\theta_n - \theta_S)} \quad (3.31)$$

donde en este caso  $B_{\bar{k}_\theta}$  es un factor de equalización dado por:

$$B_{\bar{k}_\theta}(kR) = -j^{\bar{k}_\theta} J_{\bar{k}_\theta}(kR) \quad (3.32)$$

cuando el *beamformer* está apuntando a la dirección de la fuente. Nótese que los pesos  $w_n$  han sido puestos a 1 ya que todos los micrófonos tienen igual peso. En general y para una posición de fuente desconocida el *beamformer* debe buscar en todas las posiciones de la fuente posibles, es decir,  $0 \leq \theta \leq 2\pi$ . En este caso, la expresión general viene dada por

$$G_{DAS}(kR, \theta) = AP_0 \sum_{n=1}^N \bar{P}(kR, \theta) \sum_{\bar{k}_\theta = -L}^L B_{\bar{k}_\theta}^*(kR) e^{-j\bar{k}_\theta(\theta_n - \theta)} \quad (3.33)$$

Como se ha mencionado previamente, el máximo en la salida se alcanza cuando el *beamformer* apunta en la dirección de la fuente. A diferencia con el *beamforming* CHB, donde la salida del *beamformer* venía dada por la división de los coeficientes aproximados con los teóricos, Ecuación (3.22), el *beamformer* DAS presenta una multiplicación de estos términos. En el caso de aperturas como las utilizadas en esta tesis, las singularidades que puede presentar en el método CHB debido a los ceros de los coeficientes de Fourier, en este caso con un *beamformer* DAS no se producen y ese inconveniente está solventado.

### 3.7. Comparativa entre CHB y DAS

Las Figuras 3.11 y 3.12 muestran una comparación entre los *beam-patterns* en banda ancha resultado de aplicar el método convencional DAS [44], el método regularizado CHB y el método CHB sin regularización basado en la selección de modos descrito en [41] para un array compuesto por  $N = 13$  y  $R = 0,12$  m apuntando a una dirección acimut de  $\theta_s = \pi$ . El factor de regularización escogido ha sido  $\beta = 6,5 \cdot 10^{-4}$  como en [42]. Destacar, que el método CHB con regularización proporciona un *beam-pattern* más estrecho que el método DAS aunque el efecto de los ceros de Bessel pueden observarse de forma clara en la respuesta como líneas verticales distorsionadas alrededor de los valores de 1100 Hz, 1750 Hz, 2350 Hz, etc. de la Figura 3.11. Con respecto al método CHB sin regularizar, comentar que tiene un comportamiento similar al espectro de CHB regularizado, aunque es menos robusto y menos suavizado. Además, CHB regularizado mejora la ganancia de ruido blanco, facilitando esta regularización la implementación y aplicación de forma más sencilla dentro del marco de procesado de tiempo-frecuencia. En lo sucesivo, se considera el método CHB con regularización, que es el que hemos aplicado y desarrollado en esta tesis.

Además y con el fin de ilustrar la comparativa entre los métodos expuestos en este capítulo, se va a proceder a analizar dos arrays cuyas características vienen reflejadas en la Tabla 3.1 de nuevo mediante el método CHB y mediante el método DAS.

Las Figuras 3.13, 3.14 y 3.15 muestran en este caso los *beam-patterns*

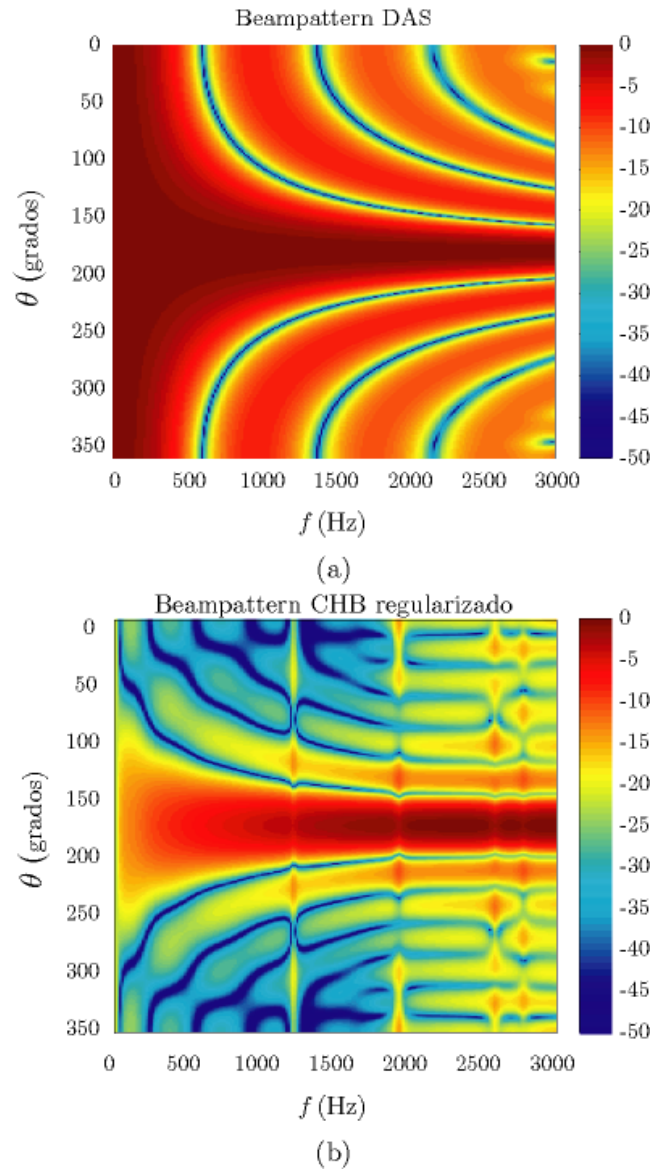
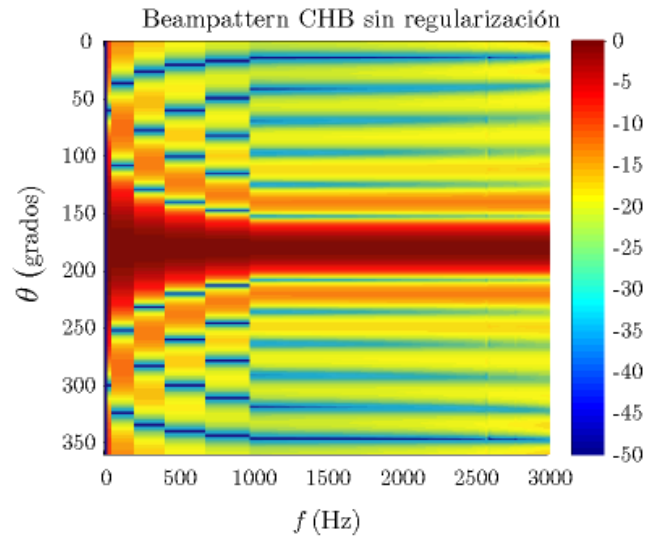
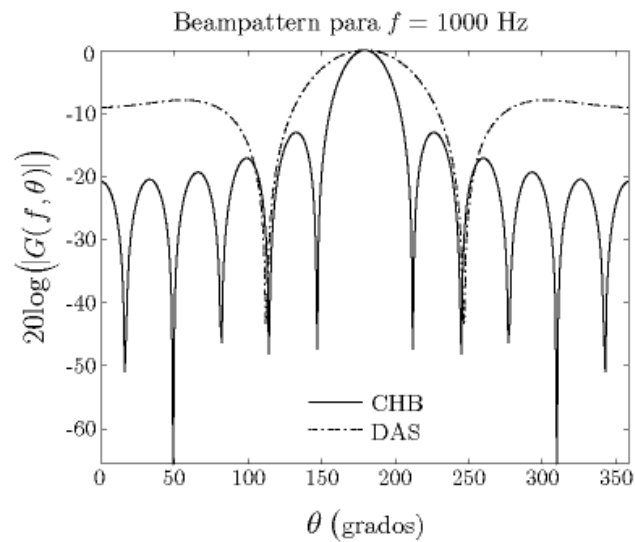


Figura 3.11: Beampatterns para banda ancha con apuntamiento en la dirección  $\theta_s = \pi$  utilizando un UCA con  $N = 13$  y  $R = 0,12$  m. (a) Delay and sum (DAS) (b) CHB beamforming regularizado.





(c)



(d)

Figura 3.12: Beampatterns para banda ancha con apuntamiento en la dirección  $\theta_s = \pi$  utilizando un UCA con  $N = 13$  y  $R = 0,12$  m. (c) CHB beamforming sin regularizar, (d) Sección transversal para  $f = 1000$  Hz.

correspondientes a esos dos métodos y para los arrays descritos. Como puede comprobarse la dependencia del radio del array, del número de modos seleccionados, del número de micrófonos y de la frecuencia máxima permitida viene reflejado en el resultado de los *beampatterns* obtenidos.

De nuevo y para todos los casos, el método CHB con regularización proporciona un *beampattern* más estrecho, aunque el efecto de los ceros de Bessel pueden observarse de forma clara en la respuesta como líneas verticales distorsionadas. Con respecto al método DAS de forma general, aunque éste muestra ser más robusto que CHB, éste último método obtiene mejor resolución angular y mejores características de lóbulo lateral que el clásico método de *beamforming*. Como se comprobará en el Capítulo 6, también se obtendrán mejores resultados con la técnica analizada en esta tesis (CHB) para localización de fuentes que con respecto a otras técnicas tratadas dentro del mismo ámbito de procesado.

	R[cm]	$L_{\max}$	N	$f_{\max}$ [Hz]
Array 1	5	3	7	3275
Array 2	10	5	11	2729
Array 3	20	10	21	2729

Tabla 3.1: Características de tres arrays simulados donde el número de micrófonos es el mínimo requerido.

### 3.8. Aplicación del método CHB en la obtención de ecogramas

Como se demostrará en el Capítulo 6, el método CHB puede aplicarse de forma eficaz para localizar fuentes sonoras. Sin embargo también permite la obtención de los correspondientes ecogramas de salas para poder analizarlos de forma similar a como se realizaba previamente al método de descomposición de ondas planas. De tal forma, este método proporciona una alternativa para el análisis del campo acústico con poco coste computacional y correctos resultados.

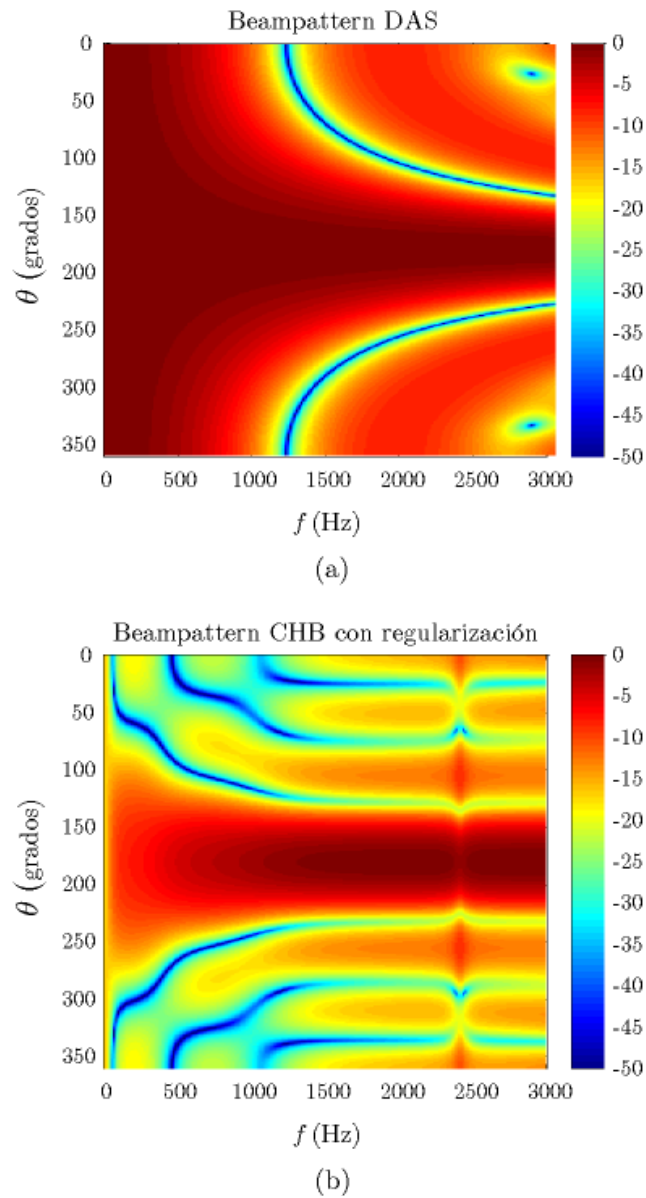


Figura 3.13: Salidas normalizadas de dos beamformers para las características del Array 1 de la Tabla 3.1 mediante el método DAS (a), y el método CHB (b).

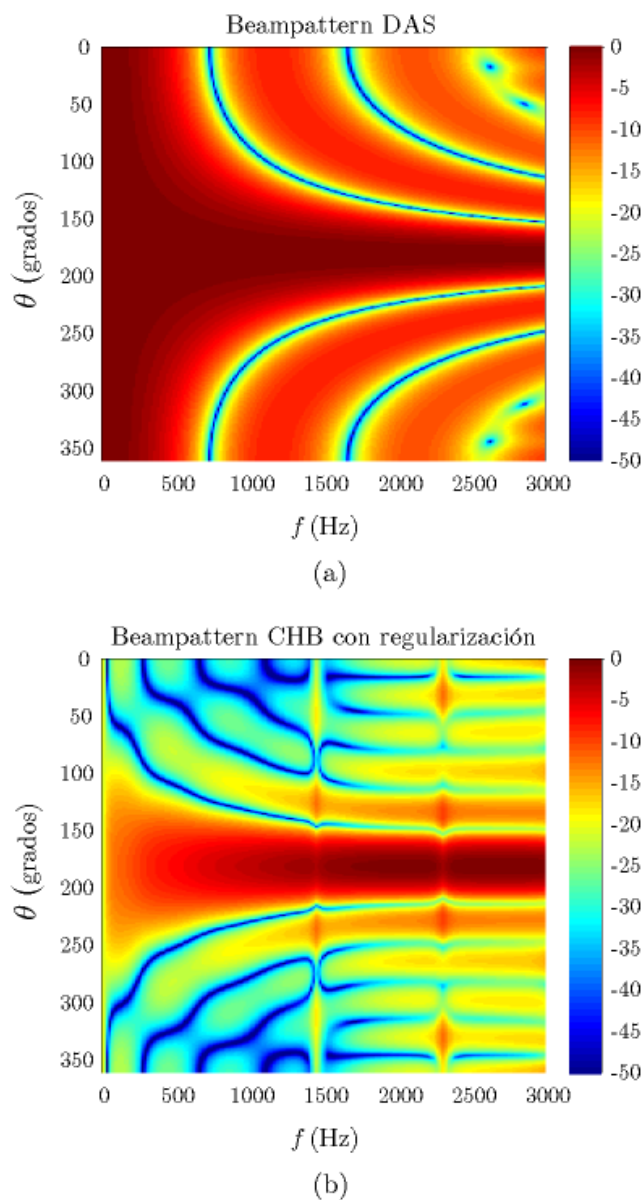


Figura 3.14: Salidas normalizadas de dos beamformers para las características del Array 2 de la Tabla 3.1 mediante el método DAS (a), y el método CHB (b).

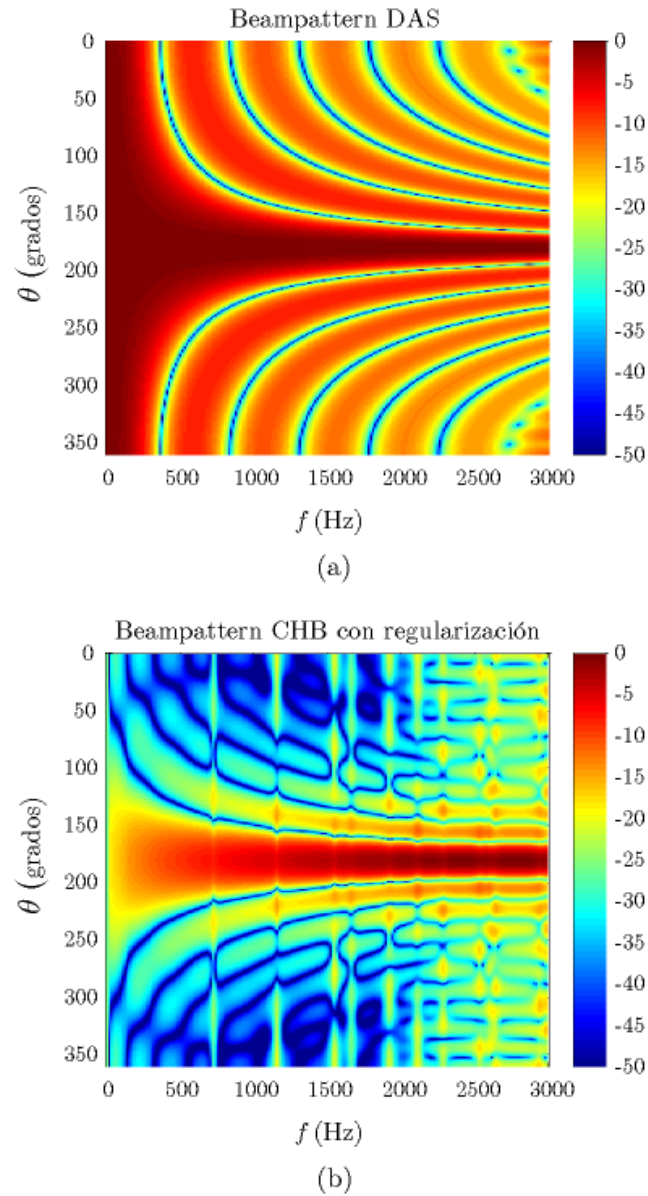


Figura 3.15: Salidas normalizadas de dos beamformers para las características del Array 3 de la Tabla 3.1 mediante el método DAS (a), y el método CHB (b).

A continuación se va a proceder a la obtención de ecogramas tanto de forma simulada como real con este método de *beamforming*, demostrando la buena operatividad en las medidas y corroborando su validez en un posterior análisis de salas.

### 3.8.1. Ecograma simulado

Para comparar el resultado de las huellas obtenidas con el método CHB, y antes de utilizarlo en medidas reales, es interesante llevar a cabo la simulación de una sala donde las condiciones están bajo control, y demostrar la validez del método. La simulación en cuestión, consiste en una sala vacía y rectangular, donde se ha ubicado un UCA en una determinada posición y la fuente se ha posicionado en otra. Las señales captadas por los micrófonos pueden ser obtenidas utilizando un software de modelado de acústica de salas basado en un modelo de imágenes. La selección del software acústico a utilizar para la generación de las RIRs, puede estar basado en la elección de una técnica basada en resolver numéricamente la ecuación de onda, o bien en el análisis de la acústica geométrica. Dependiendo del método, el modelado en la generación de respuestas impulsivas será diferente. En este caso, se ha elegido un programa de modelización basado en técnicas de imagen. ROOMSIM [45] se trata de una aplicación desarrollada por miembros del CIPIC para el entorno de desarrollo Matlab. Permite caracterizar un habitáculo con gran cantidad de parámetros (dimensiones, posición en la habitación del oyente, materiales y características de las paredes, impedancia característica del medio de propagación...). Además también incluye la posibilidad de auralizar una o más fuentes sonoras simultáneamente generando un archivo de audio en el que también se reflejan las características acústicas de la sala en cuestión [46].

Las medidas se realizaron para una sala vacía donde tanto array como fuente se posicionaron tal y como muestra la Figura 3.16. Las características que se introdujeron en el software ROOMSIM vienen dadas en la Tabla 3.2.

Después de calcular las señales por medio del software ROOMSIM y tras aplicar el método CHB, se obtuvo el ecograma mostrado en la Figura 3.17. Como se puede observar, el ecograma conseguido por el

Tabla 3.2: Características y parámetros introducidos en la simulación de una sala mediante el software ROOMSIM.

Características	Parámetros ROOMSIM
	11
Dimensiones (m)	30
	3
Absorción	Sí
Humedad	50 %
Temperatura	20 %
Frq. bandas (Hz)	125-4000
	Yeso
Superficie	Cemento
	Panel
Scattering	0
	0,75
SP (m)	2,6
	1,1
SO	0,0,0
ST	Omnidireccionales
RP (m)	10,6-5,5-1,1
RO	0-360°
RT	Cardioides

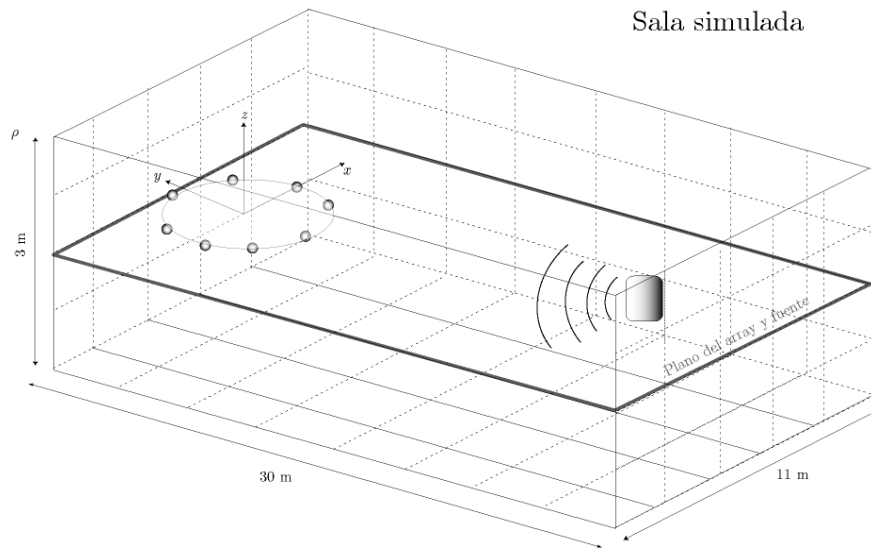


Figura 3.16: Sala simulada donde tanto la fuente como el array están localizados en el plano  $xy$ .

método CHB es similar al original obtenido directamente (a partir de las señales RIRs, y las correspondientes posiciones de los micrófonos del array). De hecho, las primeras y segundas reflexiones tienen casi los mismos valores en ambos ejes de tiempos y ángulos (ver Tablas 3.3 and 3.4).

Las pequeñas diferencias que se aprecian en la intensidad de potencia pueden ser debidas a los diferentes criterios de directividad que utiliza el método directo y el método CHB. Sin embargo esta sutil diferencia no tiene gran importancia en el caso de querer estimar la dirección de llegada de las diferentes reflexiones de la sala.

### 3.8.2. Ecograma real

En el anterior punto, se ha demostrado la validez del método CHB como alternativa a la obtención de ecogramas de salas a través de una simulación. En este apartado, se aplica de nuevo el método de *beamforming* modal pero para una sala real. En este caso se ha trabajado con una sala de usos múltiples localizada en el Pabellón Polideportivo



Tabla 3.3: Diferencias de tiempo del método CHB.

Posición	Ecog. original $t$ (ms)	CHB $t$ (ms)	Diferencia $t$
1	20	23	3
2	20	23	3
3	85	88	3
4	100	103	3
5	85	88	3
6	108	112	4

Tabla 3.4: Diferencias angulares del método CHB.

Posición	Ecog. original $\theta$ ( $^{\circ}$ )	CHB $\theta$ ( $^{\circ}$ )	Diferencia $\theta$ ( $^{\circ}$ )
1	-180	-175	5
2	180	177	3
3	-135	-135	0
4	0	3	3
5	130	133	3
6	0	3	3

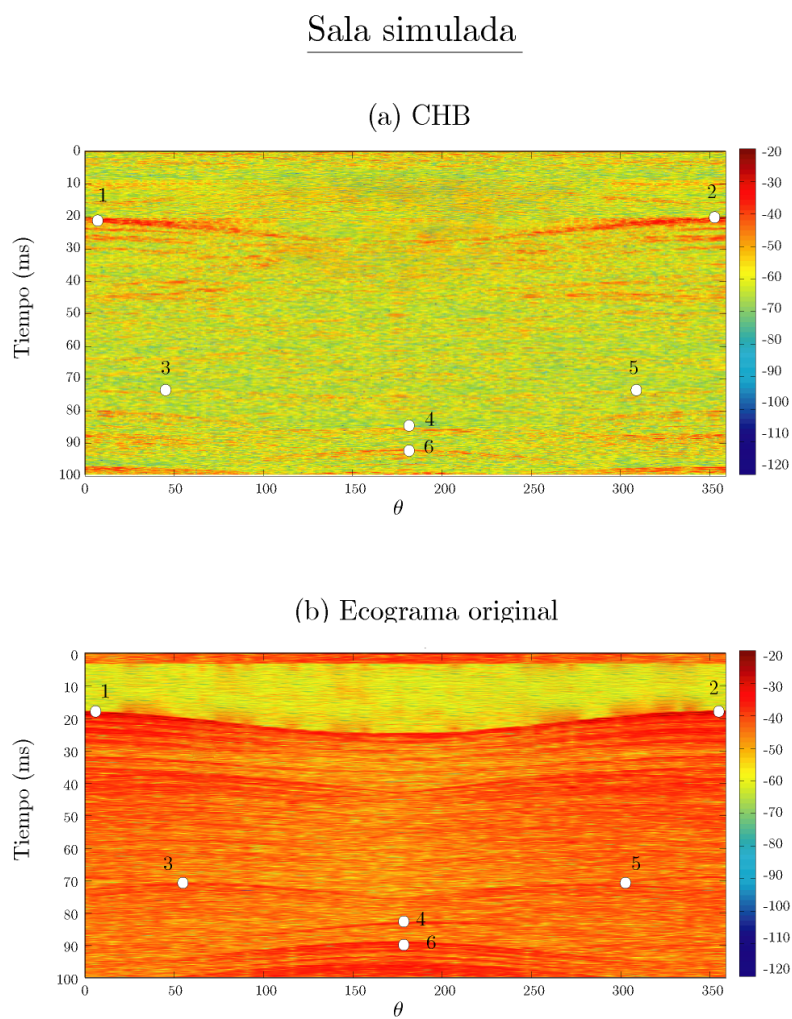


Figura 3.17: Ecogramas obtenidos por el método CHB y de forma directa.

Tabla 3.5: Posición de fuente y array en la sala real medida.

Distancias ( $m$ ) a	Posición de la fuente	Posición del array
Suelo	1,2	1,6
Pared derecha	11,1	5
Pared izquierda	0,3	5
Pared trasera	19,9	3
Pared frontal	3,5	19

de la Universidad de Castilla-La Mancha, en Cuenca, y cuya fotografía y plano vienen mostrados en la Figura 3.18.

La toma de las RIRs se llevaron a cabo en la sala vacía con una posición lateral de la fuente, y un array circular posicionado en el centro del área. En la Tabla 3.5, se detallan las posiciones de la fuente y del array.

El array circular, formado por dos micrófonos cardioides tomó 72 RIRs bajo una fuente que emitía sweeps de 2 segundos de duración. Una vez obtenidas esas señales se obtuvo el ecograma directo y el correspondiente después de aplicar el método CHB, Figura 3.19. A pesar de la compleja estructura del campo sonoro en la sala, debido a interferencias y fenómenos de difracción, muchos de los eventos reflejados y difractados pudieron ser fácilmente reconocibles en las representaciones gráficas finales.

Las reflexiones de primer orden desde el techo y las reflexiones de segundo orden a través de las paredes laterales y el techo deben ser las contribuciones no horizontales más importantes del campo sonoro. Tomando en cuenta las geometrías de las salas, muchos frentes de onda reflejados y difractados se pueden identificar como se indica en los ecogramas presentados. Además, pueden distinguirse más frentes de ondas hiperbólicos como las reflexiones con las paredes, el suelo y otros obstáculos que están marcadas con 2, 3, etc, que tienen sus vértices en tiempos de llegada mayores y con una forma más aplanada.

En la Figura 3.19 las formas coseonoidales representan los frentes de onda reflejados donde se puede comprobar que el origen no

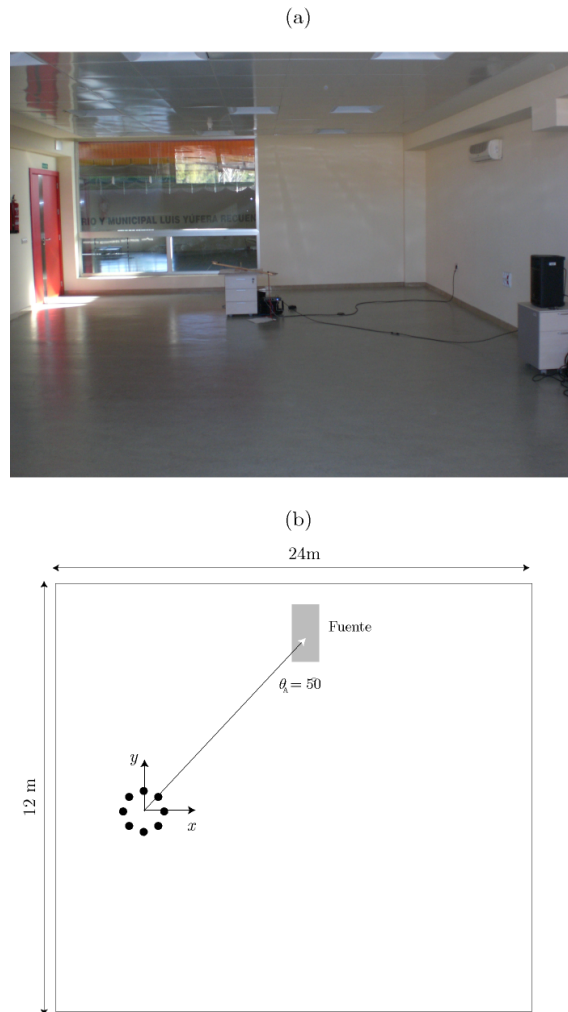


Figura 3.18: (a) Sala de usos múltiples utilizada para las medidas. (b) Ubicación de fuente y array en la misma.

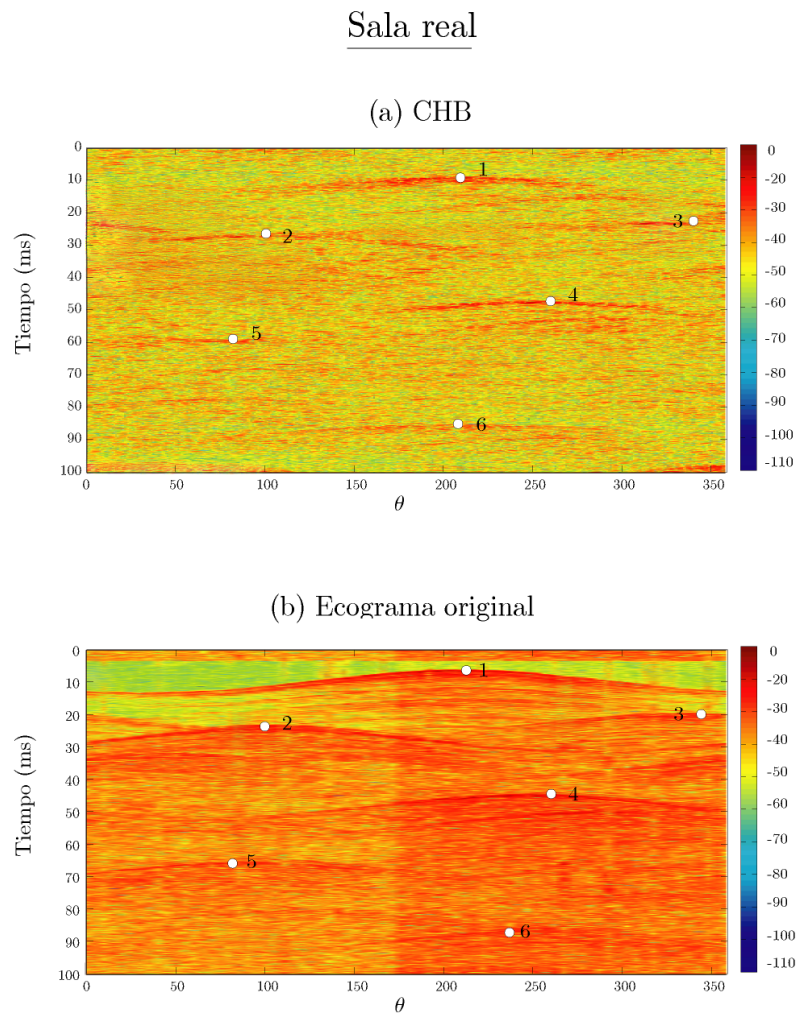


Figura 3.19: Ecogramas obtenidos de forma directa y mediante el método CHB.

Tabla 3.6: Diferencias de tiempo del método CHB.

Posición	Ecog. original $t$ (ms)	CHB $t$ (ms)	Diferencia $t$ (ms)
1	5,23	7,59	2,36
2	19,98	22,77	2,79
3	17,01	19,93	2,92
4	39,05	41,54	2,49
5	57,44	59,82	2,38
6	76,17	78,07	1,9

Tabla 3.7: Diferencias angulares del método CHB.

Posición	Ecog. original $\theta$ (°)	CHB $\theta$ (°)	Diferencia $\theta$ (°)
1	215	205	10
2	98	100	2
3	340	337	3
4	257	260	3
5	80	73	7
6	240	210	30

corresponde a la posición cero del array debido a las posiciones dentro de la sala de la fuente y del array. Comparando ambos ecogramas, se obtienen unos resultados con mínimas diferencias entre sí tanto en tiempo como en ángulo de las diferentes reflexiones marcadas (Tablas 3.6 y 3.7).

Aunque en este ejemplo no se ha llevado a cabo la extrapolación de distancias a las paredes, suelo y techo de las diferentes reflexiones (lo cuál será analizado con profundidad en el Capítulo 5), los ecogramas obtenidos demuestran que el método CHB obtiene satisfactoriamente unos ecogramas de salas reales que permiten el análisis y la obtención de información acústica adicional importante en la acústica de salas.

Con este capítulo se finaliza la base teórica de *beamforming* a partir

de la cuál se analizará en los próximos Capítulos 5 y 6 las medidas experimentales y localización de fuentes en salas reales.

# Capítulo 4

## Método de localización de ondas

---

<b>4.1. Detección de reflexiones tempranas . . . . .</b>	<b>104</b>
<b>4.2. Algoritmo de detección de ondas en ecogramas . . .</b>	<b>105</b>
4.2.1. Umbralización . . . . .	105
4.2.2. Correlación cruzada . . . . .	109
4.2.3. Selección de regiones . . . . .	110
<b>4.3. Algoritmo de detección de ondas en representaciones     PWD . . . . .</b>	<b>112</b>
4.3.1. Umbralización . . . . .	112
4.3.2. Selección de regiones . . . . .	115

---

*En el Capítulo 2 se describieron las principales representaciones gráficas del campo sonoro obtenidas a partir de las respuestas al impulso capturadas con un array circular. A partir de éstas, es necesario identificar de forma automática las ondas correspondientes a los eventos producidos en una sala. Con este fin, en este capítulo se ha desarrollado un método basado en morfología de imagen que permite discriminar los eventos de onda más representativos a partir de los ecogramas y de las descomposiciones de onda plana.*



## 4.1. Detección de reflexiones tempranas

Las propiedades de las reflexiones tempranas o primeras reflexiones juegan un papel importante en la acústica de una sala. El conocimiento de la localización de estas reflexiones es de interés por ejemplo en aplicaciones de auralización. Su detección en medidas de RIRs ha sido ampliamente estudiada en la literatura por diversos autores y con diversos enfoques. Usher [47] propuso un análisis basado en la energía para determinar el tiempo de comienzo de las primeras reflexiones en las RIRs sin tener en cuenta la información espacial. También se han considerado en esta línea las RIRs binaurales. En este contexto, Vesa y Lokki [48] desarrollaron un método basado en el análisis de las señales del canal auditivo por medio de una transformada *Wavelet* continua. Por otra parte, también se han utilizado arrays de micrófonos para el análisis de reflexiones tempranas sin desarrollar métodos automáticos para detectar eventos de onda [49]. En este contexto, Tervo et al. [50] propusieron un método automático de detección basado en un array de micrófonos esféricos estudiando la dirección de llegada y la difusividad del campo sonoro.

Sin embargo, existen pocos trabajos en la literatura que describan el uso de herramientas de procesamiento de imagen para el análisis del campo acústico. Hulsebos [14] propuso el uso de la transformada Random para caracterizar la descomposición en onda plana del campo acústico captado con un UCA, dando lugar a unos resultados no muy precisos.

En la Sección 2.8 dentro del Capítulo 2, se vieron algunas de las representaciones más útiles del campo sonoro obtenidas a partir de la captura de RIRs mediante arrays circulares. Estas representaciones pueden analizarse de forma adicional para identificar automáticamente los diferentes eventos de onda que ocurren dentro de una sala medida. En las secciones de este capítulo, se presentará una metodología de procesamiento basado en análisis morfológico de imágenes para la detección automática de ondas. Este algoritmo se aplicará a representaciones espacio-tiempo y ecogramas y mediante ejemplos simulados se comprobará su buena operatividad y precisión.

## 4.2. Algoritmo de detección de ondas en ecogramas

En la Sección 2.6, se comprobó cómo las ondas planas capturadas por medio de un UCA tenían una forma de coseno en su representación espacio-tiempo. Aunque estas formas coseno que representan las reflexiones producidas en la sala de medida pueden discernirse claramente en medidas ideales, en las RIRs captadas en salas reales son difíciles de separar e identificar ya que en sus correspondientes ecogramas aparecen numerosas imperfecciones derivadas del ruido de medida, limitaciones espaciales, canales u otros fenómenos acústicos. De esta forma, una vez obtenidas las RIRs, éstas deberían ser de nuevo procesadas de alguna forma para eliminar lo máximo posible estos efectos y realzar la presencia de las reflexiones tempranas. En esta tesis se ha propuesto un método de procesamiento basado en imagen que consta de tres pasos diferenciados: umbralización, correlación y selección de regiones.

Aunque en el próximo capítulo este algoritmo se aplicará directamente sobre las capturas tomadas en las diferentes salas reales, en este apartado se ilustrará cada uno de esos tres pasos por medio de un ejemplo simulado. Se ha tomado en cuenta una sala rectangular con dimensiones  $15 \times 10 \times 3$  m y un coeficiente de reverberación de 0,5 que caracteriza un volumen medio de los que se medirán en las salas del siguiente capítulo. El UCA aplicado en esta simulación y cuya geometría se adapta al de la Figura 2.9, tiene  $N = 101$  micrófonos y un radio de  $R = 0,7$  m. La frecuencia de muestreo para este ejemplo se ha tomado de  $f_s = 16000$  Hz.

### 4.2.1. Umbralización

En primer lugar se obtiene el ecograma obtenido por medio de la captura de las respuestas al impulso  $\mathbf{H}$  para la sala simulada y que viene mostrado en la Figura 4.1. A partir de el anterior ecograma y como primer paso en este algoritmo de detección de ondas se va a aplicar un suavizado al valor absoluto de ese ecograma utilizando un filtro paso bajo [51]. Esta operación de filtrado se

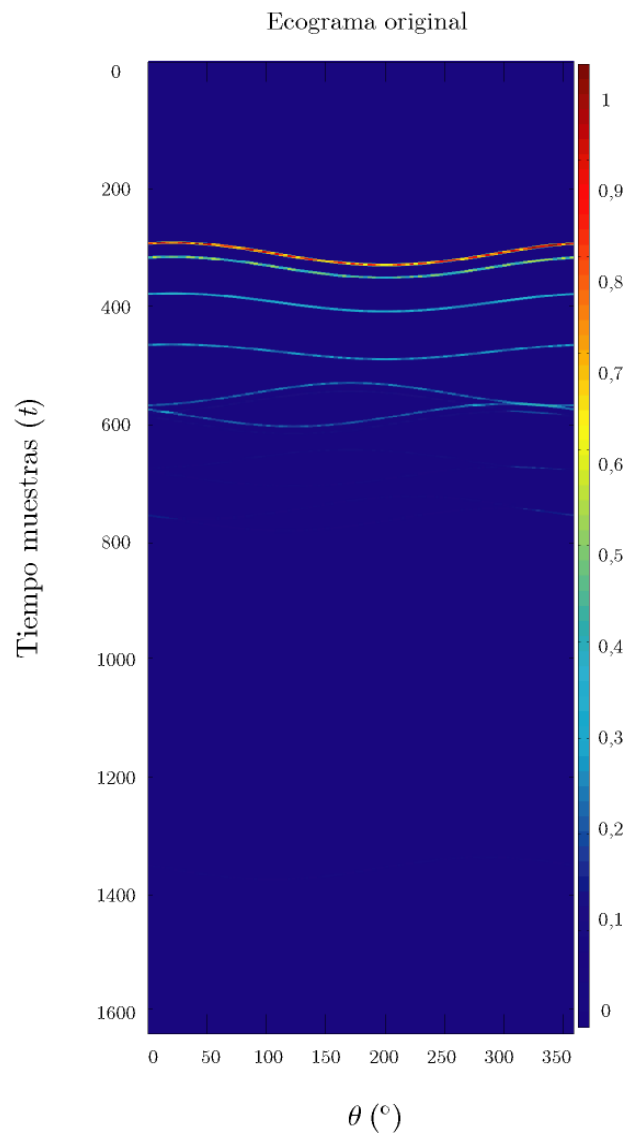


Figura 4.1: Ecograma obtenido para el ejemplo simulado y a partir del cuál se realizará el algoritmo de detección de onda.

realiza en torno al eje vertical, lo cual es equivalente a realizar un filtrado a cada RIR  $|h_n(t)|$  de forma individual e independiente. El objetivo de este filtro es atenuar las altas frecuencias de las RIRs. Debido a que estas componentes normalmente degradan el resultado del proceso de binarización, se suele implementar un filtrado paso bajo antes de la operación de umbralización [52]. Sin embargo, su longitud debe ser lo suficientemente pequeña para evitar efectos de enmascaramiento entre las reflexiones próximas y cercanas entre sí. De manera experimental, se ha determinado que una longitud apropiada para este filtro es de 0,5 ms. Una vez realizado este filtrado que enfatiza las reflexiones tempranas, la umbralización de imagen se aplica sobre la anterior matriz resultante  $\mathbf{H}'(t, n)$  para seleccionar los puntos significativos correspondientes a las reflexiones más importantes en forma de coseno (ondas planas). El umbral  $\mathbf{T}(t, n)$  es elegido como

$$T(t, n) = \begin{cases} 1 & \text{si } H'(t, n) \geq \gamma \\ 0 & \text{resto} \end{cases} \quad \forall t, n. \quad (4.1)$$

El umbral  $\gamma$  es elegido como [51]

$$\gamma = 0,1 \frac{1}{N} \sum_{n=1}^N \max(\mathbf{h}'_n(t)), \quad (4.2)$$

donde  $\mathbf{h}'_n(t)$  es la columna  $n$ -ésima de  $\mathbf{H}'(t, n)$ , es decir, la versión filtrada de  $\mathbf{h}_n(t)$ . Es importante ver que el umbral es una fracción de la media de los valores máximos encontrados en las respuestas medidas, correspondientes generalmente a las trayectorias directas del sonido.

Este proceso de umbralizado devuelve una imagen binaria con los puntos elegidos correspondientes a las formas coseno detectadas. Además, esta umbralización permitirá enfatizar aquellas reflexiones con poca energía de forma que podrán ser detectadas y analizadas en los siguientes pasos dentro de este algoritmo. La imagen binarizada a partir del ecograma original para el ejemplo simulado se muestra en la Figura 4.2.

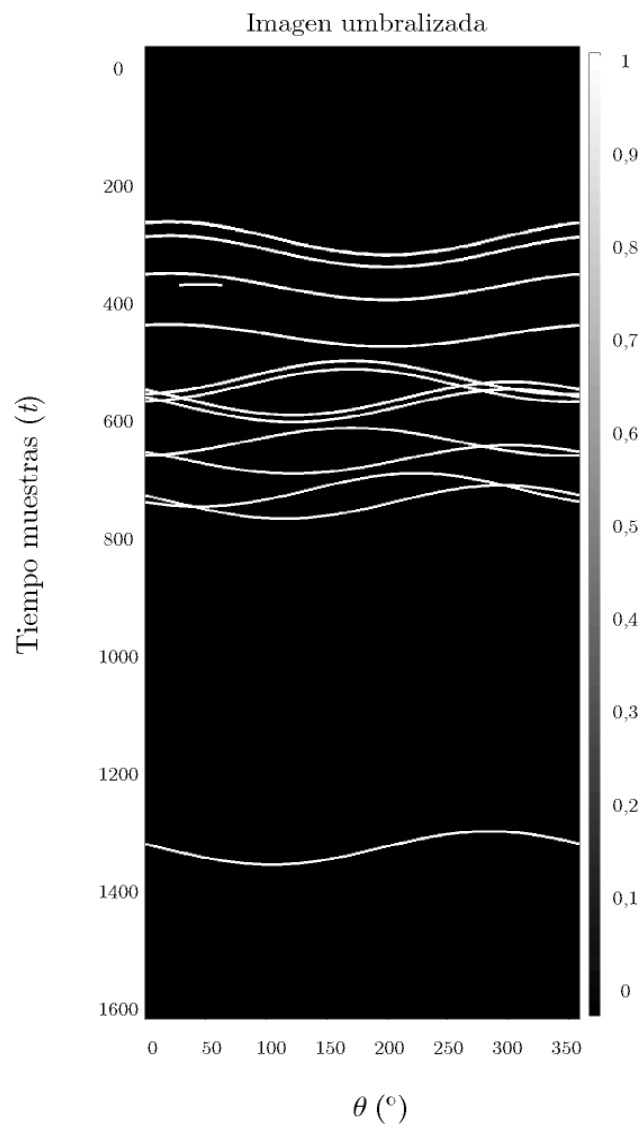


Figura 4.2: Imagen umbralizada después de haber filtrado el ecograma inicial por un filtro Gaussiano.

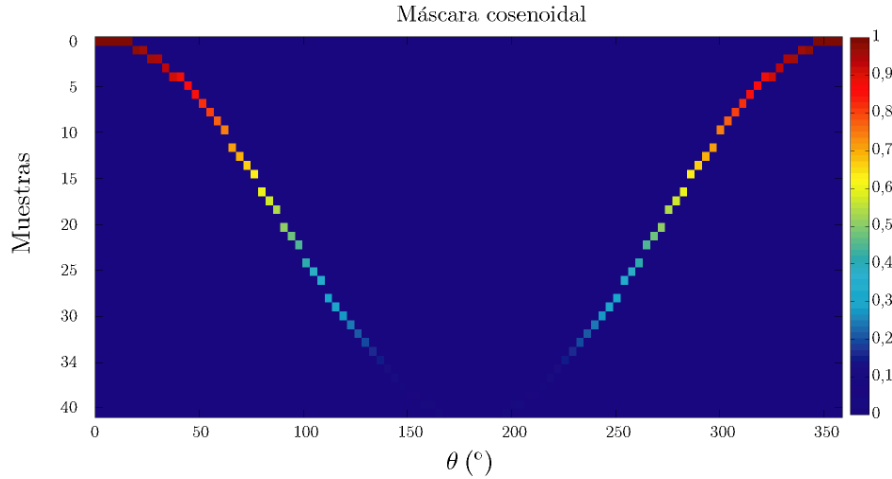


Figura 4.3: Máscara particularizada a los parámetros del ejemplo simulado y que se utilizará para realizar la correlación cruzada.

#### 4.2.2. Correlación cruzada

La detección de las curvas con forma de coseno, se basa en correlar circularmente una máscara coseno, como se muestra en la Figura 4.3, con la imagen binaria umbralizada resultante del anterior paso [51]. Esta máscara de correlación debe de coincidir con las propiedades de las curvas que se encuentran en el ecograma o representación espacio-tiempo. Para este fin, se hace uso de los parámetros del array para construir la forma de la máscara coseno que se adaptará a los parámetros de procesado. Si se considera una incidencia horizontal (es decir,  $\phi_i = 0^\circ$ ) las curvas encontradas en el ecograma vienen dadas por:

$$t_n(\tau_i, \theta_i) = \left[ f_s \tau_i - f_s \frac{R}{c} \cos(\theta_n - \theta_i) \right], \quad (4.3)$$

donde  $t_n(\tau_i, \theta_i)$  representa el instante de tiempo (en muestras) en el que una onda plana con DOA  $\theta_i$  y tiempo de llegada  $\tau_i$  alcanza a cada micrófono  $n$ . Además, y puesto que se utilizan micrófonos cardioides tanto en la simulación como en las medidas reales llevadas a cabo en el próximo capítulo, la amplitud captada por cada micrófono es

$$A(n, \theta_i) = 0,5 + 0,5 \cos(\theta_n - \theta_i). \quad (4.4)$$

De esta forma, la máscara de correlación  $\mathbf{M}$  tendrá unas dimensiones  $[\lceil f_s \frac{2R}{c} \rceil, N]$  y se completará de la siguiente forma [51]:

$$M(t, n) = \begin{cases} A_n(0) & \text{si } t = \lceil f_s \frac{R}{c} (1 - \cos \theta_i) \rceil + 1 \\ 0 & \text{resto} \end{cases} \quad \forall t, n. \quad (4.5)$$

A continuación, la anterior máscara es desplazada circularmente y correlada de forma cruzada con la imagen binaria obtenida de la anterior umbralización en dirección vertical para encontrar los valores máximos en cuyos tiempos y direcciones de llegada corresponden las principales reflexiones. El valor del pixel  $(u, v)$  vendrá dado por

$$C(u, v) = \sum_{t, n} T(t, n) \cdot M^{(v-1)}(t - u - 1, n), \quad (4.6)$$

donde  $M^{(v-1)}$  es la plantilla circular desplazada en el eje horizontal:

$$M^{(v-1)}(t, n) = M(t, (n - v + 1) \bmod N). \quad (4.7)$$

El resultado de esta operación es una nueva matriz  $\mathbf{C}$ , la cuál para el ejemplo simulado se muestra en la Figura 4.4, donde las zonas brillantes corresponden a los valores donde se ha obtenido unos mayores valores en la correlación.

### 4.2.3. Selección de regiones

La imagen resultante de la correlación cruzada anterior  $\mathbf{C}$  es de nuevo umbralizada para obtener aquellas regiones con una puntuación de máxima similitud. El valor umbral utilizado en este caso es seleccionado como [51]:

$$T_c(t, n) = \begin{cases} 1 & \text{si } C(t, n) \geq \gamma_c \\ 0 & \text{resto} \end{cases} \quad \forall t, n. \quad (4.8)$$

De forma análoga a la Ecuación (4.5), el valor umbral utilizado para la selección es [51]:

$$\gamma_c = 0,8 \frac{1}{N} \sum_{n=1}^N \text{máx}(\mathbf{C}_n), \quad (4.9)$$

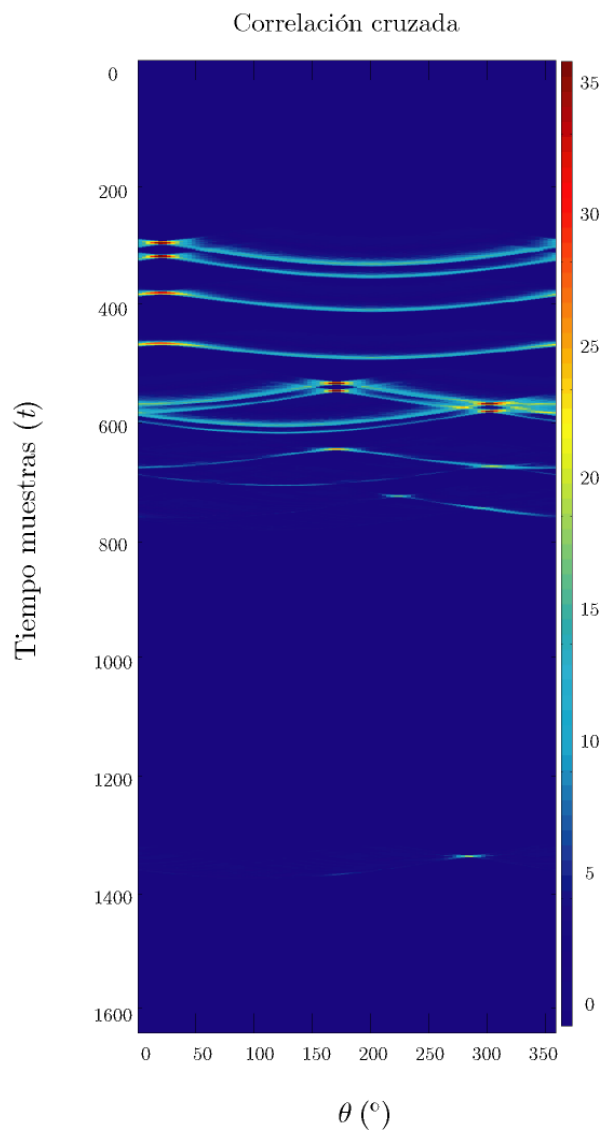


Figura 4.4: Resultado obtenido después de haber correlado la imagen umbralizada con una máscara cosenoidal.



donde  $C_n$  es la  $n$ -ésima columna de  $C$ . El etiquetado de regiones se realiza con ocho píxeles vecinos y se seleccionan aquellas regiones con un considerable número de componentes. De esta forma, se genera una región siempre y cuando se cumpla la existencia de esta premisa. La selección se lleva a cabo para eliminar regiones de pequeño tamaño. El número mínimo de componentes deseadas es la mitad de la media del número de componentes en todas las regiones. Para el ejemplo llevado a cabo, las regiones que permanecen después de esta selección vienen mostradas en la Figura 4.5.

Para finalizar, los picos correspondientes a las reflexiones de ondas planas mostradas en el ecograma inicial, son aquellos que corresponden a los valores máximos en las regiones seleccionadas finales. Estos picos se muestran en la Figura 4.6.

### 4.3. Algoritmo de detección de ondas en representaciones PWD

De manera similar al anterior método propuesto para la detección de ondas en ecogramas, es posible procesar las representaciones PWD tiempo-espacio como imágenes para seleccionar los “picos” correspondientes a los eventos de ondas planas. En las representaciones PWD de tiempo-espacio, se aplicará una versión simplificada del método de imagen para detectar ondas. En este caso la correlación cruzada con una máscara cosenoidal no es necesaria debido a que ahora aparecen cúmulos de energía en determinadas posiciones y no formas cosenoidales. De esta forma, sólo se aplicará una umbralización en amplitud, y una selección de regiones sobre la representación PWD (pasos 1 y 3 de la Sección 4.2).

#### 4.3.1. Umbralización

En este caso, la imagen  $s^t$  es binarizada utilizando el criterio de umbralización dado en la Ecuación (4.1) donde el umbral puede determinarse según la Ecuación (4.2), pero utilizando las columnas PWD  $\varrho_n^t$ .

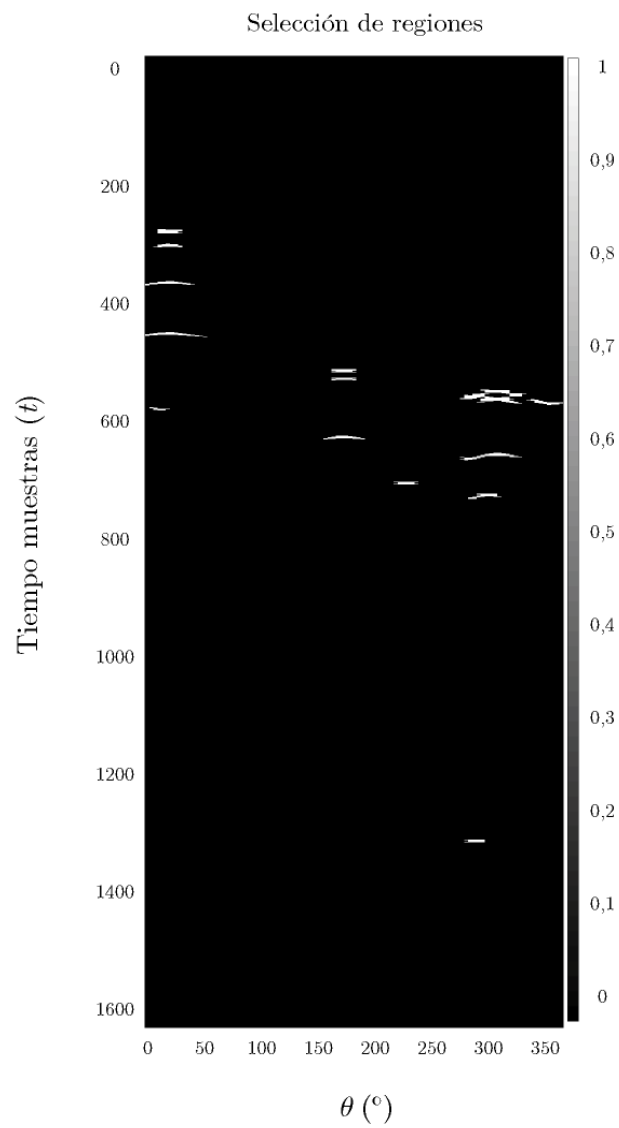


Figura 4.5: Representación de las regiones que se han seleccionado en el ejemplo simulado.

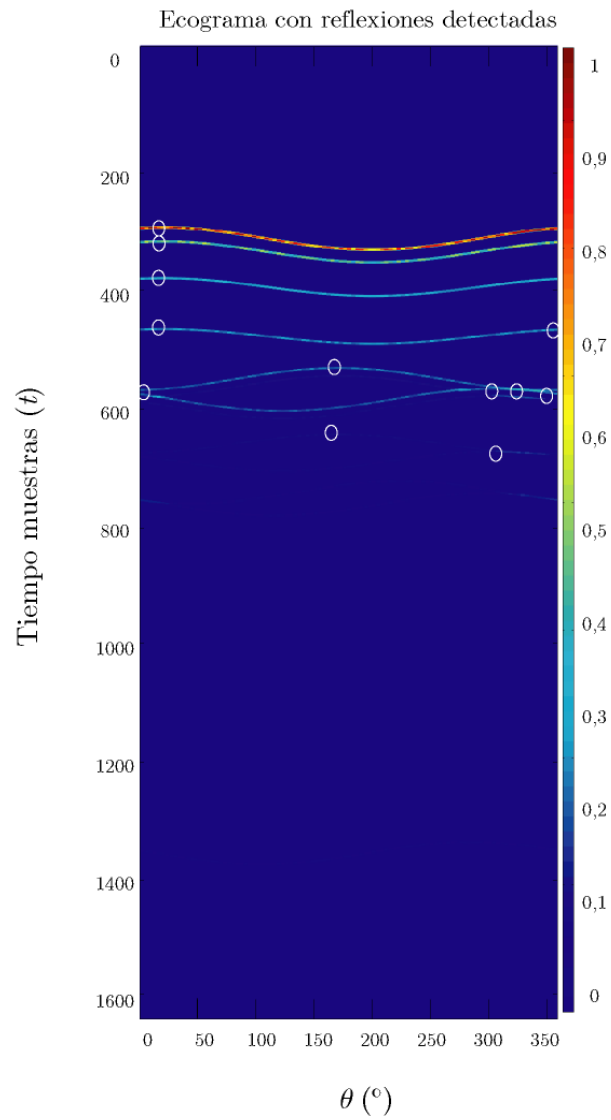


Figura 4.6: Detección de picos de aquellas reflexiones más representativas después de haber aplicado el algoritmo de detección de ondas.

### **4.3.2. Selección de regiones**

El etiquetado de regiones y selección de máximos, se realiza sobre la imagen binaria resultante después de haber aplicado el anterior paso de umbralización y de forma análoga a lo indicado en la Sección 4.2.3.



# Capítulo 5

## Aplicación de la descomposición de ondas planas para el análisis de salas

---

<b>5.1. Salas analizadas</b>	<b>119</b>
5.1.1. Sala de Reuniones	119
5.1.2. Paraninfo	119
5.1.3. Auditorio	121
<b>5.2. Método de medida</b>	<b>121</b>
<b>5.3. Análisis de resultados</b>	<b>125</b>
5.3.1. Descripción espacial del campo sonoro	125
5.3.2. Detección de ondas	133
<b>5.4. Conclusiones</b>	<b>136</b>

---

*En este capítulo se presenta la metodología que se llevó a cabo para medir la acústica de diferentes salas reales por medio del método de análisis de descomposición por ondas planas. Las medidas fueron realizadas en tres salas de diferentes tamaños y formas con el fin de validar la efectividad y precisión del método en diferentes escenarios. Además, el algoritmo basado en morfología de imagen explicado en el*

*capítulo anterior se ha aplicado para la detección de las principales reflexiones obtenidas en ecogramas y en representaciones de descomposición de onda. Como se verá en este capítulo, los resultados obtenidos por los métodos implementados fueron consistentes en todas las medidas realizadas. Este capítulo está organizado en dos partes. En primer lugar se describen las salas reales analizadas, y posteriormente se presenta la metodología llevada a cabo para la medida de las mismas.*

## 5.1. Salas analizadas

Con el fin de comprobar la efectividad del método de ondas planas en salas reales, se llevaron a cabo diferentes medidas en tres salas con diferentes propiedades acústicas, geometrías y necesidades. En concreto se analizaron un Paraninfo, un Auditorio y una Sala de Reuniones localizadas en el Campus de Cuenca, perteneciente a la Universidad de Castilla-La Mancha. La variabilidad de las salas en cuanto a forma y tamaño fue escogida precisamente para validar la correcta precisión del método bajo diferentes escenarios. En todas ellas, se realizaron diferentes medidas donde la ubicación en la sala de la fuente como del array de micrófonos variaba. El objetivo de estos cambios de localización era poder contrastar el método de descomposición de ondas planas y poder analizar objetivamente los resultados obtenidos. Como se comprobará en este capítulo, los ecogramas resultantes fueron coherentes para todas las configuraciones que se llevaron a cabo.

### 5.1.1. Sala de Reuniones

Esta primera sala correspondía a una sala de reuniones localizada en la Facultad de Bellas Artes. Las paredes estaban cubiertas por diferentes materiales: las dos que aparecen en la Figura 5.1 estaban cubiertas de paneles de madera y cortinas, y las otras dos estaban lucidas con cemento y pintadas. Sus medidas eran 11,12 m de largo, 7,6 m de ancho y 3,2 m de altura. Su tiempo aproximado de reverberación para la sala vacía era de 0,65 segundos. En la figura antes reseñada se muestra una fotografía del interior así como un plano donde se indica las posiciones donde se ubicaron tanto la fuente como el array de micrófonos.

### 5.1.2. Paraninfo

La segunda sala bajo estudio era un paraninfo recientemente construido en la Facultad de Magisterio del Campus de Cuenca, el cuál servía como salón principal de conferencias y aunque su principal



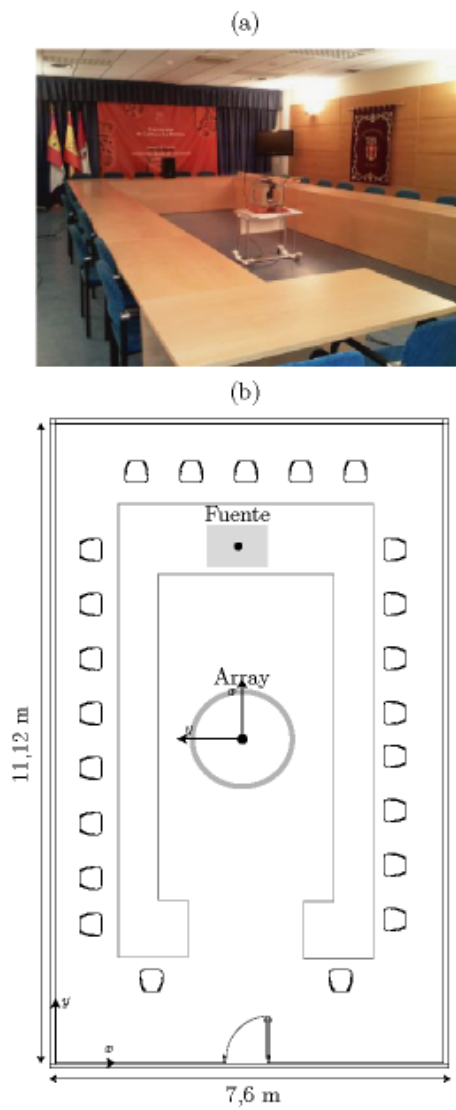


Figura 5.1: (a) Fotografía de la Sala de Reuniones, (b) plano y ubicación respectivas de la fuente y del array de micrófonos en la sala.

funcionalidad es la de auditorio, también es utilizado como sala de conciertos. A diferencia de las otras dos salas, su planta tiene forma trapezoidal, y su volumen aproximado de  $10000 \text{ m}^3$  es sustancialmente mayor con un tiempo de reverberación de 1,41 segundos. La Figura 5.2 muestra el interior del paraninfo así como el plano con las posiciones de fuente y array. En este caso, se colocó la fuente en el escenario mientras que el array de micrófonos se posicionó en la parte central del pasillo que dividía ambos lados de los asientos.

### 5.1.3. Auditorio

La tercera sala era un auditorio sensiblemente más pequeño que el paraninfo descrito anteriormente. Estaba localizado al igual que la sala de reuniones, en la Facultad de Bellas Artes. Su forma era aproximadamente cuadrada y sus dimensiones eran 22 m de largo, 14,4 m de ancho, 5 m de alto y tenía un volumen de unos  $1584 \text{ m}^3$ . Su tiempo de reverberación era de 1,23 segundos y a diferencia de la primera sala, sus paredes estaban hechas con madera y mármol. De nuevo en la Figura 5.3 se muestra la sala y el plano de ubicación de fuente y array. La colocación de la fuente de nuevo fue en la parte central del escenario pero el array se situó en un lateral de la zona derecha de las butacas o asientos con el propósito de analizar también cómo influía en esta nueva colocación.

## 5.2. Método de medida

Con el objetivo de realizar las medidas de las anteriores salas en iguales condiciones, éstas fueron llevadas a cabo con el mismo altavoz en el centro del escenario y en diferentes posiciones dentro del área de audiencia. Las salas estaban vacías y no sufrieron alteración alguna mientras se realizaron las medidas experimentales. El array de micrófonos se situó en diferentes posiciones dentro de esas salas, de acuerdo a las recomendaciones de [53]. La Tabla 5.1 muestra las ubicaciones tanto del array como de la fuente con respecto al eje de coordenadas que fue ubicado para el análisis de todas las salas en la esquina inferior izquierda de las mismas.

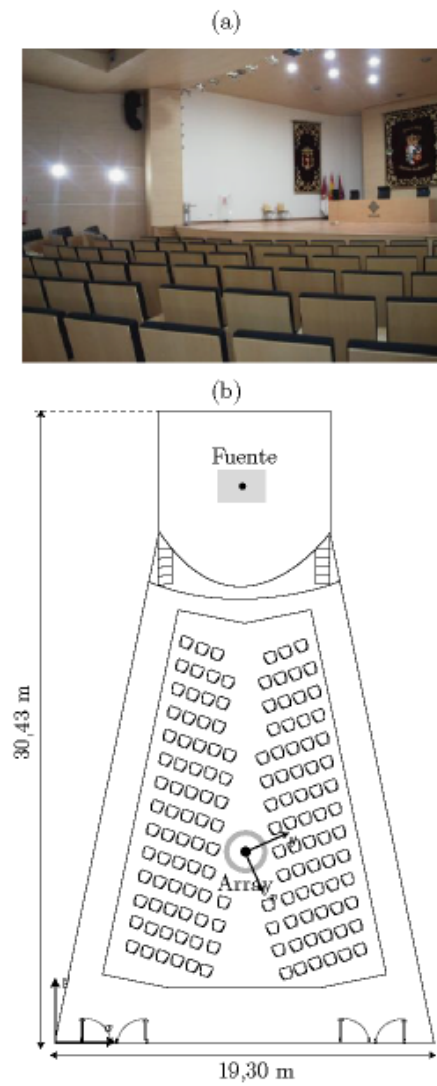


Figura 5.2: (a) Fotografía de Paraninfo, (b) plano y ubicación de la fuente y del array de micrófonos en la sala.

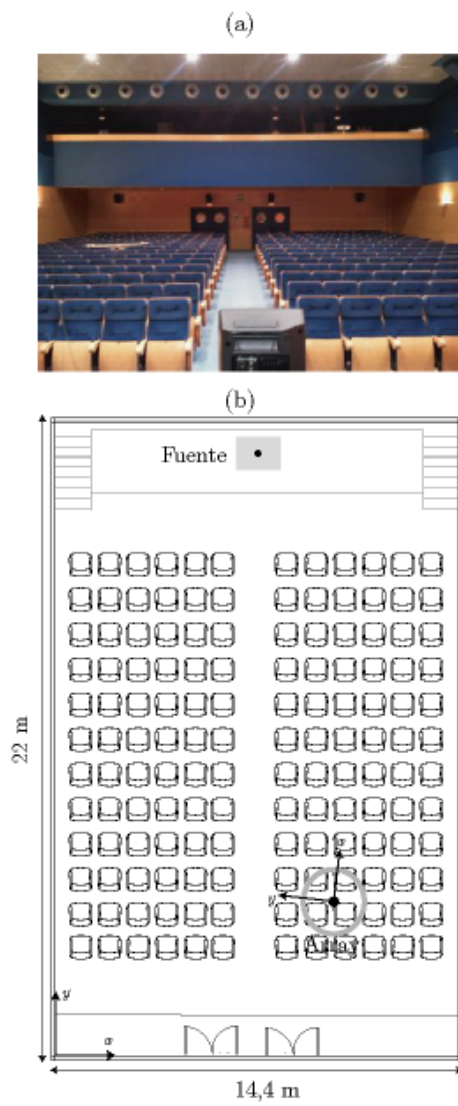


Figura 5.3: (a) Fotografía de Auditorio, (b) plano y ubicación de la fuente y del array de micrófonos en la sala.

Distancia a (m)	Distancia horizontal ( $x$ )	Distancia vertical ( $y$ )
Fuente (H1)	3,80	9,60
Array (H1)	3,80	5,40
Fuente (H2)	7,36	19,20
Array (H2)	10,80	3,90
Fuente (H3)	9,50	28,92
Array (H3)	9,50	14,22

Tabla 5.1: Relación de las distancias en las salas, donde H1 (Sala Reuniones, H2 (Paraninfo) y H3 (Auditorio).

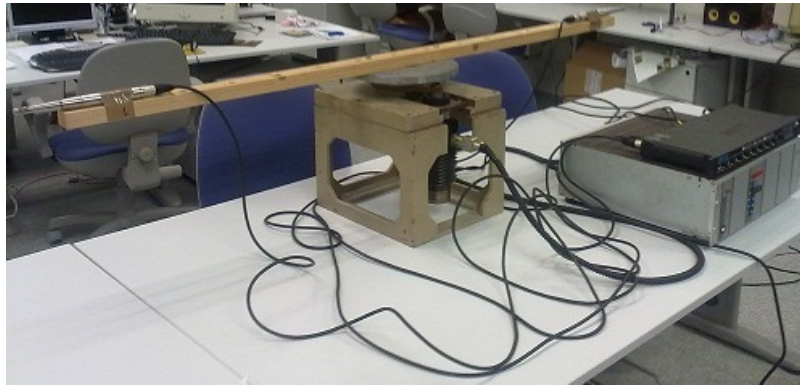


Figura 5.4: Detalle de la disposición de los micrófonos cardioides sobre el motor giratorio.

Tal y como se demostró en la Sección 2.3, utilizar micrófonos cardioides en el array de medidas era suficiente para analizar el campo sonoro con el método de descomposición en ondas planas. Esto era debido a que el término de velocidad  $V(k_\theta, \omega)$  para un micrófono de figura en ocho combinado con el término de presión  $P$  resultaba el patrón de un micrófono cardioide, el cuál era idóneo desde una perspectiva práctica [27]. De tal forma, el array estaba compuesto por dos micrófonos cardioides de la marca Earthworks SR 30 sujetos en el final de una barra de 2 metros de longitud y anclada en un motor giratorio formando entre sí  $180^\circ$  como se aprecia en la Figura 5.4. Para cada sesión de medida, se muestrearon 72 posiciones espaciales por medio de pasos rotatorios de  $5^\circ$ . Como las condiciones

de temperatura y humedad no cambiaron significativamente durante la serie de medidas en la salas, los resultados conseguidos fueron los mismos que si se hubieran tomado con un array completo de  $N = 72$  micrófonos. Para evitar oscilaciones mecánicas provocadas por la rotación de la estructura, se tuvo que tener cuidado de esperar a que los micrófonos se estabilizaran y no oscilaran de una rotación a otra. Aunque este retraso incrementó el tiempo total de medida, se aseguró no tener distorsiones entre muestras acústicas consecutivas. Las medidas del array se capturaron de forma automática y repetitiva para todas las posiciones uniformemente distribuidas sobre el array de 2 m de diámetro dentro del área de escucha. Se utilizaron secuencias MLS [54] como señales de excitación y una frecuencia de muestreo de 44100 Hz. La fuente de señal estaba controlada por medio de un ordenador portátil conectado a una tarjeta de sonido *M-Audio Fast Track Pro audio interface* y el procesado se realizó con un programa matemático basado en Matlab. Finalmente, las respuestas al impulso obtenidas en cada sala fueron analizadas por medio del método de PWD y del algoritmo de detección de ondas explicados en los Capítulos 2 y 4 respectivamente.

### 5.3. Análisis de resultados

A continuación se va analizar la capacidad del método PWD y del algoritmo de detección de ondas para detectar las reflexiones más importantes dentro de las diferentes salas reales medidas. Inicialmente se va a presentar una descripción general de la estructura espacial del campo sonoro a partir de las representaciones tiempo-espacio y PWD tiempo-espacio obtenidas para cada sala. A continuación se analizará la comparativa entre los métodos de detección de onda aplicados a las anteriores representaciones.

#### 5.3.1. Descripción espacial del campo sonoro

Las Figuras 5.5-5.10 muestran las representaciones gráficas más importantes del campo sonoro capturado para las diferentes salas analizadas (representaciones tiempo-espacio o ecogramas, PWD tiempo-

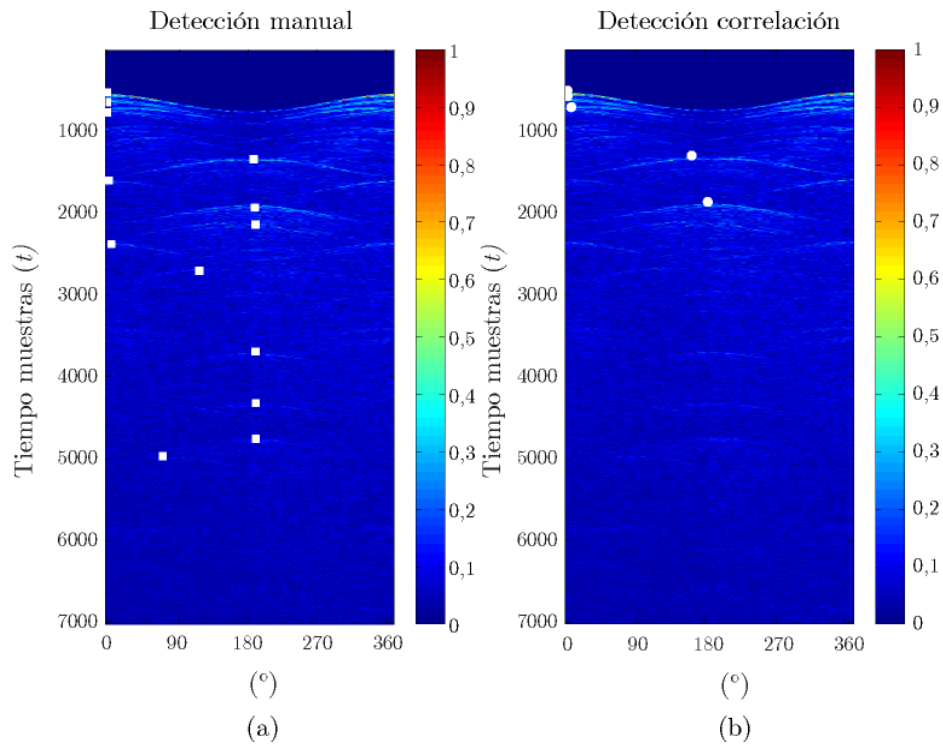


Figura 5.5: Ecograma de la Sala de Reuniones con las principales reflexiones detectadas mediante inspección visual (a) y mediante el algoritmo de detección de ondas (b).

espacio y PWD frecuencia-espacio) así como la detección en ellas de los principales eventos de onda. Como se puede comprobar, las Figuras 5.5 (a), 5.7 (a) y 5.9 (a), muestran la representación conjunta de las RIRs captadas por los micrófonos del array circular, revelando las diferentes estructuras espaciales del campo de onda en cada caso. Puede comprobarse también el carácter ondulatorio de los campos acústicos de las salas sin aplicar ningún tipo de procesamiento posterior a los resultados obtenidos. De forma general, la identificación individual de las reflexiones a partir de las medidas reales no es una tarea tan simple como en el caso simulado mostrado en la Figura 4.6. Sin embargo y a pesar del hecho de que el campo sonoro presenta una estructura compleja debido a efectos de difracción e interferencia, muchas de las

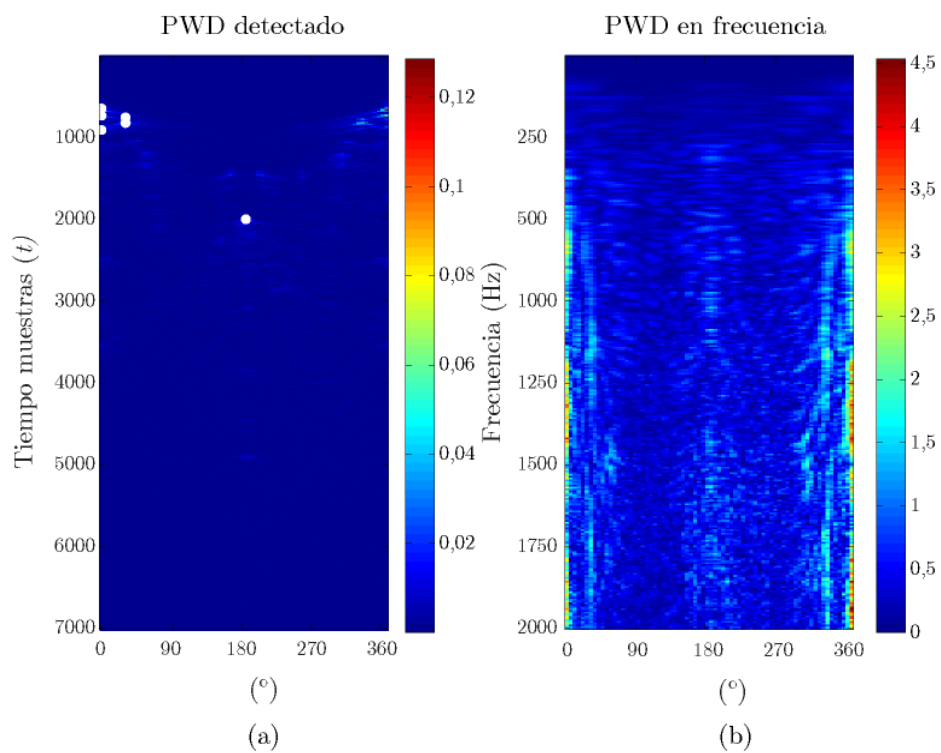


Figura 5.6: Representaciones de PWD en tiempo-espacio con detección de ondas (a), y en frecuencia-espacio (b) para la Sala de Reuniones.



principales reflexiones pueden ser discriminadas. Por otro lado, las representaciones de PWD tiempo-espacio mostradas en las Figuras 5.6 (a), 5.8 (a) y 5.10 (a), proporcionan una alta compactación de energía en las posiciones correspondientes a las reflexiones localizadas. En todos los casos y debido a que las medidas se realizaron utilizando el mismo array circular, el resultado de sus correspondientes ecogramas y descomposición de ondas planas no permite la discriminación en el plano de elevación. Es decir, los frentes de onda desde techo y suelo son proyectados en el mismo instante de llegada. Dado que la sensibilidad del oído humano en la localización vertical no es muy elevada, este efecto no tendría en este caso mayor transcendencia desde un punto de vista perceptual. No obstante, la amplitud de las reflexiones con forma coseno observadas en las representaciones tiempo-espacio, podrían ser un indicador en la elevación de ángulos de las ondas planas reflectadas.

La representación PWD frecuencia-espacio, representada para cada sala en las Figuras 5.6 (b), 5.8 (b) y 5.10 (b), proporciona una impresión general de la distribución espacial de energía para una posición determinada. Como puede comprobarse en las anteriores figuras, la energía se concentra mayoritariamente en la dirección del sonido directo (localización angular del máximo del primer coseno). Por ejemplo, se puede observar cómo la mayoría de energía en la Figura 5.8 (b) viene de la dirección del sonido directo ( $285^\circ$ ). Sin embargo, dependiendo de las características de la sala (materiales, geometría, dimensiones...), se puede observar cómo la energía del sonido a diferentes frecuencias se propaga espacialmente con mayor o menor facilidad. Se puede observar por ejemplo, que la Sala Auditorio (5.10 (b)) tiene una mayor dispersión espacial que las otras dos salas, debido probablemente por el aumento del efecto de difusión debido al tratamiento acústico en sus paredes. De hecho, la determinación de la dispersión de frecuencia en el espacio puede ser una cuestión importante a considerar en el diseño de nuevas salas o en el estudio de sus características para un fin determinado. La evolución con el tiempo de las propiedades espaciales del sonido puede observarse tanto en las representaciones tiempo-espacio como en las PWD tiempo-espacio. Tanto las primeras como las últimas reflexiones pueden identificarse

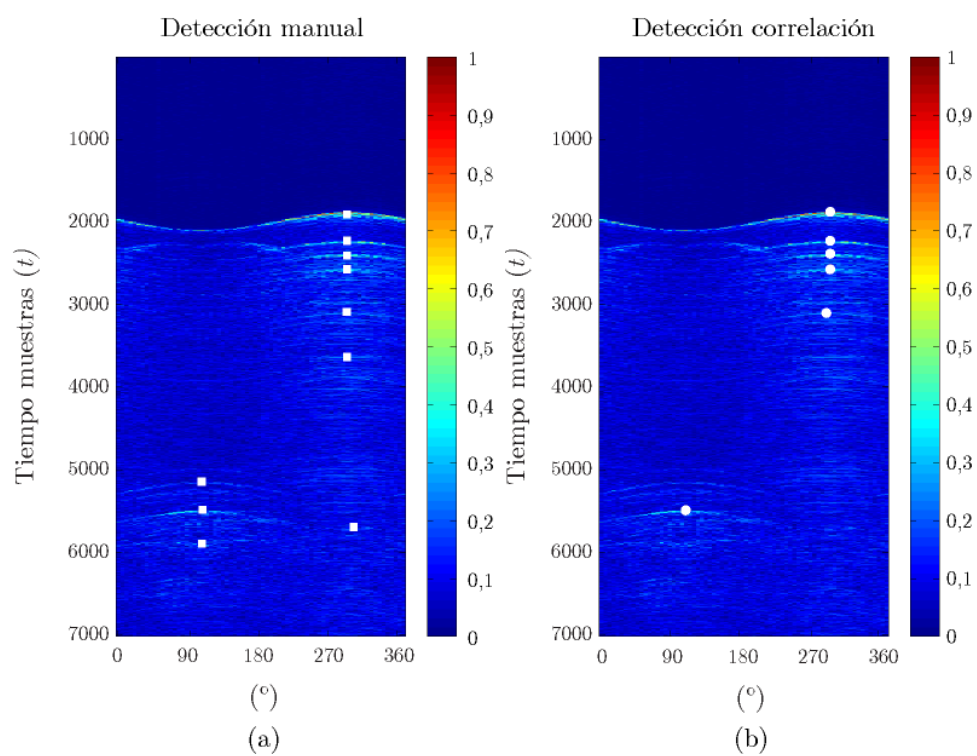


Figura 5.7: Ecograma de la sala Paraninfo con las principales reflexiones detectadas mediante inspección visual (a) y mediante el algoritmo de detección de ondas (b).

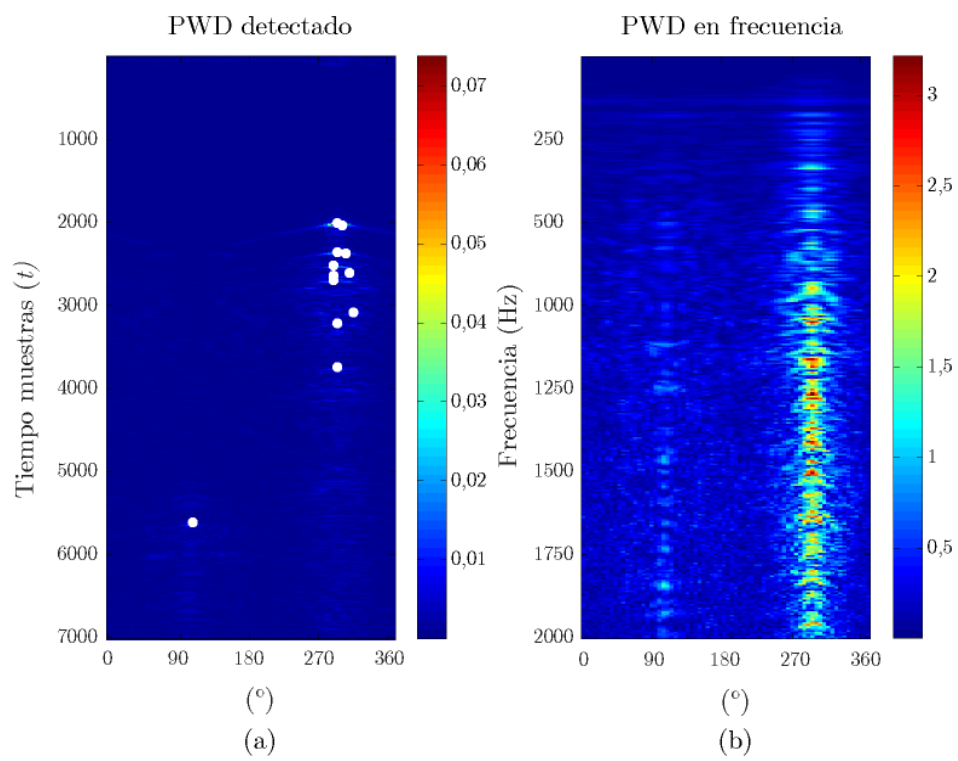


Figura 5.8: Representaciones de PWD en tiempo-espacio con detección de ondas (a), y en frecuencia-espacio (b) para el Paraninfo.

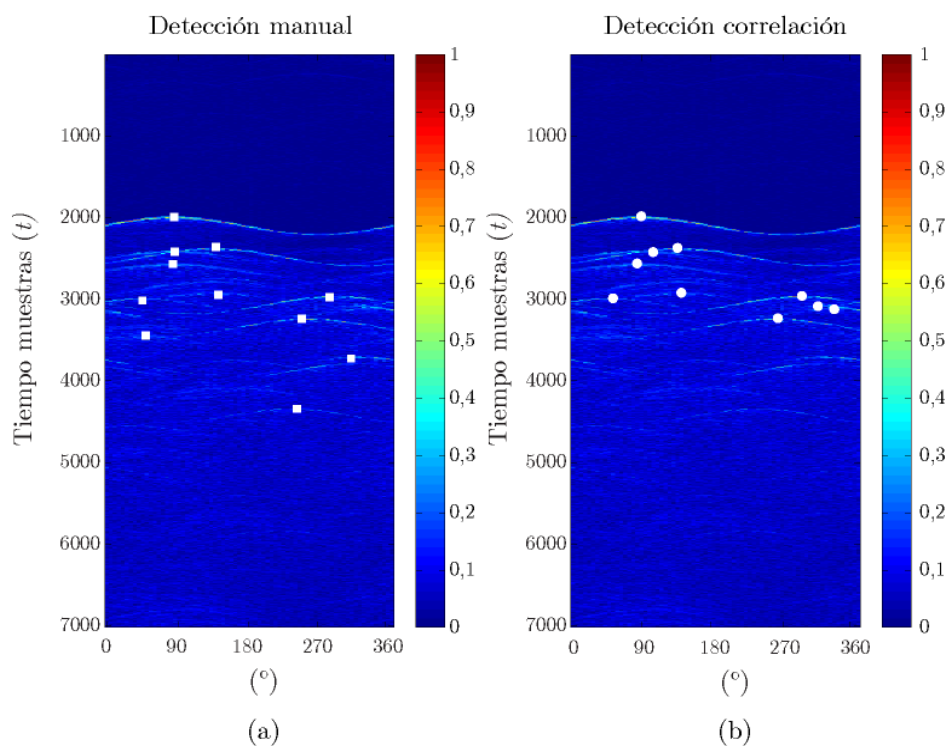


Figura 5.9: Ecograma de la sala Auditorio con las principales reflexiones detectadas mediante inspección visual (a) y mediante el algoritmo de detección de ondas (b).

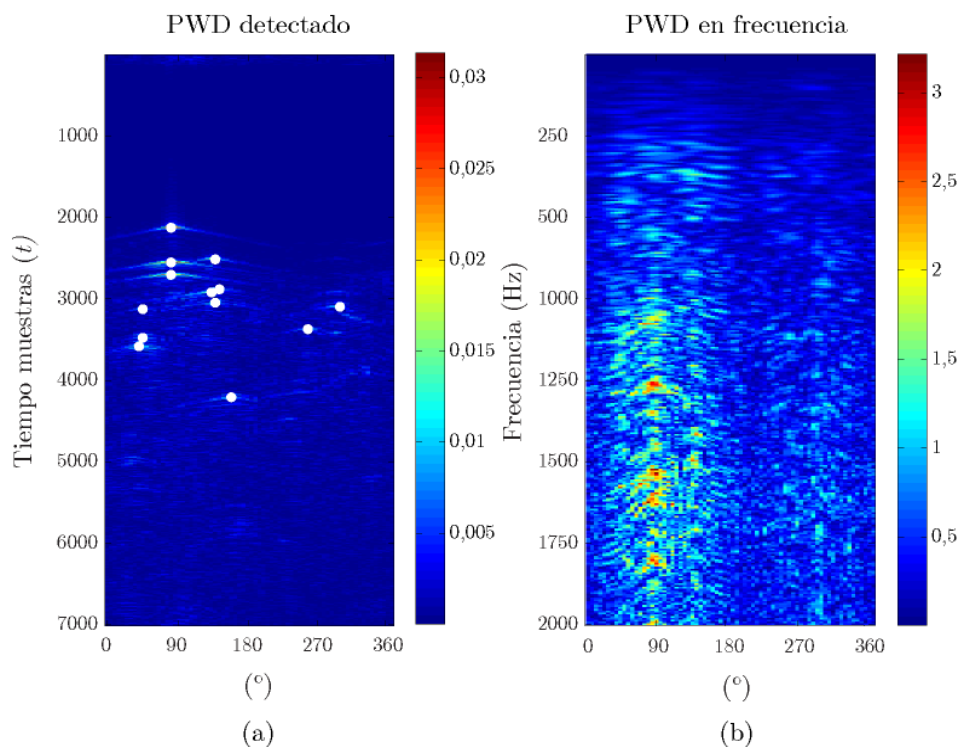


Figura 5.10: Representaciones de PWD en tiempo-espacio con detección de ondas (a), y en frecuencia-espacio (b) para el Auditorio.

en estas representaciones observando que la densidad de estas ondas y su dirección de propagación se incrementa en el tiempo. Este fenómeno cualitativamente apoya el supuesto habitual que dice que el campo sonoro reverberante forma una distribución isotrópica de ondas planas.

### 5.3.2. Detección de ondas

La detección de las principales reflexiones a partir de los campos sonoros captados en las salas, se ha llevado a cabo aplicando los diferentes métodos de detección de ondas:

- *Detección manual*: Este método de detección está basado en un análisis visual de la representación conseguida con las RIRs para identificar los principales eventos de ondas planas ocurridos en el interior de las diferentes salas. Para realizar la selección, se les preguntó a cinco personas diferentes familiarizadas con el procesamiento de imagen la identificación visual de los picos más representativos que pudieran encontrar en los ecogramas. Los cuadros blancos en las Figuras 5.5 (a), 5.7 (a) y 5.9 (a), muestran las reflexiones finales seleccionadas, habiendo sido elegidas de forma común por al menos cuatro personas de las preguntadas. Debido a que esta selección es principalmente experimental, no hay una verdadera exactitud o criterio para definir qué frentes de ondas tienen que ser considerados como los más importantes. Sin embargo, para el resto del análisis de las medidas, estas ondas detectadas van a ser consideradas como las ondas correctas a partir de las cuáles se van a comparar los otros métodos automáticos de detección de ondas.
- *Detección basada en correlación cruzada*: En este caso, las ondas son detectadas de forma automática siguiendo el procedimiento basado en morfología de imagen descrito en el Capítulo 4. Las reflexiones obtenidas mediante este método vienen representadas por círculos blancos en las Figuras 5.5 (b), 5.7 (b) y 5.9 (b).
- *Detección en representaciones PWD*: En las representaciones PWD de tiempo-espacio y como se ha explicado en la Sección 4.3 se

Sala de Reuniones			
	Manual	Ecograma	PWD
Reflexiones totales	12	5	6
Reflexiones correctas	-	5	4
Falsos positivos	-	0	2
Falsos negativos	-	7	7
$MAE_t$	-	5,8	34,0
$MAE_\theta$	-	6,0	0,0

Tabla 5.2: Resultados para la Sala de Reuniones.

aplicará una versión simplificada del método de imagen para detectar ondas. En este caso sólo se aplicará una umbralización en amplitud, y una selección de regiones sobre la representación PWD. De forma similar a la anterior detección, las reflexiones detectadas por este método vienen reflejadas con círculos blancos en las Figuras 5.6 (a), 5.8 (a) y 5.10 (a).

Los parámetros para la detección en todos los casos, fueron constantes y se mantuvieron inalterados. El valor de los umbrales fueron calculados a partir de las Ecuaciones (4.2) y (4.9). La operatividad y funcionamiento del método basado en correlación cruzada sobre las representaciones tiempo-espacio, y del método de detección para las representaciones PWD se estudió mediante la comparación del número de falsos negativos (reflexiones no detectadas) y falsos positivos (reflexiones detectadas incorrectamente) con respecto a los eventos seleccionados manualmente.

Para el análisis de resultados se ha considerado como reflexión detectada correctamente a aquel evento cuya distancia a su correspondiente obtenido de forma manual sea menor a  $20^\circ$  en ángulo y a 100 muestras en tiempo. En ese análisis, también se ha calculado el error medio absoluto (MAE) en tiempo y en ángulo para las reflexiones correctamente detectadas. Estos resultados vienen mostrados en la Tabla 5.2 para la Sala de Reuniones, en la Tabla 5.3 para el Paraninfo y en la Tabla 5.4 para el Auditorio.

En general, puede observarse en la Tabla 5.5, que tanto el método basado en correlación como el método de detección aplicado a las re-

Paraninfo			
	Manual	Ecograma	PWD
Reflexiones totales	9	6	12
Reflexiones correctas	-	6	7
Falsos positivos	-	0	5
Falsos negativos	-	3	2
$MAE_t$	-	1,0	14,8
$MAE_\theta$	-	0,8	2,5

Tabla 5.3: Resultados para la sala Paraninfo.

Sala Auditorio			
	Manual	Ecograma	PWD
Reflexiones totales	11	10	13
Reflexiones correctas	-	8	9
Falsos positivos	-	2	4
Falsos negativos	-	3	2
$MAE_t$	-	1,5	25,6
$MAE_\theta$	-	1,8	6,6

Tabla 5.4: Resultados para la sala Auditorio.



Resultados medios		
	Ecograma	PWD
Reflexiones correctas	59,4 %	62,5 %
Falsos positivos	6,3 %	34,4 %
Falsos negativos	40,6 %	34,4 %
$MAE_t$	2,8	24,8
$MAE_\theta$	2,9	3,0

Tabla 5.5: Resultados globales obtenidos a partir de los resultados particulares de cada sala analizada.

presentaciones PWD son capaces de detectar de forma automática la mayoría de las principales ondas con buena precisión (aproximadamente un 60 % para ecogramas y un 62,5 % para PWD). Aunque el algoritmo de detección en representaciones tiempo-espacio detectan un menor número de reflexiones, la detección es más robusta que para PWD. De hecho, la cantidad de falsos positivos para todas las salas es bastante menor cuando se utiliza el método basado en correlación, (6,3 %). Por otro lado, excepto para la Sala de Reuniones, el número de falsos negativos en PWD es menor que el número de falsos positivos. Como resultado, aunque PWD parece que proporciona mayor número de reflexiones detectadas, muchas de ellas deben de ser descartadas (34,4 %). Con respecto al error medio en ángulo acimut, ambos métodos de detección obtienen precisión similar (aproximadamente  $3^\circ$ ), sin embargo la desviación en tiempo para el método PWD es mayor que para el basado en correlación.

## 5.4. Conclusiones

Después de aplicar el método PWD para analizar los campos sonoros en diferentes salas reales es posible obtener a partir de las RIRs capturadas por el array, las representaciones gráficas que describen correctamente la naturaleza espacial de la acústica en cada sala. Además y mediante la aplicación de los algoritmos de detección de onda se ha podido identificar automáticamente y con buena precisión

---

aquellas reflexiones más significativas para su posterior análisis. Una vez aplicado el algoritmo de detección sobre los resultados del método PWD, se ha comprobado cómo la detección en representaciones tiempo-espacio es más robusta que en representaciones PWD tiempo-espacio. Además, ha sido posible describir la distribución energética espacial de frecuencias y la facilidad de su dispersión en las salas a partir de representaciones PWD frecuencia-espacio, lo cuál representa un elemento importante para el diseño y el estudio de características de nuevos recintos. El análisis y resultados obtenidos en este capítulo, ponen de manifiesto la utilidad del método analizado para investigar sistemáticamente los beneficios e inconvenientes del fenómeno de reflexión dentro de las salas, abriendo un nuevo marco de trabajo para acústica virtual y reproducción de sonido espacial 3D.



# Capítulo 6

## Aplicación de Circular Harmonic Beamforming en localización de fuentes acústicas

---

<b>6.1. Introducción</b>	<b>141</b>
<b>6.2. Algoritmo tiempo-frecuencia aplicado a CHB</b>	<b>142</b>
6.2.1. Modelo de señal	143
6.2.2. CHB tiempo-frecuencia	145
6.2.3. Estimación del DOA	146
<b>6.3. Experimentos</b>	<b>146</b>
6.3.1. Medidas simuladas	147
6.3.2. Comparación con otros métodos	156
6.3.3. Medidas reales	158
<b>6.4. Discusión</b>	<b>165</b>

---

*En esta parte de la tesis se presenta un método de localización de fuentes basado en el procesamiento combinado de CHB y T-F donde CHB se aplica sobre cada punto T-F estimando el DOA de la fuente más dominante. Se han llevado a cabo diferentes experimentos que serán analizados en este capítulo considerando diferente número de*

*fuentes, micrófonos, niveles de reverberación y SNR. Los resultados mostrarán que el método propuesto es capaz de localizar con alta precisión múltiples fuentes sonoras en ambientes muy reverberantes y ruidosos.*

## 6.1. Introducción

La localización de fuentes acústicas utilizando arrays de micrófonos es uno de los temas más activos de investigación en el procesado de señal multicanal, teniendo aplicaciones en múltiples y diferentes áreas como ambientes inmersivos, interfaces hombre-máquina, teleconferencia o audición artificial [55–59]. Sin embargo, la localización de fuentes en banda ancha bajo alto nivel de ruido y reverberación todavía sigue siendo una tarea difícil. En los últimos años, se han desarrollado nuevos algoritmos para hacer frente a este problema, haciendo uso de diferentes geometrías de array y estrategias de localización [60–62]. En este contexto, ha recibido mucha atención el procesado modal utilizando arrays circulares uniformes [42]. Los métodos basados en *beamforming* con armónicos circulares han demostrado que proporcionan mejores resultados en la localización que los obtenidos con métodos clásicos de *beamforming*. De hecho, como se ha comentado en la Sección 3.3, CHB pertenece a un tipo de método más reciente de los conocidos como *eigenbeamforming* [40,63,64]. Como se discute en el Capítulo 3, en [41] se muestra también que CHB alcanza mejor resolución y propiedades en los niveles medios de los lóbulos laterales que el método de *Delay and Sum beamforming*, gracias a una determinada selección de *phase modes* o términos espaciales de Fourier.

Además de los métodos de localización basados en *beamforming*, se han desarrollado recientemente algoritmos que trabajan en el dominio tiempo-frecuencia, T-F, [65–67], [68]. Debido a las propiedades de escasez o *sparsity* de la voz en este dominio de T-F, estos métodos son capaces de localizar múltiples fuentes sonoras activas en ambientes reales, incluso en aquellos casos donde el número de fuentes excede al número de micrófonos. Para este fin, se analizan diferencias entre canales de las señales de los micrófonos para estimar la dirección de llegada de la fuente sonora dominante en cada punto T-F. La dirección de las fuentes finalmente se estiman ajustando un modelo específico a la distribución de la dirección de llegada o DOA observada.

En este capítulo, se analiza un método de localización de fuentes basado en el procesamiento combinado de CHB y T-F. En contraste con los métodos de localización basados en diferencias interaurales de

fase (IPD), los DOA estimados se obtienen tomando la dirección donde se presenta el CHB máximo, proporcionando una alta precisión de estimación de DOA bajo altos niveles de reverberación y en situaciones de baja relación SNR. Se mostrarán las simulaciones y experimentos que se han realizado considerando diferente número de fuentes, micrófonos, niveles de reverberación y SNR. Los resultados avalarán el método propuesto como técnica capaz de localizar múltiples fuentes sonoras en ambientes ruidosos y muy reverberantes. Por último se comparará con otros métodos de localización de fuentes, demostrando su mejor comportamiento con respecto a éstos.

## 6.2. Algoritmo tiempo-frecuencia aplicado a CHB

Los campos de aplicación de las representaciones tiempo-frecuencia (T-F-R) son cada vez más amplios, pues se ha comprobado que mejoran los resultados de los métodos espectrales y temporales clásicos al ser capaces de reflejar cambios en frecuencia con respecto al tiempo (también llamados transitorios espectrales) y como en un análisis espectral clásico estos cambios no pueden detectarse, se mejora la clasificación o detección de determinadas propiedades de la señal analizada. Análogamente, los métodos basados en propiedades temporales no consiguen detectar características esenciales de la señal que son las que muestran con certeza su naturaleza. Por ello, un uso combinado de ambos dominios, resulta provechoso para así realizar análisis fiables. Inicialmente se aplicó en la detección por radar y en reconocimiento del habla, pero hoy en día se aplica en casi todos los campos del tratamiento digital de señales.

Por otro lado, el procesado con arrays circulares ha demostrado ser eficaz a la hora de localizar fuentes acústicas alrededor de  $360^\circ$  [69]. De manera específica, los métodos de *beamforming* basados en CHB han atraído la atención de muchos investigadores en los últimos años, siendo el procesado de señal modal en arrays, un tema muy activo, en cuanto a la investigación se refiere. Además, los métodos de localización de fuentes basados en el dominio de tiempo-frecuencia,

han demostrado ser capaces también de localizar varias fuentes de forma simultánea con alta precisión. De tal forma, en este capítulo se estudia una técnica de localización donde el método CHB se aplica sobre cada intervalo o segmento tiempo-frecuencia, proporcionando una precisa localización bajo muy adversas condiciones acústicas. La selección de modos y la localización basada en las características del habla son definidas para estimar la dirección de llegada de múltiples fuentes que inciden simultáneamente sobre un UCA.

### 6.2.1. Modelo de señal

Se considera el siguiente modelo de señal con  $M$  fuentes  $s_m$  localizadas en campo lejano en las direcciones  $\theta_m$ . Por simplificación, se asumirá un ambiente anecoico:

$$x_n(t) = \sum_{m=1}^M s_m(t - \delta_{nm}), \quad n = 0, \dots, N - 1, \quad (6.1)$$

donde  $\delta_{nm}$  es el retraso de tiempo correspondiente al camino acústico entre la fuente  $m$  y el micrófono  $n$ .

En el dominio de la transformada STFT, el anterior modelo de señal se expresa como

$$X_n(v, l) = \sum_{m=1}^M S_m(v, l) e^{-j2\pi f_v \delta_{nm}}, \quad (6.2)$$

donde  $X_n(v, l)$  y  $S_m(v, l)$  son las representaciones T-F de las señales de los micrófonos y fuentes respectivamente,  $(v, l)$  son los índices correspondientes a los intervalos de frecuencia y de tiempo y  $f_v$  es la frecuencia analógica correspondiente al índice frecuencial  $v$ .

- *Propiedad de Sparsity de la voz en el dominio T-F*

Las señales como el habla y la música han mostrado ser dispersas (o *sparse* en inglés) en el dominio T-F [70]. La función de densidad de probabilidad de una fuente dispersa tiene forma estrecha o puntiaguda, lo cual es debido a que la señal en la mayoría de



sus puntos T-F, tiene un valor próximo a cero y alcanza mayores valores en escasas y puntuales ocasiones. Esta propiedad ha sido ampliamente utilizada y aplicada en numerosos trabajos relacionados con localización y separación de fuentes [71], [72], [73], [74]. Sin embargo, una fuente dispersa sola no tiene utilidad si las fuentes se solapan en un elevado número de grados. La disyunción o *disjointness* de una mezcla de fuentes puede definirse como el grado de no solapamiento de señales mezcladas. A la medida objetiva de este parámetro se le conoce como *W-Disjoint Orthogonality* (WDO) [75], [76].

El solapamiento espectral depende no sólo de la dispersión de la fuente, sino también de las relaciones mutuas entre señales. Las señales de voz a menudo se solapan de forma aleatoria y sin correlación, como ocurre por ejemplo en el efecto de *cocktail party* [77]. Además, las propiedades de disyunción de la voz y de señales musicales son dependientes del tamaño de ventana que afecta en su análisis al número de bandas frecuenciales [78]. Se debe reseñar que las propiedades de *sparsity* y *disjointness* de señales de audio se ven afectadas por ambientes reverberantes. La energía de la respuesta al impulso se ve alterada y estropeada tanto en tiempo como en frecuencia aumentando el solape espectral de las diferentes fuentes en el dominio T-F debido al fenómeno de la reverberación. A pesar de este efecto, se ha demostrado que la premisa de tener fuentes no solapadas es todavía útil en aplicaciones basadas en la propiedad de escasez o *sparsity*, como por ejemplo en aplicaciones de separación de fuentes [65].

Asumiendo que sólo hay una fuente dominante en cada intervalo T-F (premisa WDO), el modelo de señal puede simplificarse todavía más:

$$X_n(v, l) = S_{\check{m}(v,l)}(v, l)e^{-j2\pi f_v \delta_{n\check{m}(v,l)}}, \quad (6.3)$$

donde  $\check{m}$  corresponde al índice de la fuente dominante en el punto T-F  $(v, l)$ .

### 6.2.2. CHB tiempo-frecuencia

Para estimar la dirección de la señal correspondiente a la fuente dominante en cada punto T-F, se debe de realizar un *beamforming* basado en armónicos circulares sobre cada elemento T-F tal y como se describe a continuación. Primeramente, las señales de los micrófonos son transformadas al dominio de la frecuencia para cada punto T-F  $(v, l)$  por medio de la DFT de  $\mathbf{x}(v, l) = [X_0(v, l), \dots, X_{N-1}(v, l)]^T$ :

$$\mathbf{c}(v, l) = \text{DFT} \{ \mathbf{x}(v, l) \}. \quad (6.4)$$

El vector de coeficientes debe de acomodarse a la siguiente estructura:

$$\mathbf{c}(v, l) = [\tilde{C}_0(f_v), \dots, \tilde{C}_L(f_v), \tilde{C}_{-L}(f_v), \dots, \tilde{C}_{-1}(f_v)]^T, \quad (6.5)$$

donde los coeficientes  $\tilde{C}_p(f_v)$  corresponden a los de una apertura circular muestreada en la Ecuación (3.17). Teniendo en cuenta la condición  $N \geq 2L + 1$ , el coeficiente de la DFT cuya posición fuera la de  $(L + 1)$  debería de descartarse en el caso de tener un número par de sensores.

A continuación, se define la siguiente matriz de apuntamiento, formada por  $Q$  vectores diferentes de pesos que cubriendo todo el rango acimutal  $\theta_q \in [0, 2\pi]$ ,  $q = 1 \dots Q$ :

$$\mathbf{W} = [\mathbf{h}(\theta_1), \dots, \mathbf{h}(\theta_Q)]. \quad (6.6)$$

donde

$$\mathbf{h}(\theta_q) = [e^{j0\theta_q}, \dots, e^{jL\theta_q}, e^{j(-L)\theta_q}, \dots, e^{j(-1)\theta_q}]^T. \quad (6.7)$$

La matriz  $\mathbf{W}$  define el rango angular que escaneará espacialmente con el fin de localizar las fuentes activas. El vector de coeficientes de ecualización para cada frecuencia vendrá dado por:

$$\mathbf{b}(v) = [B'_0(f_v), \dots, B'_L(f_v), B'_{-L}(f_v), \dots, B'_{-1}(f_v)]^T, \quad (6.8)$$

donde los elementos  $B'_p(f_v)$  son calculados según la Ecuación (3.27) utilizando la relación  $kR = (2\pi f_v/c)R$ .

Los coeficientes ecualizados de Fourier son calculados de la siguiente forma

$$\bar{\mathbf{c}}(v, l) = \mathbf{c}(v, l) \circ \mathbf{b}(v), \quad (6.9)$$

donde  $\circ$  corresponde al operador producto *Hadamard*.

Finalmente, la salida del *beamformer* para cada ángulo escaneado  $\mathbf{y}(v, l) = [Y(v, l, \theta_1), \dots, Y(v, l, \theta_Q)]^T$  es calculado como

$$\mathbf{y}(v, l) = \frac{1}{2L+1} \mathbf{W}^T \bar{\mathbf{c}}(v, l). \quad (6.10)$$

### 6.2.3. Estimación del DOA

Como se ha comentado antes y asumiendo que sólo hay una fuente dominante para cada punto T-F, la salida del *beamformer* tendrá máxima potencia en la dirección de llegada. De esta forma, el ángulo DOA para cada intervalo T-F puede ser estimado como

$$\hat{\theta}(v, l) = \arg \max_{\theta_q} \{|Y(v, l, \theta_q)|^2\}, \quad (6.11)$$

donde

$$\hat{\theta}(v, l) \approx \theta_{\check{m}(v, l)}, \quad (6.12)$$

siendo  $\theta_{\check{m}(v, l)}$  la DOA real correspondiente a la fuente dominante en  $(v, l)$ .

Puesto que en diferentes puntos T-F, hay varias fuentes simultáneas dominantes, el histograma del DOA estimado sobre el plano T-F mostrará picos que corresponderán a la localización de diferentes fuentes. Aunque en esta sección se ha asumido un modelo de señal anecoico, a continuación se comprobará cómo este método funciona de forma muy robusta bajo condiciones acústicas adversas, incluyendo una alta reverberación así como una baja relación SNR.

## 6.3. Experimentos

Esta sección es la encargada de estudiar el comportamiento del método propuesto (denotado a continuación como TF-CHB) considerando diferentes configuraciones de array y bajo diversas condiciones acústicas [79]. Se implementaron algunas simulaciones acústicas basadas en el método de fuente de imagen [80] para analizar la influencia en los resultados de las reflexiones de la sala, del

ruido, del número de fuentes y del número de micrófonos del array. A continuación se ha comparado este método con otras técnicas ampliamente utilizadas en este ámbito de localización de fuentes como CHB convencional en el dominio del tiempo, DAS, y EB-ESPRIT [81], [40], [79]. Además, el método de localización fue también analizado y evaluado con grabaciones reales obtenidas de una base de datos pública y disponible en internet.

### 6.3.1. Medidas simuladas

El uso de grabaciones simuladas permite analizar mejor cómo el ruido y la reverberación afectan en la precisión de la localización para una configuración de array determinado. Además, las grabaciones sintéticas hacen más sencillo observar las mejoras en el comportamiento cuando se incrementa el número de micrófonos en un ambiente multifuente. Como señales fuente, se utilizaron fragmentos hablados de mujeres y hombres de 4 segundos de longitud, provenientes de la base de datos 'Dev2' dentro del estudio de evaluación (SiSEC) [82]. La frecuencia de muestreo utilizada fue  $f_s = 8$  kHz, y por lo tanto, todos los arrays simulados fueron diseñados para trabajar con una frecuencia de *aliasing* de  $f_{al} = 4$  kHz.

Se consideraron tres configuraciones de array:

- Array 1:  $N = 5$ ,  $R = 3,6$  cm,
- Array 2:  $N = 9$ ,  $R = 6,3$  cm,
- Array 3:  $N = 21$ ,  $R = 14,4$  cm.

Con respecto a la disposición de las fuentes, se consideraron dos casos complejos multifuente:

- Caso 1:  $M = 4$ ,  
 $\theta_m \in \{20^\circ, 110^\circ, 200^\circ, 290^\circ\}$ .
- Caso 2:  $M = 8$ ,  
 $\theta_m \in \{20^\circ, 65^\circ, 110^\circ, 155^\circ, 200^\circ, 245^\circ, 290^\circ, 335^\circ\}$ .

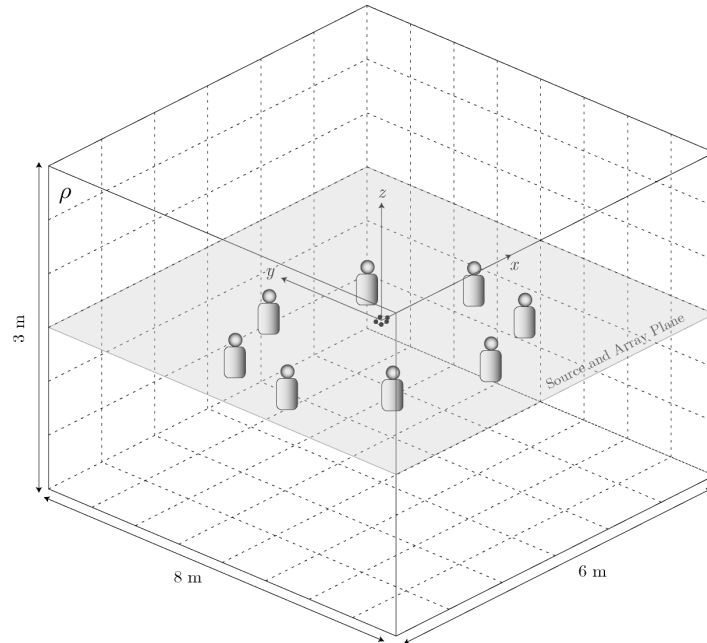


Figura 6.1: Sala simulada con dimensiones de  $6\text{ m} \times 8\text{ m} \times 3\text{ m}$  y factor de reflexión en paredes  $\rho$ . Las fuentes y el array, están localizados en el plano  $xy$ .

Para evaluar la influencia de las reflexiones de la sala, se consideró una habitación simulada con forma rectangular de dimensiones  $6\text{ m} \times 8\text{ m} \times 3\text{ m}$ . Las fuentes estaban localizadas a  $2\text{ m}$  del dentro del array tal y como muestra la Figura 6.1. La influencia de la reverberación es controlada por medio del factor de reflexión en cada pared de  $\rho$  [4] y que se supone que es igual para todas las paredes de la sala. Se utilizaron tres factores diferentes de reflexión:  $\rho = 0$  (caso anecoico),  $\rho = 0,5$  (reverberación moderada) y  $\rho = 0,9$  (alta reverberación). Además, se introdujo un ruido aditivo Gaussiano con potencia  $\sigma_n^2$  a cada señal de micrófono para obtener de tal forma diferentes valores SNR:

$$SNR = 10 \log \left( \frac{\left(\frac{1}{T}\right) \sum_t \sum_{m=1}^M s_{mn}^2(t)}{\sigma_n^2} \right), \quad (6.13)$$

donde  $s_{mn}(t)$  son las señales fuente originales convolucionadas con las

respuestas al impulso simuladas fuente a micrófono, es decir,  $s_{mn}(t) = s_m(t) * h_{nm}(t)$ . Los valores SNR considerados han sido SNR=  $\infty$  dB (libre de ruido), SNR= 10 dB (ambiente ruidoso) and SNR= 0 dB (ambiente muy ruidoso).

Con respecto a los parámetros de procesado, se computaron STFTs utilizando ventanas de Hamming de 512 muestras de longitud y un 50 % de solape. Se ha comprobado en [70] que estos valores son apropiados para señales habladas. El factor de regularización fue de  $\beta = 6,5 \times 10^{-4}$ . El espacio acimutal escaneado fue de 360 ángulos uniformemente espaciados ( $Q = 360$ ), con una resolución angular de un grado.

Los histogramas que se obtuvieron de los DOAs correspondientes a  $M = 4$  fuentes, se muestran en las Figuras 6.2, 6.3, y 6.4, para diferentes valores de SNR y  $\rho$ . Los histogramas obtenidos de los DOAs correspondientes a  $M = 8$  fuentes, se muestran a su vez en las Figuras 6.5, 6.6 y 6.7 para los mismos valores de SNR y  $\rho$  que en el caso anterior.

Los puntos negros en el eje  $\theta$  denotan las localizaciones exactas de las fuentes. Para facilitar el procedimiento de detección de picos, se aplicó a los histogramas un filtro paso-bajo que eliminaba los picos espúreos, suavizando la curva final. El DOA final de las fuentes puede ser obtenida utilizando varias alternativas como por ejemplo con técnicas de detección de picos (*peak-picking*) u otros métodos basados en modelado. En esta tesis se ha utilizado la función *findpeaks* de Matlab para la obtención de los máximos locales en los histogramas suavizados. En este caso se ha tomado el número de fuentes  $M$  como conocido y una distancia de separación mínima entre picos de  $10^\circ$ . Los valores de RMSE obtenidos para cada configuración de array también se presentan en la Tabla 6.1. Como era de esperar, en el caso anecoico y libre de ruido (Figura 6.2 y Figura 6.5), los histogramas muestran de forma clara los picos localizados en las posiciones reales de las fuentes. La anchura de los picos es muy dependiente del número de micrófonos, siendo considerablemente más estrechos cuantos más elementos se utilizan en el array. Cuando el ruido y la reverberación están presentes en las señales de entrada, los histogramas empiezan a empeorar y la anchura media de pico se incrementa substancialmente (Figuras 6.2, 6.4 para 4 fuentes y 6.6 y 6.7 para 8 fuentes). Aunque

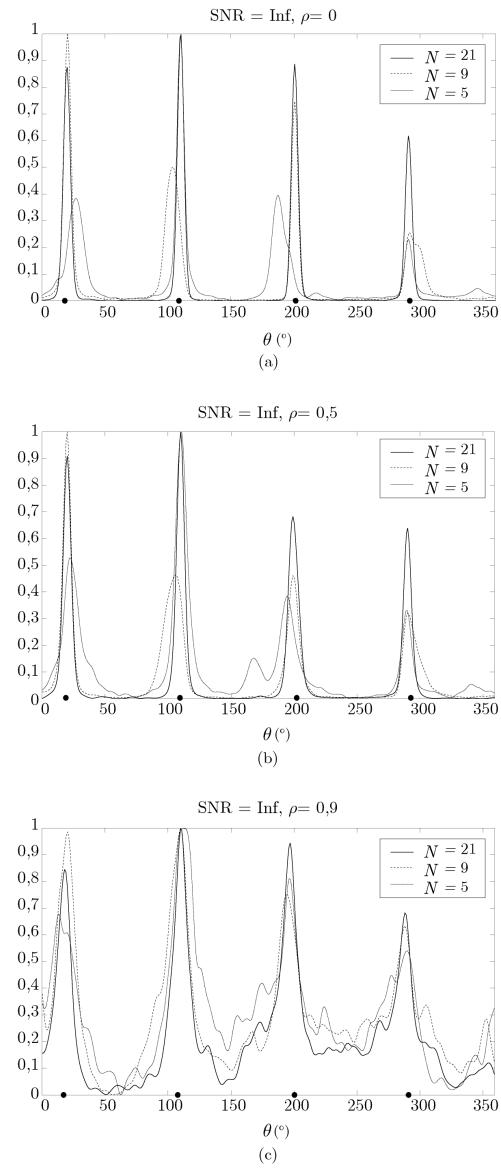


Figura 6.2: Histogramas obtenidos para las simulaciones con  $M = 4$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos para  $\text{SNR} = \infty$ : (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$ .

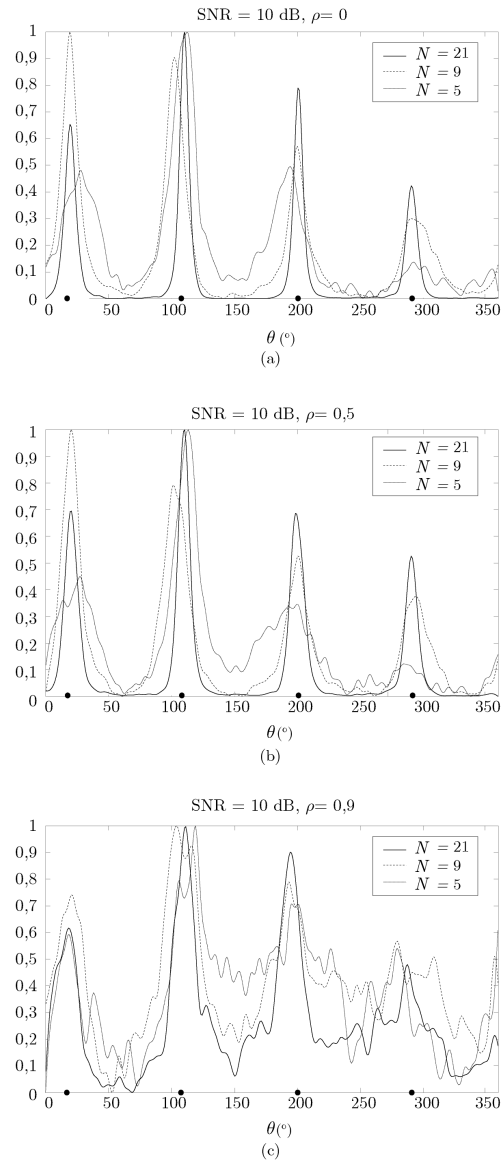


Figura 6.3: Histogramas obtenidos para las simulaciones con  $M = 4$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos para SNR=10 dB: (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$ .



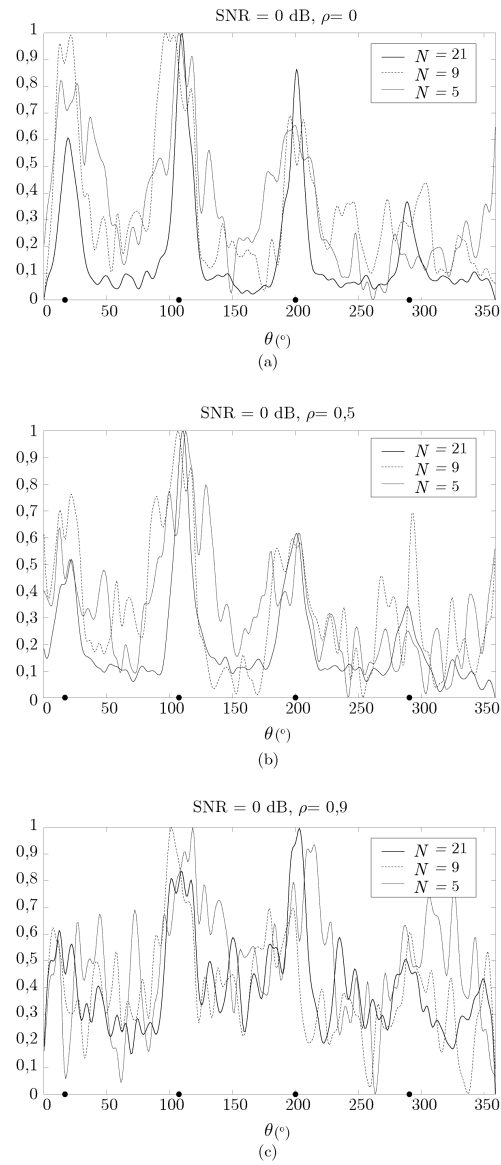


Figura 6.4: Histogramas obtenidos para las simulaciones con  $M = 4$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos para SNR=0 dB: (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$ .

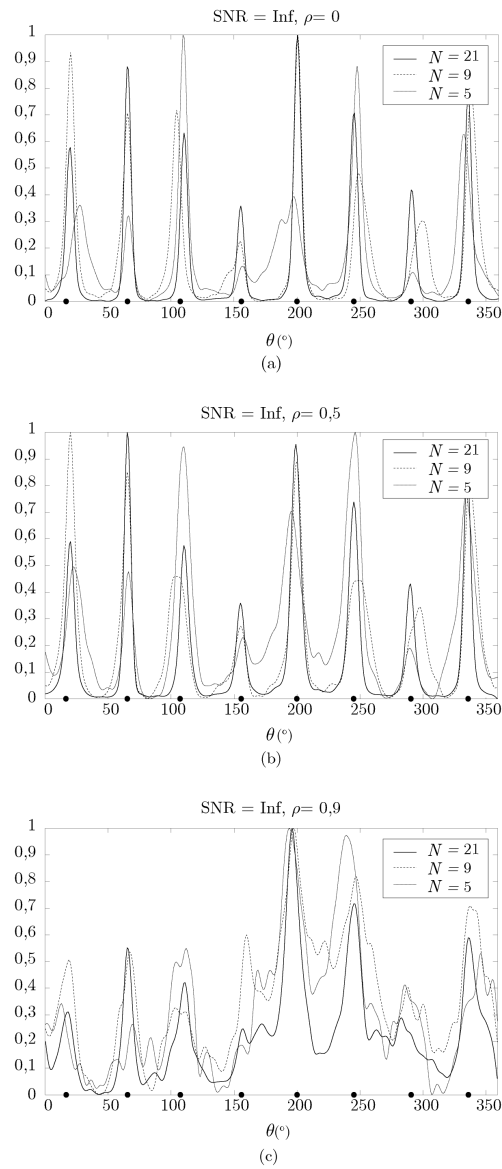


Figura 6.5: Histogramas obtenidos para las simulaciones con  $M = 8$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos para SNR= Inf: (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$ .

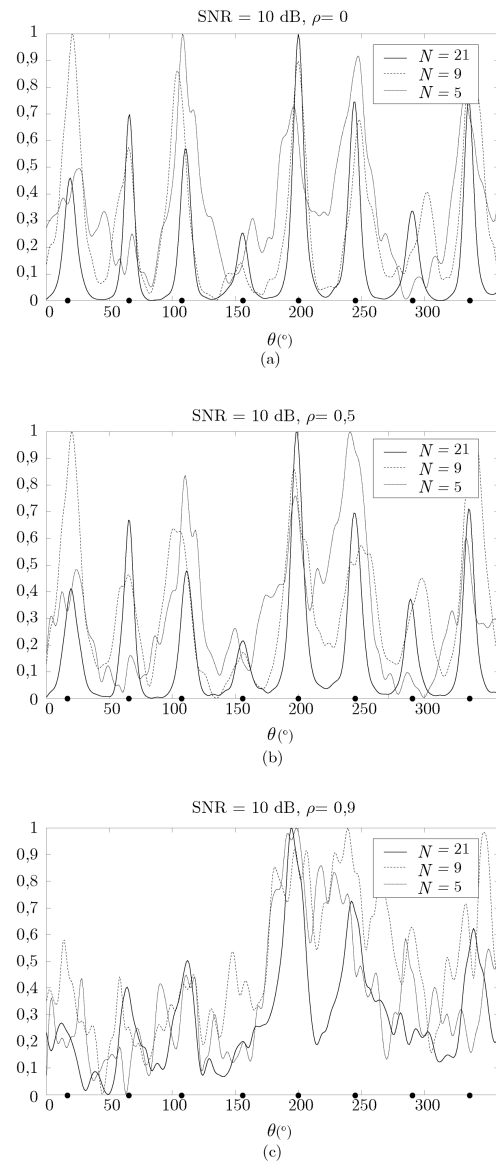


Figura 6.6: Histogramas obtenidos para las simulaciones con  $M = 8$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos para SNR=10 dB: (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$ .

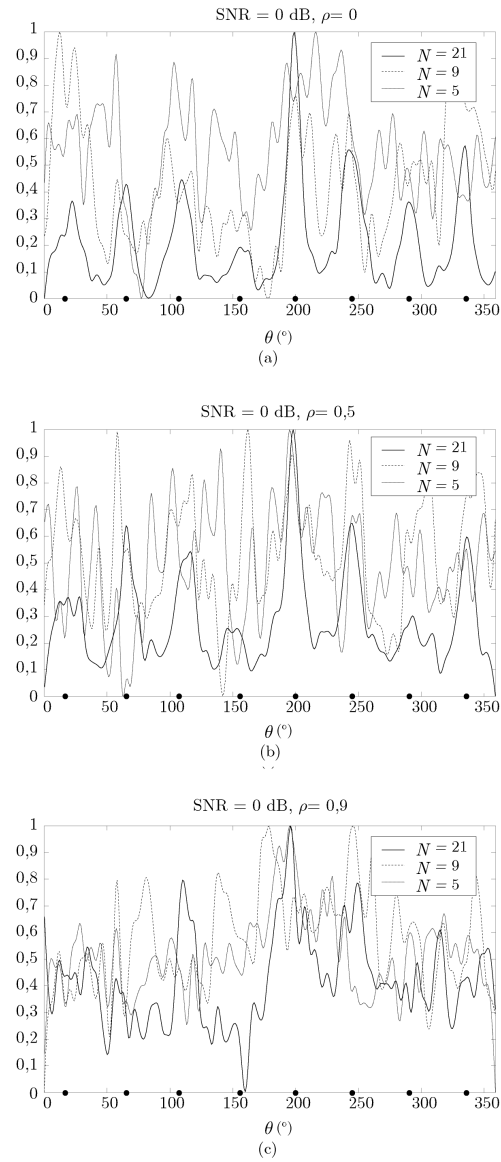


Figura 6.7: Histogramas obtenidos para las simulaciones con  $M = 8$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos para SNR=0 dB: (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$ .

se puede aplicar un filtro paso bajo a los histogramas para eliminar picos espúreos y observar mejor dónde se ubican los máximos, se ha considerado más ilustrativo mostrar los histogramas sin ningún tipo de procesamiento de las estimaciones de los DOA reales. El efecto del ruido y la reverberación en el rendimiento puede ser crítico en función del número de fuentes simultáneas y su separación angular. Debido al incremento de la anchura de pico en condiciones acústicas adversas, las fuentes de alta potencia situadas muy cerca de fuentes de baja potencia pueden enmascarar a estas últimas. No obstante, dado que añadiendo más micrófonos las anchuras de los picos se vuelven más estrechas, la exactitud de localización puede ser siempre mejorada utilizando un mayor número de micrófonos.

Como se observa en la Tabla 6.1, la precisión de la localización para todos los casos simulados depende de los factores anteriormente descritos. Incluso la localización de 8 fuentes simultáneas pueden estimarse con un error medio de  $20^\circ$ . Debe de nuevo enfatizarse que el caso peor ( $\rho=0,9$  y  $SNR=0$  dB), es un caso extremo en la práctica, puesto que tener 2 o 3 hablantes de forma simultánea en un ambiente de comunicación (como teleconferencia o sala de reuniones) no suele ser habitual [83]. En cualquier caso, la media de error en la localización con unos valores medios de reverberación y ruido, están entorno a un grado para el "Array 3" y 8 grados para el "Array 2" en el caso de tener 4 hablantes simultáneamente hablando.

### 6.3.2. Comparación con otros métodos

En esta sección, se compara la aplicación del método TF-CHB con otras técnicas conocidas en localización de fuentes. Estas técnicas son DAS [84], CHB [41] en el dominio del tiempo y EB-ESPRIT [40]. Todos estos métodos se emplean ampliamente con arrays circulares. Los métodos DAS y CHB se comparan en [41], donde se muestra que CHB obtiene mejor resolución angular y características de lóbulo que el método DAS a pesar de ser menos robusto en presencia de ruido. Las capacidades de localización sonora de arrays modales han sido también examinados [40], donde el algoritmo ESPRIT se aplica sobre señales en el dominio temporal para localizar varias fuentes. Las Figuras 6.8, 6.9 y 6.10 muestran los resultados correspondientes

Array 1

SNR (dB)	$M = 8$			$M = 4$		
	$\infty$	10	0	$\infty$	10	0
$\rho = 0,0$	6,60	9,35	14,31	5,39	9,22	11,76
$\rho = 0,5$	11,29	13,86	25,86	8,60	10,25	12,38
$\rho = 0,9$	19,45	27,65	33,54	11,84	14,76	19,72

Array 2

SNR (dB)	$M = 8$			$M = 4$		
	$\infty$	10	0	$\infty$	10	0
$\rho = 0,0$	6,01	8,71	14,06	4,12	8,02	8,67
$\rho = 0,5$	8,19	13,82	22,09	6,08	8,06	9,44
$\rho = 0,9$	10,51	23,64	32,75	7,02	10,74	12,38

Array 3

SNR (dB)	$M = 8$			$M = 4$		
	$\infty$	10	0	$\infty$	10	0
$\rho = 0,0$	0,00	1,41	6,04	0,00	1,00	3,35
$\rho = 0,5$	1,00	2,45	8,25	1,00	1,00	3,35
$\rho = 0,9$	8,61	14,56	26,08	3,74	6,48	9,91

Tabla 6.1: RMSE para las diferentes configuraciones de array.

a la potencia angular de salida obtenidos después de comparar entre estos métodos y TF-CHB para diferentes configuraciones de SNR y  $\rho$ . Debido a que EB-ESPRIT proporciona las direcciones estimadas de las fuentes de forma directa, los resultados para este método son representados por líneas verticales en esas direcciones estimadas. Las diferentes figuras muestran los resultados para el "Caso 1" y el "Array 3" ( $M = 4$ ,  $N = 21$ ), en las mismas condiciones acústicas que las presentadas en las Figuras 6.2, 6.2, y 6.4. Se puede observar claramente que la anchura de los lóbulos son más estrechos en CHB que en DAS, proporcionando por lo tanto mejor resolución angular. Sin embargo se observa que DAS es más robusto que CHB en presencia de ruido blanco, como muestra la Figura 6.10 (c). EB-ESPRIT demuestra ser un buen método con niveles de reverberación y ruido moderados, pero su comportamiento se degrada severamente en condiciones extremas. Las Tablas 6.2, 6.3, y 6.4 muestran los valores de RMSE de estos métodos para  $M = 4$  y  $M = 8$  fuentes. En aquellos casos en los que las fuentes no pueden ser detectadas, no se proporcionan los valores RMSE. El método TF-CHB supera a los otros métodos especialmente cuando hay un gran número de fuentes sonoras activas. Además, también puede observarse cómo TF-CHB se mantiene robusto bajo condiciones acústicas muy adversas aunque como es de esperar, la precisión en la localización, decrece significativamente.

### 6.3.3. Medidas reales

Los datos reales recogidos y de libre distribución fueron los correspondientes a la base de datos AV16.3 [85]. Estos datos se utilizaron para testear y comprobar la efectividad del método propuesto con señales capturadas de un UCA real en una sala de reuniones con tres personas hablando simultáneamente (ver Figura 6.11). Esta base de datos ha sido ampliamente utilizada en muchos trabajos relacionados con el procesamiento del habla [86], [87]. De forma específica, las señales utilizadas en esta tesis corresponden al conjunto de grabaciones tituladas 'seq37-3p-0001', usando 9 de los 32 segmentos hablados en los que se dividió la grabación completa (3 fragmentos en cada caso,  $M = 1$ ,  $M = 2$  y  $M = 3$ ). Las grabaciones fueron tomadas en el IDIAP *Smart Meeting Room* [88], cuyas dimensiones eran  $3,6 \text{ m} \times 8,2 \text{ m} \times 2,4 \text{ m}$

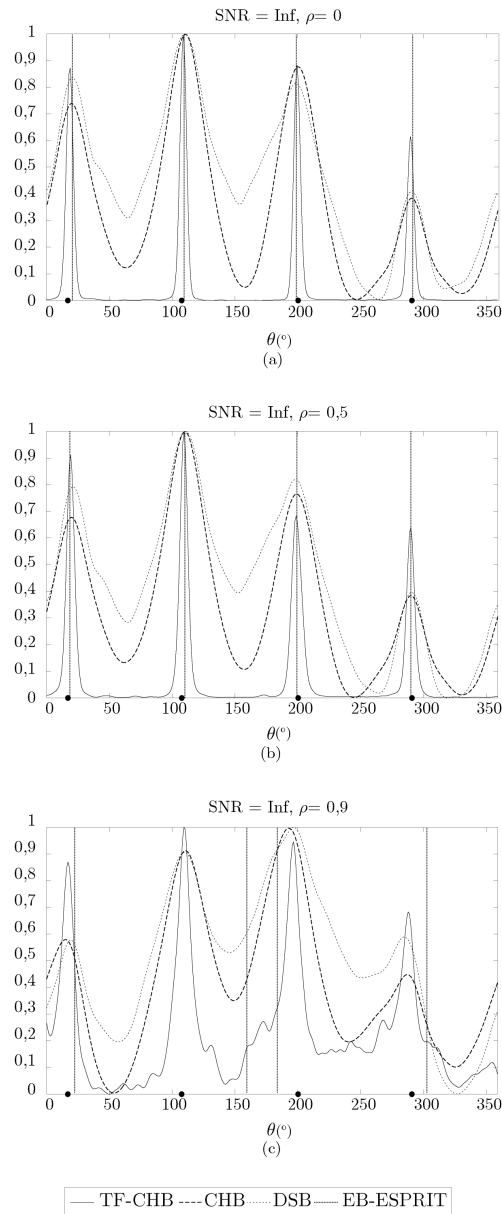


Figura 6.8: Histogramas normalizados para grabaciones simuladas con  $M = 4$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos. (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$  para  $\text{SNR} = \text{Inf}$ .



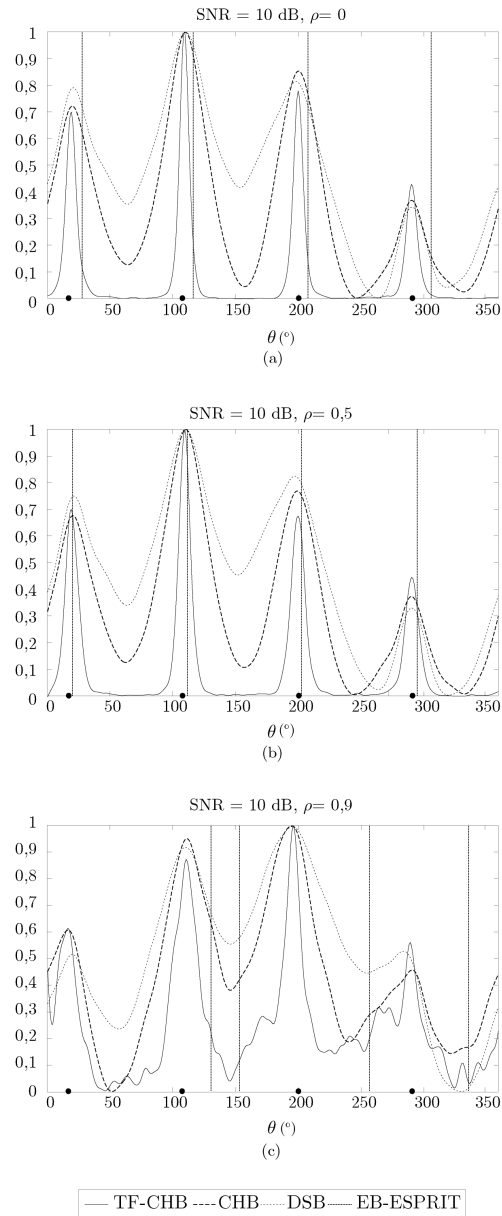


Figura 6.9: Histogramas normalizados para grabaciones simuladas con  $M = 4$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos. (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$  para SNR= 10 dB.

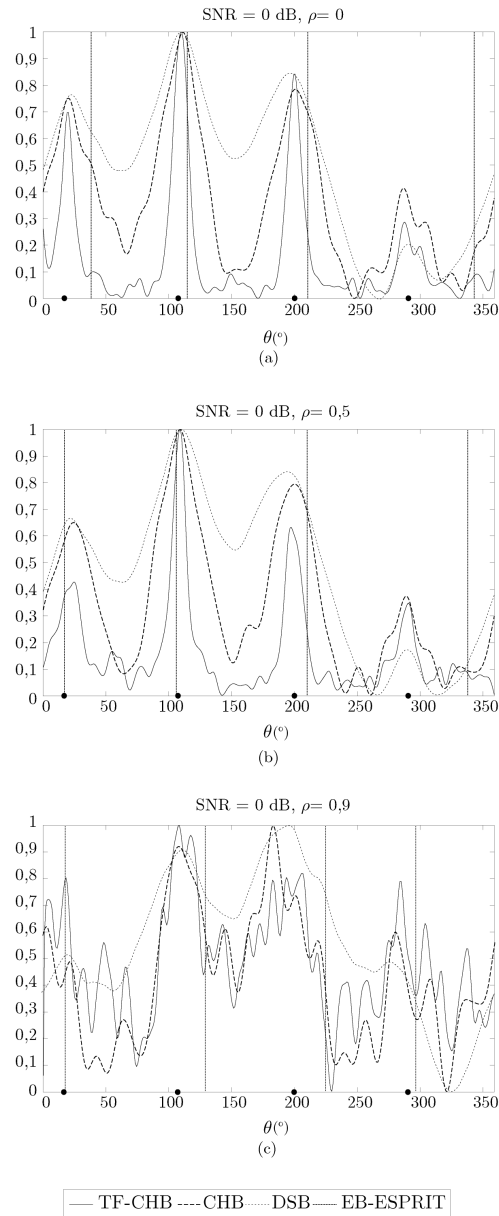


Figura 6.10: Histogramas normalizados para grabaciones simuladas con  $M=4$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos. (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$  para SNR= 0 dB.

$M = 4$				
SNR = $\infty$				
	TF-CHB	CHB	DSB	EB-ESPRIT
$\rho = 0,0$	0,00	0,00	2,06	1,42
$\rho = 0,5$	1,00	1,00	2,24	1,82
$\rho = 0,9$	3,74	5,20	4,03	43,30

$M = 8$				
SNR = $\infty$				
	TF-CHB	CHB	DSB	EB-ESPRIT
$\rho = 0,0$	0,00	5,01	4,24	23,80
$\rho = 0,5$	1,00	6,33	7,94	30,71
$\rho = 0,9$	8,61	-	-	50,40

Tabla 6.2: RMSE para diferentes métodos de localización utilizando UCAs para SNR=Infinito.

$M = 4$				
SNR = 10 dB				
	TF-CHB	CHB	DSB	EB-ESPRIT
$\rho = 0,0$	1,00	0,00	2,24	2,54
$\rho = 0,5$	1,00	1,00	3,16	4,32
$\rho = 0,9$	6,48	8,79	5,83	43,39

$M = 8$				
SNR = 10 dB				
	TF-CHB	CHB	DSB	EB-ESPRIT
$\rho = 0,0$	1,41	6,09	5,91	31,13
$\rho = 0,5$	2,45	11,36	9,01	35,38
$\rho = 0,9$	14,56	-	-	52,82

Tabla 6.3: RMSE para diferentes métodos de localización utilizando UCAs para SNR=10 dB.

$M = 4$				
SNR = 0 dB				
	TF-CHB	CHB	DSB	EB-ESPRIT
$\rho = 0,0$	3,35	2,06	4,12	6,95
$\rho = 0,5$	3,35	2,24	5,50	21,98
$\rho = 0,9$	9,91	14,08	14,76	53,52

$M = 8$				
SNR = 0 dB				
	TF-CHB	CHB	DSB	EB-ESPRIT
$\rho = 0,0$	6,04	13,21	12,89	65,27
$\rho = 0,5$	8,25	25,31	-	73,08
$\rho = 0,9$	26,08	-	-	73,59

Tabla 6.4: RMSE para diferentes métodos de localización utilizando UCAs para SNR=0 dB.

y cuyo tiempo de reverberación aproximado fue de  $T_{60} = 0,2$  s. Un UCA con  $N = 8$  micrófonos y radio  $R = 10$  cm fue el que se utilizó para capturar las señales habladas provenientes de los tres hablantes localizados en diferentes posiciones, tal y como muestra la Figura 6.11.

La frecuencia de muestreo de las señales originales era  $f_s = 16$  kHz, sin embargo, éstas fueron remuestreadas a 8 kHz con el fin de poder trabajar con los mismos parámetros de procesamiento que se han descrito en la Sección 6.3.1. El RMSE para cada conversación simultánea se presenta en la Tabla 6.5. Es relevante comentar que las fuentes tienden a cambiar de forma ligera la posición de la cabeza mientras hablan al ser personas reales y al contrario de cuando se utilizan altavoces como fuentes. De esta forma y debido a estos ligeros cambios, el RMSE obtenido no sólo refleja el método de localización sino también es consecuencia del efecto secundario de la hipótesis de aplicación real.

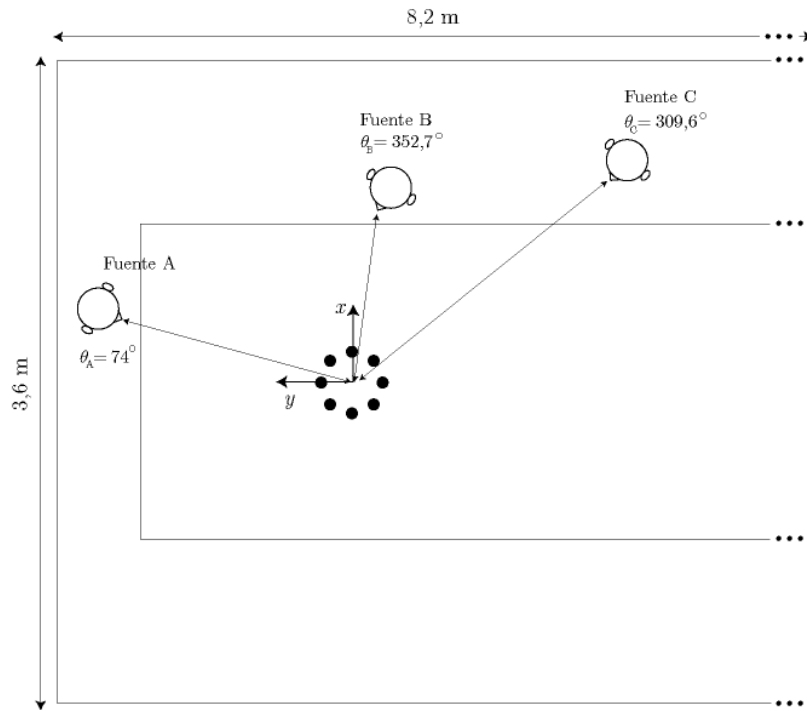


Figura 6.11: Sala real con dimensiones de  $3,6\text{ m} \times 8,2\text{ m} \times 3\text{ m}$  y factor de reflexión de pared  $\rho$ . Las fuentes y el array, están localizados en el plano  $xy$ .

	$M = 1$	$M = 2$	$M = 3$
	Fuente B	Fuentes B, C	Fuentes A, B, C
RMSE	1,18	2,0	2,98

Tabla 6.5: RMSE para grabaciones reales

## 6.4. **Discusión**

La evaluación del proceso llevado a cabo en la última sección muestra cómo la precisión en la localización con el método propuesto depende del diseño del array, de las condiciones acústicas y de la distribución de las fuentes. Con respecto al diseño del array, hay que tener en cuenta dos factores importantes: el espaciado entre micrófonos y su número. Mientras que el primer factor limita la frecuencia máxima de trabajo (frecuencia de *aliasing*), el segundo determina la robustez del método bajo condiciones acústicas adversas. Básicamente, la precisión en la localización es posible en la mayoría de situaciones prácticas siempre y cuando el número de micrófonos utilizados sea suficientemente alto. El número óptimo de micrófonos depende del escenario en el que se quiera aplicar el método de localización: el número de fuentes simultáneas, la precisión requerida y necesaria, el ruido y/o características de reverberación de la sala. Por ejemplo, se ha comprobado que con un array de 8 micrófonos utilizados en los experimentos con grabaciones reales, los resultados obtenidos han sido muy buenos en un ambiente común de reuniones.

Es también importante hacer constar que la presente aproximación puede utilizarse junto a otras técnicas de modelado basado en mezcla de distribuciones. Como se ha explicado en la Sección 6.3, aunque en este trabajo sólo se ha utilizado una técnica de detección de picos para la estimación de los DOA de las fuentes, también pueden aplicarse métodos más sofisticados tal como el modelado de mezcla Gaussiana, GMM [89] o el modelado de mezcla Laplaciana, LMM [90] para incrementar la robustez del método cuando en el histograma aparecen picos no fácilmente distinguibles.



# Capítulo 7

## Conclusiones, aportaciones y líneas futuras

---

7.1. Conclusiones generales de la tesis . . . . .	168
7.2. Contribuciones de la tesis . . . . .	170
7.3. Trabajo futuro . . . . .	172
7.4. Listado de publicaciones . . . . .	173

---

*Tras la presentación de los métodos propuestos y los resultados obtenidos, este capítulo final aborda las conclusiones extraídas de la valoración de los puntos anteriores. También se realiza un breve repaso de las aportaciones generadas por esta investigación, así como las líneas futuras de actuación derivadas de las conclusiones expuestas.*



## 7.1. Conclusiones generales de la tesis

El análisis y evaluación de la acústica del campo sonoro depende además de la geometría de la sala en cuestión, de la formalización de una serie de pautas y mecanismos que permiten describir su comportamiento acústico de forma completa. La captura de las respuestas al impulso, permite dibujar una huella acústica que necesariamente ha de estar sustentada en técnicas más complejas para poder obtener espacialmente y de forma completa todos los aspectos geométricos, sonoros y espaciales de las salas. La motivación de esta tesis viene precisamente derivada de la necesidad de un estudio exhaustivo experimental en la acústica de salas por medio de técnicas basadas en arrays circulares.

Como se introdujo en el Capítulo 1, en esta tesis se ha realizado un profundo estudio experimental sobre análisis de la acústica de salas, utilizando técnicas de descomposición de ondas planas y *beamforming* modal mediante arrays circulares de micrófonos, pudiendo caracterizar con más detalle las reflexiones de las salas individualmente, lo cuál no es posible con métodos que utilizan una única muestra microfónica.

Los Capítulos 2 y 3 han descrito la base matemática de ambos métodos, que aunque diferentes en planteamiento, han demostrado poder aplicarse de manera similar al análisis de salas.

Por un lado, en el Capítulo 2 se ha comprobado cómo la técnica PWD facilita la identificación de frente de ondas en campos de onda complejos mediante las integrales de Kirchhoff-Helmholtz y Rayleigh. Así mismo, se han introducido los armónicos cilíndricos como un método de análisis de gran alcance dentro de aquellos orientados a la identificación de campos de onda. Su aplicación permite un estudio acústico cuantitativo de las propiedades direccionales del campo de onda en cuestión. Además, los armónicos cilíndricos permiten una descomposición total del campo de onda en subcampos que pueden ser similares a los límites reflectantes principales del recinto. El análisis de las medidas con arrays muestran que en situaciones prácticas, las difracciones con los contornos o bordes, juegan un papel importante. El estudio de estos fenómenos de difracción, pueden indicar

cómo perfeccionar los actuales algoritmos de modelización numérica. Se ha comprobado también que en el caso de utilizar en las capturas de RIRs de arrays circulares compuestos por micrófonos cardioides, la descomposición de onda plana se simplifica. Este tipo de arrays permite eliminar restricciones implícitas y derivadas de las anteriores integrales obteniendo en sus resultados ecogramas completos y representaciones de descomposición de onda tanto en tiempo como en frecuencia que permiten además de distinguir las diferentes reflexiones de la sala, analizar otras características añadidas de dispersión de energía.

En el Capítulo 3 se ha comparado las dos técnicas más importantes de *beamforming* aplicadas a arrays circulares, la técnica DAS y CHB. Mediante una serie de simulaciones, se ha comprobado cómo el método CHB, el cuál está basado en la combinación selectiva de frecuencia de las señales de entrada en el dominio de armónicos circulares, confiere unas mejores resoluciones angulares y características de lóbulo lateral que el clásico método de *beamforming*. Además, el método CHB ha proporcionado un mejor criterio de selección de modos basados en la amplitud de las componentes de Bessel. Este criterio evita el uso de modos que se hacen cero y que conducen a unos filtros de enorme compensación lo cuál distorsionaría la respuesta del *beamformer*.

En el Capítulo 4 se ha comprobado tanto de forma simulada como experimental, una técnica de procesado de imagen para la detección de máximos en los ecogramas obtenidos y en las representaciones de onda plana después de aplicar el método PWD. Además, y bajo la misma morfología de imagen basada en un modelo cosenoidal (forma en la que aparecen las distintas reflexiones en los ecogramas), es posible aplicar esta técnica de detección sobre las representaciones de ondas planas tiempo-espacio obteniendo una buena precisión en su localización.

En el Capítulo 5 se ha llevado a cabo la medida de tres salas diferentes tanto en aplicación como en geometría y se ha puesto de manifiesto el buen funcionamiento de los métodos PWD y de localización de on-

das a partir de las medidas capturadas con un array circular con micrófonos cardioideos.

Por último, en el Capítulo 6 y motivados por el efecto de escasez (*sparsity*) dentro de una mezcla de señales acústicas, se ha propuesto un algoritmo de localización de varias fuentes acústicas simultáneas bajo condiciones acústicas adversas. Para ello, se ha aplicado un *beam-forming* de armónicos circulares con una regularización de Tikhonov a cada uno de los segmentos tiempo-frecuencia, donde el array se dirige hacia un conjunto de ángulos que cubren el plano de acimut. Aunque otras técnicas de localización trabajan en el dominio T-F, el método propuesto en esta tesis se apoya en la geometría de arrays circulares dentro del campo de trabajo del procesado modal. Varios experimentos relevantes con medidas simuladas y reales han conducido a una evaluación del comportamiento del método bajo diferentes condiciones acústicas y diferentes configuraciones de arrays. Los resultados han demostrado que pueden alcanzarse localizaciones precisas en la mayoría de situaciones prácticas experimentadas.

## 7.2. Contribuciones de la tesis

En el punto anterior se ha dado una visión general del trabajo realizado en esta tesis y de sus conclusiones generales, las cuales se pueden particularizar en las siguientes contribuciones al campo de estudio de la acústica de salas:

- Mediante múltiples medidas de salas se ha validado el método de análisis propuesto que hace uso de arrays circulares de micrófonos y su posterior descomposición en ondas planas. Aunque en trabajos previos de otros autores ya se ha tratado este tema, en esta tesis se ha profundizado y analizado el suficiente aparato experimental como para contribuir en el establecimiento y validación de este método como una herramienta sólida de análisis.
- Se ha realizado una contribución al estudio de la acústica de salas presentando una estrategia completa de análisis que incluye

una serie de herramientas para la localización de las reflexiones en las salas e identificación con su geometría. En concreto estas herramientas permiten identificar en los ecogramas y descomposiciones de ondas las reflexiones de la sala más significativas desarrollando un método específico de búsqueda de máximos locales basado en técnicas de morfología de imagen. Este algoritmo de detección se ha basado específicamente en una máscara cosenoidal que, junto a mecanismos de filtrado y umbralización, ha permitido alcanzar con buena precisión los puntos máximos de esas reflexiones con independencia de la aplicación, tanto sobre ecogramas de salas como en representaciones de onda plana.

- Se ha implementado un algoritmo alternativo de análisis de las capturas basado en técnicas de *beamforming* modal para arrays circulares. Este tipo de *beamforming* está basado en la descomposición del campo sonoro en series de armónicos y se ha comprobado con respecto a métodos clásicos de *beamforming* como el *Delay and Sum*, que además de proporcionar patrones más estrechos consigue mejores resoluciones angulares y mejores características de lóbulo lateral que el clásico método de *beamforming*.
- Se ha realizado el estudio del método de PWD y de localización de ondas a través del análisis de numerosas medidas en diferentes salas, destacando las ventajas e inconvenientes de cada uno de ellos e indicando qué método es más conveniente según la aplicación o el parámetro a analizar.
- Finalmente, como objetivo adicional de valor añadido en el análisis de la acústica de salas, también se ha aplicado el análisis modal para la localización de fuentes acústicas bajo condiciones adversas de las salas, es decir, con alto coeficiente de reverberación, y baja relación señal a ruido.

### 7.3. Trabajo futuro

La diversidad de trabajos en el procesado del campo acústico, así como la innovación de algoritmos y herramientas en la localización de fuentes sonoras, dan una idea de la complejidad que conlleva su estudio y de la imposibilidad de abarcar en un solo trabajo de tesis doctoral todos los aspectos posibles y derivados de la misma. Por ello, todos los métodos y técnicas propuestas en esta tesis, constituyen tan sólo una aportación que permitirá desarrollar un marco de conocimiento que mejore significativamente el procesado acústico de salas. A continuación se presentan algunas líneas futuras de trabajo que han surgido en el desarrollo de la investigación en esta tesis:

- Modificar la acústica de las salas de forma activa, mediante la utilización eficiente de altavoces y sistemas de síntesis de ondas planas como por ejemplo WFS y con altavoces planos de modos distribuidos (DML). El uso de un gran número de altavoces dinámicos tradicionales implica tener diversas complicaciones estéticas. Por esta razón se emplearía una nueva generación de transductores conocida como altavoces de modos distribuidos que presentan forma de panel y por ello disponen de enormes posibilidades de variación en cuanto a forma y tamaño, además de su integración estética.
- Realizar una comparativa entre el comportamiento de una sala real y auralizada definiendo una función de coste que permitiera minimizar el error entre las dos señales mediante la modificación de las respuestas al impulso generadas por los altavoces. El objetivo sería poder disponer de respuestas al impulso fieles a las capturadas en las salas de interés.
- Desarrollar nuevos modelos de arrays esféricos para poder extraer información de elevación y evaluar las capturas con métodos como los descritos en esta tesis.
- Implementar algoritmos de detección de ondas alternativos que mejoren la localización de reflexiones obtenidas en el análisis de salas.

- Desarrollar nuevos métodos de estimación de DOAs para la localización de fuentes, con algoritmos más sofisticados como *Gaussian Mixture Modelling*, o *Laplacian Mixture Modelling*.

Cada uno de los citados puntos anteriores constituye un nuevo trabajo con entidad suficiente para justificar una investigación más profunda, así como el inicio de nuevas tesis doctorales en el estudio del campo sonoro mediante herramientas de procesamiento avanzado.

## 7.4. Listado de publicaciones

Las principales aportaciones derivadas del trabajo de investigación presentado son las siguientes:

- A.M. Torres, J.J. López, B. Pueo, "Validation of Room Plane Wave Decomposition as a Tool for Spatial Ecogram Analysis of Rooms", 132th Audio Engineering Society Convention, 978-0-937803-86-8, Budapest, Hungría, Abril 2012.
- A.M. Torres, J.J. López, B. Pueo, "Experiments on room plane wave decomposition for virtual acoustics", Proceedings of EEA, 978-960-89639-1-7, Patrás, Grecia, Septiembre 2011.
- A.M. Torres, M. Cobos, B. Pueo, J.J. López, "Robust Acoustic Source localization Based on modal Beamforming and Time-Frequency Processing Using Circular Microphone Arrays", J. Acoust. Soc. Am., vol. 132, nº 3, pp. 1511-1520, September 2012.
- A.M. Torres, J.J. López, B. Pueo, M. Cobos "Room Acoustics Analysis Using Circular Arrays: An Experimental Study Based on Sound Field Plane-Wave Decomposition", J. Acoust. Soc. Am., November 2012, in review.



# Índice de Tablas

3.1. Características de tres arrays simulados donde el número de micrófonos es el mínimo requerido. . . . .	89
3.2. Características y parámetros introducidos en la simulación de una sala mediante el software ROOMSIM. . . . .	94
3.3. Diferencias de tiempo del método CHB. . . . .	96
3.4. Diferencias angulares del método CHB. . . . .	96
3.5. Posición de fuente y array en la sala real medida. . . . .	98
3.6. Diferencias de tiempo del método CHB. . . . .	101
3.7. Diferencias angulares del método CHB. . . . .	101
5.1. Relación de las distancias en las salas, donde H1 (Sala Reuniones, H2 (Paraninfo) y H3 (Auditorio). . . . .	124
5.2. Resultados para la Sala de Reuniones. . . . .	134
5.3. Resultados para la sala Paraninfo. . . . .	135
5.4. Resultados para la sala Auditorio. . . . .	135
5.5. Resultados globales obtenidos a partir de los resultados particulares de cada sala analizada. . . . .	136
6.1. RMSE para las diferentes configuraciones de array. . . . .	157
6.2. RMSE para diferentes métodos de localización utilizando UCAs para SNR=Infinito. . . . .	162
6.3. RMSE para diferentes métodos de localización utilizando UCAs para SNR=10 dB. . . . .	162
6.4. RMSE para diferentes métodos de localización utilizando UCAs para SNR=0 dB. . . . .	163
6.5. RMSE para grabaciones reales . . . . .	164





# Índice de Figuras

2.1. Geometría utilizada para el segundo teorema de Green: un volumen $V$ rodeado por una superficie $S$ . . . . .	14
2.2. Geometría utilizada para las integrales de Rayleigh y la función de Green. $R'$ es el punto espejo de $R$ en la superficie $S_0$ . . . . .	16
2.3. Síntesis del campo de onda por una distribución de monopolos en $z=z_s$ donde la velocidad en el plano es conocida. . . . .	17
2.4. Síntesis del campo de onda por una distribución de dipolos en $z=z_s$ donde la velocidad en el plano es conocida. . . . .	18
2.5. Geometría de un array circular. . . . .	22
2.6. Reconstrucción de una grabación circular utilizando las integrales de Kirchhoff-Helmholtz directa (a), inversa (b) y suma (c). . . . .	23
2.7. Transformada 2 D de Fourier de una onda plana simulada mediante un array circular con (b) micrófonos omnidireccionales, (c) de figura en ocho, (d) cardioides. Estas respuestas son comparadas con la descomposición en armónicos cilíndricos presentada en (a). . . . .	28
2.8. Diagrama del sistema de auralización, consistente en una parte de grabado, otra de procesado y una parte de reproducción. . . . .	31
2.9. Geometría de un array circular donde incide una onda plana. . . . .	32

2.10. Ejemplo simulado para tres ondas planas incidiendo en un array circular con direcciones $45^\circ$ , $245^\circ$ y $200^\circ$ . (a) Grabación del array circular. (b) Descomposición de onda plana frecuencia-espacio. (c) Descomposición de onda plana tiempo-espacio. . . . .	34
2.11. Descomposición en armónicos cilíndricos, donde el eje $y$ representa la frecuencia y el eje $x$ representa el número de armónicos. . . . .	38
2.12. Grabación de la presión sonora captada por el array. (a) Posición de la fuente en $45^\circ$ , (b) en $180^\circ$ , y (c) $300^\circ$ . . . . .	42
2.13. Expansión de armónicos cilíndricos para diferentes ángulos de acimut. . . . .	43
2.14. Descomposición de onda plana en frecuencia para diferentes ángulos de acimut. . . . .	44
2.15. Descomposición de onda plana en tiempo para diferentes ángulos de acimut. . . . .	45
2.16. Ecogramas para diferentes frecuencias de corte del ancho de banda de la respuesta impulsiva. . . . .	47
2.17. Expansión de armónicos cilíndricos para diferentes frecuencias de corte del ancho de banda de la respuesta impulsiva. . . . .	48
2.18. Descomposición de onda plana en frecuencia para diferentes frecuencias de corte del ancho de banda de la respuesta impulsiva. . . . .	49
2.19. Descomposición de onda plana en tiempo para diferentes frecuencias de corte del ancho de banda de la respuesta impulsiva. . . . .	50
2.20. Presiones captadas por el array para diferente número de micrófonos. . . . .	52
2.21. Expansión en armónicos cilíndricos para diferente número de micrófonos. . . . .	53
2.22. Descomposición de onda plana en frecuencia. . . . .	54
2.23. Descomposición de onda plana en tiempo. . . . .	55
2.24. Ecogramas obtenidos para diferentes tamaños de arrays. . . . .	57
2.25. Expansión en armónicos cilíndricos obtenidos para diferentes tamaños de arrays. . . . .	58

2.26. Descomposición de onda plana en frecuencia. . . . .	59
2.27. Descomposición de onda plana en tiempo. . . . .	60
2.28. Montaje experimental con las posiciones del altavoz con respecto al eje del array. . . . .	61
2.29. Detalle del montaje llevado a cabo, donde se aprecia el parábán que tendrá especial relevancia en la descomposición de ondas planas. . . . .	62
2.30. Vista de la barra sujeta al motor giratorio. . . . .	63
2.31. Descomposición de onda plana en la sala GTAC para medidas reales. . . . .	64
3.1. Convenio de representación en coordenadas esféricas. . .	71
3.2. Patrón de directividad para un array lineal con $f=1$ kHz, longitud=0,5 y número de elementos variable. . . .	72
3.3. Patrón de directividad para un array lineal con $f=1$ kHz, $N=5$ y longitud variable. . . . .	73
3.4. Patrón de directividad sin aliasing espacial (a) y con aliasing (b). . . . .	75
3.5. Cálculo de la resolución para un beampattern determinado. . . . .	76
3.6. Salida del beamformer con dos fuentes activas, cuya distancia angular es igual a la resolución, pudiéndose distinguir cada fuente correctamente. . . . .	76
3.7. Salida del beamformer con dos fuentes activas, cuya distancia angular es menor a la resolución, no pudiéndose distinguir en este caso, las fuentes correctamente. . .	77
3.8. Cálculo del NMLL para una salida de beamformer determinada. . . . .	78
3.9. Geometría de un UCA con $N$ elementos. . . . .	79
3.10. Magnitud de los cuatro órdenes más bajos de las funciones de Bessel de primer orden. . . . .	83
3.11. Beampatterns para banda ancha con apuntamiento en la dirección $\theta_s = \pi$ utilizando un UCA con $N = 13$ y $R = 0,12$ m. (a) Delay and sum (DAS) (b) CHB beamforming regularizado. . . . .	87

3.12. Beampatterns para banda ancha con apuntamiento en la dirección $\theta_s = \pi$ utilizando un UCA con $N = 13$ y $R = 0,12$ m. (c) CHB beamforming sin regularizar, (d) Sección transversal para $f = 1000$ Hz. . . . .	88
3.13. Salidas normalizadas de dos beamformers para las características del Array 1 de la Tabla 3.1 mediante el método DAS (a), y el método CHB (b). . . . .	90
3.14. Salidas normalizadas de dos beamformers para las características del Array 2 de la Tabla 3.1 mediante el método DAS (a), y el método CHB (b). . . . .	91
3.15. Salidas normalizadas de dos beamformers para las características del Array 3 de la Tabla 3.1 mediante el método DAS (a), y el método CHB (b). . . . .	92
3.16. Sala simulada donde tanto la fuente como el array están localizados en el plano $xy$ . . . . .	95
3.17. Ecogramas obtenidos por el método CHB y de forma directa. . . . .	97
3.18. (a) Sala de usos múltiples utilizada para las medidas. (b) Ubicación de fuente y array en la misma. . . . .	99
3.19. Ecogramas obtenidos de forma directa y mediante el método CHB. . . . .	100
4.1. Ecograma obtenido para el ejemplo simulado y a partir del cuál se realizará el algoritmo de detección de onda. .	106
4.2. Imagen umbralizada después de haber filtrado el ecograma inicial por un filtro Gaussiano. . . . .	108
4.3. Máscara particularizada a los parámetros del ejemplo simulado y que se utilizará para realizar la correlación cruzada. . . . .	109
4.4. Resultado obtenido después de haber correlado la imagen umbralizada con una máscara cosenoidal. . . . .	111
4.5. Representación de las regiones que se han seleccionado en el ejemplo simulado. . . . .	113
4.6. Detección de picos de aquellas reflexiones más representativas después de haber aplicado el algoritmo de detección de ondas. . . . .	114

5.1. (a) Fotografía de la Sala de Reuniones, (b) plano y ubicación respectivas de la fuente y del array de micrófonos en la sala. . . . .	120
5.2. (a) Fotografía de Paraninfo, (b) plano y ubicación de la fuente y del array de micrófonos en la sala. . . . .	122
5.3. (a) Fotografía de Auditorio, (b) plano y ubicación de la fuente y del array de micrófonos en la sala. . . . .	123
5.4. Detalle de la disposición de los micrófonos cardioides sobre el motor giratorio. . . . .	124
5.5. Ecograma de la Sala de Reuniones con las principales reflexiones detectadas mediante inspección visual (a) y mediante el algoritmo de detección de ondas (b). . . . .	126
5.6. Representaciones de PWD en tiempo-espacio con detección de ondas (a), y en frecuencia-espacio (b) para la Sala de Reuniones. . . . .	127
5.7. Ecograma de la sala Paraninfo con las principales reflexiones detectadas mediante inspección visual (a) y mediante el algoritmo de detección de ondas (b). . . . .	129
5.8. Representaciones de PWD en tiempo-espacio con detección de ondas (a), y en frecuencia-espacio (b) para el Paraninfo. . . . .	130
5.9. Ecograma de la sala Auditorio con las principales reflexiones detectadas mediante inspección visual (a) y mediante el algoritmo de detección de ondas (b). . . . .	131
5.10. Representaciones de PWD en tiempo-espacio con detección de ondas (a), y en frecuencia-espacio (b) para el Auditorio. . . . .	132
6.1. Sala simulada con dimensiones de $6\text{ m} \times 8\text{ m} \times 3\text{ m}$ y factor de reflexión en paredes $\rho$ . Las fuentes y el array, están localizados en el plano $xy$ . . . . .	148
6.2. Histogramas obtenidos para las simulaciones con $M = 4$ fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos para $\text{SNR} = \infty$ : (a) $\rho = 0$ , (b) $\rho = 0,5$ , (c) $\rho = 0,9$ . . . . .	150

- 6.3. Histogramas obtenidos para las simulaciones con  $M = 4$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos para SNR=10 dB: (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$ . . . . . 151
- 6.4. Histogramas obtenidos para las simulaciones con  $M = 4$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos para SNR=0 dB: (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$ . . . . . 152
- 6.5. Histogramas obtenidos para las simulaciones con  $M = 8$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos para SNR= Inf: (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$ . . . . . 153
- 6.6. Histogramas obtenidos para las simulaciones con  $M = 8$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos para SNR=10 dB: (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$ . . . . . 154
- 6.7. Histogramas obtenidos para las simulaciones con  $M = 8$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos para SNR=0 dB: (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$ . . . . . 155
- 6.8. Histogramas normalizados para grabaciones simuladas con  $M = 4$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos. (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$  para SNR=Inf. . . . . 159
- 6.9. Histogramas normalizados para grabaciones simuladas con  $M = 4$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos. (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$  para SNR= 10 dB. . . . . 160
- 6.10. Histogramas normalizados para grabaciones simuladas con  $M=4$  fuentes utilizando diferentes configuraciones de array en diversos ambientes acústicos. (a)  $\rho = 0$ , (b)  $\rho = 0,5$ , (c)  $\rho = 0,9$  para SNR= 0 dB. . . . . 161
- 6.11. Sala real con dimensiones de  $3,6\text{ m} \times 8,2\text{ m} \times 3\text{ m}$  y factor de reflexión de pared  $\rho$ . Las fuentes y el array, están localizados en el plano  $xy$ . . . . . 164

# Bibliografía

- [1] W. C. Sabine, *Collected Papers on Acoustics*, P. Publishing, Ed. Los Altos, U. S.: Trade Cloth, 1993, no. LCCN: 93-085708.
- [2] *Acoustics-Measurement of reverberation time of rooms with reference to other acoustical parameters*, ISO Std. 3382, 1997, International Organization for Standardization, ISO 3382:1997.
- [3] T. Peltonen, T. Lokki, B. Goutarbés, J. Merimaa, and M. Karjalainen, "A system for multi-channel and binaural room response measurements," in *Proceedings of the AES 110th Convention*, Amsterdam, The Netherlands, May 2001, Paper No. 5289.
- [4] H. Kuttruff, *Room acoustics*, S. Press, Ed. Abingdon, Oxford, UK: Taylor & Francis, October 2000, 368 pages.
- [5] S. Neely and J. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, vol. 66, pp. 165–169, 1979.
- [6] D. Rife and J. Vanderkoy, "Transfer-function measurement with maximum-length sequences," *J. Acoust. Soc. Am.*, vol. 37 (6), pp. 419–443, 1989.
- [7] A. Farina, "Simultaneous measurement of impulse response and distortion with a sweeps-sine technique," in *Proceedings of the AES 108th Convention*, Paris, France, 2000, Paper No. 5093.
- [8] C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (HRTF's): Representations of HRTF's in time, frequency, and space," *J. Acoust. Soc. Am.*, vol. 49, pp. 231–249, April 2001.



- 
- [9] C. P. Brown and R. O. Duda, "An efficient HRTF model for 3-D sound," in *Proceedings of the IEEE Workshop on ASSP*, New Paltz, NY, USA, 1997.
- [10] E. Hulsebos, "Improved microphone array configurations for auralization of sound fields by wave field synthesis," in *Proceedings of the AES 110th Convention*, The Netherlands, Amsterdam, May 2001, Paper No. 0779.
- [11] D. Vries and J. Baan, "Auralization of sound fields by wave field synthesis," in *Proceedings of the AES 106th Convention*, Munich, Germany, May 1999, Paper No. 4927.
- [12] A. Farina and R. Avalon, "Recording concert hall acoustics for posterity," in *Proceedings of the AES 24th International Conference on Multichannel Audio*, Industrial Engineering Department, University of Parma, Italy, 2005.
- [13] T. Heinz, *Modal array signal processing: principles and applications of acoustic wavefield decomposition*, M. Thoma and M. Morari, Eds. Springer, 2007, 255 pages.
- [14] E. Hulsebos, "Auralization using wave field synthesis," Ph.D. dissertation, Delft University of Technology, The Netherlands, October 2004.
- [15] A. Berkhout, *Applied Seismic Wave Theory*, Elsevier, Ed. Amsterdam: Science, 1987, 377 pages.
- [16] I. Bronstein, K. Semendjajew, G. Musiol, and H. Mühlig, *Taschenbusch der Mathematik*, A. Lager, Ed. Verlag Harri Deutsch, 1995.
- [17] E. Verheijen, "Sound reproduction by wave field synthesis," Ph.D. dissertation, Utrecht, 2010.
- [18] E. G. Williams, *Fourier Acoustics. Sound Radiation and Nearfield Acoustic Holography*, A. Press, Ed., 1999, 307 pages.

- 
- [19] P. Peretti, S. Cecchi, L. Romoli, and F. Piazza, *Numerical Analysis and Scientific Computing*. InTech Open, 2011, ch. Performance Evaluation of Adaptive Algorithms for Wave Field Analysis/Synthesis Using Sound Field Simulations, pp. 543–560.
- [20] E. Hulsebos, D. Vries, and E. Bourdillat, “Improved microphone array configurations for auralization of sound fields by wave field synthesis,” *J. Acoust. Soc. Am.*, vol. 50, pp. 779–790, October 2002.
- [21] J. Escolano, B. Pueo, S. Bleda, and J. López, “An approach to discrete-time modelling auralization for wave field synthesis applications,” *Proceedings of 118th AES Convention, Barcelona, Spain*, May, 2005.
- [22] P. Vogel, “Application of wave field synthesis in room acoustics,” Ph.D. dissertation, Delft University of Technology, The Netherlands, 1993.
- [23] S. Bleda, J. López, J. Escolano, and B. Pueo, “Design and implementation of a compatible wave field synthesis authoring tool,” in *Proceedings of the AES 118th Convention, Barcelona, Spain*, May 2005.
- [24] E. Bourdillat, “Auralization of sound fields in auditoria using wave field synthesis,” Master’s thesis, University of Technology, Delft, University of Technology, 2001.
- [25] E. Start, “Direct sound enhancement by wave field synthesis,” Ph.D. dissertation, Delft, University of Technology, The Netherlands, 1997.
- [26] B. Pueo, J. Escolano, and M. Cobos, “Control de la directividad en arrays lineales de altavoces,” in *Proceedings of the Simposium de la Union Científica Internacional de Radio, La Coruña, Spain*, September 2003.
- [27] D. Vries, L. Horchens, and P. Grond, “Extraction of 3D information from circular array measurements for auralization with wave field synthesis,” *Eurasip Journal on Advances in Signal Processing*, vol. 1, pp. 190–190, January 2007.

- [28] R. Kennedy, T. Abhayapala, and D. Ward, "Broadband nearfield beamforming using a radial beampattern transformation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 46, pp. 2147–2156, 1998.
- [29] J. Hald, *Handbook of Signal Processing in Acoustics*. Springer New York, 2008, 1889 pages.
- [30] J. Benesty, Y. Huang, and M. Sondhi, *Handbook of Speech Processing*. Springer, December, 2007, 1175 pages.
- [31] D. Martínez, "Técnicas de procesado en array para mejora de la calidad de voz," Ph.D. dissertation, Universidad Politécnica de Valencia, 2008.
- [32] McCowan, "Microphone arrays : A tutorial," Idiap Institute, Tech. Rep., April 2005.
- [33] A. Ramos, S. Holm, S. Gudvangen, and R. Otterlei, "Delay-and-sum beamforming for direction of arrival estimation applied to gunshot acoustics," in *Proceedings of the SPIE 8019, Sensors, and Command, Control, Communications and Intelligence (C31)*, vol. 80190U, May 2011, p. 10.1117/12.886833.
- [34] E. Tiana-Roig, "Beamforming techniques for environmental noise," Ph.D. dissertation, Technical University of Denmark, 2009.
- [35] H. Teutsch and W. Kellermann, "Eigen-beam processing for direction-of-arrival estimation using spherical apertures," *Multimedia Communications and Signal Processing University of Erlangen-Nuremberg*, Tech. Rep., 2009.
- [36] M. Seltzer, "Microphone array processing fo robust speech recognition," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, July 2003.
- [37] J. Jourdan and M. Marschall, "Comparison of beamforming techniques on a sphere," *Acoustic Technology-DTU Elektro*, Tech. Rep., October 2008.

- 
- [38] J. Christensen and J. Hald, "Technical review," Brüel & Kjaer, Tech. Rep., 2004.
- [39] D. E.Ñ. Davies, *Circular Arrays*, A. W. Rudge, K. Milne, A. D. Olver, and P. Knight, Eds. London: Peter Peregrinus Ltd., 1983, vol. 2.
- [40] H. Teutsch and W. Kellermann, "Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays," *J. Acoust. Soc. Am.*, vol. 120, no. 5, pp. 2724–2736, 2006.
- [41] E. Tiana-Roig, F. Jacobsen, and E. Fernandez Grande, "Beamforming with a circular microphone array for localization of environmental noise sources," *J. Acoust. Soc. Am.*, vol. 128, no. 6, pp. 3535–3542, December 2010.
- [42] A. Parthy, N. Epain, A. van Schaik, and T. J. Craig, "Comparison of the measured and theoretical performance of a broadband circular microphone array," *J. Acoust. Soc. Am.*, vol. 130, no. 6, pp. 3827–3837, December 2011.
- [43] H. Teutsch, "Wavefield decomposition using microphone arrays and its application to acoustic scene analysis," Ph.D. dissertation, Friedrich-alexander Universität Erlangen-Nürnberg, 2006.
- [44] G. Elko and J. Meyer, "Microphone arrays," in *Springer Handbook of Speech Processing*, J. Benesty, M. Sondhi, and Y. Huang, Eds. Berlin, Germany: Springer-Verlag, 2008, ch. 50, pp. 1021–1042.
- [45] D. R. Campbell, "Roomsim: a MATLAB simulation shoebox room acoustics," 2007, <http://media.paisley.ac.uk/campbell/Roomsim>.
- [46] M. Hodgson, "Evidence of diffuse surface reflections in rooms," *J. Acoust. Soc. Am.*, vol. 89, pp. 765–761, 1990.
- [47] J. Usher, "An improved method to determine the onset timings of reflections in an acoustic impulse response," *J. Acoust. Soc. Am.*, vol. 127, pp. 172–177, April 2010.

- [48] S. Vesa and T. Lokki, "Detection of room reflections from a binaural room impulse response," in *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx06)*, Montreal, Canada, September 2006, pp. 215–220.
- [49] X. Falourd, L. Rohr, M. Rossi, and H. Lissek, "Spatial echogram analysis of a small auditorium with observations on the dispersion of early reflections," in *Proceedings of the 39th International Congress and Exposition on Noise Control Engineering (INTER-NOISE)*, Lisboa, Portugal, 2010.
- [50] S. Tervo, T. Korhonen, and T. Lokki, "Estimation of reflections from impulse responses," in *Proceedings of the International Symposium on Room Acoustics (ISRA)*, 2010.
- [51] A. M. Torres, M. Cobos, B. Pueo, and J. Lopez, "Room acoustics analysis using circular arrays: an experimental study based on sound field plane-wave decomposition," *J. Acoust. Soc. Am.*, vol. Under review, p. Under review, Sending November 2012.
- [52] R. Gonzalez and R. E. Woods, *Digital Image Processing*. Boston, USA: Addison-Wesley Longman Publishing Co., 2001, 976 pages.
- [53] L. Tronchin and A. Farina, "Room acoustics measurement and auralization," in *International Advanced Course on Musical Acoustics (IACMA)*, Bologna, Italy, July 2005.
- [54] T. D. Kite, "Techniques in audio and acoustic measurement," *J. Acoust. Soc. Am.*, vol. 114, no. 4, pp. 2381–2381, 2003.
- [55] C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *J. Acoust. Soc. Am.*, vol. 116, pp. 3075–3089, 2004.
- [56] C. Liu, B. C. Wheeler, R. C. Bilger, C. R. Lansing, and A. S. Feng, "Localization of multiple sound sources with two microphones," *J. Acoust. Soc. Am.*, vol. 108, no. 4, pp. 1888–1905, 2000.
- [57] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays: Signal*

- Processing Techniques and Applications*, M. S. Brandstein and D. Ward, Eds. Berlin, Germany: Springer-Verlag, 2001, ch. 8, pp. 157–180.
- [58] J.-M. Valin, F. Michaud, and J. Rouat, “Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering,” *Robotics and Autonomous Systems Journal*, vol. 55, no. 3, pp. 216–228, 2007.
- [59] N. Madhu and R. Martin, *Advances in Digital Speech Transmission*. New York, NY, USA: Wiley, 2008, ch. Acoustic Source Localization with Microphone Arrays, pp. 135–166.
- [60] S. Haohai, H. Teutsch, E. Mabande, and W. Kellermann, “Robust localization of multiple sources in reverberant environments using EB-ESPRIT with spherical microphone arrays,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 22–27 May 2011, pp. 117–120.
- [61] M. Cobos, A. Marti, and J. J. Lopez, “A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling,” *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 71–74, January 2011.
- [62] S. Araki, H. Sawada, R. Mukai, and S. Makino, “DOA estimation for multiple sparse sources with arbitrarily arranged multiple sensors,” *J. Adv. Sig. Proc. Syst.*, vol. 63, no. 3, pp. 265–275, 2009.
- [63] B. Rafaely, “Plane-wave decomposition of the sound field on a sphere by spherical convolution,” *J. Acoust. Soc. Am.*, vol. 116, no. 4, pp. 2149–2157, October 2004.
- [64] H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*. Berlin, Germany: Springer, 2007.
- [65] M. Cobos and J. J. Lopez, “Two-microphone separation of multiple speakers based on interclass variance maximization,” *J. Acoust. Soc. Am.*, vol. 127, pp. 1661–1673, 2010.

- [66] W. Zhang and B. D. Rao, "A two microphone-based approach for source localization of multiple speech sources," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 8, pp. 1913–1928, November 2010.
- [67] S. Mohan, M. E. Lockwood, M. L. Kramer, and D. L. Jones, "Localization of multiple acoustic sources with small arrays using a coherence test," *J. Acoust. Soc. Am.*, vol. 123, no. 4, pp. 2136–2147, 2008.
- [68] J. Escolano, "Contributions to discrete-time domain methods in room acoustic simulations," Ph.D. dissertation, Universidad Politécnica de Valencia, 2008.
- [69] Y. Rui, D. Florêncio, W. Lam, and J. Su, "Sound source localization for circular arrays of directional microphones," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 3, 2005, pp. iii/93–iii/96.
- [70] J. J. Burred and T. Sikora, "On the use of auditory representations for sparsity-based sound source separation," in *Proceedings of the 5th International Conference on Information, Communications and Signal Processing (ICICS 2005)*, Bangkok, Thailand, December 2005.
- [71] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Performance evaluation of sparse source separation and DOA estimation with observation vector clustering in reverberant environments," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, France, 2006.
- [72] S. Rickard and F. Dietrich, "DOA estimation of many w-disjoint orthogonal sources from two mixtures using DUET," in *Proceedings of the 10th IEEE Workshop on Statistical Signal and Array Processing (SSAP2000)*, Pocono Manor, PA, August 2000, pp. 311–314.
- [73] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, July 2004.

- 
- [74] P. Bofill and M. Zibulevski, "Underdetermined blind source separation using sparse representations," *J. Adv. Sig. Proc.*, vol. 81, pp. 2353–2362, 2001.
- [75] S. Rickard and O. Yilmaz, "On the w-disjoint orthogonality of speech," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, Florida, USA, May 2002, pp. 529–532.
- [76] A. Jourjine, S. Richard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: Demixing  $n$  sources from 2 mixtures," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 5, Turkey, 2000, pp. 2985–2988.
- [77] H. Asari, B. A. Pearlmutter, and A. M. Zador, "Sparse representations for the cocktail party problem," *Journal of Neuroscience*, vol. 26, pp. 7477–7490, 2006.
- [78] J. J. Burred, "From sparse models to timbre learning: New methods for musical source separation," Ph.D. dissertation, Technical University of Berlin, 2008.
- [79] A. M. Torres, M. Cobos, B. Pueo, and J. Lopez, "Robust acoustic source localization based on modal beamforming and time-frequency processing using circular microphone arrays," *J. Acoust. Soc. Am.*, vol. 132 (3), pp. 1511–1520, September 2012.
- [80] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.
- [81] H. Teutsch and W. Kellermann, "Eb-esprit: 2d localization of multiple wideband acoustic sources using eigenbeams," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 3, Philadelphia, Pa, USA, 2005, pp. 89–92.



- [82] E. Vincent, S. Araki, and P. Bofill, "The 2008 signal separation evaluation campaign: A community-based approach to large-scale evaluation," *Lecture Notes in Computer Science. Independent Component Analysis and Signal Separation*, vol. 5441/2009, pp. 734–741, 2009.
- [83] J. Skowronek, A. Raake, K. Hoeldtke, and M. Geier, "Speech recordings for systematic assessment of multi-party conferencing," in *Proceedings of Forum Acusticum 2011*. Aalborg, Denmark: European Acoustics Association, June 2011, pp. 111–116.
- [84] H. Krim and M. Viberg, "Two decades of array signal processing research, the parametric approach," *IEEE Signal Processing Magazine*, pp. 67–94, 1996.
- [85] G. Lathoud, J.-M. Odobez, and D. Gatica-Perez, "AV16.3: an audio-visual corpus for speaker localization and tracking," in *Proceedings of First International Workshop on Machine Learning for Multimodal Interaction (MLMI 2004)*, Martigny, Switzerland, 21-23 June 2004, pp. 192–195.
- [86] G. Lathoud and M. Magimai-Doss, "A sector-based, frequency-domain approach to detection and localization of multiple speakers," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 3, Philadelphia, PA, USA, 18-23 March 2005, pp. 265–268.
- [87] A. Plinge, M. H. Hennecke, and G. A. Fink, "Robust neuro-fuzzy speaker localization using a circular microphone array," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, August 2010.
- [88] D. C. Moore, "The IDIAP smart meeting room," IDIAP, Tech. Rep. IDIAP-COM 02-07, 2002, available at <http://glat.info/ma/av16.3/com02-07.pdf>.
- [89] N. Madhu and R. Martin, "A scalable framework for multiple speaker localization and tracking," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, WA, USA., 14-17 September 2008.

- 
- [90] M. Cobos, J. J. Lopez, and D. Martinez, "Two-microphone multi-speaker localization based on a laplacian mixture model," *Digital Signal Processing*, vol. 21, pp. 66–76, 2011.



