

Grado en Ingeniería de Sonido e Imagen

2020-2021

Trabajo Fin de Grado

“Simulación y propuesta de un sistema de sonorización espacial para el Teatro Buero Vallejo”

Óscar Campuzano Demetrio

Tutor: Luis Antonio Azpicueta-Ruiz

Leganés (Madrid), 2021

AGRADECIMIENTOS

Me gustaría dar las gracias a todas aquellas personas que han hecho posible este proyecto. En especial a Luis A. Azpicueta-Ruiz, por darme la oportunidad de realizarlo, guiarme en el proceso y ofrecerme sus conocimientos y sugerencias. También darle las gracias por sus clases, que sin duda me han aportado mucho.

Gracias a los responsables del Teatro Buero Vallejo de Guadalajara por permitirnos llevar a cabo las mediciones. Por supuesto también a Antonio Pedrero González, por permitirme participar en ellas y por enseñarnos su metodología. Agradecer también a Alejandro y Javier su colaboración, así como a Daniel de la Prida Caballero por ayudarnos el día de las mediciones, por esas prácticas en el laboratorio de acústica, y por estar siempre disponible para resolver dudas de manera activa.

Gracias a mis compañeros de carrera, en especial a Álex, Sergio y Dani, por acompañarme estos años y estar siempre ahí para echar una mano. Os deseo suerte.

Por último y no menos importante, gracias a mi familia por darme su apoyo y animarme a estudiar una carrera, y a Mariaje por todos estos años juntos y los que quedan.

RESUME

In this project a tridimensional audio playback system is proposed for the auditorium of the Teatro Buero Vallejo, located in Guadalajara, using the Ambisonics spatial sonorization technology based on scenes, which is the most widely used nowadays.

For that purpose, at the beginning of this document, an investigation into some of the traditional spatial sonorization technologies is made. Subsequently, this project performs a deep research about Ambisonics, covering from its operating principle and possible formats, to its codification and decodification processes. It also studies the most common sound field capture techniques, as well as some signal processing for manipulating the spatial content and the pros and cons of this technology over other traditional sound systems.

The motivation of implementing this technology is its ability for recreating any real or simulated sound field around the listener, offering the possibility to perform real time signal processing for a manipulation of the spatial content, with a low computational cost. This technology also allows a non-strict speaker distribution, which is an advantage that facilitate dealing with geometrical limitations when designing the speaker placement in rooms with irregular forms.

In summary, Ambisonics offers an immersive experience that may be interesting for some kinds of events in the auditorium, such as theatrical works or shows with live music that may find useful introducing tridimensional audio effects in their performances.

For evaluating the proposed speaker system, it is necessary to perform computational simulations before installing it. For that purpose, in this project a tridimensional model of the auditorium is built using the software EASE (Enhanced Acoustic Simulator for Engineers) from the AFMG company. Firstly, the auditorium's geometry is built for the model, and then the materials are assigned to its faces.

Since this model contains approximations due to the complexity of the actual geometry of the theater and the uncertainty of the acoustic properties of the materials, it is required to calibrate it: after taking measurements *in situ* that enable us to characterize the acoustic of the auditorium, the absorption and scattering coefficients of the assigned materials are adjusted until the simulations performed in EASE and the taken measurements are similar. Once the error is considered small enough, it is deemed that the model is finished and serves as a valid replica of the actual auditorium, so the sonorization system can be evaluated on it.

Taking into account the dimensions of the auditorium and the previous investigation into Ambisonics technology, this project proposes an adapted audio system for third order ambisonic signals, with a speaker distribution suitable for the room. Since a decodification of the ambisonic signals is also needed for its playback through the system, this project also offers an Ambisonics decoder programmed on Matlab. This decoder is scalable in order (up to three) and number of speakers, as well as in the speaker and sweet spot positions. This enables the possibility of implementing, with this same decoder, alternative speaker distributions that the one proposed in this project, as well as designing other Ambisonic systems in different situations aside the chosen auditorium for this project.

Once the sonorization system, the test signals and the decodification are finished, a set of auralizations are made in EASE. Playing these audio files through headphones enables us to perceive what a listener at a certain point of the theater would approximately hear if the decoded signal test were played through the speakers. In this way, it is possible to evaluate the proposed sonorization system in approximate terms. Finally, this project evaluates the sonorization and clarity estimated for the proposed system.

At the end of the document, the regulatory framework, the estimated budget and the social and economic impact of this project on the theater and the city are exposed.

ÍNDICE

1.	Introduction and motivations	1
2.	Tecnologías de sonorización espacial	3
2.1.	Historic context	3
2.2.	Psychoacoustic mechanisms for localization of sound sources.....	4
2.3.	Sistemas estereofónicos	5
2.3.1.	Principio de funcionamiento.....	5
2.3.2.	Técnicas de captación estereofónica	5
2.3.3.	Localización	12
2.4.	Sistemas <i>surround</i>	13
2.4.1.	Principio de funcionamiento.....	13
2.4.2.	Localización	14
2.5.	VBAP.....	15
2.5.1.	Principio de funcionamiento.....	15
2.5.2.	Localización	16
3.	<i>Ambisonics</i>	17
3.1.	Principio de funcionamiento.....	17
3.2.	Formatos principales.....	21
3.2.1.	Formato A	21
3.2.2.	Formato B.....	22
3.2.3.	Formato C (UHJ)	22
3.3.	Técnicas de captación del campo sonoro	22
3.3.1.	FOA 2D	23
3.3.2.	FOA 3D	25
3.3.3.	HOA	28
3.4.	Codificación.....	29
3.5.	Procesamiento	29
3.5.1.	Efecto espejo.....	29
3.5.2.	Rotación	30
3.5.3.	Ecuilización y compresión	30
3.6.	Decodificación y reproducción	31

3.6.1.	Reproducción en 2D.....	31
3.6.2.	Reproducción en 3D.....	35
4.	Construcción del modelo 3D.....	37
4.1.	Contexto del Teatro Buero Vallejo.....	37
4.2.	Descripción del auditorio.....	38
4.2.1.	Patio de butacas.....	39
4.2.2.	Anfiteatro.....	39
4.2.3.	Foso de orquesta.....	39
4.2.4.	Escenario.....	39
4.3.	Proceso de construcción.....	40
4.3.1.	Lectura de los planos en AutoCAD.....	40
4.3.2.	Construcción en EASE Room Editor.....	41
4.4.	Geometría final del modelo.....	42
4.5.	Áreas de audiencia.....	43
4.6.	Visualización en EASE Vision.....	44
5.	Medición <i>in situ</i> y caracterización del recinto.....	47
5.1.	Respuesta al impulso, $h(t)$	47
5.1.1.	Método de la señal impulsiva.....	48
5.1.2.	Método MLS.....	48
5.1.3.	Método del doble canal FFT.....	49
5.1.4.	Método del barrido tonal.....	49
5.2.	Tiempo de reverberación, RT.....	50
5.2.1.	Método del ruido interrumpido.....	51
5.2.2.	Método de la respuesta al impulso integrada.....	52
5.3.	Otros parámetros acústicos.....	53
5.3.1.	<i>Early Decay Time</i> , EDT.....	53
5.3.2.	Claridad, $Ct0$	54
5.3.3.	Tiempo central, T_s	54
5.4.	Situación del teatro.....	55
5.5.	Equipo y configuración.....	55
5.6.	Muestreo espacial.....	58
5.7.	Procedimiento e incidencias.....	60

5.8.	Postprocesado en Room EQ Wizard	62
5.9.	Resultados.....	64
6.	Calibración del modelo	67
6.1.	Procedimiento.....	67
6.2.	Diseño de materiales	68
6.3.	Comparación con los resultados de la medición in situ.....	72
7.	Sistema de reproducción <i>Ambisonics</i>	79
7.1.	Distribución de altavoces.....	79
7.2.	Decodificador programado en Matlab.....	81
7.2.1.	Funcionamiento del decodificador	81
7.2.2.	Creación de señales de test en Nuendo.....	83
7.3.	Auralización en EASE <i>Ears</i>	87
7.4.	SPL directo y C80 en el área de audiencia	88
8.	Marco regulador e impacto socio-económico	89
8.1.	Marco regulador	89
8.2.	Presupuesto	89
8.3.	Impacto socio-económico.....	90
9.	Conclusión y trabajos futuros	91
	REFERENCIAS.....	93

1. INTRODUCTION AND MOTIVATIONS

The most common sonorization systems installed in theaters like the Buero Vallejo in Guadalajara are stereo. While they are simple to implement and allow to cover completely an audience area using reinforcement speakers, these systems offer a relatively poor immersive experience, inasmuch using just two channels L and R limits the spatial capability.

The tridimensional audio expands the possibilities and offers to the user a worthy immersive experience. There are technologies of tridimensional spatial sonorization such as stereo surround systems, Dolby Atmos or VBAP. However, one of the most used technologies is Ambisonics, which even though it was created in the 70s, is booming nowadays due to its ability to recreate sound fields around the listener and processing the spatial content of the signal with a low computational cost, which is really useful in applications such virtual reality videos and the videogames industry.

Since it may be interesting to implement a tridimensional audio system in an auditorium, being able to recreate a sound field that offers an immersive experience to the audience in theatrical shows, either for recreating a scene or experimenting with the spatiality of the sound using creative effects, this project proposes a sonorization system based on Ambisonics technology for the Teatro Buero Vallejo.

2. TECNOLOGÍAS DE SONORIZACIÓN ESPACIAL

2.1. Historic context

The first record of an audio signal was made back in 1860 with a device called phonautograph, patented by Lean Scott three years before. This invention achieved for the first time the transcription of the vibrations of the air particles. However, the first audio playback would not take place until 1877, when Thomas Edison invented the phonograph, which not only was able to record short audio samples but also to play it. Ten years later, Emilie Berliner would patent the gramophone, which improved the storage capability of Edison's phonograph. Its operating principle consisted of an acoustic-mechanical transduction that ended in a needle that recorded the vibrations on a disc, which in the 40s would lead to the invention of the vinyl disc by Columbia Records company. That same decade, the coil tape recorder began to spread. It was based on a model invented in Germany in the 30s, and was able to store electric signals like those coming from the output of a microphone.

At that time, most audio signals contained just one channel: they were monophonic. Even though de stereophonic sound (signals with two channels L and R) arose in 1931 from the hand of the British engineer Alan Blumlein, who performed the first stereophonic record, it did not start to boom until the mid-60s. It was at this point when sound engineers started to experiment with the spatial perception of the sound in the cinemas, the radio and the music.

In the process of experimenting with the placement of the microphones, techniques for the spatial sound capture started to come up. Some of them were based on already existing techniques at that moment, such as the Blumlein one proposed by the inventor of the stereophonic sound.

The engineers tried to emulate the depth perception in the stereo signals using the distance between the microphone and the source. In this way, they positioned the sound elements on virtual layouts through the low frequency boost due to the proximity effect of the microphone (as long as this was not omnidirectional), the attenuation, the distance level losses in high frequency and the direct sound upon reverberant sound proportion, as well as using delays to emulate the timing offsets caused by sound propagation. These techniques could be applied either at the recording or post production phases.

However, over time, engineers needed to delve even further in the depth perception, especially in the cinematographic field. It was then when surround systems showed up: by using multichannel signals, they achieved recreating the sound field in two dimensions. These systems were based on the first spatial playback system, Fantasound, created by the Walt Disney company, which was designed for providing sound spatiality to the flight of a blowfly. This system was formed by a set of horns around the audience.

Until this moment, the sonorization systems were based on channels: the spatiality was achieved on the codification process employing, in most of the cases, level differences between the channels of the signal.

With the rise of the 3D cinema, more complex surround systems started to be proposed, growing the number of the channels in order to achieve a more realistic spatiality, according to the visual immersion provided by the image.

Later on, and in the line of tridimensional cinema, spatial sonorization systems based on objects arose, such as Dolby Atmos or VBAP. Unlike the previous systems, the panning process was in the decoder: depending on the point of the space where the engineer wanted to position a source, the decodification would process the signal by using different gains for each channel. Finally, the speaker system would play the decoded signal and the listener would be able to locate the source at the determined position.

Once the internet revolution and the videogame industry started to push virtual reality, spatial sonorization technologies based on scenes started to boom. This time, instead of recreating isolated sources, the objective was to recreate the full sound field itself. Although there are some of these technologies like Wave Field Synthesis, based on the Huygens principle, the most extended one is Ambisonics, created by the British sound engineer Michael Gerzon, which was developed in the 70s by the National Research Development Corporation of the United Kingdom.

Therefore, nowadays there are different types of technologies for the spatial sonorization, which can be classified in three main groups attending to their operating principles: those based on channels (like stereo and stereo surround systems), those based on objects (like VBAP) and those based on scenes (Ambisonics). Throughout this chapter, these first three are briefly explained, while the technology chosen for this project is studied in depth in the next one.

2.2. Psychoacoustic mechanisms for localization of sound sources

The Duplex Theory of Rayleigh explains how humans are able to locate sound sources interpreting the signal differences between both ears. The employed mechanism depends on the relation between the wavelength and the interaural separation, so the brain processes the information in a different way for different frequency ranges:

- **Low frequency range, below 800 Hz:** The location is mainly achieved due to the interaural time differences (ITD).
- **High frequency range, between 800 and 5000 Hz:** The interaural separation is sufficiently greater than the wavelength to produce a phase shift big enough to make the ITD localization confused. Therefore, at this frequency range the localization is mainly achieved by interaural level differences (ILD). There is a transition between 300 and 800 Hz where the mechanism is a hybrid ITD/ILD (Haas effect).
- **Range above 5 kHz:** This is a problematic range because there is a heavy spectral coloration that strongly depends on the morphology of each person.

In summary, when a human hears a signal with spatial content, this spatial sensation is perceived because of spectral, time and level interaural differences. This is also complemented by the visual information and a learning process.

2.3. Sistemas estereofónicos

2.3.1. Principio de funcionamiento

Son sistemas basados en canales. Las señales estéreo están compuestas por dos: canal izquierdo (L, *Left*) y canal derecho (R, *Right*). Estas pueden obtenerse a partir de una captación estereofónica del campo sonoro o mediante un proceso de panoramización de la fuente, y si se reproducen de manera adecuada, permiten recrear la espacialidad del sonido.

Para que la percepción de la imagen estéreo resulte satisfactoria, es necesario tomar en consideración la distribución y orientación de los altavoces. Es por ello que, cuando surgieron los sistemas de reproducción estereofónica, se comenzó a dar mayor importancia a la configuración del sistema de reproducción y a la posición del receptor, cobrando especial relevancia el concepto de *sweet spot* o punto dulce: punto focal de los altavoces que forman el sistema; es la posición óptima de escucha, aquella en el que la imagen estéreo se reproduce con mayor precisión.

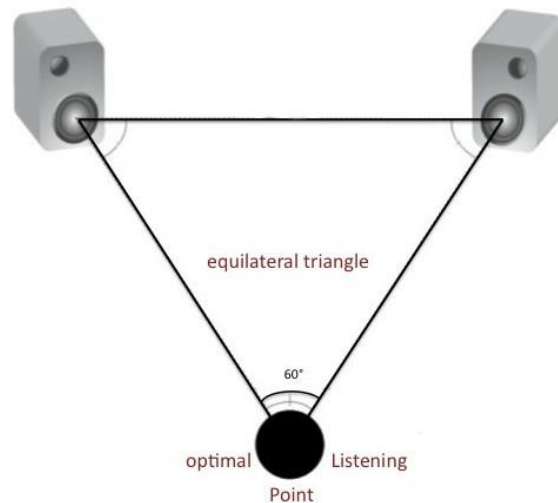


Fig. 2.1 Esquema de un sistema óptimo de reproducción estéreo mediante altavoces

El sistema más simple de reproducción estéreo es un par de auriculares. En el caso de un sistema de altavoces, ambos de igual modelo, el *sweet spot* se sitúa ligeramente detrás de la cabeza del oyente, de manera que los rayos en el eje de los altavoces vayan a parar directamente a los oídos del receptor.

2.3.2. Técnicas de captación estereofónica

Apoyándose en los mecanismos que utiliza el cerebro para la localización de fuentes sonoras, y a raíz del nacimiento y la popularización del sonido espacial, comenzaron a surgir técnicas para la grabación estereofónica que buscaban captar estas diferencias en nivel y tiempo mediante arreglos de micrófonos. Algunas de estas técnicas, y que a día de hoy se siguen utilizando, son la XY, el par espaciado AB, la técnica MS o la captación binaural.

2.3.2.1. XY

Existen diferentes variantes. La primera surgió a principio de los años 30 de la mano del inventor del sonido estereofónico, Alan Blumlein. Esta técnica consiste en un arreglo formado por dos micrófonos bidireccionales coincidentes en un mismo punto del espacio y angulados 90° entre sí, de manera que cada uno registra la información de los canales L y R respectivamente. En consecuencia, la imagen estéreo se produce por una diferencia espectral y de niveles entre ambos canales, debido a sus directividades.

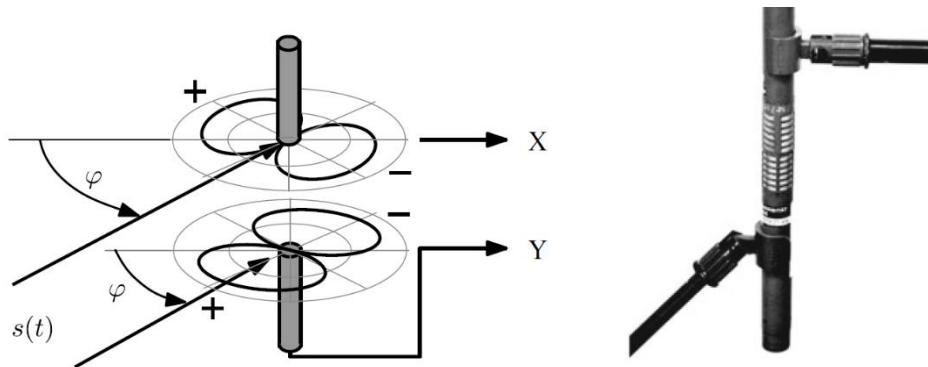


Fig. 2.2 Configuración de un array Blumlein XY con dos cápsulas bidireccionales

Si se posiciona el eje x en el eje de simetría entre ambas cápsulas y el eje positivo de las y hacia la izquierda, el azimut del micrófono X, encargado de registrar la información del canal R, es de - 45°, mientras que el del micrófono Y, encargado de registrar la del L, es de 45°. Por tanto, recordando cuál es la función que describe la directividad de las cápsulas de gradiente de presión, la directividad estereofónica del array vendrá definida por la siguiente expresión vectorial:

$$D_{LR}(\varphi) = \begin{bmatrix} D_L(\varphi) \\ D_R(\varphi) \end{bmatrix} = \begin{bmatrix} \cos(\varphi - 45) \\ \cos(\varphi + 45) \end{bmatrix} \quad (2.1)$$

Una variante más moderna, sustituye estas cápsulas por otras cardioides de primer orden. Esta técnica es más adecuada en situaciones en las que la fuente sonora se sitúe siempre en el eje positivo de las x, ya que se pondera menos la parte trasera, lo que reduce la eficiencia en la incidencia de energía aleatoria (REE) del conjunto y, en consecuencia, se logra una mayor relación sonido directo sobre reverberante.

Recordando la función de directividad de las cápsulas de presión y gradiente de presión de orden uno, la directividad del arreglo es la siguiente:

$$D_{LR}(\varphi) = \begin{bmatrix} \frac{1}{2} + \frac{1}{2} \cdot \cos(\varphi - 45) \\ \frac{1}{2} + \frac{1}{2} \cdot \cos(\varphi + 45) \end{bmatrix} \quad (2.2)$$

Debido al realce en baja frecuencia por el efecto proximidad, propio de estos dos tipos de cápsulas, estas técnicas funcionan mejor a distancias cortas.

2.3.2.2. ORTF

Muy similar a la XY, la técnica ORTF (creada por la cadena *Office de Radiodiffusion Télévision Française*) consiste en dos micrófonos cardioides de primer orden angulados 110° y separados 17 cm. Debido a esta separación entre las cápsulas, la imagen estéreo registrada, además de a una diferencia de niveles, se debe a una diferencia de tiempos. Si bien emula en cierta manera la posición de los oídos, no tiene en cuenta el impacto acústico de la cabeza y las orejas.



Fig. 2.3 Configuración microfónica ORTF

Existen otras variantes como la técnica DIN, que emplea una separación de 20 cm y una angulación de 90° , ponderando aún más el frente que los lados.

2.3.2.3. Par espaciado AB

Consiste en dos micrófonos paralelos separados entre sí una determinada distancia, captando cada uno las señales L y R respectivamente. La idea es captar la imagen estéreo producida por el desplazamiento de una fuente sonora entre los puntos A y B. Tanto el diagrama polar de las cápsulas como su separación pueden variar. Una configuración común es la de dos micrófonos omnidireccionales separados 50 cm.



Fig. 2.4 Configuración microfónica AB

Una variante es el *Decca Tree*, que emplea una separación de dos metros y añade otro micrófono de refuerzo en el centro adelantado un metro, cuya información va a para a ambos canales L y R. Las cápsulas, todas omnidireccionales, se posicionan en paralelo creando un arreglo en forma de "T".

Esta técnica es empleada mayoritariamente para la grabación de orquesta, y ofrece buenos resultados sin la necesidad de emplear micrófonos individuales para cada músico.



Fig. 2.5 Configuración microfónica Decca Tree

2.3.2.4. MS

Consiste en situar en un mismo punto un micrófono omnidireccional perpendicular a otro bidireccional; mientras la cápsula omnidireccional registra la información central (*mid o M*), la figura de ocho registra la lateral (*side o S*).

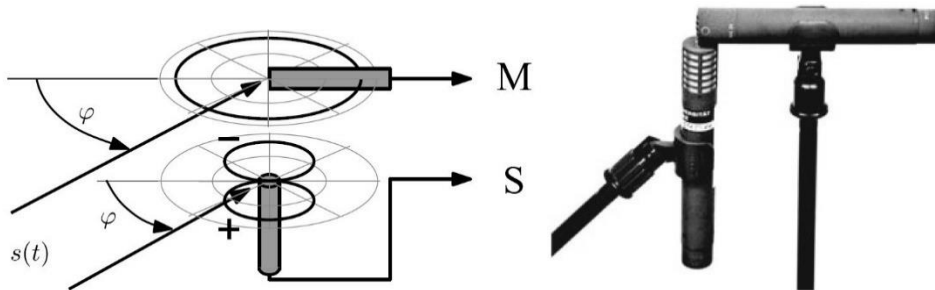


Fig. 2.6 Configuración de un array MS con una cápsula omnidireccional y otra bidireccional

$$D_{MS}(\varphi) = \begin{bmatrix} D_M(\varphi) \\ D_S(\varphi) \end{bmatrix} = \begin{bmatrix} 1 \\ \cos(\varphi - 90) \end{bmatrix} = \begin{bmatrix} 1 \\ \text{sen } \varphi \end{bmatrix} \quad (2.3)$$

Otra forma de implementarlo es mediante dos cápsulas cardioides de primer orden enfrentadas y alineadas en el eje de las y. El inconveniente de este método frente al anterior, es que se pierde cobertura en el eje de las x.

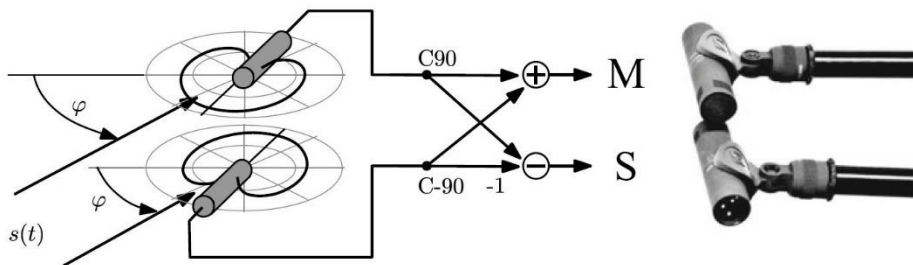


Fig. 2.7 Configuración de un array MS con dos cápsulas cardioides enfrentadas

En este caso, es necesario aplicar una codificación en la que la señal M sea reconstruida como una suma de la señal que registran ambos micrófonos, mientras que la señal S se obtenga mediante una resta, emulando la polaridad invertida en el eje negativo de las y:

$$D_{MS}(\varphi) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} + \frac{1}{2} \cdot \cos(\varphi - 90) \\ \frac{1}{2} + \frac{1}{2} \cdot \cos(\varphi + 90) \end{bmatrix} = \begin{bmatrix} 1 \\ \text{sen } \varphi \end{bmatrix} \quad (2.4)$$

Una tercera variante consiste en un micrófono bidireccional orientado al eje positivo de las y , encargado de captar la señal S , coincidente con otro cardioide de primer orden orientado al eje positivo de las x , encargado de registrar la señal M .

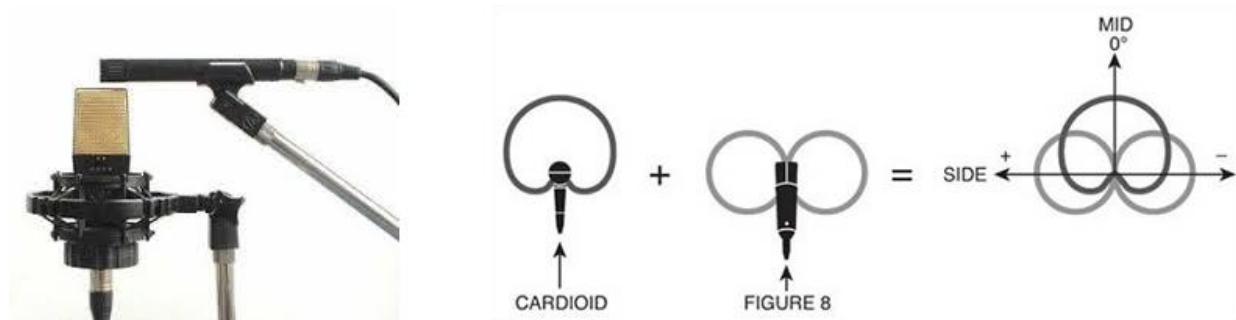


Fig. 2.8 Configuración de un array MS con una cápsula bidireccional y otra cardioide

De nuevo, al igual que pasaba con la variante de la técnica XY que empleaba cápsulas cardioide, esta técnica es más adecuada en situaciones en las que la fuente sonora se sitúe en el eje positivo de las x , ya que este arreglo pondera menos la parte trasera. No necesita codificación, por lo que la directividad de sus canales es la siguiente:

$$D_{MS}(\varphi) = \begin{bmatrix} \frac{1}{2} + \frac{1}{2} \cdot \cos \varphi \\ \cos(\varphi - 90) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} + \frac{1}{2} \cdot \cos \varphi \\ \text{sen } \varphi \end{bmatrix} \quad (2.5)$$

Cabe mencionar que la técnica *mid side* no es sólo microfónica, ya que es ampliamente usada en procesos de mezcla y masterización para controlar de manera independiente en una señal estéreo el contenido *mid* (toda aquella información de la señal que es igual en L y R) y el contenido *side* (toda aquella información de la señal que es distinta en L y R). La codificación necesaria para obtener las componentes MS se realiza a partir de una suma y una diferencia entre los canales estéreo, dividiendo entre dos para compensar la suma de señales:

$$\begin{bmatrix} M \\ S \end{bmatrix} = \frac{1}{2} \cdot \begin{bmatrix} L + R \\ L - R \end{bmatrix} \quad (2.6)$$

Un ejemplo muy común de procesamiento *mid-side* consiste en disminuir el nivel del *mid* para enfatizar la imagen estéreo, con cuidado de no aplicar ajustes drásticos que pudieran dar lugar a cancelaciones entre ambos canales si estos se reprodujesen en mono ($L + R$) o en un sistema estéreo de altavoces con una separación entre L y R muy estrecha.

La decodificación necesaria para reproducir una señal MS en un sistema estéreo es la siguiente:

$$\begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} M + S \\ M - S \end{bmatrix} \quad (2.7)$$

2.3.2.5. Biaural

Las señales biaurales se basan en la HRTF (*Head-Related Transfer Function*), que es la respuesta en frecuencia del conjunto formado por la cabeza humana, las orejas y los conductos auditivos, a veces tomando también en cuenta el torso. Debido a nuestra morfología y al posicionamiento de los oídos, esta respuesta en frecuencia varía con la dirección de llegada del sonido.

El cerebro recibe las señales registradas por cada oído, las compara e interpreta sus diferencias tanto en tiempo (ITD), nivel (ILD) como en coloración espectral. Entonces, lleva a cabo un proceso de decodificación que permite posicionar las fuentes en el campo sonoro. Como el camino y la dirección de llegada son distintos para cada oído, se distingue entre las HRTF de hemisferio izquierdo (L-HRTF) y derecho (R-HRTF).

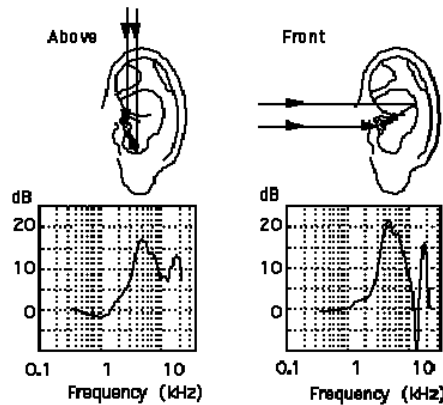


Fig. 2.9 HRTF para distintas direcciones de llegada del sonido

Visto en el dominio temporal, si se convolucionase una señal mono con la respuesta al impulso de la cabeza (*Head-Related Impulse Response*, HRIR) para ambos oídos y una posición fuente-receptor determinada, se obtendría la señal biaural que habría percibido una persona situada en el lugar del receptor, siendo la señal en cuestión emitida desde la posición de la fuente. Por esta razón, las señales biaurales son muy utilizadas en el ámbito de la realidad virtual.

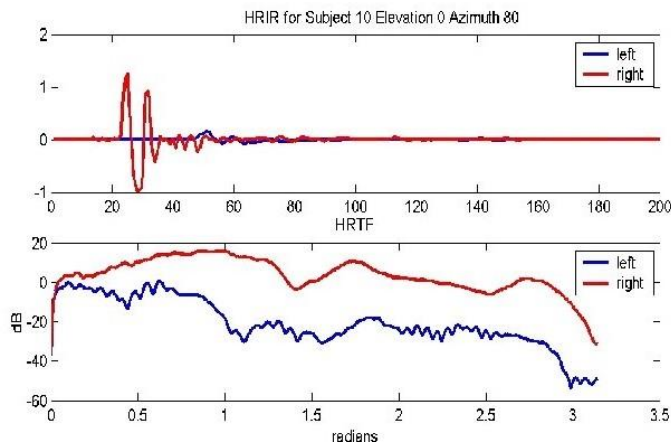


Fig. 2.10 HRIR y HRTF para cada hemisferio, para una fuente en una localización determinada

Las señales binaurales son señales estéreo que se registran con micrófonos especializados. Los más habituales son los integrados en los oídos de la cabeza de un maniquí que emula la HRTF de un humano promedio (Fig. 2.11). La reproducción de estas señales en auriculares permite recrear con precisión y de manera relativamente sencilla la percepción espacial registrada en la posición del maniquí.



Fig. 2.11 Neumann KU-100 Dummy-Head

Otro tipo de micrófonos diseñados para registrar señales en formato binaural, consiste en dos cápsulas omnidireccionales integradas en una estructura similar a un auricular *in ear* (Fig. 2.12). La ventaja frente al *dummy-head* es la movilidad: el usuario puede desplazarse con facilidad por el campo sonoro y registrarlo de manera ágil y sencilla. Sin embargo, debido a que dichas cápsulas no se sitúan con la suficiente profundidad en el conducto auditivo, no logran registrar del todo la HRTF, por lo que el resultado no es tan realista.



Fig. 2.12 Roland CS-10EM

La mejor solución son los micrófonos intraurales: se sitúan a una profundidad suficiente del conducto auditivo como para registrar el campo sonoro con la HRTF del propio usuario. El problema de estos micrófonos es que resulta más complicado encontrar soluciones comerciales, y que hay usuarios que rechazan su uso por miedo y resultar demasiado invasivos.

Además de la reproducción en auriculares, existen sistemas de reproducción binaural basados en altavoces. Sin embargo, estos presentan un gran inconveniente: el sonido que el usuario percibe, contiene la HRTF del maniquí más la de su propia cabeza, así como una diafonía entre canales debido a que, a diferencia de la reproducción binaural en auriculares, no existe un aislamiento interaural en el usuario.

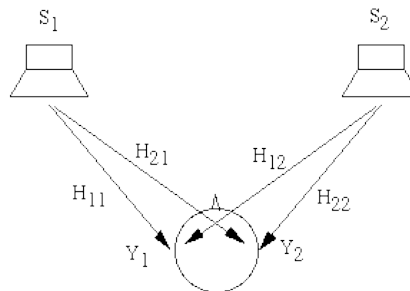


Fig. 2.13 Crosstalk producido en la reproducción de señales binaurales en un sistema de altavoces

Existen filtros orientados a eliminar la HRTF del usuario y la diafonía, monitorizando la posición del receptor mediante sensores en forma de diadema (*head-trackers*). Sin embargo, el funcionamiento resulta complejo y el sistema tiene un *sweet spot* muy limitado.

2.3.3. Localización

En un sistema de reproducción estéreo convencional, la localización de una fuente se produce gracias a la panoramización. Si bien el siguiente desarrollo matemático puede implementarse para el caso de una panoramización mediante retardos (ITD), solo considera la diferencia de niveles entre los canales L y R (ILD), ya que es el método más común y el más sencillo de implementar.

Para una pareja de altavoces iguales, correctamente orientados y equidistantes al receptor, la localización en el *sweet spot* se produce de la siguiente manera:

$$r = (1 - q) \cdot \theta_L + q \cdot \theta_R \quad (2.8)$$

Siendo:

- r = Vector de dirección de la localización percibida por el usuario para una fuente sonora.
- θ_L = Vector unitario de dirección del altavoz izquierdo desde el *sweet spot*.
- θ_R = Vector unitario de dirección del altavoz derecho desde el *sweet spot*.
- q = Ganancia de proporcionalidad L/R que determina la localización de la fuente.

Una localización perfecta de la fuente implicaría $|r| = 1$, como podría ocurrir en la situación de un sistema de altavoces en un campo libre ideal; en cambio, una localización indeterminada debido a una equiprobabilidad para todas las direcciones de llegada de la onda implicaría $|r| = 0$, como podría ser el caso de un campo difuso ideal.

El parámetro q se maneja mediante las ganancias (g_L, g_R) aplicadas a cada canal. Para lograr que la sonoridad de la fuente sea independiente de su panoramización, es necesario llevar a cabo una normalización de dichas ganancias:

$$q = \frac{|g_R|^\gamma}{|g_L|^\gamma + |g_R|^\gamma} \quad (2.9)$$

El parámetro γ es el *slope*, y es quien determina el comportamiento del sistema a la hora de compensar la sonoridad al llevar a cabo una panoramización: un $\gamma = 1$ implicaría una normalización basada en una suma coherente de las ganancias de cada canal; en cambio, un $\gamma = 2$ implicaría una normalización basada en una suma energética de estas ganancias. Si se integra este parámetro en la ecuación de la localización, se obtiene la siguiente expresión:

$$r_\gamma = \frac{|g_L|^\gamma \cdot \theta_L + |g_R|^\gamma \cdot \theta_R}{|g_L|^\gamma + |g_R|^\gamma} \quad (2.10)$$

Puesto que en la práctica el receptor no siempre está perfectamente centrado en el *sweet spot*, y puesto que aún en el caso de que la señal que salga por ambos altavoces sea idéntica suelen existir reflexiones que afectan a la coherencia entre ambas señales, es más habitual y suele dar mejores resultados emplear un valor de *slope* de $\gamma = 2$ (suma incoherente de ganancias), definiéndose el vector de localización como r_E . En aquellos casos en los que el *slope* es fijado en $\gamma = 1$, el vector de localización se define como r_V .

2.4. Sistemas *surround*

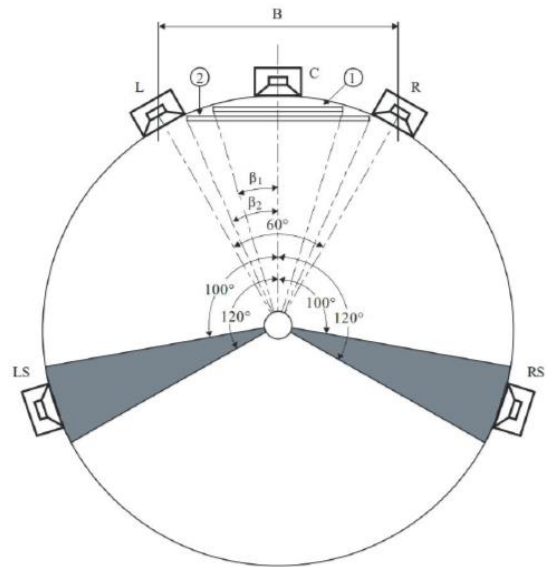
2.4.1. Principio de funcionamiento

Los sistemas *surround* añaden más canales a la señal para obtener más información de espacialidad y reproducir la profundidad del sonido. Siguen una nomenclatura I, J , siendo I el número de altavoces de alta, media y media-baja frecuencia (han de ser de igual modelo) y J el número de altavoces de baja frecuencia. Un sistema de reproducción estereofónico sería un sistema 2.0, mientras que un estereofónico con refuerzo de un subgrave sería un sistema 2.1. Los más comunes son el 5.1 y el 7.1:

- **Sistema 5.1**

Compuesto por cinco canales con contenido en media y alta frecuencia (reproducidos respectivamente por cinco altavoces cuya disposición óptima se muestra a la derecha), más un canal LFE de refuerzo con contenido en graves (emitido por un subgrave, cuya posición no es tan importante). Estos canales son:

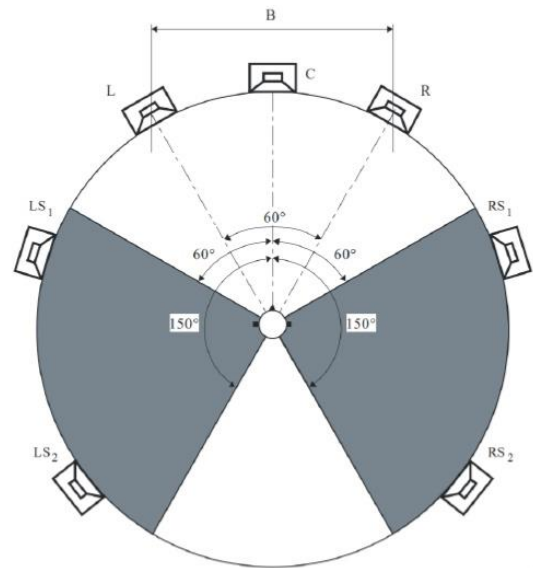
- *C, Center*
- *L, Left*
- *R, Right*
- *LS, Left Surround*
- *RS, Right Surround*
- *LFE, Low Frequency Effects*



- **Sistema 7.1**

Añade dos canales LR *surround* para rellenar puntos muertos, envolver aún más al usuario y mejorar la experiencia de espacialidad. Los canales de los que está compuesto son los siguientes:

- *C, Center*
- *L, Left*
- *R, Right*
- *LS1, Left Surround 1*
- *RS1, Right Surround 1*
- *LS2, Left Surround 2*
- *RS2, Right Surround 2*
- *LFE, Low Frequency Effects*



2.4.2. Localización

En un sistema *surround*, el canal de subgrave tiene un escaso impacto en la localización de la fuente, debido a la omnidireccionalidad en baja frecuencia y al mayor tiempo de reverberación que suele haber en los recintos en este rango. Por tanto, son el resto de canales quienes colaboran en mayor medida a la localización percibida por el usuario.

Para un sistema *surround* de L canales, siendo reproducidos por un subgrave y $L-1$ altavoces iguales, el vector de localización viene determinado por una generalización de la expresión definida en el apartado de localización para sistemas estereofónicos:

$$r_\gamma = \frac{\sum_{l=1}^{L-1} |g_l|^\gamma \cdot \theta_l}{\sum_{l=1}^{L-1} |g_l|^\gamma} \quad (2.11)$$

Esta expresión es válida cuando el receptor se encuentra en el *sweet spot*. Sin embargo, cuando el usuario no está centrado en el sistema de reproducción, se produce un desequilibrio de niveles entre los altavoces ajeno a la panoramización llevada a cabo en el codificador, y que es debido a las diferencias de distancia entre el usuario y cada altavoz. Por tanto, en este caso es necesario incluir estas pérdidas en la expresión, ya que tienen un impacto en la localización percibida:

$$r_\gamma = \frac{\sum_{l=1}^{L-1} |G_{d,l} \cdot G_{t,l} \cdot g_l|^\gamma \cdot \theta_l}{\sum_{l=1}^{L-1} |G_{d,l} \cdot G_{t,l} \cdot g_l|^\gamma} \quad (2.12)$$

El parámetro $G_{d,l}$ hace referencia a la pérdida de energía de la señal emitida por un altavoz l debido a la distancia d en metros que lo separa del usuario:

$$G_{d,l} = \frac{1}{d} \quad (2.13)$$

El parámetro $G_{t,l}$ hace referencia a la pérdida de energía de la señal emitida por un altavoz l debido al tiempo en segundos que tarda en llegar al usuario, y se debe esencialmente a un filtro peine (*comb filter*) que se produce al sumarse coherentemente con la señal emitida por otro altavoz situado a una distancia diferente. Esta pérdida de energía depende de la coherencia, la frecuencia, la diferencia de caminos entre los altavoces y la diferencia de niveles entre ambas señales a la hora de interactuar. Por la necesidad de fijar un criterio, experimentalmente se determinó que estas pérdidas pueden aproximarse a 1 dB/4 ms.

$$G_{t,l} = 10^{-\frac{1/0,004}{20} \cdot t_l} \quad (2.14)$$

Aunque estas pérdidas debidas a la desviación del usuario respecto al *sweet spot* no fueron definidas en los sistemas estereofónicos, ocurren de igual manera.

2.5. VBAP

2.5.1. Principio de funcionamiento

VBAP (*Vector-Base Amplitude Panning*) es una tecnología de sonorización basada en objetos. Trata al espacio que rodea al oyente como a una esfera construida a partir de triángulos, en cuyos vértices sitúa los altavoces, todos con la misma respuesta, equidistantes al usuario y orientados a él.

La señal está compuesta por tantos canales como altavoces tenga el sistema. Entre dichos canales se produce de manera intencionada diferencias de amplitud que permiten al usuario percibir ILD y localizar por triangulación la fuente sonora en el punto p premeditado. De esta manera se logra envolver al usuario con un conjunto infinito de altavoces “fantasma”.

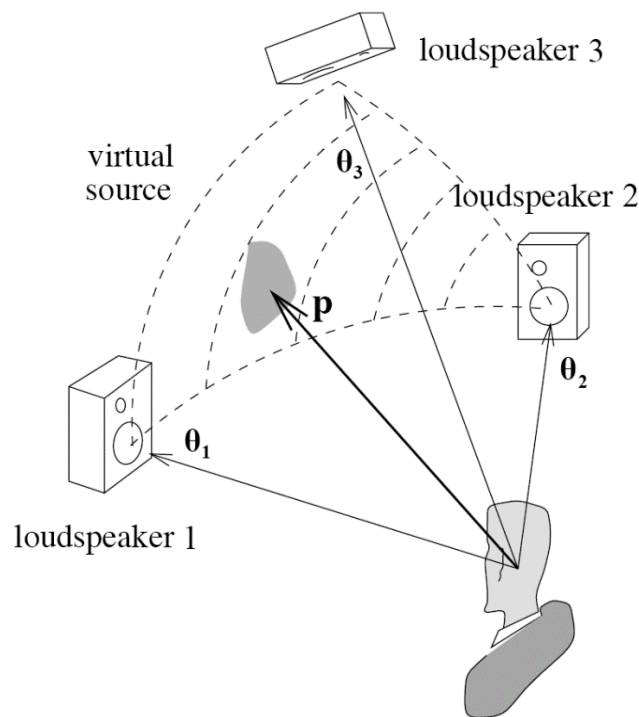


Fig. 2.16 Esquema del funcionamiento de un sistema VBAP de tres canales

El número mínimo de canales necesarios para un sistema VBAP es de tres, con sus tres altavoces correspondientes formando un triángulo, como puede verse en la figura.

Este triángulo no ha de ser necesariamente equilátero. Sin embargo, para evitar que se produzcan huecos o puntos muertos en la localización por triangulación, se recomienda que el ángulo entre altavoces no exceda los 90° . Por tanto, partiendo del concepto de producto escalar y recordando que los vectores de dirección de los altavoces son unitarios, se deduce que dos altavoces cualesquiera del sistema de reproducción han de cumplir lo siguiente:

$$\arccos(\theta_1 \cdot \theta_2) \leq 90^\circ \quad (2.15)$$

2.5.2. Localización

El vector p es el vector unitario de dirección de la panoramización: apunta a la posición en la que se desea situar la fuente virtual. A diferencia de los sistemas *surround*, en VBAP este posicionamiento se lleva a cabo en el decodificador y, una vez más, se logra mediante el control de unas ganancias g aplicadas a cada canal de la señal, logrando una panoramización mediante amplitud.

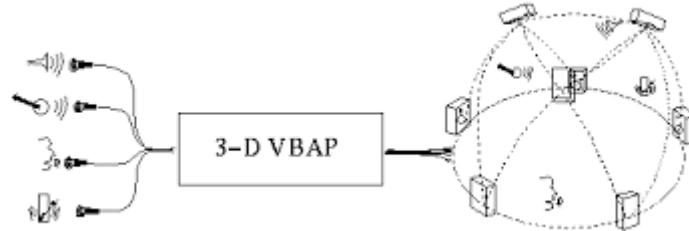


Fig. 2.17 Configuración de un sistema VBAP

Estas ganancias deben respetar la polaridad de las señales, por lo que han de cumplir siempre $g > 0$; de no hacerlo, podría producirse desfases que provoquen efectos desagradables en la reproducción, como que el usuario perciba que el sonido proviene desde su propia cabeza. Para mantener la sonoridad, estas ganancias además han de estar normalizadas:

$$p_V = [\theta_1 \quad \theta_2 \quad \theta_3] \cdot \frac{1}{\sqrt{g_1^2 + g_2^2 + g_3^2}} \cdot \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} \quad (2.16)$$

Mientras que el vector p_V determina la posición intencionada de la fuente, el vector de localización r_V definido anteriormente determina la posición percibida por el usuario. En un caso ideal, ambas direcciones son coincidentes. Sin embargo, debido a factores como la descentralización del usuario, la acústica de la sala o un desequilibrio en la respuesta de cada altavoz, pueden existir desviaciones que distorsionen la localización. En casos en los que estas desviaciones sean excesivas, puede optarse por la variante VBIP (*Vector-Base Intensity Panning*), basado en el modelo de r_E definido en el apartado 2.3.3:

$$p_E = [\theta_1 \quad \theta_2 \quad \theta_3] \cdot \frac{1}{\sqrt{g_1^2 + g_2^2 + g_3^2}} \cdot \begin{bmatrix} g_1^2 \\ g_2^2 \\ g_3^2 \end{bmatrix} \quad (2.17)$$

Para un sistema basado en VBAP con más de tres altavoces, el vector de panoramización se obtiene como una generalización:

$$p_V = \frac{\sum_{l=1}^L \theta_l \cdot g_l}{\sqrt{\sum_{l=1}^L g_l^2}} \quad (2.18)$$

3. AMBISONICS

Ambisonics, acrónimo de “*ambient*” (ambiente) y “*sonic*” (sonido), es la tecnología de sonorización espacial basada en escenas más extendida actualmente. A pesar de haber sido creada a principios de los años 70 en el Reino Unido, ha sido en los últimos años, con la popularización de los vídeos de realidad virtual, cuando ha empezado a cobrar mayor relevancia.

En este capítulo se profundiza en su principio de funcionamiento, así como en sus formatos más extendidos, sus técnicas microfónicas para la captación espacial del sonido, y otros aspectos como la codificación/decodificación, el procesamiento de la señal y su reproducción.

3.1. Principio de funcionamiento

A diferencia de los sistemas basados en canales y objetos descritos anteriormente, *Ambisonics* no busca reconstruir fuentes aisladas, sino que trata de descomponer el propio campo sonoro alrededor de un punto del espacio en una serie de armónicos esféricos de manera que, al superponerse todos entre sí, se logre reconstruir el campo sonoro en torno al oyente.

Este principio de funcionamiento se basa en el hecho de que cualquier señal $s(t, \theta)$ del campo sonoro puede descomponerse en una serie infinita de armónicos esféricos espaciales $Y_n^m(\theta)$, de orden n y grado m , ponderados cada uno de ellos por un determinado peso f_{nm} . Algo similar a lo que ocurre con las series de Fourier.

$$s(t, \theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^n Y_n^m(\theta) \cdot f_{nm} \quad (3.1)$$

Si bien una señal con infinitos armónicos esféricos reconstruiría el campo sonoro de manera ideal, por motivos prácticos resulta inevitable truncarlos, obteniendo entonces una señal ambisónica de orden N que será una aproximación de la señal original:

$$s(t, \theta) \approx \sum_{n=0}^N \sum_{m=-n}^n Y_n^m(\theta) \cdot f_{nm} \quad (3.2)$$

Por tanto, esta tecnología es escalable en N : cuanto mayor sea el orden de la señal, mayor número de canales contendrá, o lo que es lo mismo, mayor número de armónicos esféricos se superpondrán entre sí, dando lugar a una señal con mayor resolución espacial.

Los primeros 16 armónicos esféricos encargados de descomponer el espacio son los que se representan en la siguiente figura:

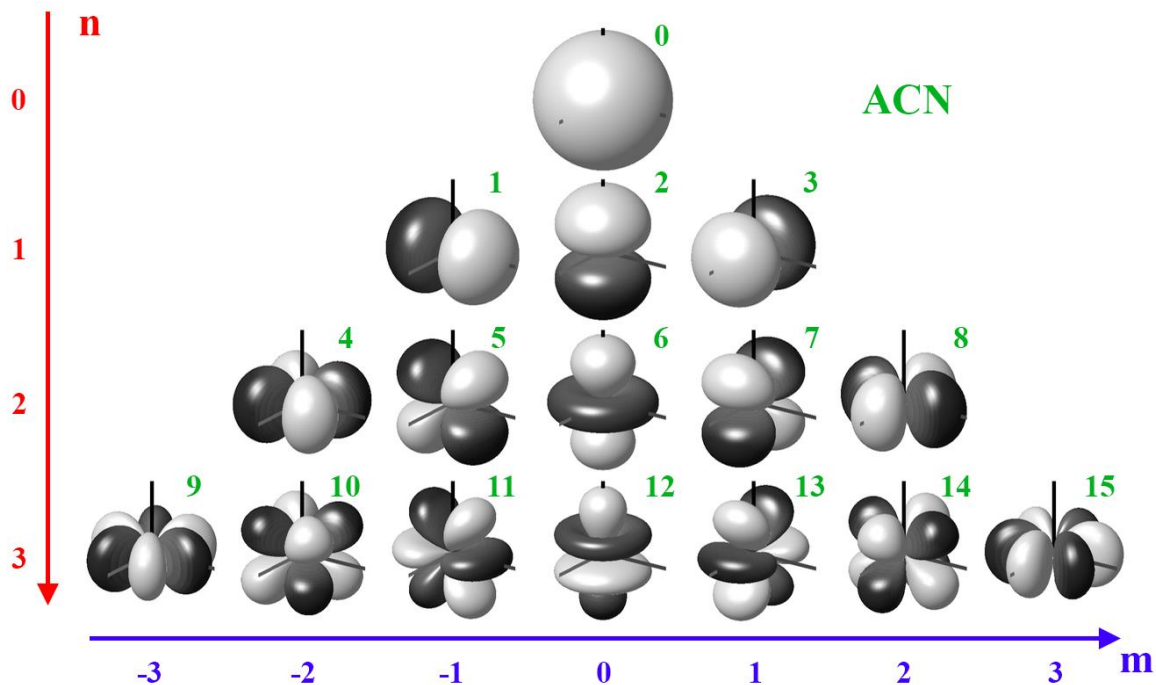


Fig. 3.1 Armónicos esféricos de Ambisonics

Cada armónico esférico representa la directividad asociada a un canal, con el color blanco indicando polaridad positiva y el negro negativa. El **orden del armónico** $n \geq 0$ se corresponde con el número de intersecciones con el centro de coordenadas, mientras que la m , acotada en $-n \leq m \leq n$, representa el **grado del armónico**. Los armónicos están ordenados por el **índice del canal** $ACN = n^2 + n + m$ (*Ambisonic Channel Number*).

Una señal ambisónica de orden $N = 0$ contiene sólo el canal fundamental ACN_0 , y es equivalente a una señal mono registrada por un micrófono omnidireccional; una señal de orden $N = 1$, denominada FOA (*First Order Ambisonics*) contiene los cuatro primeros canales ACN_0, ACN_1, ACN_2 y ACN_3 , llamados respectivamente W, Y, Z, X; una señal de orden $N = 2$ contiene los nueve primeros; una de orden $n = 3$ los 16, y así sucesivamente. Por tanto, una señal ambisónica tridimensional de orden N contiene $(N + 1)^2$ canales, y es denominada HOA (*Higher Order Ambisonics*) siempre que $N \geq 1$.

Estos armónicos tridimensionales, encargados de descomponer el espacio desde unas coordenadas esféricas, se rigen por la siguiente fórmula (siendo θ el parámetro que consiste en un vector unitario de dirección de coordenadas xyz , y siendo en coordenadas esféricas φ y ϑ el azimut y la elevación respectivamente):

$$Y_n^m(\theta) = Y_n^m(\varphi, \vartheta) = \phi_m(\varphi) \cdot N_n^{|m|} \cdot P_n^{|m|}(\cos \vartheta) \cdot (-1)^m \quad (3.3)$$

Como se observa, la función de los armónicos se compone de cuatro componentes. La primera es $\phi_m(\varphi)$, que consiste en una función dependiente del azimut encargada de describir los armónicos en el plano horizontal, y cuya expresión es la siguiente:

$$\phi_m(\varphi) = \begin{cases} \frac{1}{\sqrt{\pi}} \cdot \text{sen}(|m| \cdot \varphi), & m < 0 \\ \frac{1}{\sqrt{2\pi}}, & m = 0 \\ \frac{1}{\sqrt{\pi}} \cdot \text{cos}(m \cdot \varphi), & m > 0 \end{cases} \quad (3.4)$$

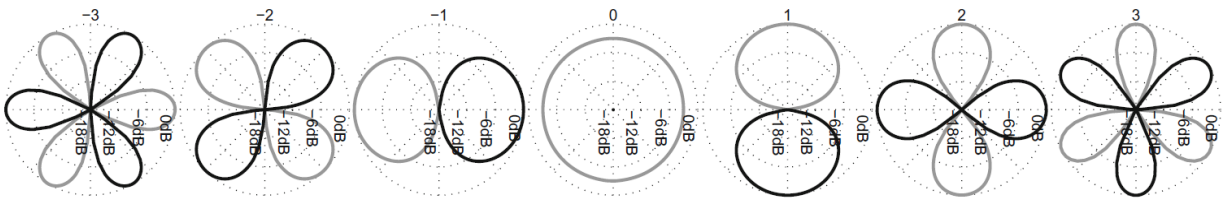


Fig. 3.2 Representación de los armónicos esféricos 2D en escala logarítmica, para $m = -3, \dots, 3$

La siguiente componente es el término de normalización $N_n^{|m|}$, encargado de igualar la ponderación de cada armónico. Aunque existen diferentes criterios, la normalización más común es la SN3D:

$$N_n^{|m|} = \sqrt{\frac{1}{2} \cdot \frac{(n - |m|)!}{(n + |m|)!}} \quad (3.5)$$

El tercer elemento $P_n^{|m|}(\cos \vartheta)$, proporciona al armónico esférico la componente de elevación. Consiste en un polinomio de Legendre dependiente del orden n y del módulo del grado m del armónico en cuestión, y que recibe como argumento el coseno de la elevación.

$P_n^{ m }$		$ m $			
		0	1	2	3
n	0	1	N/A	N/A	N/A
	1	$\cos \vartheta$	$-\text{sen } \vartheta$	N/A	N/A
	2	$\frac{1}{2}(3\cos^2 \vartheta - 1)$	$-3\cos \vartheta \text{sen } \vartheta$	$3\text{sen}^2 \vartheta$	N/A
	3	$\frac{1}{2}(5\cos^3 \vartheta - 3\cos \vartheta)$	$-\frac{3}{2}(5\cos^3 \vartheta - 1)\text{sen } \vartheta$	$15\cos \vartheta \text{sen}^2 \vartheta$	$-15\text{sen}^3 \vartheta$

Tabla 3.1 Polinomios de Legendre

Por último, el elemento $(-1)^m$ se encarga de corregir el signo de los polinomios de Legendre para que las polaridades de los armónicos se correspondan con los representados en la Figura 3.1. De esta manera, para una señal HOA de tercer orden, los ejes cartesianos se construyen tal que:

Posición	Eje	Canales con polaridad positiva	Canales con polaridad negativa
Delante	x^+	{0, 3, 8, 15}	{6, 13}
Detrás	x^-	{0, 8, 13}	{3, 6, 15}
Izquierda	y^+	{0, 1}	{6, 8, 9, 11}
Derecha	y^-	{0, 9, 11}	{1, 6, 8}
Arriba	z^+	{0, 2, 6, 12}	Ninguno
Abajo	z^-	{0, 6}	{2, 12}

Tabla 3.2 Polaridad de los canales para los ejes cartesianos

También es posible obtener una señal ambisónica en 2D, de $2N + 1$ canales, que consiste en una proyección de su equivalente 3D sobre el plano horizontal. Para ello, hay que desechar la componente de elevación, eliminando aquellos canales cuyos armónicos dependan de ella: los armónicos asociados a cada canal estarán descritos por la función $\phi_m(\varphi)$.

PROYECCIÓN 3D → 2D		
n	ACN que se eliminan	ACN que se mantienen
1	{2}	{0, 1, 3} = {W, Y, X}
2	{2, 5, 6, 7}	{0, 1, 3, 4, 8}
3	{2, 5, 6, 7, 10, 11, 12, 13, 14}	{0, 1, 3, 4, 8, 9, 15}
N	$\{n^2 + n + m\} \mid n \neq m $	$\{n^2 + n + m\} \mid n = m $

Tabla 3.3 Proyección de Ambisonics 3D sobre 2D

La ventaja que ofrece esta tecnología frente a otros sistemas es la optimización en el procesamiento espacial de la señal. Un ejemplo es la rotación del campo sonoro: mientras en sistemas basados en canales y objetos habría que repositionar cada fuente del campo sonoro mediante ILD recalculando las ganancias g para cada canal, en *Ambisonics* basta con multiplicar por una matriz. Lo mismo ocurre con otros procesos como el efecto espejo, en el que simplemente hay que invertir la polaridad de determinados canales en función de la simetría impar de sus armónicos asociados. Este bajo coste computacional facilita el procesamiento espacial en tiempo real, lo que la convierte en una tecnología idónea para VR.

Sin embargo, como se verá más adelante, otros procesamientos más básicos de la señal, como pueden ser la ecualización o la compresión de una fuente puntual, resultan más tediosos que en sistemas basados en canales u objetos.

3.2. Formatos principales

Como se ha explicado, una señal ambisónica tridimensional de orden N contiene su correspondiente número de canales. Sin embargo, el significado de la información que estos contienen depende de su formato. Existen principalmente tres: los formatos A, B y C.

3.2.1. Formato A

Una señal ambisónica de orden N en formato A, es aquella registrada por un arreglo de $(N + 1)^2$ cápsulas cardioides de primer orden dispuestas en forma de tetraedro. Este arreglo puede llevarse a cabo mediante micrófonos individuales o con micrófonos especializados que integran estas cápsulas en un mismo sistema de captación del campo sonoro (Fig. 3.3).

En un FOA de formato A, el nombre de los cuatro canales es el mismo que el de las cápsulas a las que se conectan, y hace referencia a sus posiciones en el tetraedro: *FLU* (*Front Left Up*), *FRD* (*Front Right Down*), *BLD* (*Back Left Down*) y *BRU* (*Back Right Up*).

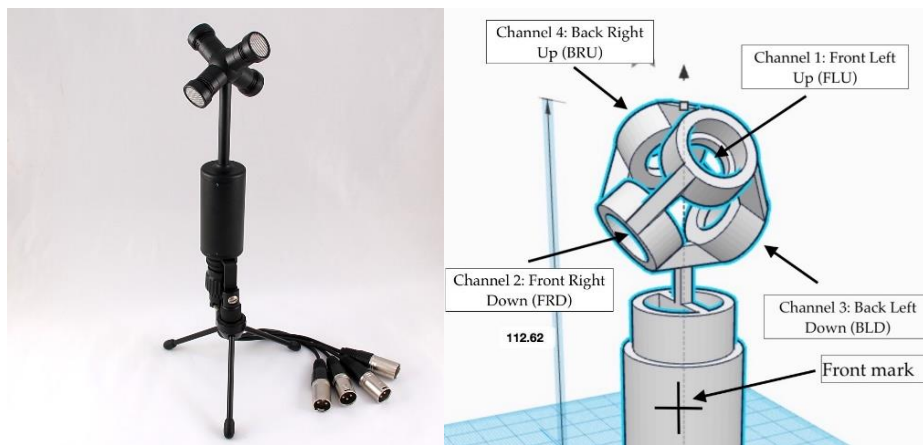


Fig. 3.3 Sistema FOA para la captación del campo sonoro en formato A

Para asegurar que la ponderación espacial sea homogénea, además de que las cápsulas más sus respectivos preamplificadores tengan la misma respuesta en frecuencia, resulta imprescindible haber calibrado previamente los micrófonos. Para ello, mediante una señal test de banda ancha, se mide el nivel registrado por cada cápsula; mediante una grabadora o interfaz de audio que cuente con un amplificador independiente dedicado a cada una de dichas cápsulas, se asegura que la diferencia de nivel entre los canales no exceda los 0'1 dB para la señal de test.

Para registrar correctamente el campo sonoro, antes de empezar a grabar, conviene evaluar cuáles son las fuentes de interés, para así orientar el array hacia ellas y posicionarlo en el punto exacto que se desea recrear en la posterior reproducción.

Una vez se obtiene la señal ambisónica en formato A, es necesario codificarla para pasarla a formato B y poder procesarla.

3.2.2. Formato B

Consiste en una codificación del formato anterior. A partir de una combinación lineal de los diagramas polares de los canales del formato A, se reconstruye los armónicos esféricos.

Para el caso de primer orden, a partir de FLU FRD, BLD y BRU, se realiza una codificación para obtener las señales que se habría grabado con un micrófono omnidireccional en el origen de coordenadas (W) y tres micrófonos bidireccionales en los ejes cartesianos (XYZ). Puesto que hacer coincidir cuatro cápsulas en un mismo punto resulta poco práctico, es habitual grabar en formato A para, posteriormente, convertir el conjunto de señales. Sin embargo, a pesar de ser menos práctico, también es posible captar el campo sonoro directamente en formato B, como se verá más adelante.

Existen dos convenciones:

- **AmbiX (Ambisonic eXchange)**: La más común. La secuencia de canales para las señales de primer orden es WYZX, y emplea normalización SN3D.
- **Fu-Ma (Furse-Malham)**: La secuencia de canales es WXYZ, y el nivel del canal W es 3 dB menor que en AmbiX.

Para que la recreación espacial sea satisfactoria, además de elegir la convención correcta en el decodificador, es imprescindible haber posicionado y orientado correctamente el micrófono en el momento de la grabación, distinguiendo entre las orientaciones *front-facing*, *upside-down*, y *end-fire*.

3.2.3. Formato C (UHJ)

Es el menos común. Fue diseñado para ser compatible con decodificadores tradicionales mono y estéreo. El FOA consta de cuatro canales LRTQ, siendo los LR obligatorios y los TQ opcionales.

Son los dos primeros canales los que ofrecen la compatibilidad con decodificadores mono (L+R) y estéreo (L, R). Si se añade información en el canal T, se obtiene una señal ambisónica 2D equivalente al WXY del formato B. Mientras que, si se añade el último canal Q, se obtiene la componente de elevación, logrando una reconstrucción 3D equivalente al WXYZ.

3.3. Técnicas de captación del campo sonoro

Las técnicas microfónicas de *Ambisonics* están inspiradas en la técnica MS explicada anteriormente. Buscan ahondar en ella y extrapolar su principio de funcionamiento para la captación del campo sonoro en los ejes xyz, de manera que los diagramas de polaridad asociados a cada canal emulen los armónicos esféricos. Por ejemplo, un FOA contiene una esfera en el centro de coordenadas más el *mid* de los ejes cartesianos, mientras que, si se sube al segundo orden, se añaden las componentes *side* a cada eje xyz.

A continuación, se expone algunas técnicas microfónicas para la captación de señales *ambisónicas* de primer orden, tanto en dos como en tres dimensiones.

3.3.1. FOA 2D

3.3.1.1. Un micrófono omnidireccional y dos bidireccionales

Consiste en tres cápsulas coincidentes en un mismo punto: una cápsula omnidireccional posicionada en el centro de coordenadas y dos bidireccionales anguladas 90° entre sí, con una apuntando al eje positivo de las x (incidencia a $\varphi = 0^\circ$) y la otra al eje positivo de las y (incidencia a $\varphi = 90^\circ$).

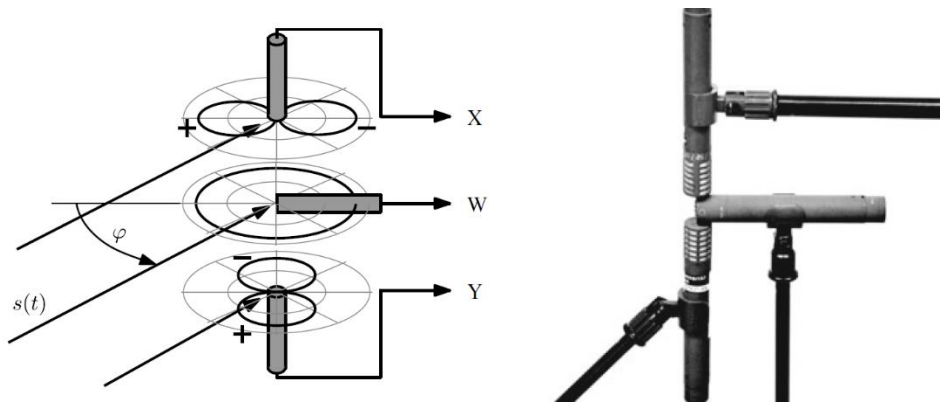


Fig. 3.4 Configuración de un array FOA 2D para la captación de señales ambisónicas en formato B

No es necesaria ninguna codificación ya que, por definición, las señales que registran los micrófonos son los canales WXY del formato B de *Ambisonics*: mientras la cápsula omnidireccional obtiene la señal W, las bidireccionales registran las señales X e Y, respectivamente, con la polaridad invertida para incidencias desde los ejes negativos. La directividad de los canales del array en el plano azimutal es la siguiente:

$$D_{WXY}(\varphi) = \begin{bmatrix} 1 \\ \cos \varphi \\ \cos(\varphi - 90) \end{bmatrix} = \begin{bmatrix} 1 \\ \cos \varphi \\ \sin \varphi \end{bmatrix} \quad (3.6)$$

3.3.1.2. Tres micrófonos cardioides

Consiste en tres micrófonos cardioides de primer orden equiangulados entre sí: mientras uno apunta al frente, los otros dos apuntan a los ángulos de llegada de 120° y -120° respectivamente.

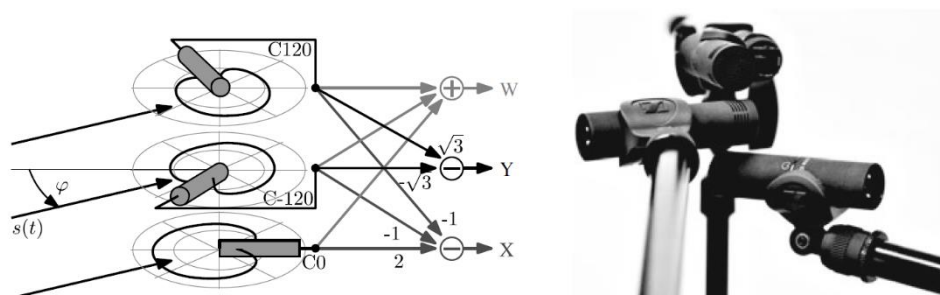


Fig. 3.5 Configuración de un array FOA 2D formado por tres micrófonos cardioides equiangulados

En este caso, es necesaria una codificación de las señales registradas por los micrófonos. Mientras el canal W consiste en una suma de las tres señales, los canales X e Y se construyen atendiendo al coseno de la diferencia entre el ángulo al que apunta cada micrófono (120° y -120°) y la dirección de los ejes cartesianos (0° para el eje x, y 90° para el eje y). Teniendo en cuenta la propiedades trigonométricas del coseno de la suma de dos ángulos, se obtiene la directividad del arreglo anterior:

$$D_{WXY}(\varphi) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} + \frac{1}{2} \cdot \cos \varphi \\ \frac{1}{2} + \frac{1}{2} \cdot \cos(\varphi + 120) \\ \frac{1}{2} + \frac{1}{2} \cdot \cos(\varphi - 120) \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ \frac{3}{4} \cdot \cos \varphi \\ \frac{3}{4} \cdot \sin \varphi \end{bmatrix} \quad (3.7)$$

Como puede verse, el proceso de codificación descrito altera los niveles de los canales. Por ello, para normalizar el resultado y mantener la energía, se multiplica el canal W por una constante compensatoria de $2/3$ (lo que implica una atenuación de 1,76 dB) y los canales XY por $4/3$ (ganancia de 1,25 dB).

Si se integra estas normalizaciones en la matriz de codificación, el resultado es el siguiente:

$$D_{WXY}(\varphi) = \begin{bmatrix} \frac{2}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{4}{3} & -\frac{2}{3} & -\frac{2}{3} \\ 0 & \frac{2\sqrt{3}}{3} & -\frac{2\sqrt{3}}{3} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} + \frac{1}{2} \cdot \cos \varphi \\ \frac{1}{2} + \frac{1}{2} \cdot \cos(\varphi + 120) \\ \frac{1}{2} + \frac{1}{2} \cdot \cos(\varphi - 120) \end{bmatrix} = \begin{bmatrix} 1 \\ \cos \varphi \\ \sin \varphi \end{bmatrix} \quad (3.8)$$

3.3.1.3. Cuatro micrófonos cardioides

Similar al anterior, con la diferencia de que esta vez las cápsulas están anguladas de manera ortogonal, por lo que la resolución en la captación del campo sonoro es ligeramente mayor.

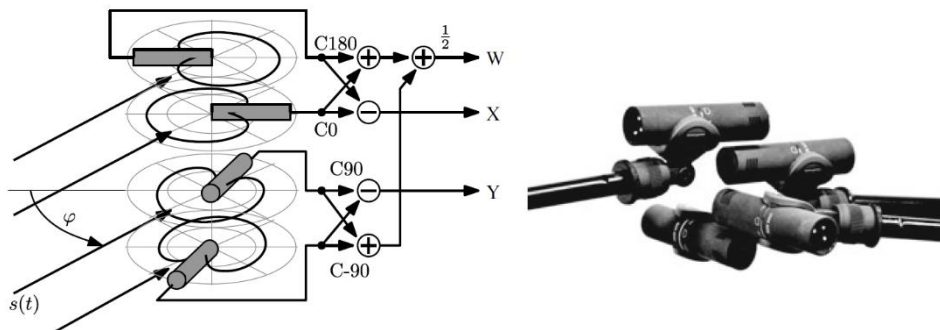


Fig. 3.6 Configuración de un array FOA 2D formado por cuatro micrófonos cardioides equiangulados

De nuevo, es necesaria una codificación. De la misma manera que en el arreglo anterior, el canal W se construye como una suma de todos los micrófonos, y los canales X e Y se construyen en función de la dirección de cada cápsula. En este caso sólo es necesario compensar el canal W, atenuándolo 3 dB.

Integrando la normalización del canal W en la matriz de codificación, se observa que el resultado es el mismo que en la técnica microfónica anterior:

$$D_{WXY}(\varphi) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} + \frac{1}{2} \cdot \cos \varphi \\ \frac{1}{2} + \frac{1}{2} \cdot \cos(\varphi + 180) \\ \frac{1}{2} + \frac{1}{2} \cdot \cos(\varphi - 90) \\ \frac{1}{2} + \frac{1}{2} \cdot \cos(\varphi + 90) \end{bmatrix} = \begin{bmatrix} 1 \\ \cos \varphi \\ \sin \varphi \end{bmatrix} \quad (3.9)$$

3.3.2. FOA 3D

Por simplicidad, en el espacio tridimensional se trabaja con vectores unitarios de dirección, tanto para la incidencia de las ondas de presión acústica como para las orientaciones de las cápsulas:

- θ_s = Vector unitario que indica la dirección de llegada de la onda al centro del array.
- θ_μ = Vector unitario que indica la dirección a la que apunta una cápsula del array.

Puesto que el módulo de ambos vectores es igual a uno, su producto escalar es igual al coseno del ángulo que forman entre sí, y debido a la simetría de revolución de la directividad de las cápsulas, este es equivalente al azimut φ :

$$\theta_\mu^T \theta_s = \cos(\varphi) \quad (3.10)$$

3.3.2.1. Un micrófono omnidireccional y tres bidireccionales (Formato B)

Consiste en un array formado por un micrófono omnidireccional en el centro de coordenadas (W), y tres bidireccionales ortogonales entre sí apuntando respectivamente a los ejes x, y, z. Puesto que esta es la definición de FOA en 3D, no es necesario llevar a cabo ninguna codificación: las cápsulas ya registran los canales WXYZ del formato B.

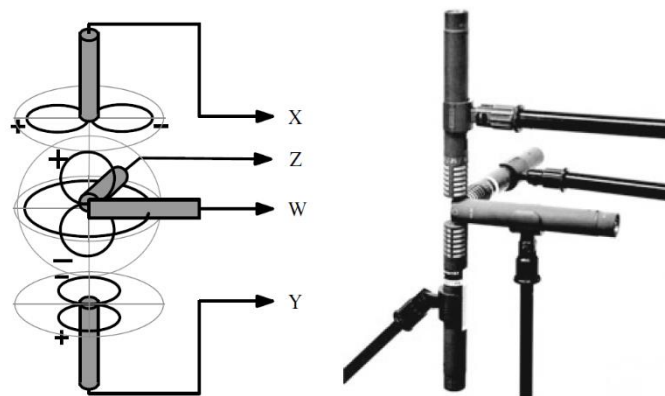


Fig. 3.7 Configuración de un array FOA 3D para la captación de señales ambisónicas en formato B

Teniendo en cuenta la función de directividad de una cápsula de gradiente de presión, y recordando el producto escalar mencionado, la directividad del arreglo de micrófonos es la siguiente:

$$D_{WXYZ}(\theta_s) = \begin{bmatrix} 1 \\ \theta_{\mu_x}^T \theta_s \\ \theta_{\mu_y}^T \theta_s \\ \theta_{\mu_z}^T \theta_s \end{bmatrix} \quad (3.11)$$

3.3.2.2. Un micrófono compuesto por cuatro cápsulas cardioides (Formato A)

Consiste en un array de cuatro cápsulas cardioides de primer orden, orientadas de forma tetraédrica y equianguladas 120° entre sí. Permite captar señales ambisónicas en formato A y, si bien puede llevarse a cabo mediante un arreglo de micrófonos, existen sistemas de captación constituidos por cuatro cápsulas no coincidentes que agilizan el proceso de grabación del campo sonoro. Los vectores unitarios de dirección de dichas cápsulas son los siguientes:

$$\theta_{Array} = \begin{bmatrix} \theta_{FLU}^T \\ \theta_{FRD}^T \\ \theta_{BLD}^T \\ \theta_{BRU}^T \end{bmatrix} = \frac{1}{\sqrt{3}} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix} \quad (3.12)$$

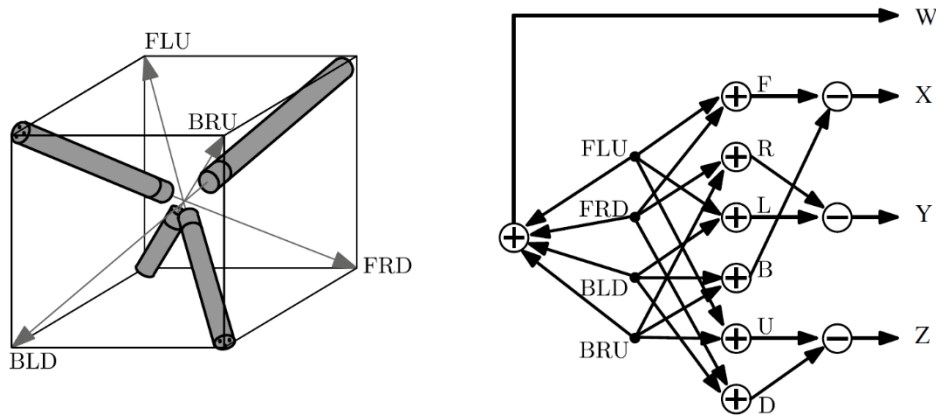


Fig. 3.8 Configuración de un array FOA 3D para captación de señales ambisónicas en formato A, y su conversión a formato B

Como se comentó anteriormente, es necesaria una codificación para pasar el cuarteto de canales del formato A al formato B. Una vez más, el canal W se construye mediante la suma de la señales registradas por cada una de las cápsulas, mientras que los canales XYZ se construyen a partir del producto escalar de los ejes cartesianos con los vectores de dirección de cada cápsula.

En la convención *Fu-Ma* la secuencia de canales es WXYZ y el canal W es atenuado 3 dB, por lo que para obtener este último se realiza la semisuma de las cápsulas FLU, FRD, BLD y BRU:

$$D_{WXYZ}(\theta_s) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} + \frac{1}{2} \cdot \theta_{FLU}^T \theta_s \\ \frac{1}{2} + \frac{1}{2} \cdot \theta_{FRD}^T \theta_s \\ \frac{1}{2} + \frac{1}{2} \cdot \theta_{BLD}^T \theta_s \\ \frac{1}{2} + \frac{1}{2} \cdot \theta_{BRU}^T \theta_s \end{bmatrix} \quad (3.13)$$

En el caso de la convención *AmbiX*, que es la más habitual, el orden de los canales es WYZX y el canal W no sufre ninguna atenuación en el proceso de codificación:

$$D_{WYZX}(\theta_s) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} + \frac{1}{2} \cdot \theta_{FLU}^T \theta_s \\ \frac{1}{2} + \frac{1}{2} \cdot \theta_{FRD}^T \theta_s \\ \frac{1}{2} + \frac{1}{2} \cdot \theta_{BLD}^T \theta_s \\ \frac{1}{2} + \frac{1}{2} \cdot \theta_{BRU}^T \theta_s \end{bmatrix} \quad (3.14)$$

Si se emplea un sistema de cuatro cápsulas en lugar de un arreglo de micrófonos coincidentes, gracias a esta codificación se logra obtener una señal ambisónica en formato B sin errores de fase entre canales producidos por la imposibilidad física de hacer coincidir exactamente en un mismo punto del espacio más de una cápsula de micrófono.

3.3.3. HOA

En el caso de la captación de señales ambisónicas de mayor orden, las técnicas microfónicas son conceptualmente similares a los arreglos de micrófonos descritos anteriormente: en última instancia, tratan de obtener diagramas polares que emulen los armónicos esféricos asociados a cada uno de sus canales.

Sin embargo, aunque dichas técnicas de captación espacial pueden realizarse con arreglos de micrófonos, estos resultan excesivamente tediosos y poco prácticos. Es por ello que se recurre a sistemas constituidos por varias cápsulas (Fig. 3.9), y que están especializados para la captación de *Ambisonics* en formato A para, posteriormente, codificar los canales y obtener el formato B.



Fig. 3.9 Micrófono para la captación de Ambisonics de tercer orden en Formato A

En la práctica, rara vez se emplean arrays de micrófonos. Mayoritariamente se opta por estos sistemas multicápsula, ya que facilitan y agilizan significativamente el proceso de captación del campo sonoro.

3.4. Codificación

En el apartado anterior se ha explicado algunas técnicas para grabar el campo sonoro. Sin embargo, también es posible simularlo a partir de la codificación de un conjunto de C señales mono $s_c(t)$, similar a cómo se hace en los sistemas de sonorización basados en objetos.

$$\chi(t) = \sum_{c=1}^C y_N(\theta_c) \cdot s_c(t) = \sum_{c=1}^C \begin{bmatrix} Y_0^0(\theta_c) \\ Y_1^{-1}(\theta_c) \\ Y_1^0(\theta_c) \\ \vdots \\ Y_n^n(\theta_c) \end{bmatrix} \cdot s_c(t) = \begin{bmatrix} ACN_0(t) \\ ACN_1(t) \\ ACN_2(t) \\ \vdots \\ ACN_{N-1}(t) \end{bmatrix} \quad (3.15)$$

Para ello, se asigna a cada señal un vector de dirección θ_c que determine la incidencia a simular; se calcula el valor de los armónicos para dicha incidencia y se multiplican por $s_c(t)$, obteniéndose la señal codificada en *Ambisonics*; por último, se suma todas las señales codificadas, obteniéndose la señal ambisónica $\chi(t)$ que emula el campo sonoro deseado.

3.5. Procesamiento

Existe una gran variedad de procesos y efectos que se pueden realizar sobre una señal ambisónica. Algunos de ellos son los siguientes:

3.5.1. Efecto espejo

Invirtiendo la polaridad de determinados canales, se puede aplicar sobre el campo sonoro un efecto espejo respecto a cada eje de coordenadas. Para ello, simplemente hay que multiplicar la señal ambisónica de orden N , compuesta por $K = (N + 1)^2$ canales, por una matriz diagonal $K \times K$.

$$\tilde{\chi}(t) = \text{diag}\{c_{nm}\} \cdot \chi(t) \quad (3.16)$$

El valor de estos coeficientes depende de n , m y de la dimensión del espacio que se desea invertir:

Conmutar	Canales que hay que invertir	c_{nm}
Delante/Detrás	Aquellos cuyo armónico sea simétrico impar respecto al plano $x = 0$.	$(-1)^{m+(m<0)}$
Izq./Dcha.	Aquellos cuyo armónico sea simétrico impar respecto al plano $y = 0$.	$(-1)^{m<0}$
Arriba/Abajo	Aquellos cuyo armónico sea simétrico impar respecto al plano $z = 0$.	$(-1)^{n+m}$

Tabla 3.4 Efectos espejo sobre una señal ambisónica

3.5.2. Rotación

La rotación del campo sonoro se lleva a cabo mediante un producto matricial. Dada una matriz de rotación respecto a los ejes cartesianos x (φ), y (ϑ), z (γ):

$$\mathbf{R}(\varphi, \vartheta, \gamma) = \begin{bmatrix} \cos \varphi & -\operatorname{sen} \varphi & 0 \\ \operatorname{sen} \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos \vartheta & 0 & -\operatorname{sen} \vartheta \\ 0 & 1 & 0 \\ \operatorname{sen} \vartheta & 0 & \cos \vartheta \end{bmatrix} \cdot \begin{bmatrix} \cos \gamma & -\operatorname{sen} \gamma & 0 \\ \operatorname{sen} \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.17)$$

La señal rotada se obtiene modificando las proporciones de cada canal:

$$\tilde{\mathbf{X}}(t) = \mathbf{y}_N(\theta) \cdot \mathbf{y}_N^T(\mathbf{R}^T \theta) \cdot \mathbf{X}(t) \quad (3.18)$$

3.5.3. Ecuación y compresión

Ambisonics es sensible a los retardos, cambios de nivel y distorsiones de fase ya que, como la información espacial del campo sonoro se encuentra fragmentada entre todos los canales, alterar el equilibrio de todos ellos puede suponer una distorsión no deseada del campo sonoro.

Si se desea llevar a cabo una ecualización sobre una señal ambisónica, el filtro ha de ser el mismo para todos los canales, de manera que todos sufran la misma coloración tanto en amplitud como en fase.

De igual manera, si se desea comprimir la señal, la atenuación llevada a cabo por el VCA del compresor ha de ser común para todos los canales: la entrada en el circuito detector (*key*) y los parámetros del compresor han de ser los mismos. Además, para que la compresión dependa sólo de la energía media de la señal y no de su información espacial, se debe emplear la técnica *side-chain* con el canal fundamental como señal de *key*, ya que su armónico es el único que pondera por igual todas las direcciones del campo sonoro. De esta manera se logra controlar la sonoridad de la señal sin afectar a su información espacial.

Si se desease ecualizar o comprimir exclusivamente una fuente sonora del campo sonoro, habría que llevar a cabo un proceso de decodificación para extraerla:

$$s_c(t) = \mathbf{y}_N^T(\theta_c) \cdot \mathbf{X}(t) \quad (3.19)$$

Hecho esto, se procesaría por separado y luego se codificaría y sumaría de nuevo con el campo sonoro.

3.6. Decodificación y reproducción

Para reproducir los canales de la señal ambisónica, antes es necesario llevar a cabo una decodificación que dependerá del orden, el número de dimensiones a recrear y el sistema de reproducción.

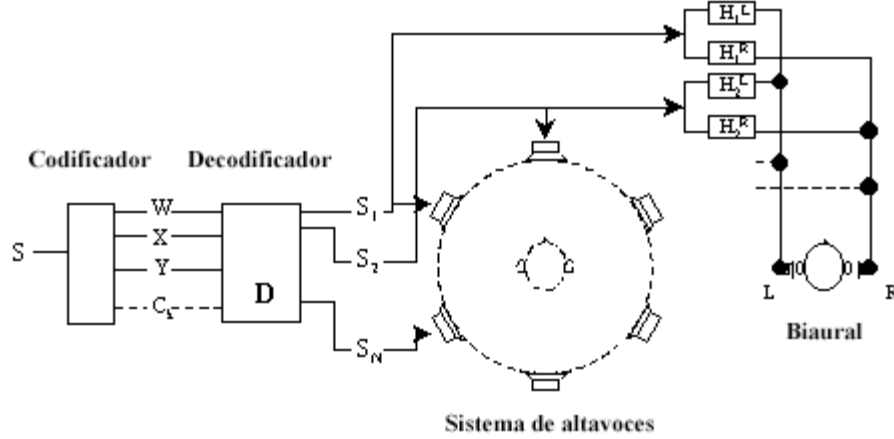


Fig. 3.10 Diagrama de bloques de la codificación y decodificación en Ambisonics

3.6.1. Reproducción en 2D

3.6.1.1. Sistema de altavoces

Sea una onda sonora $s(t)$ con un ángulo de incidencia φ_s en el centro del array de micrófonos, y sea $\chi_{2D}(t)$ la señal ambisónica de $2N + 1$ canales registrada en formato B. Si se obvia las transducciones acústico-mecánica y mecánico-eléctrica, asumiéndose ideales e idénticas, la relación entre las señales es descrita por la directividad asociada a los canales: los armónicos bidimensionales $[\phi_{-n}(\varphi_s), \dots, \phi_n(\varphi_s)]$.

$$\chi_{2D}(t) = \begin{bmatrix} \phi_{-n}(\varphi_s) \\ \vdots \\ \phi_n(\varphi_s) \end{bmatrix} \cdot s(t) \quad (3.20)$$

Dado un sistema de $L \geq 2N + 2$ altavoces iguales, distribuidos uniformemente en el plano azimutal, equidistantes al *sweet spot* y situados respectivamente en los azimuts $\{\varphi_1, \dots, \varphi_L\}$, cada altavoz ha de reproducir una señal $S_l(t)$ construida a partir de unas proporciones de los canales de $\chi_{2D}(t)$. Estas proporciones dependen de la ubicación de cada altavoz y de unos coeficientes $a_{|n|}$:

$$\begin{bmatrix} S_1(t) \\ \vdots \\ S_L(t) \end{bmatrix} = \sqrt{\frac{2\pi}{L}} \cdot \begin{bmatrix} \phi_{-n}(\varphi_1) & \cdots & \phi_n(\varphi_1) \\ \vdots & \ddots & \vdots \\ \phi_{-n}(\varphi_L) & \cdots & \phi_n(\varphi_L) \end{bmatrix} \cdot \text{diag}\{a_{|n|}\} \cdot \chi_{2D}(t) \quad (3.21)$$

Si se sustituye la señal ambisónica $\chi_{2D}(t)$ por su relación con $s(t)$, se llega a la siguiente expresión:

$$\begin{bmatrix} S_1 \\ \vdots \\ S_L \end{bmatrix} = \sqrt{\frac{2\pi}{L}} \cdot \begin{bmatrix} \sum_{m=-n}^n a_{|m|} \phi_m(\varphi_1) \phi_m(\varphi_s) \\ \vdots \\ \sum_{m=-n}^n a_{|m|} \phi_m(\varphi_L) \phi_m(\varphi_s) \end{bmatrix} \cdot s(t) \quad (3.22)$$

Agrupando los sumatorios, se puede extraer la función de panoramización $g(\varphi, \varphi_s)$, dependiente de la dirección φ_s de la fuente sonora y del parámetro azimuth φ resultante de la superposición de todos los altavoces del sistema de reproducción:

$$g(\varphi, \varphi_s) = \frac{\{S_l(t)\}}{s(t)} = \sum_{m=-n}^n a_{|m|} \phi_m(\varphi) \phi_m(\varphi_s) \quad (3.23)$$

Si se particulariza para FOA ($n = 1$) y unos coeficientes $a_0 = 2\pi$ y $a_1 = a\pi$, y recordando la propiedad trigonométrica del coseno de la diferencia de ángulos, se obtiene la función de un patrón de directividad cardioide de primer orden apuntando a φ_s , con el parámetro a ponderando la componente bidireccional:

$$g(\varphi, \varphi_s) = 1 + a \cdot \cos(\varphi - \varphi_s) \quad (3.24)$$

La conclusión es que, al recrear el campo sonoro en torno al usuario, el sonido $s(t)$ es reconstruido de tal manera que la respuesta conjunta de los L altavoces genera un patrón virtual de directividad cardioide centrado en el *sweet spot* y alineado con la fuente.

Como ya se mencionó, a mayor orden de *Ambisonics* mayor resolución espacial. Esto es porque el orden de la señal se corresponde con el del patrón cardioide $g(\varphi, \varphi_s)$: un orden infinito implicaría una directividad infinita o, lo que es lo mismo, una delta alineada con la fuente (panoramización exacta):

$$g(\varphi, \varphi_s) = \sum_{m=-\infty}^{\infty} \phi_m(\varphi) \phi_m(\varphi_s) = \delta(\varphi - \varphi_s) \quad (3.25)$$

Una señal con infinitos canales no es viable; se emplea un orden finito junto a unos coeficientes $a_{|m|}$ encargados de maximizar la localización. Como a diferencia de los sistemas basados en canales y objetos, no hay un conjunto de ganancias g sino una función continua, el vector de localización r_γ ya no se calcula como un sumatorio sino como una integral. Si se fija $\varphi_s = 0$ para simplificar los cálculos, se tiene que:

$$r_\gamma = \frac{\int_0^{2\pi} |g(\varphi)|^\gamma \cdot \cos \varphi \, d\varphi}{\int_0^{2\pi} |g(\varphi)|^\gamma \, d\varphi} \quad (3.26)$$

Existen diferentes modelos de decodificación. Uno muy popular es el $max-r_E$, que busca maximizar el vector de localización basado en el modelo energético, descrito en el apartado [2.3.3](#).

Para ello, se fija un slope $\gamma = 2$ y se sustituye en la expresión del vector de localización (3.26) la función de panoramización (3.24), particularizando para una incidencia en el eje $\varphi_s = 0$ para simplificar los cálculos. El resultado es la función a maximizar, que depende del coeficiente a que pondera la componente bidireccional del patrón de directividad:

$$r_E = \frac{\int_0^{2\pi} (1 + a \cdot \cos \varphi)^2 \cdot \cos \varphi \, d\varphi}{\int_0^{2\pi} (1 + a \cdot \cos \varphi)^2 \, d\varphi} = \frac{2a}{a^2 + 2} \quad (3.27)$$

Recordando que una localización perfecta implicaría $|r_E| = 1$ y que una localización indeterminada implicaría $|r_E| = 0$, el modelo $max-r_E$ trata de buscar los coeficientes que maximicen la expresión anterior y que, en consecuencia, otorguen la mejor localización dentro de las limitaciones provocadas por acotar el orden de la señal ambisónica. En este caso concreto, el coeficiente a que maximiza la localización resulta en:

$$a = \arg \max_a \{ |r_E| \} = \arg \max_a \left\{ \frac{2a}{a^2 + 2} \right\} = \sqrt{2} \quad (3.28)$$

La conclusión es la siguiente: incluso en el mejor de los casos (campo libre ideal, ruido de fondo nulo y altavoces idénticos, repartidos de manera homogénea y equidistantes al usuario), la mejor localización que puede otorgar un FOA bidimensional es $|r_E| = 0,707$. Por tanto, en un caso práctico siempre sucederá que $|r_E| < 0,707$; para “empujar” este límite y mejorar la localización más allá de esa cota, resulta inevitable tener que aumentar el orden de *Ambisonics*.

3.6.1.2. Auriculares

La reproducción de *Ambisonics* en auriculares requiere de sensores de movimiento para rastrear en todo momento la orientación de la cabeza del usuario en el plano horizontal, de manera que se aplique los procesos de rotación oportunos a la señal sobre los canales WXY, antes de la decodificación. De esta manera se logra que el oyente experimente una interacción con el campo sonoro recreado.

La decodificación binaural emplea la decodificación FOA 2D para mapear la señal en un set de altavoces virtuales, obteniendo la señal binaural mediante una HRTF y estas señales decodificadas.

Sea un hipotético set de $L = 4$ altavoces FLBR distribuidos en cruz en torno a un *sweet spot*:

- **F, Front.** Situado a $\varphi_F = 0^\circ$, reproduce la señal S_F .
- **L, Left.** Situado a $\varphi_L = 90^\circ$, reproduce la señal S_L .
- **B, Back.** Situado a $\varphi_B = 180^\circ$, reproduce la señal S_B .
- **R, Right.** Situado a $\varphi_R = 270^\circ$, reproduce la señal S_R .

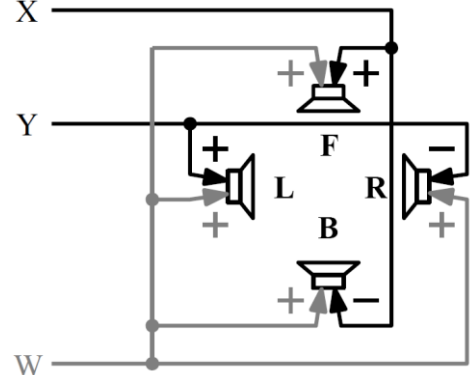


Fig. 3.11 Configuración de un sistema FLBR

$$\begin{bmatrix} S_F \\ S_L \\ S_B \\ S_R \end{bmatrix} = \frac{1}{2} \cdot \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} W \\ X \\ Y \end{bmatrix} \quad (3.29)$$

Los canales LR de la señal binaural se construyen respectivamente a partir de la convolución de las señales $\{S_l\}$ con las HRIR izquierda (LHRIR) y derecha (RHRIR) para los ángulos de incidencia correspondientes a los ángulos de los altavoces del sistema FLBR:

$$\begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} LHRIR_{0^\circ} * S_F + LHRIR_{90^\circ} * S_L + LHRIR_{180^\circ} * S_B + LHRIR_{270^\circ} * S_R \\ RHRIR_{0^\circ} * S_F + RHRIR_{90^\circ} * S_L + RHRIR_{180^\circ} * S_B + RHRIR_{270^\circ} * S_R \end{bmatrix} \quad (3.30)$$

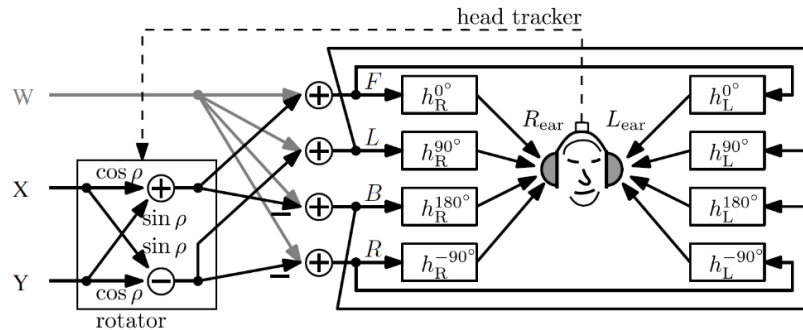


Fig. 3.12 Configuración de un convertor FOA 2D/Binaural con rastreo de la orientación de la cabeza del usuario

3.6.2. Reproducción en 3D

3.6.2.1. Sistema de altavoces

La decodificación a realizar para la reproducción tridimensional se basa en los mismos principios que en el caso bidimensional, con la diferencia de que ahora se introduce la componente de altura a través de la elevación ϑ .



Fig. 3.13 Sistema de altavoces para la reproducción de HOA 3D

Sea una onda sonora $s(t)$ con una incidencia sobre el centro del array de micrófonos determinada por el vector de dirección θ_s , y sea $\chi_{3D}(t)$ la señal ambisónica de $K = (N + 1)^2$ canales registrada en formato B. De nuevo, la relación entre ambas señales estará determinada por la directividad asociada a cada canal. Puesto que se trata de una reproducción 3D, esta vez estas directividades se expresan mediante los armónicos esféricos $Y_n^m(\theta_c)$:

$$\chi_{3D}(t) = \begin{bmatrix} ACN_0(t) \\ ACN_1(t) \\ ACN_2(t) \\ \vdots \\ ACN_{K-1}(t) \end{bmatrix} = \begin{bmatrix} Y_0^0(\theta_c) \\ Y_1^{-1}(\theta_c) \\ Y_1^0(\theta_c) \\ \vdots \\ Y_N^N(\theta_c) \end{bmatrix} \cdot s(t) \quad (3.31)$$

Se desea reproducirla en un sistema de $L \geq (N + 1)^2$ altavoces distribuidos uniformemente y de manera equidistante en torno al *sweet spot*, siendo sus vectores de dirección $\{\theta_1, \dots, \theta_L\}$. Cada altavoz reproduce una señal $S_l(t)$, construida de manera análoga al caso bidimensional:

$$\begin{bmatrix} S_1(t) \\ \vdots \\ S_L(t) \end{bmatrix} = \sqrt{\frac{4\pi}{L}} \cdot \begin{bmatrix} Y_0^0(\theta_1) & Y_1^{-1}(\theta_1) & Y_1^0(\theta_1) & \dots & Y_N^N(\theta_1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_0^0(\theta_L) & Y_1^{-1}(\theta_L) & Y_1^0(\theta_L) & \dots & Y_N^N(\theta_L) \end{bmatrix} \cdot \text{diag}\{a_{|n|}\} \cdot \chi_{3D}(t) \quad (3.32)$$

De nuevo, sustituyendo $\chi_{3D}(t)$ por su relación con $s(t)$:

$$\begin{bmatrix} S_1 \\ \vdots \\ S_L \end{bmatrix} = \sqrt{\frac{4\pi}{L}} \cdot \begin{bmatrix} \sum_{n=0}^N \sum_{m=-n}^n |a_n| Y_n^m(\theta_1) Y_n^m(\theta_s) \\ \vdots \\ \sum_{n=0}^N \sum_{m=-n}^n |a_n| Y_n^m(\theta_L) Y_n^m(\theta_s) \end{bmatrix} \cdot s(t) \quad (3.33)$$

La función de panoramización que se obtiene es la siguiente:

$$g(\theta, \theta_s) = \sum_{n=0}^N \sum_{m=-n}^n |a_n| Y_n^m(\theta) Y_n^m(\theta_s) \quad (3.34)$$

La siguiente fórmula es una particularización para FOA. Como no se puede describir los armónicos $Y_n^m(\theta)$ en función de vectores, se descompone el vector θ en azimut y altura (φ, ϑ) :

$$g(\theta, \theta_s) = \frac{|a_0|}{4\pi} + \frac{|a_1|}{4\pi} (\cos \vartheta \cos \vartheta_s + \sin \vartheta \sin \vartheta_s \cos(\varphi - \varphi_s)) \quad (3.35)$$

Se puede comprobar que la función de panoramización es de nuevo un patrón cardioide apuntando a la fuente, pero esta vez en el espacio tridimensional: al igualar $\varphi = \varphi_s$ se obtiene una función cardioide con una componente bidireccional $\cos(\vartheta - \vartheta_s)$, lo que indica que la elevación también apunta a la fuente.

En el caso 3D, el decodificador *max-r_E* emplea unos coeficientes dependientes de los polinomios de Legendre, tal que $a_n = P_n(\cos \frac{137,9}{n+1,51})$.

3.6.2.2. Auriculares

Esta decodificación está muy extendida en VR. El principio de funcionamiento es equivalente al del caso bidimensional, con la salvedad de que se introduce la componente de elevación (canal Z) tanto en la decodificación como en los procesamientos llevados a cabo sobre la señal: se añaden filtros basados en la HRIR para elevaciones distintas de 0°, y se aplican las rotaciones convenientes para lograr que el usuario interactúe con el campo sonoro no solo moviendo la cabeza a los lados, sino también arriba y abajo.

4. CONSTRUCCIÓN DEL MODELO 3D

4.1. Contexto del Teatro Buero Vallejo

El Teatro Buero Vallejo de Guadalajara, situado en el centro de la ciudad, concretamente en la Calle de Cifuentes, número 30, es el más grande de la ciudad. Junto al Teatro Moderno, acoge los principales eventos culturales de la localidad. Toma su nombre del dramaturgo alcarreño Antonio Buero Vallejo, y organiza eventos musicales, teatrales, de danza, conferencias y monólogos, entre otros.



Fig. 4.1 Foto exterior del acceso principal al edificio (izq.) y foto interior del hall (dcha.)

El edificio es de obra pública y fue construido entre los años 1998 y 2002, suponiendo un coste total de 9.738.316 €. Suele acoger una media de 38.000 espectadores cada año, y realiza en torno a 50 espectáculos por temporada.

Su atractivo principal es el auditorio, el cual cuenta con la capacidad de albergar eventos de diversa índole, y con un aforo de hasta 1.003 espectadores.

Este teatro acoge eventos de toda clase: desde conciertos de música de diversos géneros, como rock, pop, jazz, flamenco o música clásica; hasta zarzuelas y representaciones de obras teatrales de todo tipo, pasando por eventos de danza, exhibiciones de magia y monólogos de comedia.



Fig. 4.2 Evento teatral con música en vivo en el auditorio

Además del auditorio, el edificio cuenta también con una sala más pequeña orientada a la realización de exposiciones. Cuenta con un aforo de hasta 133 localidades y una superficie de 180 m².

En ella se suele llevar a cabo conferencias relacionadas con la pintura, la historia, la lectura o el arte en general; presentaciones de libros y otro tipo de eventos de la mano de determinadas asociaciones culturales.



Fig. 4.3 Asociación "Libros y Más" en sala de exposiciones

4.2. Descripción del auditorio

El teatro consta de un total de 1.003 butacas repartidas en siete áreas de audiencia: cuatro de ellas en la platea y las otras tres en el anfiteatro. Cuenta con un pequeño foso de orquesta de profundidad configurable, y un escenario bastante amplio de 465 m².



Fig. 4.4 Fotografía desde el escenario del área de audiencia del auditorio

Las paredes laterales cuentan con acoplamientos para poder colocar hasta doce cortinas triangulares con las que reducir el tiempo de reverberación en alta frecuencia. Estas parecían ser bastante ligeras y de un grosor no muy excesivo, por lo que posiblemente su uso sea principalmente estético.



Fig. 4.5 Fotografía del escenario del auditorio desde el anfiteatro

4.2.1. Patio de butacas

La platea consta de dos patios de butacas principales situados en el centro, y dos secciones laterales ligeramente elevadas y encerradas por un pequeño muro de madera. Cada patio central de butacas consta de 17 filas y 244 asientos, mientras que cada patio lateral está formado por 15 filas y 73 butacas, habiendo un total de 634 butacas en la platea.

Hay unas escaleras que dividen los dos patios centrales y que están parcialmente cubiertas por una moqueta verde de fieltro de escaso grosor. Tanto las escaleras situadas entre las paredes y los patios laterales como el suelo que conecta arriba con ellas, están completamente cubiertas por la misma moqueta. Al fondo de la platea se encuentra el acceso principal al teatro, al cual se accede desde el hall.

4.2.2. Anfiteatro

Consta de una grada principal situada en el centro y dos gradas laterales a los lados. La grada principal cuenta con 10 filas y 209 asientos, mientras que cada grada lateral está formada por nueve filas y 80 butacas. El número total de asientos en el anfiteatro es de 369.

Tanto las escaleras que dividen las gradas, como las que se sitúan entre las gradas laterales y las paredes, así como el suelo del fondo del anfiteatro, están formados por el mismo material que el suelo que conecta el patio de butacas con el foso de la orquesta.

En el fondo del anfiteatro hay cinco ventanas sin acristalar que dan acceso a cuatro pequeñas habitaciones. También hay un acceso formado por tres puertas dobles.

4.2.3. Foso de orquesta

Tiene una superficie de unos 55 m², una anchura de 19 metros y su profundidad es modulable en un rango de 4,60 metros.

El material de las paredes que lo cubren es similar al de los muros de madera que separan los patios laterales de butacas de los centrales, y su suelo es similar al de las escaleras de las gradas.

4.2.4. Escenario

La superficie de la planta del escenario es de aproximadamente 465 m². Su altura al techo llega a los 20 metros y cuenta con seis bambalinas que en su día se encontraban recogidas.

Cuenta con un pequeño proscenio que termina en punta y que conecta con el foso de orquesta.

El material del suelo aparece especificado en la página oficial del teatro: doble contrachapado de okumen sobre tablero de DM (fibra de densidad media).

4.3. Proceso de construcción

4.3.1. Lectura de los planos en AutoCAD

Tras contactar con el Ayuntamiento de Guadalajara y realizar una solicitud en su sede electrónica, aportando la documentación pertinente, se obtuvo los planos del auditorio en formato PDF.

A partir de este formato, es posible importar los planos en un software de diseño 2D como AutoCAD. Si se ajusta la escala de los planos importados a 1:1 y se fija el origen de coordenadas en el mismo punto que en el módulo *Room Editor* de EASE, se puede ir reconstruyendo punto a punto el modelo 3D, a partir de las coordenadas y las mediciones de distancia que proporciona AutoCAD.

En este caso, el origen de coordenadas se situó en el centro del escenario, a ras de suelo, a 1,80 metros de la punta del proscenio que conecta con el foso de orquesta, orientando el eje y positivo hacia el público, el eje x positivo hacia la derecha y el eje z positivo hacia el techo. De esta manera, para cada punto del auditorio se puede obtener las coordenadas (x, y) a partir de los planos en planta, y la coordenada z a partir de los planos en sección, replicando su localización en EASE.

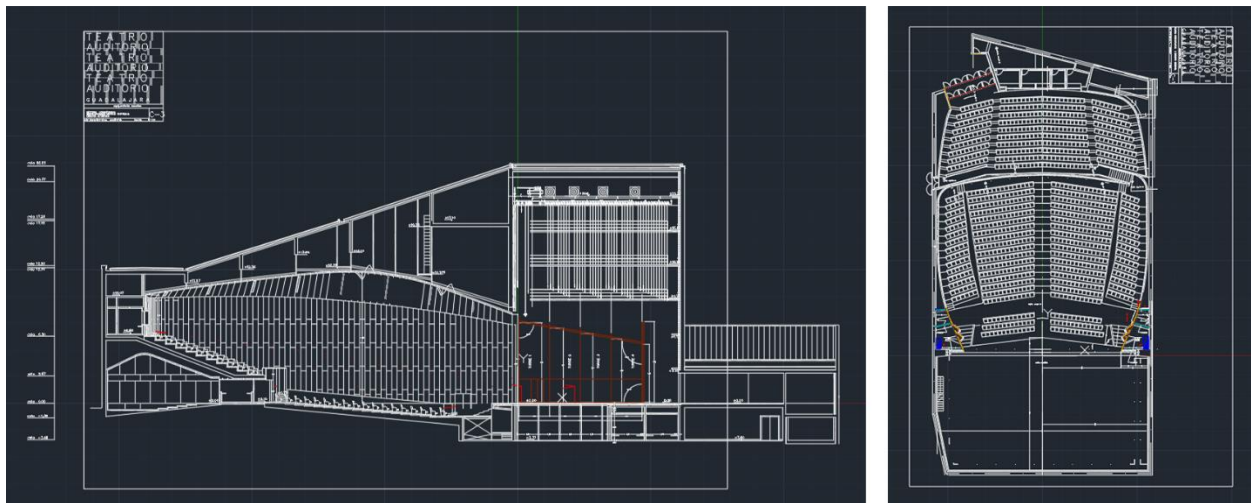


Fig. 4.6 Planos del auditorio importados en AutoCAD: vistas en sección (izq.) y planta (dcha.)

Cabe mencionar que el foso de orquesta no queda fielmente reflejado en los planos; en su lugar, aparecen dos secciones de tres filas de butacas muy próximas al escenario que, en el auditorio real, no existen. Por tanto, fue necesario medir *in situ* las dimensiones del foso, que se encontraba bajado, para poder reconstruirlo correctamente en EASE.

De paso, y aprovechando la oportunidad de medir *in situ*, se realizaron algunas mediciones generales de las dimensiones del teatro para comprobar la fidelidad de los planos proporcionados por el Ayuntamiento: salvando el foso de orquesta, los planos eran fiables.

4.3.2. Construcción en EASE Room Editor

Para agilizar la construcción, y antes de empezar a dibujar en *Room Editor*, dentro de *Edit/Room Data* se marcó la casilla “*Room Symmetric*”, de manera que por cada elemento que se dibujase en los cuadrantes del eje positivo de las x , EASE crease automáticamente una réplica espejo respecto al eje y . Una vez obtenida una aproximación simétrica del teatro, se desmarcó dicha casilla y se procedió a dibujar aquellas partes del teatro que no eran simétricas.

El software EASE presenta una importante limitación en el diseño del modelo: las superficies han de ser siempre planas. Así, en caso de que el auditorio conste de alguna pared u objeto convexo o cóncavo, es necesario realizar una aproximación mediante varias superficies planas.

Una vez se ha dibujado en *Room Editor* los puntos que constituyen los vértices de una superficie determinada, se crea dicha superficie uniendo los puntos mediante la herramienta “*Insert Face*”. Si se unen los puntos en sentido horario, la superficie se resaltaré en amarillo (la cara expuesta al campo sonoro); por el contrario, si se unen en sentido antihorario, la superficie aparecerá resaltada en color blanco (la cara expuesta al exterior). En caso de que ambas caras de la superficie estén expuestas al campo sonoro, como puede ser el caso de una cortina o el telón del escenario, ha de especificarse en las propiedades de dicha cara, resaltándose entonces la superficie en color naranja.

A continuación, se muestra tres ejemplos sobre las posibles disposiciones de las caras en EASE: una sección del techo resaltada en blanco (izquierda), una sección de la pared resaltada en amarillo (centro) y una cortina resaltada en naranja (derecha).

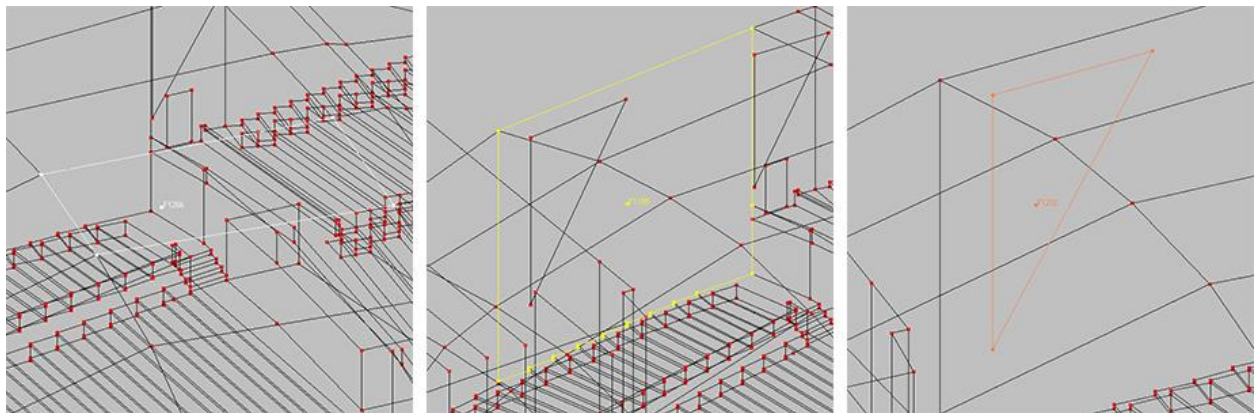


Fig. 4.7 Ejemplos de las posibles disposiciones de las caras en EASE Room Editor

Por último, y tras haber comprobado que en *Edit/Room Data* la casilla “*Room Open*” estaba desmarcada, mediante la herramienta *Tools/Check Holes* se comprobó que no existiesen fisuras ni agujeros en el modelo. Una vez que dicha herramienta dio el visto bueno, se consideró que la geometría del modelo estaba finalizada; el siguiente paso era la elección y asignación de materiales, con el correspondiente proceso de calibración para ajustar los coeficientes de absorción y *scattering* hasta aproximar las simulaciones a las medidas *in situ*.

4.4. Geometría final del modelo

Tras dedicarle mucha paciencia y un tiempo considerable, se obtuvo el modelo final del auditorio, el cual consta de un total de 1.697 vértices y 1.324 caras. El volumen total que encierra el modelo es de 17.253 m³, con una superficie interior total de 8.578 m². Si se omite el escenario, el volumen es de aproximadamente unos 8.000 m³.

Por motivos de simplificación, cualquier superficie con un relieve menor de 30 cm fue considerada plana de cara a la construcción del modelo, contrarrestando el error de aproximación mediante el factor de *scattering* en el proceso de elección de materiales. De la misma manera, y por los mismos motivos, las filas de butacas se aproximaron a escalones de igual altura y profundidad: 86 cm y 89 cm respectivamente. También se obvió toda la estructura del teatro que no estaba directamente expuesta al campo sonoro dentro del auditorio, así como la tramoya y la estructura de suspensión de elementos en el escenario.

Inicialmente se construyó una versión simple del teatro, obviando los detalles y dibujando sólo las paredes y el suelo hasta obtener la forma aproximada del recinto. Hecho esto, se dibujó las escaleras y las butacas. Posteriormente se añadieron detalles como las cortinas, las bambalinas, el telón o la moqueta. Más adelante se dibujó el techo y, por último, se desactivó la funcionalidad de simetría y se dibujó las puertas y los accesos a la platea y el anfiteatro.

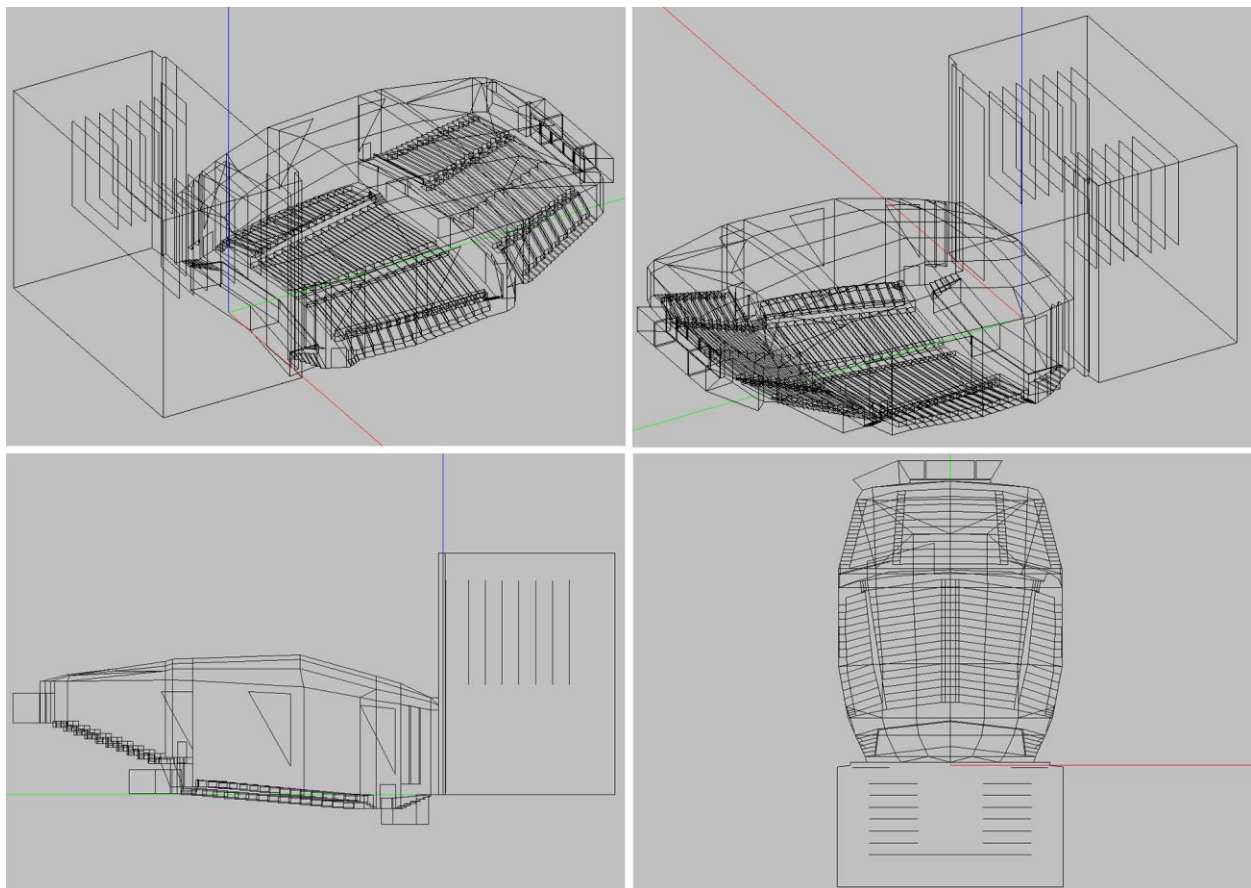


Fig. 4.8 Vista del modelo 3D en Room Editor: perspectiva (arriba), sección (abajo izq.) y planta (abajo drcha.)

4.5. Áreas de audiencia

Una vez construido el modelo, se procedió a dibujar las áreas de audiencia. Puesto que en cada área de audiencia el oído de los espectadores se encuentra en un mismo plano, se dibujó un plano de escucha común a todas las filas de butacas. Para ello, por cada área de audiencia, se creó una serie de superficies auxiliares a ras de suelo formadas por cuatro vértices cada una, de tal manera que estas cubrieran toda el área de la sección de butacas. Haciendo click derecho en cada una de ellas, se seleccionó “Area above face”, y se estableció un plano de escucha a 1,20 metros, que suele ser la altura a la que se sitúa el oído de una persona sentada. Ya creadas todas las áreas de audiencia, se eliminaron las superficies auxiliares. Puesto que también resulta interesante conocer el SPL directo existente en el escenario, también se estableció un área de audiencia en él, pero esta vez a 1’60 metros del suelo.

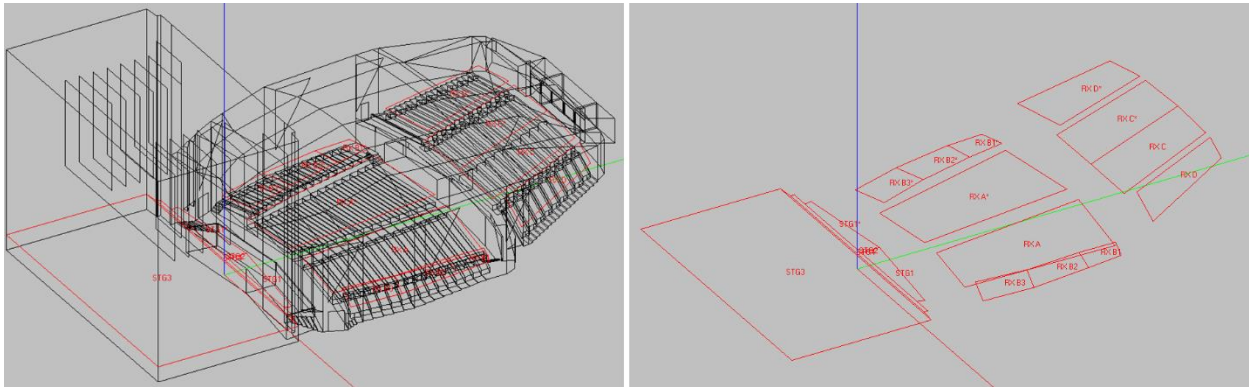


Fig. 4.9 Áreas de audiencia resaltadas en rojo, con (izq.) y sin (dcha.) las superficies reflectantes

Área	Sub áreas que la forman (cuatro vértices)	Altura respecto al suelo (m)	Descripción
RX A	<i>RX A</i>	1,20	<i>Plano de escucha del patio central derecho de butacas.</i>
RX A*	<i>RX A*</i>	1,20	<i>Plano de escucha del patio central izquierdo de butacas.</i>
RX B	<i>RX B1, RX B2, RX B3</i>	1,20	<i>Plano de escucha del patio lateral derecho de butacas.</i>
RX B*	<i>RX B1*, RX B2*, RX B3*</i>	1,20	<i>Plano de escucha del patio lateral izquierdo de butacas.</i>
RX C	<i>RX C</i>	1,20	<i>Plano de escucha de la sección central del anfiteatro (mitad derecha).</i>
RX C*	<i>RX C*</i>	1,20	<i>Plano de escucha de la sección central del anfiteatro (mitad izquierda).</i>
RX D	<i>RX D</i>	1,20	<i>Plano de escucha de la sección lateral derecha del anfiteatro.</i>
RX D*	<i>RX D*</i>	1,20	<i>Plano de escucha de la sección lateral izquierda del anfiteatro.</i>
STG (stage)	<i>STG1, STG2, STG3, STG4</i>	1,60	<i>Plano de escucha del escenario.</i>

Tabla 4.1 Áreas de audiencia

4.6. Visualización en EASE Vision

Debido a que el número de caras y líneas es bastante elevado, la visualización de la geometría del modelo en el módulo *Room Editor* puede resultar muy confusa si se hace desde determinadas perspectivas, ya que las líneas y caras se superponen unas sobre otras y la percepción de profundidad se ve comprometida.

Para solucionarlo, EASE proporciona el módulo *Vision*, que se ocupa del renderizado de las superficies de nuestro modelo para mostrarlas en escala de grises, atendiendo a sus orientaciones respecto al punto de vista del usuario. De esta manera EASE permite visualizar, fácilmente y sin confusiones, el modelo 3D y comprobar que su forma y la disposición de las superficies es correcta y fiel al auditorio real, tomando en cuenta las aproximaciones realizadas en el proceso de construcción.

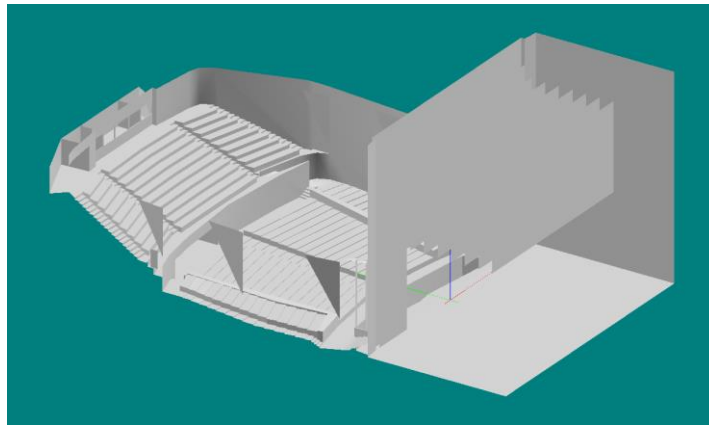


Fig. 4.10 Vista exterior del modelo 3D en perspectiva en EASE Vision

Este módulo permite además visualizar en primera persona el auditorio por dentro. Con ayuda de dos fuentes de luz, una omnidireccional situada en el centro del escenario y otra en el techo de la platea, *Vision* realiza un trazado de rayos por todo el recinto para emular una iluminación realista, lo que facilita aún más la visualización de la geometría del auditorio y la disposición de los distintos obstáculos:

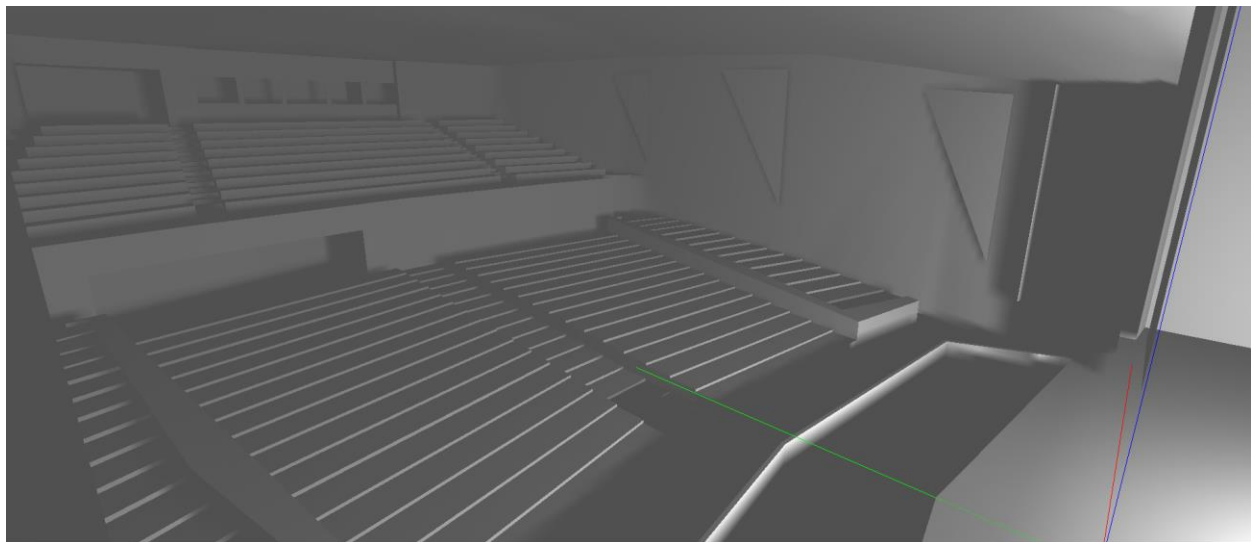


Fig. 4.11 Vista interior en Vision, desde la esquina superior izq. de la boca del escenario, con una fuente de luz en el escenario

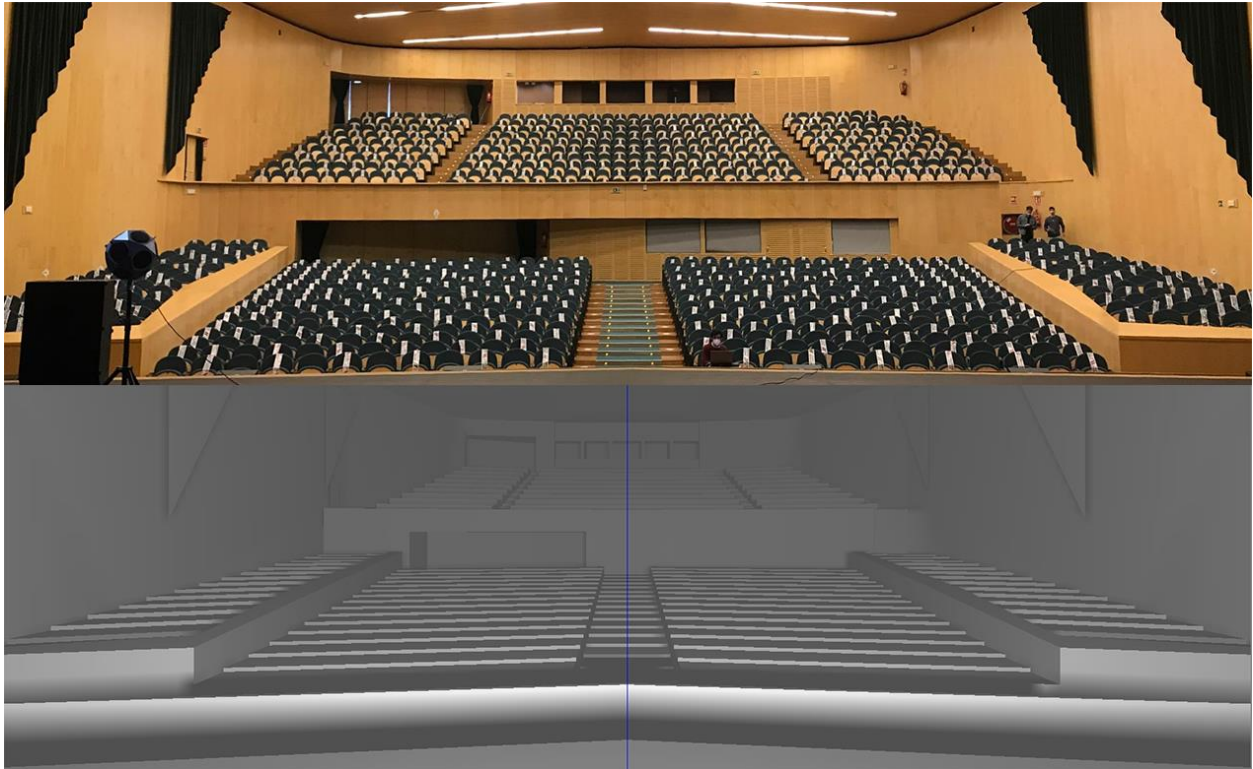


Fig. 4.12 Comparativa entre el teatro y el modelo en Vision, desde el origen de coordenadas, con fuente de luz en el escenario

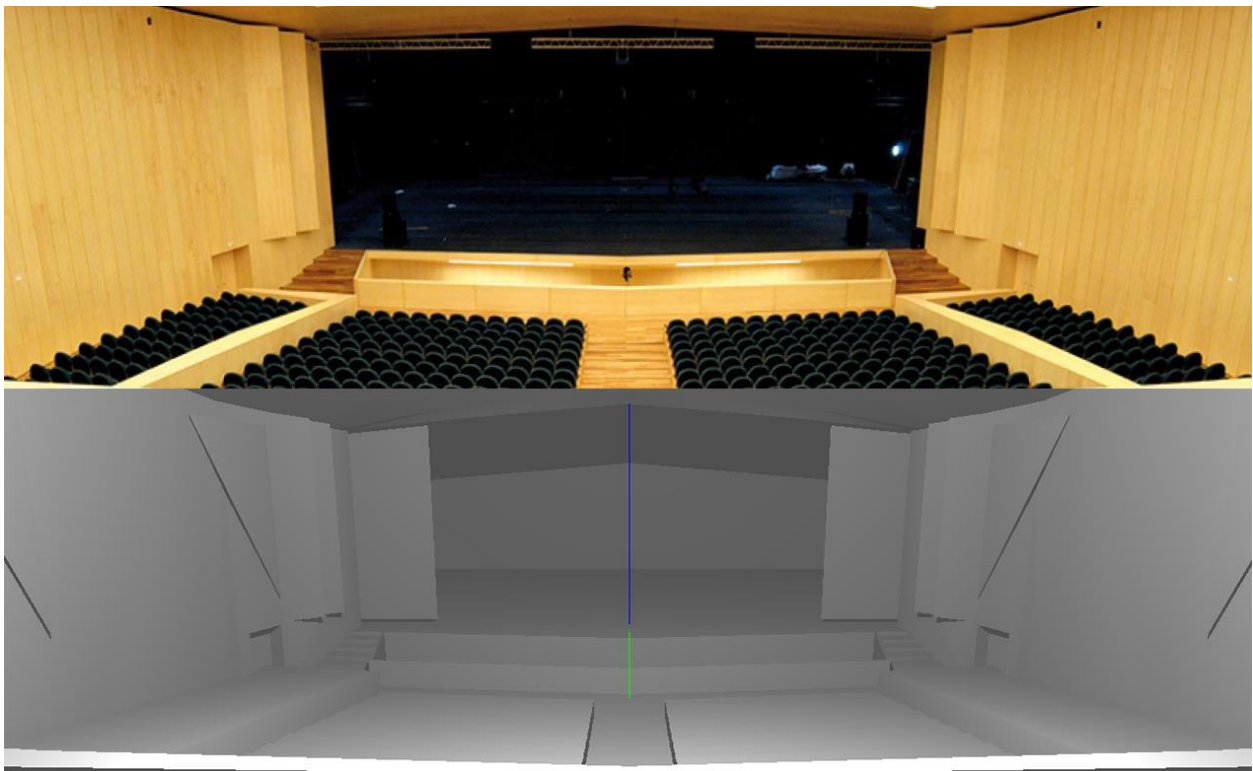


Fig. 4.13 Comparativa entre el teatro y el modelo en Vision, desde el anfiteatro con una fuente de luz en el techo de la platea

5. MEDICIÓN *IN SITU* Y CARACTERIZACIÓN DEL RECINTO

Debido a las aproximaciones realizadas en la geometría de la réplica del auditorio, así como a la incertidumbre debido al conocimiento parcial de las propiedades acústicas de sus materiales, la simple asignación de materiales a las caras del modelo puede dar lugar a un resultado que diste del real. Por ello, es necesario caracterizar el recinto para, una vez asignados los materiales, comparar las simulaciones con el recinto real para calibrar el modelo, ajustando los coeficientes de absorción y *scattering*.

El principal parámetro que permite caracterizar un recinto es el tiempo de reverberación, aunque existen otros como el tiempo de reverberación temprano (EDT), la claridad o el tiempo central (TS). Todos ellos pueden obtenerse a partir de la respuesta al impulso de la sala.

Para esta caracterización, se lleva a cabo un plan de muestreo que permita obtener *in situ* muestras del tiempo de reverberación en varios puntos estratégicos de la sala. Posteriormente, dichas muestras se agruparán atendiendo al área de audiencia a la que pertenezcan y, mediante media aritmética, se obtendrá el promedio espacial del tiempo de reverberación para cada una de las áreas. El objetivo de este promediado por zonas es facilitar el proceso de ajuste de los materiales, pues la acústica de cada área será más sensible a los materiales más próximos a ella. La finalidad es lograr una réplica virtual del recinto real en la que llevar a cabo simulaciones que permitan determinar el sistema de sonorización.

Antes de detallar el procedimiento llevado a cabo en la medición *in situ*, se expone una breve explicación de la respuesta al impulso y el tiempo de reverberación.

5.1. Respuesta al impulso, $h(t)$

La respuesta al impulso de un sistema lineal invariante en el tiempo (LTI) es la señal resultante a su salida cuando es excitado a su entrada con una delta de Dirac. Dicha delta es una función de anchura infinitesimal, altura infinita y área unitaria. Su transformada de Fourier es constante en el dominio frecuencial y, por tanto, cualquier señal convolucionada con ella es igual a la propia señal. Por ello, permite obtener la respuesta al impulso de un sistema LTI:

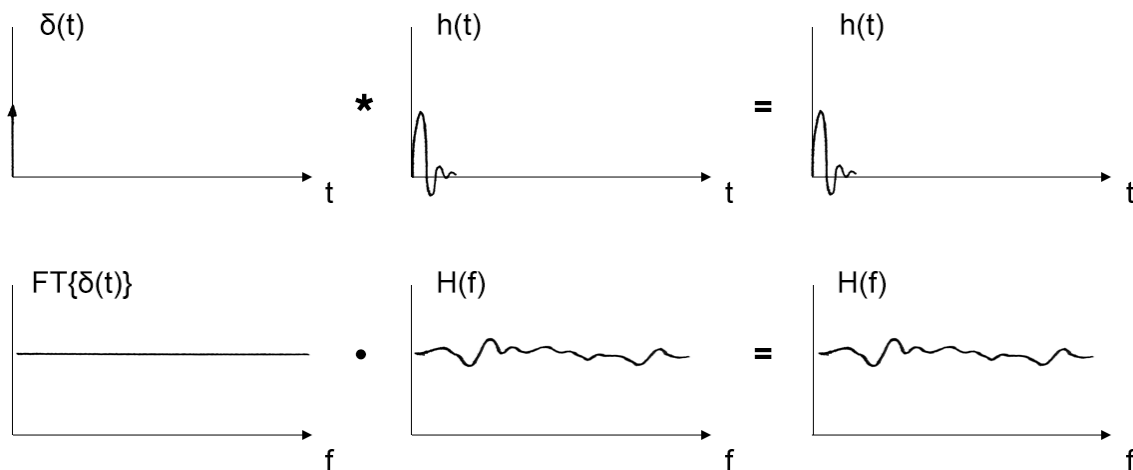


Fig. 5.1 Esquema de la obtención de la respuesta al impulso mediante una delta ideal

Dado que un recinto puede ser considerado un sistema LTI, este fenómeno permite obtener su respuesta al impulso y, a raíz de ella, los parámetros acústicos que lo caracterizan.

Existen diversos métodos para obtenerla. Algunos de ellos son los siguientes:

5.1.1. Método de la señal impulsiva

Puesto que la delta de Dirac se trata de una función ideal y teórica, es necesario recurrir a señales impulsivas que se le aproximen: un disparo, pistolas de fogueo, la explosión de un globo, etc. Mediante una de estas señales, es posible medir la respuesta al impulso $h(t)$ aproximada del recinto.

Sin embargo, recurrir a una aproximación de la delta de Dirac supone un problema: puesto que la señal de test no va a ser ideal, su respuesta no va a ser completamente plana; para que su amplitud en el dominio temporal pasase de cero a infinito de manera instantánea, necesitaría un ancho de banda ilimitado. Esto provoca que la respuesta al impulso $h'(t)$ obtenida tenga pérdidas y se vea coloreada.

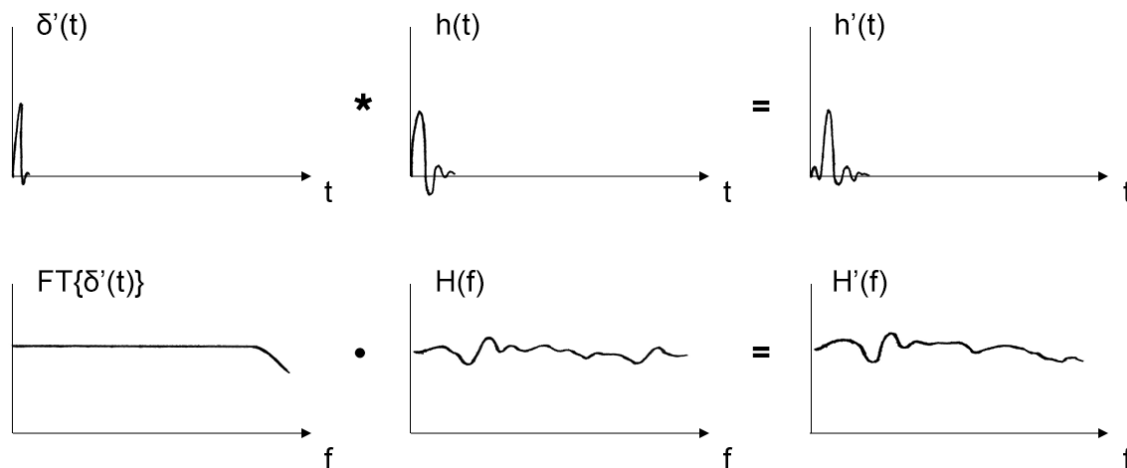


Fig. 5.2 Esquema de la obtención de la respuesta al impulso aproximada a partir de una delta real

5.1.2. Método MLS

La señal MLS (*Maximum Length Sequence*), consiste en una señal binaria pseudoaleatoria generada por un sistema compuesto de una serie de registros, y que tiene valores $\{-1, +1\}$. Su espectro es plano, a excepción de su energía en cero hercios ya que, puesto que sus símbolos son casi equiprobables, su valor medio (DC) es prácticamente nulo.

Este método obtiene la respuesta al impulso tras procesar con una transformada de Hadamard modificada la salida que devuelve el sistema LTI cuando este es excitado por una señal MLS. Resulta computacionalmente eficiente y además proporciona una buena SNR (relación señal/ruido), ya que la propia señal de excitación tiene un factor de cresta bajo (la relación entre el valor máximo y el eficaz para una MLS será muy cercana a uno).

5.1.3. Método del doble canal FFT

Este otro método consiste en excitar el sistema LTI con una señal de banda ancha $x(t)$ y obtener la salida del sistema, $y(t)$. Hecho esto, se ajusta ambas señales en amplitud y fase, se les aplica la Transformada Rápida de Fourier (FFT) y, por comparación, se obtiene la respuesta en frecuencia del sistema.

Si bien este método resulta más flexible a la hora de elegir la señal de excitación del sistema, debido a la necesidad de ejecutar la FFT resulta computacionalmente más costoso.

5.1.4. Método del barrido tonal

El *sweep sine* o barrido tonal, es una señal de test que consiste en una delta móvil en el dominio frecuencial que recorre, en un breve intervalo de tiempo, todas las frecuencias comprendidas en un determinado rango (f_{in} , f_{out}), trazando una respuesta plana en dicho intervalo. La energía de esta señal es constante en el tiempo y se concentra siempre en la delta: a diferencia de los métodos anteriores, en el método del barrido tonal las frecuencias del sistema LTI no se excitan de manera simultánea, sin llegar a convertirlo en un método más lento que los anteriores. También debido a esto, la SNR es mayor aplicando la misma energía.

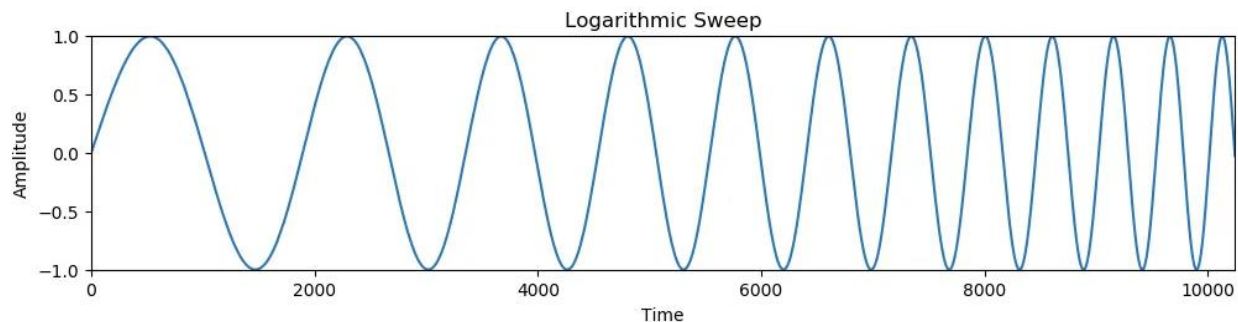


Fig. 5.3 Forma de onda de un sweep sine

El barrido de la delta puede ser lineal o exponencial. El oído humano tiene mayor resolución en baja frecuencia que en alta frecuencia. Como ejemplo, percibir la diferencia entre 100 y 105 Hz es trivial, mientras que percibir la diferencia entre 1.000 y 1.005 Hz resulta más confuso. Por tanto, un barrido exponencial se ajustará más a la resolución del oído humano.

5.2. Tiempo de reverberación, RT

El tiempo de reverberación de una sala se define como el tiempo que la energía sonora tarda en disiparse y ser absorbida por los materiales hasta verse reducida a una millonésima parte. Visto en la escala logarítmica, es el tiempo que el nivel de presión sonora tarda en caer 60 dB. En esencia, depende de cuatro factores:

- **I) Índice de absorción acústica de las superficies, α (adimensional).** Por la ley de la conservación de la energía, en una reflexión la energía incidente es igual a la suma de las energías reflejada, absorbida y transmitida:

$$p_{\text{incidente}}^2 = p_{\text{reflejada}}^2 + p_{\text{absorbida}}^2 + p_{\text{transmitida}}^2 \quad (5.1)$$

Si se divide a ambos lados entre la energía incidente, se obtiene los coeficientes de reflexión, absorción y transmisión:

$$1 = \alpha_{\text{reflexión}} + \alpha_{\text{absorción}} + \alpha_{\text{transmisión}} \quad (5.2)$$

- **II) Volumen del recinto, V (m^3).** Cuanto mayor es el volumen, mayor es la distancia entre paredes y menor es la cadencia de reflexiones: la energía tarda más en decaer.
- **III) La directividad de la fuente y el orden en el que ocurren las reflexiones.** Si la fuente es directiva y está orientada hacia una superficie de un α_1 elevado, la onda perderá la mayor parte de su energía en la primera reflexión, por lo que el RT será menor.
- **IV) La frecuencia.** Puesto que el índice de absorción acústica depende de la frecuencia, el tiempo de reverberación también. Por norma general, la energía de baja frecuencia tarda más en decaer, debido a que su absorción acústica es más complicada: para absorber la energía en baja frecuencia, es necesario recurrir a sistemas más costosos como cajas resonadoras, resonadores sintonizados (Helmholtz) o absorbentes porosos de gran anchura.

En la norma ISO 3382-2 se detalla dos métodos para obtener el tiempo de reverberación de un recinto: el método del ruido interrumpido y el método de la respuesta al impulso integrada. Tras llevar a cabo un muestreo espacial del teatro, se puede recurrir a uno de estos dos métodos para obtener el tiempo de reverberación del recinto, pudiendo obtener el tiempo de reverberación tanto en puntos de interés concretos, como un cálculo promediado de la sala.

5.2.1. Método del ruido interrumpido

Consiste en excitar un altavoz con un ruido de banda ancha e interrumpir abruptamente la señal para registrar la caída del nivel de presión sonora. Posteriormente, se ajusta una recta a dicha caída y, sobre ella, se mide el tiempo que la energía ha tardado en caer un determinado nivel. Para partir de una referencia plana en el dominio de la frecuencia, el ruido en cuestión deberá ser blanco o rosa, en función del analizador: si es de ancho de banda constante (*FFT, Fast Fourier Transform*), el ruido deberá ser blanco; si es de barras de ancho de banda porcentual (*CPB, Constant Percentage Bandwidth*), rosa:

$$d. e. p. \text{Ruido Blanco} (f) = K \xrightarrow{\text{Análisis en FFT}} \int_{f_a}^{f_b} K df = K \Delta f = cte \quad (5.3)$$

$$d. e. p. \text{Ruido Rosa} (f) = \frac{K}{f} \xrightarrow{\text{Análisis en CPB } 1/n \text{ octava}} \int_{f_a}^{2^{1/n} f_a} \frac{K}{f} df = \frac{K}{n} \ln(2) = cte \quad (5.4)$$

Debido a que la señal de test no es determinista, su aleatoriedad da lugar a pequeñas fluctuaciones en su nivel de energía. Para tener una referencia clara del nivel estacionario, se mide la caída a partir del instante en que la energía cae cinco decibelios por debajo del valor medio de dichas fluctuaciones. Por la misma razón, la caída se evalúa con un margen de 10 dB por encima del ruido de fondo.

Medir una caída de 60 dB es complicado: si se tiene en cuenta el ruido de fondo y los márgenes mencionados, se precisa de una SNR mayor de 75 dB. Si además el ruido de fondo es excesivo, posiblemente sea necesario emitir la señal de test por encima del umbral del dolor (120 dB). Es por ello que, en lugar de medir de manera directa el RT, se realiza una estimación evaluando caídas menores:

- **RT20:** Tres veces el tiempo que el SPL tarda en caer 20 dB. SNR mínima necesaria de 35 dB.
- **RT30:** Dos veces el tiempo que el SPL tarda en caer 30 dB. SNR mínima necesaria de 45 dB.

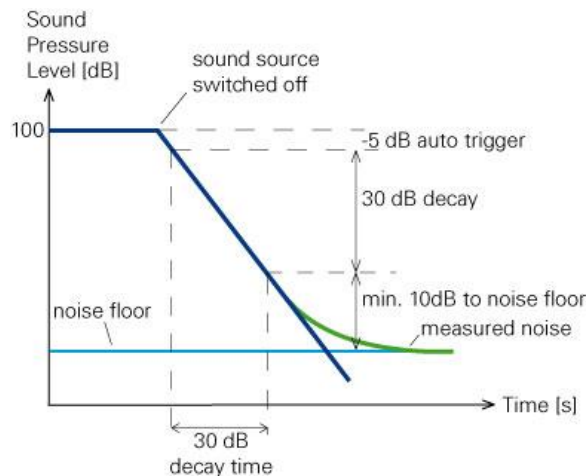


Fig. 5.4 Evolución temporal de la energía al interrumpir el ruido de banda ancha

Cuanto mayor sea el nivel de caída evaluado, más próximo a la medida ideal: el resultado será más fiable. Por otro lado, debido a la aleatoriedad de la señal, conviene evaluar varias veces la caída en el mismo punto y realizar un promedio para reducir la incertidumbre de los resultados.

5.2.2. Método de la respuesta al impulso integrada

Dada la respuesta al impulso del recinto, se pasa por un banco de filtros (de 1/1 o 1/3 oct.) obteniendo un conjunto de respuestas al impulso para las bandas de interés. Cada una de ellas, es elevada al cuadrado y procesada mediante la integración inversa de Schröder, obteniendo la ETC (*Energy - Time Curve*), que representa el decaimiento de la energía a lo largo del tiempo. De manera similar a lo que ocurría en el método anterior, sobre la ETC se ajusta una recta que permite evaluar el tiempo de caída.

Dicha operación se detalla en la norma ISO 3382-1. Consiste en la integración de la respuesta al impulso cuadrática filtrada comprendida entre el instante del transitorio y el final de la cola, siendo esta infinito en el supuesto ideal de que el ruido de fondo fuese nulo.

$$ETC(t) = \int_t^{\infty} h_f^2(\tau) d\tau \quad (5.5)$$

Una vez obtenida la ETC, se obtiene el RT evaluando el tiempo de caída de la energía. Cuanto menor sea el ruido de fondo, mayor parte de la cola aparecerá en la ETC y, por tanto, se podrá evaluar el RT para un intervalo de nivel de caída mayor. En el caso de este proyecto, se evaluó para 20 dB.

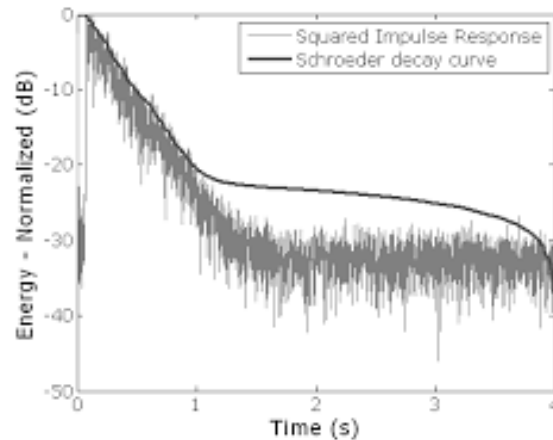


Fig. 5.5 Integral de Schröder aplicada sobre una respuesta al impulso cuadrática

A la hora de calcular la incertidumbre del RT20, la ISO 3382-2 ofrece una fórmula que, entre otros parámetros, depende del número de evaluaciones llevadas a cabo en cada punto de medida, n .

$$\sigma = 0,88 \cdot RT20 \cdot \sqrt{\frac{1 + 1,9/n}{N \cdot B \cdot RT20}} \quad (5.6)$$

- n = Número de evaluaciones promediadas en cada punto de medición.
- N = Número de combinaciones fuente-micrófono promediadas.
- B = Ancho de banda en Hz.

En la teoría, la precisión del método de la respuesta al impulso integrada equivale a la del promediado de un conjunto infinito de evaluaciones mediante el método del ruido interrumpido. Por ello, su incertidumbre es menor; la normativa especifica que, para la respuesta al impulso integrada, el valor de dicha n ha de ser diez veces el número de evaluaciones llevadas a cabo.

Debido a esta mayor precisión, se procedió a medir la respuesta al impulso mediante barrido tonal, para posteriormente extraer el tiempo de reverberación con el método de la respuesta al impulso integrada.

5.3. Otros parámetros acústicos

En la norma ISO 3382-1 se detalla otros parámetros acústicos que, si bien no son tan conocidos como el tiempo de reverberación, también pueden resultar útiles para la caracterización de recintos, especialmente en el caso de auditorios. A continuación, se explica tres de ellos: el EDT, la claridad y el tiempo central.

5.3.1. *Early Decay Time*, EDT

Expresado en segundos, el tiempo de reverberación temprano (EDT) es seis veces el tiempo que la energía acústica en un recinto tarda en verse reducida a una décima parte, o lo que es igual, seis veces el tiempo que la energía tarda en caer 10 dB.

Si se visualizase la curva de decaimiento del nivel de energía, se apreciaría cómo en el primer instante hay una caída abrupta de nivel que pronto se decelera y se estabiliza aproximándose a una recta. Mientras que parámetros como el RT10, RT20 o RT30 dejan un margen y evalúan el tiempo de caída de nivel a partir del momento en el que la energía ha caído cinco decibelios, el EDT evalúa la caída desde el primer instante, siendo por tanto más sensible a esta caída inicial, que es más notoria para el oído humano.

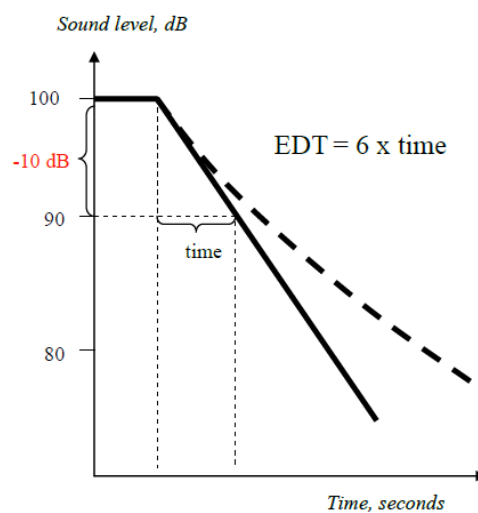


Fig. 5.6 Representación gráfica de la evaluación del EDT sobre una curva de energía sobre tiempo

En términos prácticos, el RT10, RT20 y RT30 sirven como una aproximación objetiva al tiempo de reverberación de la sala, mientras que el EDT es más adecuado a la hora de determinar el tiempo de reverberación subjetivo.

5.3.2. Claridad, C_{t_0}

Expresada en decibelios, la claridad representa la relación entre la energía acumulada en los primeros t_0 milisegundos (límite temporal precoz) de la respuesta al impulso y la que llega a partir de dicho instante. Por tanto, refleja la relación entre las reflexiones tempranas y tardías.

$$C_{t_0} = 10 \log_{10} \frac{\int_0^{t_0} p^2(t) dt}{\int_{t_0}^{\infty} p^2(t) dt} \quad (5.7)$$

Suele usarse el C_{50} para salas de conferencias, y el C_{80} (habitualmente denominado claridad) para teatros y salas de conciertos. Cuanto mayor sea su nivel, mayor será la claridad y más seca será la sala. Es por ello que está estrechamente relacionada con el tiempo de reverberación.

5.3.3. Tiempo central, T_s

Es similar al concepto de mediana en estadística. Generalmente expresado en milisegundos, representa el centro de gravedad de la respuesta al impulso; tomando el instante inicial como cero segundos, el tiempo central es el instante en el que la energía acumulada es igual a la restante. Cuanto más seca es la sala, más próxima a una delta es su respuesta al impulso y menor es el tiempo central.

$$T_s = \frac{\int_0^{\infty} t \cdot p^2(t) dt}{\int_0^{\infty} p^2(t) dt} \quad (5.8)$$

La ventaja frente a la claridad es que evita la necesidad de definir un límite temporal precoz para discriminar las energías temprana y tardía.

5.4. Situación del teatro

El día de la medición fue el miércoles 17 de febrero del 2021, entre las 10:30 y las 19:00 horas aproximadamente. La temperatura a lo largo de la jornada osciló entre 16 y 17 grados centígrados. La humedad relativa se mantuvo estable, en torno al 33 %, y la presión atmosférica fue de 942 hPa. En cuanto a la ocupación en el teatro, a lo largo del día hubo un máximo de seis personas: Antonio, Daniel, Alejandro, Javier, yo y, eventualmente, el técnico del auditorio.

Dado que en el proceso de calibración se empleará como referencia los resultados obtenidos con las mediciones *in situ*, la geometría del modelo construido en EASE trata de replicar las condiciones del auditorio el día de la medición, tanto en lo que a condiciones meteorológicas se refiere como al estado y configuración de los elementos, los cuales fueron los siguientes:

Elemento	Estado
Butacas	<i>Todas las butacas, tanto de la platea como del anfiteatro, se encontraban plegadas. Muchas de ellas estaban precintadas para garantizar las distancias de seguridad entre los espectadores, debido a la situación de pandemia.</i>
Puertas	<i>Todas las puertas del teatro se encontraban cerradas, tanto las que conectaban directamente con el auditorio como las dobles puertas del acceso al anfiteatro.</i>
Cortinas triangulares	<i>De las doce posibles posiciones, solamente seis cortinas se encontraban colocadas, tres por cada pared lateral. En cada una de las paredes, las posiciones que ocupaban las cortinas, de entre las seis posiciones posibles, eran la primera, la tercera y la quinta más próximas al escenario.</i>
Telón del escenario	<i>Asomaba aproximadamente un metro por los extremos de la boca del escenario.</i>
Bambalinas	<i>Se encontraban recogidas.</i>
Foso de orquesta	<i>Su suelo modulable se encontraba bajado a una profundidad de 2,59 metros respecto al suelo del escenario.</i>
Otros	<i>En el escenario había una escalera de plataforma, así como parte del equipamiento necesario para las mediciones, y algunas mesas con herramientas del propio teatro. Por simplicidad y bajo impacto acústico, todo esto se ha considerado transparente de cara a la simulación.</i>

Tabla 5.1 Situación del teatro el día de la medición

5.5. Equipo y configuración

El *software* utilizado, tanto para la generación de la señal de test como para la grabación, fue el desarrollado por la empresa *Brüel & Kjaer*, Dirac. Este programa está diseñado para la medición de parámetros acústicos de recintos, tales como la respuesta al impulso, así como para el postprocesado de las mediciones para extraer parámetros derivados, como el tiempo de reverberación, la respuesta en frecuencia o la claridad.

El micrófono utilizado fue el DPA 4006, que es un micrófono de presión (omnidireccional) y de condensador, por lo que requiere alimentación externa de tipo *phantom* a 48 V.

Su membrana es pequeña (1/2"), de manera que las atenuaciones producidas por directividad debido al apantallamiento acústico del propio micrófono, así como los realces por las reflexiones con la cápsula, se producen a partir de aproximadamente 10 kHz. Para adaptar la respuesta en alta frecuencia al campo sonoro, incluye diferentes rejillas que actúan como filtros compensatorios: una para incidencia aleatoria (campo difuso) y otra para incidencia en el eje (campo libre). Como la acústica del teatro era relativamente seca, se optó por la rejilla que adaptaba el micrófono para campo libre y se apuntó hacia la fuente.



Fig. 5.7 Micrófono DPA 4006

La interfaz de audio utilizada fue una *RME Fireface UFX USB*, que consta de doce entradas analógicas (cuatro combos XLR-3 hembra (mic) / Jack TRS (line) y ocho entradas de línea de Jack TRS), así como ocho salidas de línea (seis Jack TRS y 2 XLR-3 macho) más dos salidas para auriculares vía Jack TRS estéreo no balanceado. Además, cuenta con conectores digitales: dos interfaces MIDI in/out, dos ADAT in/out (hasta ocho señales por cada conector), entrada y salida AES/EBU (señales digitales estéreo), interfaz MADI, y dos conectores BNC para sincronizarse con otros equipos digitales vía *wordclock*. Para las mediciones, tan solo se necesitó una entrada de micrófono y una salida de línea para emitir la señal de test.



Fig. 5.8 Interfaz ME Fireface UFX

La fuente dodecaédrica *Brüel & Kjaer 4292-L* está compuesta por doce altavoces pasivos que operan en fase para proporcionar una emisión omnidireccional.

Se conecta vía *speakon* a un amplificador de potencia de la misma marca, modelo 2716C. Este amplificador consta de dos entradas de línea (XLR-3 hembra) y dos canales de salida CH A/B. Para lograr una mayor amplificación, y puesto que solo se necesitó un canal, se realizó una configuración *bridge*: se puentea los dos canales para que los dos amplificadores integrados suministren toda la potencia al canal A.



Fig. 5.9 Altavoz 4292-L



Fig. 5.10 Amplificador de potencia 2716C

Por tanto, la interconexión de estos equipos para la medición fue la que se detalla en el siguiente diagrama de bloques:

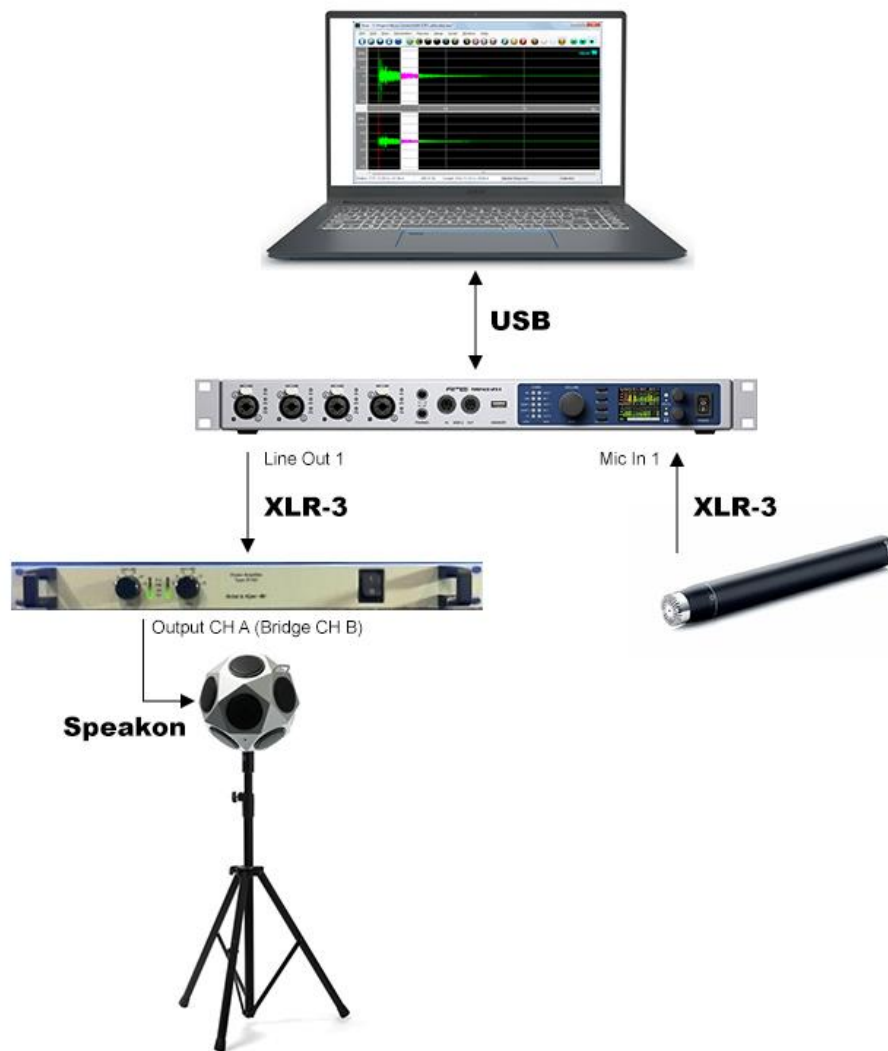


Fig. 5.11 Diagrama de bloques de la configuración empleada

5.6. Muestreo espacial

La distribución de los puntos sigue el criterio de la norma ISO 3382-1 que regula las mediciones acústicas en auditorios. Se buscó distribuir los puntos de medición de manera homogénea, especialmente por el área de audiencia, evitando establecer posiciones simétricas, alineadas en el mismo eje o muy cercanas a alguna superficie u objeto reflectante que pudiese influir en exceso en la medición.

En la siguiente página se detalla en un plano aéreo el muestreo espacial. Se estableció un total de 35 puntos de **micrófono** (marcados en azul y con una M) y tres posiciones de **fuentes** (marcados en rojo y con una F), con un total de $N = 105$ combinaciones y asegurando una distancia mínima de dos metros entre posiciones de micrófono, y de un metro entre micrófono y fuente.

En todos los puntos de micrófono la altura de la cápsula respecto al suelo era de 1,20 metros, a excepción de las posiciones M35, M36 y M37, situadas en el escenario, que fue de 1,50 metros. En el caso de las posiciones de fuente, la altura del centro del dodecaedro respecto al suelo se fijó a 1,50 metros.

Cabe mencionar que, si bien son 35 los puntos de micrófono especificados, la nomenclatura de estos va desde M1 hasta M37, no apareciendo reflejadas las posiciones M33 y M34 que fueron utilizadas para el plan de muestreo para *Ambisonics*.

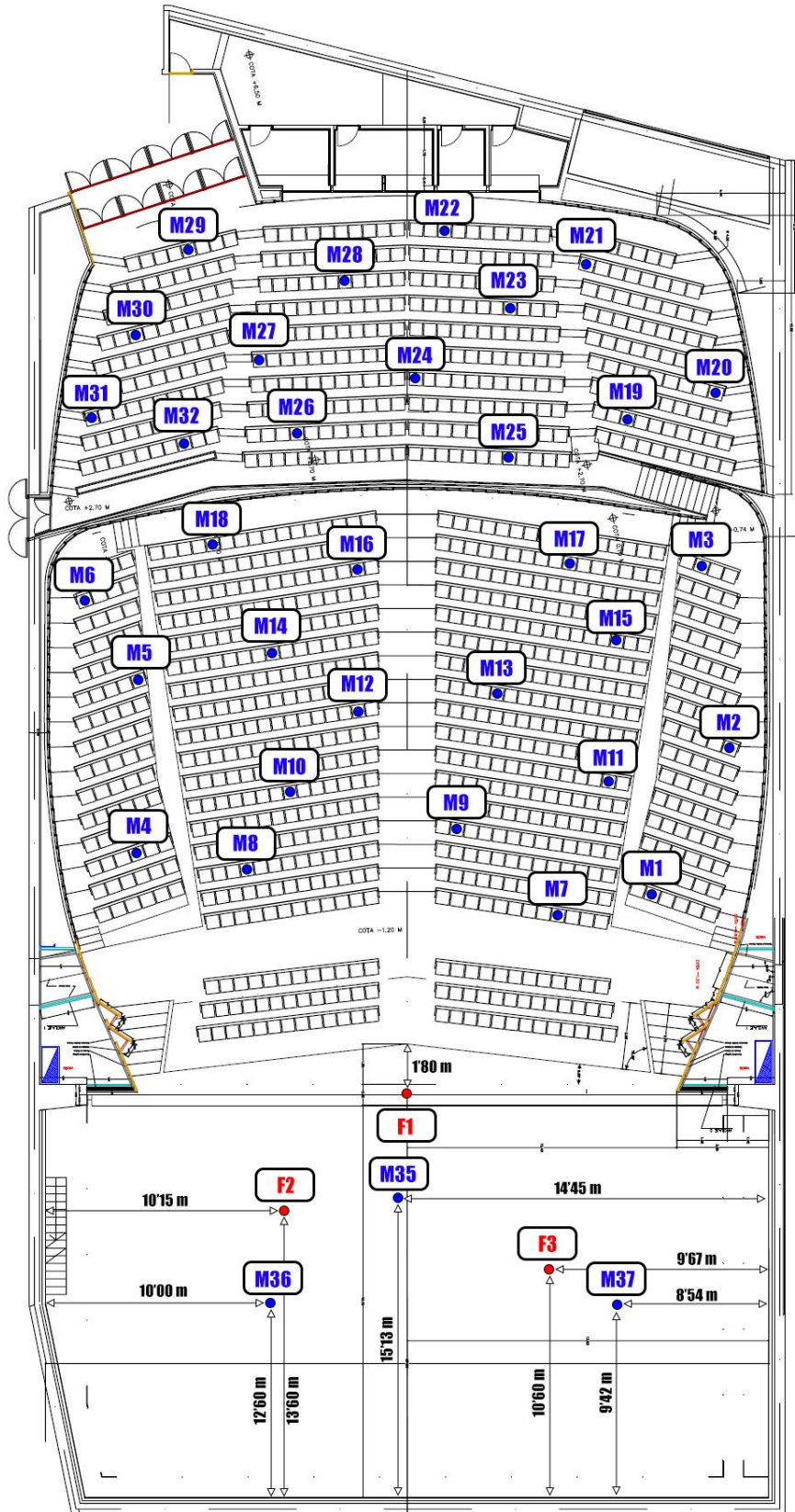


Fig. 5.12 Plan de muestreo espacial: Fuente dodecaédrica y micrófono omnidireccional

5.7. Procedimiento e incidencias

Por cada combinación de micrófono y fuente, se realizó cuatro barridos tonales exponenciales en el rango de 20 Hz a 20 kHz, con una duración de tres segundos cada uno y comprobando que existía la relación señal a ruido suficiente (35 dB) como para poder calcular la RT20 mediante la integral inversa de Schröder en el postprocesado. Tras los barridos, el *software* de Dirac registraba la respuesta al impulso resultante para dicha combinación micrófono-fuente.

Debido a que la acústica del teatro era relativamente seca, fue necesario apuntar la cápsula hacia la fuente para evitar posibles pérdidas en alta frecuencia debidas a un apantallamiento acústico producido por las propias dimensiones del micrófono de medición, logrando además una respuesta plana gracias a la compensación frente a realces proporcionada por la rejilla, que funciona como un filtro paso bajo sobre amortiguado.

En lo relativo al reparto de tareas, Javier y Antonio se encargaron de ir alternando la fuente dodecaédrica entre las tres posiciones del escenario, mientras que Daniel llevó a cabo la tarea de grabar y almacenar las muestras. Por otro lado, Alejandro y yo nos encargábamos de desplazar el pie de micrófono, llevar a cabo un registro de las medidas e incidencias que se produjesen, y comunicar a Daniel después de cada grabación el nombre de la muestra, siguiendo la nomenclatura F_iM_j , haciendo i referencia a la posición de la fuente y j a la de micrófono.

El procedimiento para los puntos de medida situados en las butacas fue el siguiente: fijando el micrófono en un punto j determinado, se medía secuencialmente las respuestas al impulso para las tres posiciones de fuente, registrando las muestras $F1M_j$, $F2M_j$ y $F3M_j$. Posteriormente, se desplazaba el pie de micrófono hasta otro punto de medida cercano y se repetía el proceso, comenzando por la última posición de la fuente.

Para los puntos de medida situados en el escenario, el procedimiento fue al revés: fijando la fuente en un punto i , se medía secuencialmente las respuestas al impulso F_iM35 , F_iM36 , F_iM37 . Hecho esto, se desplazaba la fuente y se repetía el proceso, comenzando por la última posición de micrófono.

El registro llevado a cabo para las muestras tomadas con el micrófono omnidireccional, así como las incidencias ocurridas durante el proceso, se detalla a continuación:

Secuencia	Trío de muestras	Incidencia
1	F1M29, F2M29, F3M29	Ninguna
2	F3M22, F1M22, F2M22	
3	F2M21, F1M21, F3M21	
4	F3M23, F1M23, F2M23	
5	F2M28, F1M28, F3M28	
6	F3M30, F1M30, F2M30	
7	F2M27, F1M27, F3M27	
8	F3M24, F1M24, F2M24	
9	F2M20, F1M20, F3M20	
10	F3M19, F1M19, F2M19	<i>Debido a la falta de espacio entre las filas para colocar el pie de micrófono, hubo que medir en la butaca de atrás. El plan de muestreo mostrado en la memoria tiene en cuenta esta modificación.</i>
11	F2M25, F1M25, F3M25	Ninguna
12	F3M26, F1M26, F2M26	
13	F2M31, F1M31, F3M31	<i>Hubo que repetir la medida F3M31 debido a una SNR insuficiente (menor de 35 dB de cara a calcular el RT20). La posición M31 se sitúa cerca de una puerta de acceso al auditorio, y se coló ruido de fondo en ese momento.</i>
14	F3M32, F1M32, F2M32	Ninguna
15	F2M03, F1M03, F3M03	
16	F3M02, F1M02, F2M02	
17	F2M01, F1M01, F3M01	
18	F3M17, F1M17, F2M17	
19	F2M16, F1M16, F3M16	Ninguna
20	F3M18, F1M18, F2M18	
21	F2M14, F1M14, F3M14	
22	F3M15, F1M15, F2M15	
23	F2M13, F1M13, F3M13	
24	F3M12, F1M12, F2M12	<i>Hubo que volver a registrar la muestra F2M12 debido a una SNR insuficiente.</i>
25	F2M10, F1M10, F3M10	Ninguna
26	F3M11, F1M11, F2M11	
27	F2M09, F1M09, F3M09	
28	F3M07, F1M07, F2M07	
29	F2M08, F1M08, F3M08	<i>Hubo que volver a registrar la muestra F1M08 debido a una SNR insuficiente.</i>
30	F3M06, F1M06, F2M06	Ninguna
31	F2M05, F1M05, F3M05	
32	F3M04, F1M04, F2M04	
33	F1M36, F1M35, F1M37	<i>Las posiciones de escenario M35, M36 y M37 son las más cercanas a la fuente, por lo que se atenuó seis decibelios en el amplificador de la fuente para evitar saturar en la grabación.</i>
34	F3M37, F3M35, F3M36	
35	F2M36, F2M37, F2M35	

Tabla 5.2 Registro de las muestras tomadas

5.8. Postprocesado en Room EQ Wizard

Room EQ Wizard (REW) es un software gratuito escrito en Java que permite la grabación y/o postprocesado de respuestas al impulso. Se usó para la obtención de los tiempos de reverberación en 1/3 octava mediante el método de la respuesta al impulso integrada.

Para ello, REW aplica un banco de filtros a la señal y realiza la integral inversa de Schröder descrita en la norma ISO 3382-1, obteniendo la ETC (*Energy-Time Curve*). A partir de dicha curva, extrae los RT para un intervalo de caída determinado, así como otros parámetros adicionales como EDT, TS o C80.

Al tratarse de una señal digital, la energía se representa en dBFS (*Full Scale*). Si bien la cuantización empleada depende del convertor A/D, por lo general el nivel nominal +4 dBu corresponde a -18 dBFS, siendo 0 dBFS el máximo nivel de voltaje posible a registrar.

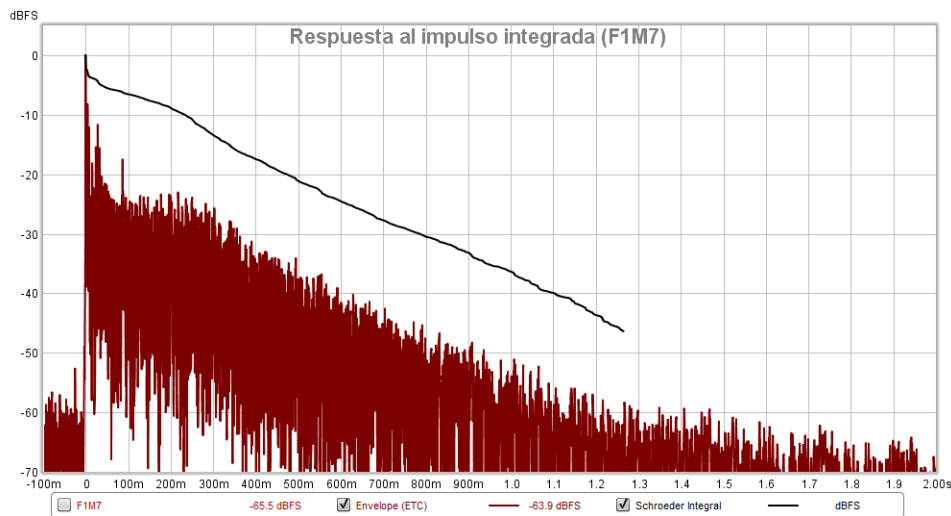


Fig. 5.13 Resposta al impulso integrada de la muestra F1M7 en REW

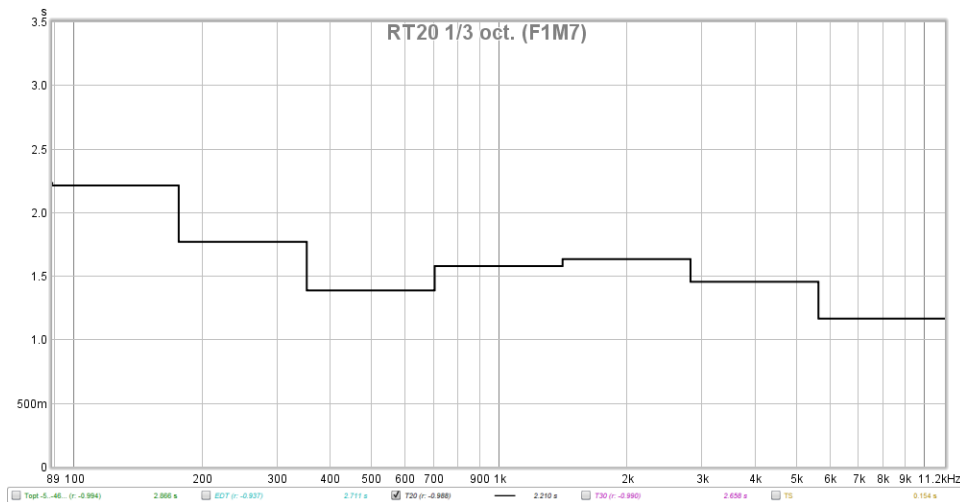


Fig. 5.124 RT20 en 1/3 octava de la muestra F1M7 en REW

Otro de los parámetros acústicos que permite extraer, es el espectro de la respuesta al impulso, representando la energía en dB SPL. Puesto que no se midió *in situ* el SPL en los puntos de medición, la siguiente gráfica no está calibrada; el nivel representado no es fiable.

Sin embargo, la gráfica *waterfall* resulta útil a la hora de visualizar el equilibrio frecuencial, así como la evolución del espectro a lo largo del tiempo, partiendo del instante inicial del transitorio de la respuesta al impulso hasta el final de la cola. Esta gráfica, por tanto, permite visualizar la evolución de la energía de la respuesta al impulso en los dominios temporal y frecuencial de manera simultánea:

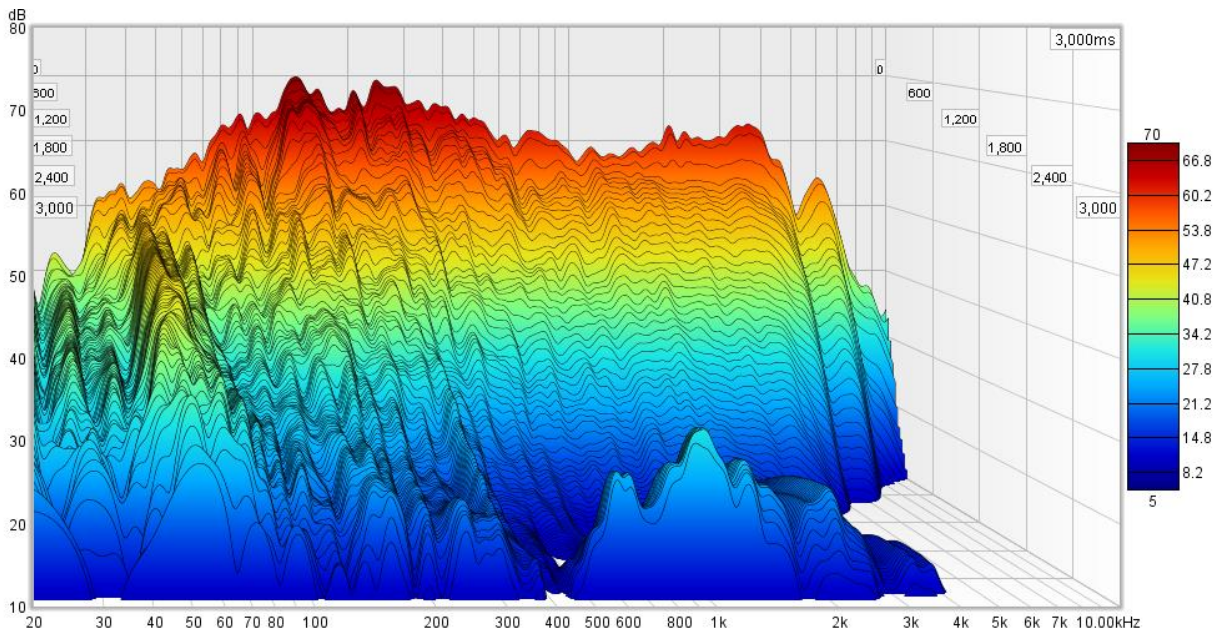


Fig. 5.135 Waterfall de la muestra F1M7 en REW

Como puede verse, la energía en alta frecuencia decae significativamente antes que la de baja frecuencia, como era de esperar. Además, se observa un modo propio en torno a 60 Hz que persiste en el tiempo, así como un *flutter* (o eco debido a dos superficies reflectantes paralelas) en el rango (500, 2500) Hz que se percibió el día de las mediciones al evaluar la acústica de forma aproximada mediante palmadas.

5.9. Resultados

Una vez extraídos los RT20 en 1/3 octava de cada una de las 105 muestras, estos fueron agrupados por áreas de audiencia para, mediante media aritmética, calcular el promedio espacial para cada una de ellas, facilitando la posterior elección de materiales.

Además, se realizó el promediado del RT20 de todos los puntos (que no áreas) para hallar el tiempo de reverberación global medio del teatro. Conociendo el volumen del modelo y la superficie total expuesta al campo sonoro, se calculó además los coeficientes de absorción medios del recinto según los modelos de Sabine y Eyring.

RT20 IN SITU: RESUMEN												Absorción, α	
Freq. 1/3 oct	RX A	RX A*	RX B	RX B*	RX C	RX C*	RX D	RX D*	STG	Global	σ	Sabine	Eyring
100	2.88	2.74	2.67	2.80	2.84	3.44	3.34	2.69	2.91	0.029	0.11	0.11	
125	2.52	2.74	2.61	2.84	2.70	2.53	2.90	2.20	2.64	0.022	0.12	0.12	
160	2.47	2.66	2.43	2.64	2.42	2.16	2.51	2.16	2.45	0.020	0.13	0.12	
200	2.42	2.46	2.42	2.38	2.24	2.22	2.26	2.23	2.34	0.019	0.14	0.13	
250	2.17	2.31	2.17	2.00	2.11	2.07	2.01	2.26	2.15	0.018	0.15	0.14	
315	1.81	1.88	1.89	1.86	1.77	1.85	1.86	1.94	1.84	0.015	0.18	0.16	
400	1.76	1.75	1.71	1.74	1.71	1.72	1.71	1.91	1.75	0.013	0.19	0.17	
500	1.66	1.56	1.60	1.69	1.65	1.65	1.66	1.74	1.64	0.011	0.20	0.18	
630	1.61	1.60	1.64	1.64	1.60	1.63	1.60	1.60	1.61	0.009	0.20	0.18	
800	1.63	1.64	1.71	1.70	1.61	1.65	1.66	1.56	1.64	0.008	0.20	0.18	
1,000	1.66	1.62	1.71	1.67	1.60	1.66	1.60	1.44	1.62	0.007	0.20	0.18	
1,250	1.62	1.64	1.65	1.65	1.57	1.65	1.60	1.44	1.60	0.006	0.20	0.18	
1,600	1.61	1.62	1.64	1.67	1.58	1.58	1.62	1.39	1.59	0.005	0.20	0.18	
2,000	1.60	1.59	1.63	1.65	1.57	1.56	1.59	1.37	1.57	0.004	0.21	0.19	
2,500	1.56	1.56	1.61	1.61	1.52	1.51	1.52	1.36	1.53	0.004	0.21	0.19	
3,150	1.46	1.48	1.47	1.52	1.42	1.43	1.44	1.32	1.45	0.004	0.22	0.20	
4,000	1.38	1.41	1.37	1.41	1.32	1.31	1.31	1.20	1.35	0.003	0.24	0.21	
5,000	1.27	1.27	1.29	1.27	1.20	1.19	1.19	1.09	1.23	0.003	0.26	0.23	
6,300	1.06	1.05	1.12	1.05	0.99	1.00	1.00	0.93	1.03	0.002	0.32	0.27	
8,000	0.91	0.91	0.96	0.91	0.87	0.87	0.86	0.84	0.89	0.002	0.36	0.30	
10,000	0.67	0.66	0.74	0.68	0.67	0.63	0.62	0.63	0.66	0.001	0.49	0.39	
(Hz)	(segundos)										(adimensional)		

Tabla 5.3 RT20 obtenidos de la medición in situ

La incertidumbre del RT20 global se calculó siguiendo la fórmula (2) de la ISO 3382-2. Puesto que, como describe la norma, la precisión del método de la respuesta integrada equivale a la de evaluar diez caídas mediante el método del ruido interrumpido, y puesto que se llevó a cabo cuatro barridos por cada posición, la sigma se calculó para $n = 40$.

En el Anexo A de esta memoria se especifica con mayor detalle los tiempos de reverberación obtenidos para todas las muestras, así como los parámetros EDT, TS y C80 obtenidos.

Además de la anterior tabla, la siguiente gráfica en Matlab muestra una comparativa visual de los RT20 promedios superpuestos. Si bien todas las gráficas son razonablemente similares, los resultados en el escenario son los que más se alejan de la media global; esto era de esperar, ya que los materiales del escenario son bastante diferentes de los de la platea y el anfiteatro.

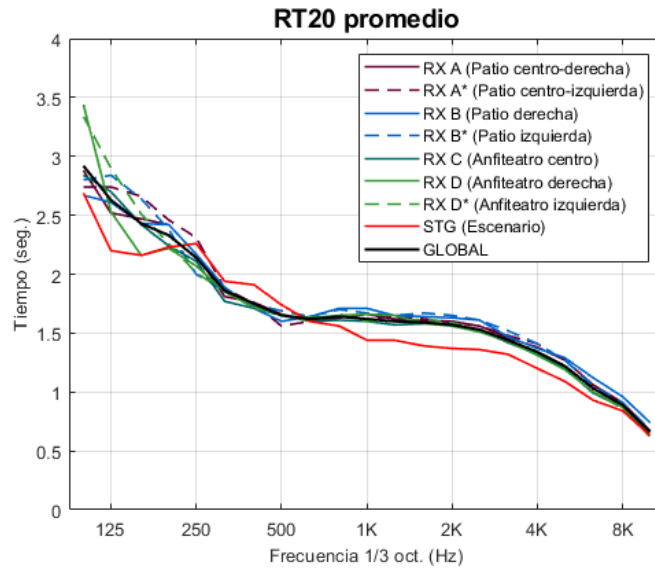


Fig. 5.146 RT20 promedios obtenidos

En cuanto a la incertidumbre, cae con la frecuencia: tal y como establece la fórmula empleada, a mayor ancho de banda, menor es la sigma. Además, como era de esperar, cuanto mayor es el número de muestras promediadas (N), menor es la incertidumbre resultante.

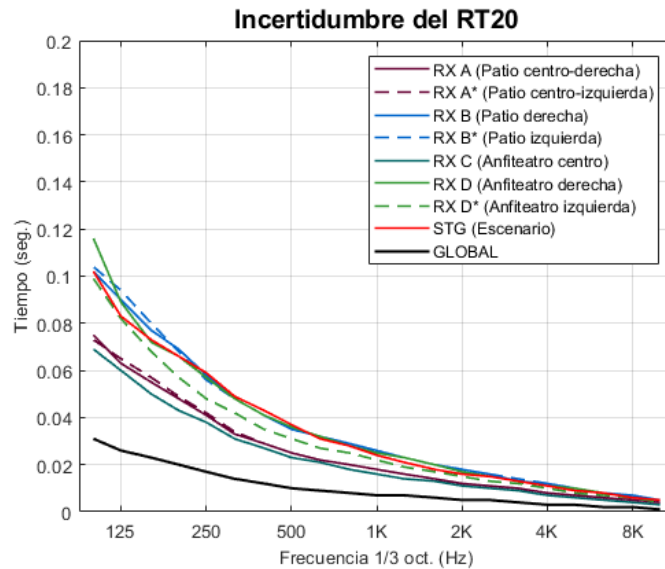


Fig. 5.17 Incertidumbre de los RT20 promediados

6. CALIBRACIÓN DEL MODELO

6.1. Procedimiento

Lo primero es posicionar en el modelo los puntos de fuente (altavoz omnidireccional o “*SPHERE*”) y los puntos de micrófono (asientos o “*seats*”) detallados en el plan de muestreo.

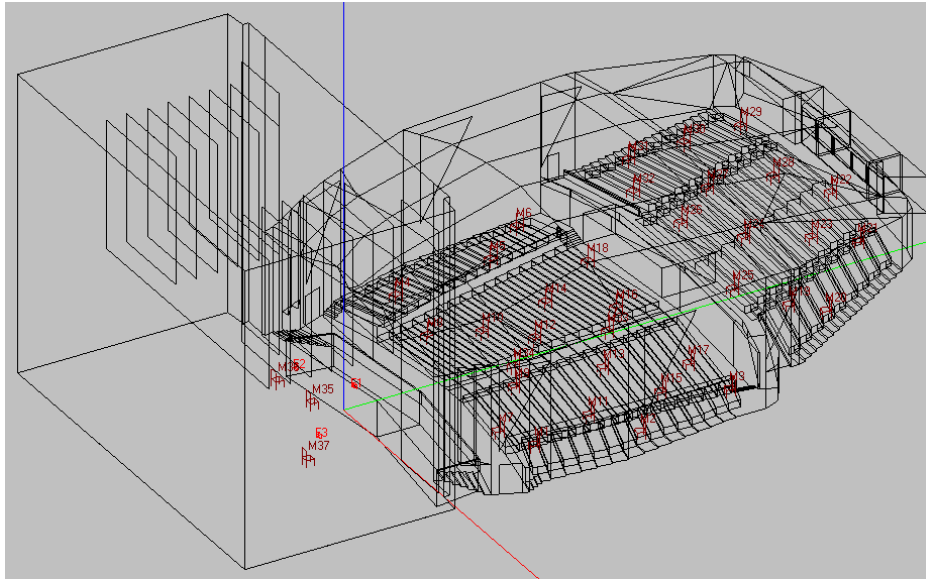


Fig. 6.1 Modelo en Room Editor con los puntos de fuente y micrófono dibujados

Hecho esto, mediante el módulo *AURA* de *EASE*, se lleva a cabo tres simulaciones de *raytracing* (una por fuente) con una alta densidad de partículas y el *scattering* por defecto, creciente en frecuencia desde 10 % hasta 40 % con *crossover* en 1 kHz para aquellas superficies sin *scattering* definido. Como resultado, *AURA* genera una respuesta al impulso para cada asiento, así como un informe con parámetros como el RT20, RT30, C80, EDT o %Alcons expresados en tercios de octava. Estos parámetros se verán afectados por los coeficientes de absorción y *scattering* de los materiales asignados a las caras del modelo.

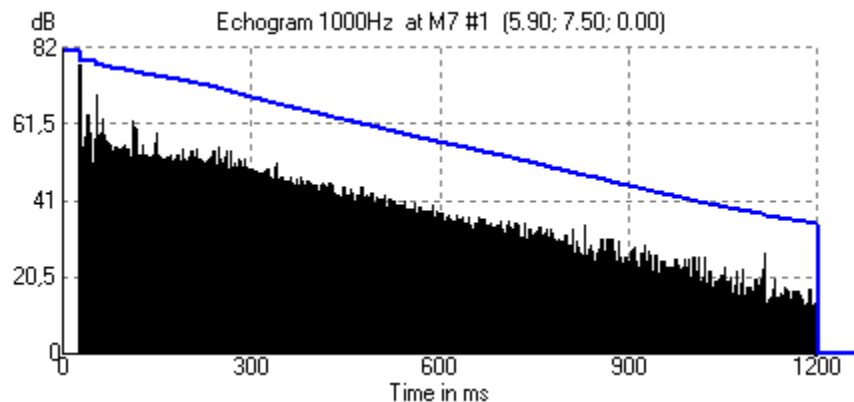


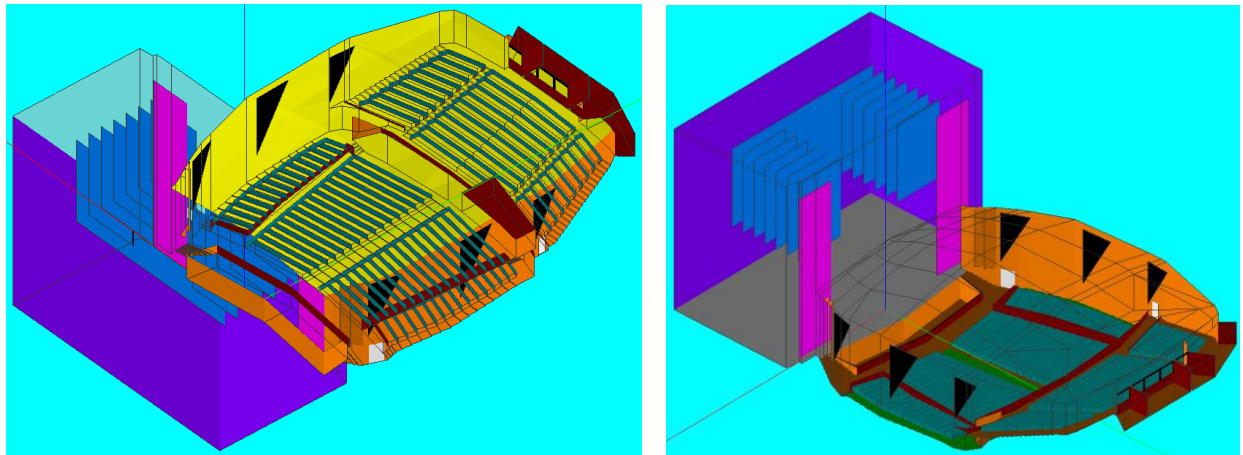
Fig. 6.2 Ecograma de alta densidad generado por *AURA*, con la integral de Schröder resaltada en azul

6.2. Diseño de materiales

Las caras del modelo se agruparon en trece grupos atendiendo a sus materiales. Estos fueron editados partiendo de la base de datos *Generic* de EASE, y su distribución y descripción, así como el porcentaje de área total que cubre cada uno, son las siguientes:

Material	Área %	Distribución	Comentarios
Tela 450 gr/m2	19,48	Bambalinas	Era de esperar absorción en alta frecuencia.
Varios	19,28	Paredes escenario	La distribución de materiales era irregular: algunas partes eran de contrachapado y otras estaban formadas por bloques suavizados de lo que parecía cemento. Además, había estructuras para la sujeción de elementos. Por simplificar, se estableció un solo material con un <i>scattering</i> del 30 %.
Butacas semitapizadas	14,41	Platea y anfiteatro	Butacas de contrachapado semitapizadas en tela. Suele ser el elemento más absorbente.
Contrachapado 15 mm	10,89	Paredes de platea y anfiteatro	A una distancia de 20 cm, y orientado hacia la pared, dije la palabra "test" varias veces. La reflexión era brillante y apreciable, por lo que parecía tener una absorción muy moderada.
Contrachapado 18 mm	10,56	Techo de platea y anfiteatro	
Doble contrachapado de okumen sobre DM	5,30	Suelo del escenario	Sonaba hueco, por lo que en términos de absorción acústica es de esperar un comportamiento similar al de una caja resonadora (absorción en baja frecuencia).
Techo con tramoya	4,97	Techo del escenario	Para contrarrestar el haber obviado la tramoya, se aplicó un <i>scattering</i> del 40 %.
Contrachapado	4,72	Varias	Muy similar a las paredes y techo de la platea y anfiteatro, de espesor ligeramente distinto.
Telón	3,88	Telón	Pesado y de gran densidad. Es de esperar absorción en media y alta frecuencia.
Madera genérica	3,50	Suelo de la platea, anfiteatro y foso de orquesta	El tipo de madera del suelo era diferente a lo largo del teatro (en general roble). Como en conjunto el área que constituía era reducida, se agrupó en una única madera genérica.
Tela ligera	1,28	Cortinas triangulares	Escaso grosor y densidad. No estaban tensadas.
Moqueta de fieltro	1,25	Escaleras platea y acceso desde el hall	De escaso grosor; seguramente solo absorba en alta frecuencia.
Puertas	0,48	Varias	En lo que a aislamiento acústico se refiere eran bastante efectivas. Sin embargo, en términos de absorción su impacto es despreciable debido a la poca área que cubren. Al encerrar una cavidad entre puerta y puerta, se espera absorción a baja frecuencia.

Tabla 6.1 Materiales del modelo
















 Tela 450 gr/m ²	 Varios	 Butacas semitapizadas	 Contrachapado 15 mm
 Contrachapado 18 mm	 Okumen sobre DM	 Techo con tramoya	 Contrachapado
 Telón	 Madera genérica	 Tela ligera	 Moqueta de fieltro
 Puertas			

Fig. 6.3 Distribución de materiales en EASE Eyes. Vista en perspectiva contrapicado (izq.) y picado (drcha.)

Inicialmente, y partiendo de algunos de los materiales de la base de datos *Generic* de EASE, así como de otro proyecto que realizó un estudio acústico del teatro [15], se asignó a cada material unos coeficientes de absorción provisionales o genéricos. Para aquellos grupos de caras que constituían un área total despreciable (menos del dos por ciento) dichos coeficientes fueron definitivos, mientras que los materiales de mayor impacto acústico fueron reajustados atendiendo a las simulaciones.

Para agilizar el proceso, se creó una plantilla de Hojas de Cálculo de Google, constituida por seis hojas:

- **AURA F1, AURA F2, AURA F3:** Contienen los informes generados por *AURA* tras las simulaciones, para las 35 posiciones de micrófono y las fuentes F1, F2 y F3 respectivamente.
- **DATOS SIMULACIÓN:** Lee los parámetros RT20, C80, EDT y TS de los informes, agrupa las muestras por áreas y calcula los promedios espaciales, así como la absorción media según Sabine y Eyring.
- **DATOS IN SITU:** Contiene los parámetros RT20, C80, EDT y TS calculados con *Room EQ Wizard* a partir de las muestras obtenidas el día de la medición *in situ*.
- **GRÁFICAS COMPARATIVAS:** Lee los parámetros promediados y los compara uno a uno.

De esta manera, copiando y pegando en sus hojas correspondientes los informes generados tras las simulaciones, se obtiene al instante una comparativa visual que permite tomar decisiones a la hora de modificar los coeficientes de absorción de los distintos materiales.

Inicialmente se ajustó las áreas más separadas entre sí: RX C/C* (más sensible a las butacas, el suelo y el contrachapado de la zona del público) y STG (más sensible a los materiales del escenario). Una vez logrado un RT20 medio similar al obtenido *in situ*, se tuvo en cuenta las demás áreas, siendo editados los coeficientes de absorción de los materiales mediante un proceso iterativo hasta lograr un equilibrio en todas las áreas de audiencia del auditorio.

Para poder tener una idea del error cometido en los coeficientes de absorción de los materiales asignados a las caras del modelo, además del apoyo visual que ofrecieron las gráficas comparativas, se calculó también el error del tiempo de reverberación de la simulación respecto a las mediciones in situ para todas las bandas y áreas de audiencia, expresado como:

$$Error (\%) = 100 \cdot \frac{|RT20_{in\ situ} - RT20_{simulación}|}{RT20_{in\ situ}} \quad (6.1)$$

A continuación, se muestra la representación gráfica de los errores superpuestos de todas las áreas, lo que facilita determinar cuán acotado está dicho error:

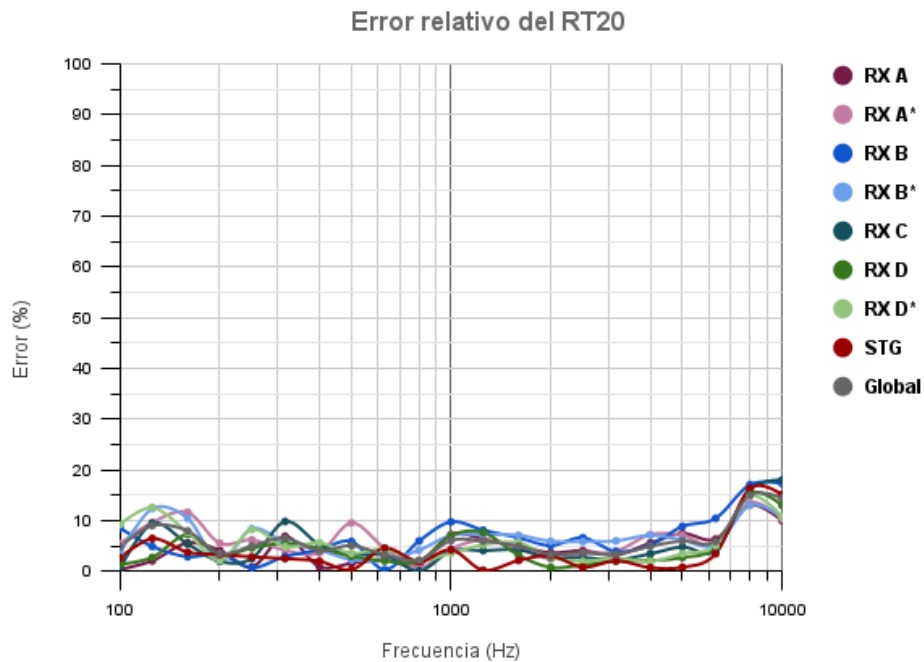


Fig. 6.4 Error relativo del RT20 cometido para todas las áreas de audiencia

Para todas las áreas de audiencia promediadas, el error obtenido se sitúa siempre por debajo del 18,01 % para todas las bandas, siendo además menor del 12,61 % en todas aquellas cuya frecuencia central es igual o inferior a 6.300 Hz. Es en la zona de alta frecuencia en la que mayor error se observa.

Se intentó ajustar los coeficientes de absorción de los distintos materiales situados en la zona del público para reducir este error. Sin embargo, editar los materiales de una cara tiene un impacto sobre el RT20 de todas las áreas. Puesto que se llegó a un punto en el que, lejos de reducir este error, lo único que se conseguía era trasladarlo de un área de audiencia a otra, y puesto que se consideró que este ya era razonablemente pequeño y puntual, se optó por dar por finalizado el proceso de ajuste de materiales. En consecuencia, se considera que el modelo está calibrado, y que sirve para la realización de simulaciones que permitan evaluar, de manera muy cercana a la realidad, distintas propuestas de sistemas de sonorización.

Finalmente, los coeficientes de absorción definitivos para los materiales del recinto fueron los siguientes:

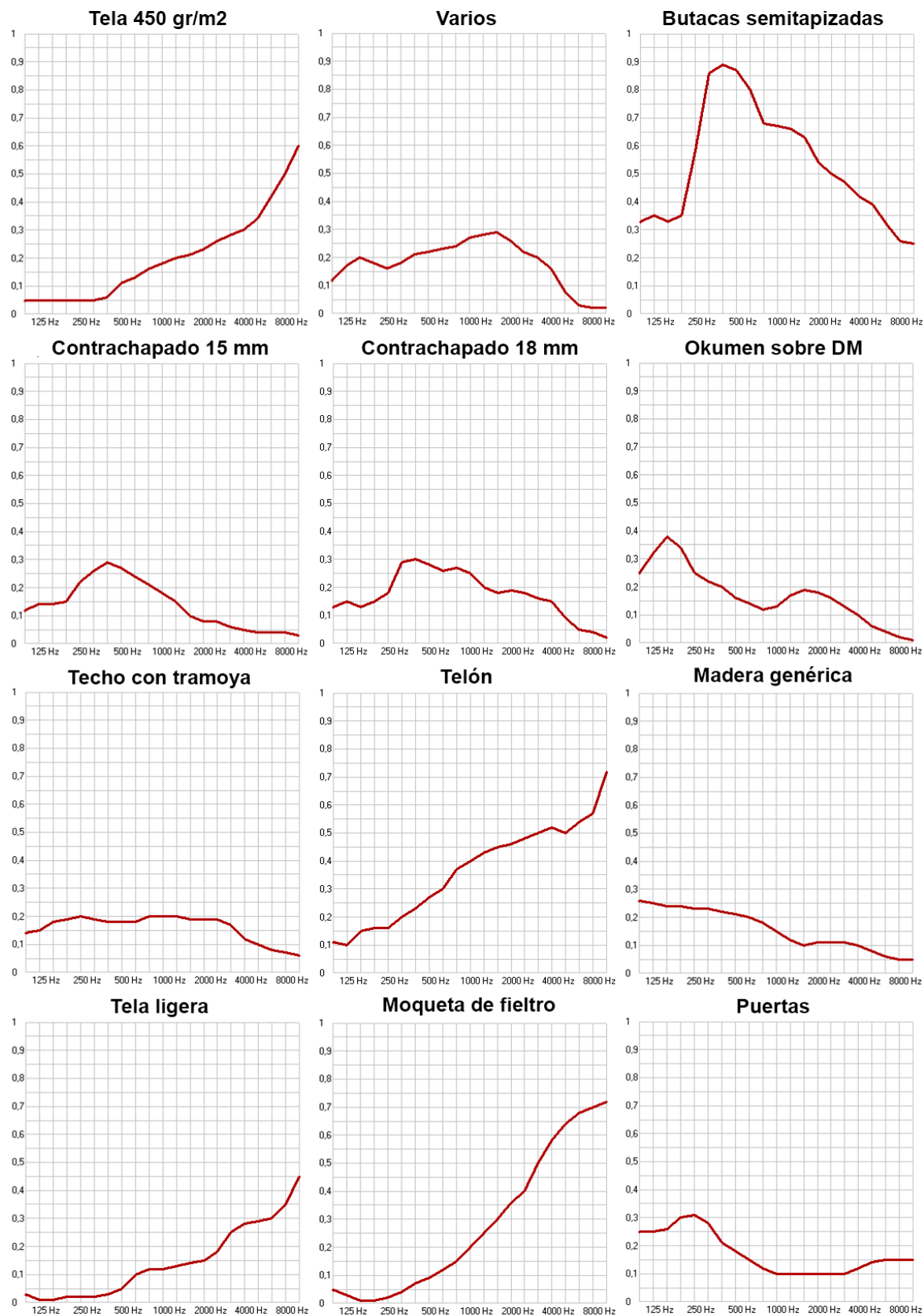


Fig. 6.5 Coeficientes de absorción de los materiales en tercios de octava

6.3. Comparación con los resultados de la medición in situ

En las gráficas de las siguientes hojas se muestra una comparativa por cada área de audiencia entre la simulación realizada en el modelo y los resultados de las mediciones *in situ* para los parámetros RT20, EDT, TS y C80 anteriormente descritos.

En la simulación, el RT20 no se diferencia demasiado del EDT, lo que indica que el modelo no logra replicar la caída inicial de energía que se comentó anteriormente. Sin embargo, en las mediciones realizadas *in situ* sí se observa cierta diferencia entre estos dos parámetros: en términos globales, el RT20 supera al EDT fuera del rango del habla humana (500 Hz – 4 kHz). Esto indica que, mientras que para una voz humana la reverberación percibida en el recinto sería próxima a la reverberación objetiva, para un programa con contenido musical (con más información en baja frecuencia) subjetivamente la acústica del teatro parecería más seca de lo que en realidad es, especialmente en el escenario.

Cabe mencionar que el RT20 global calculado, tanto para *in situ* como para la simulación, es más representativo del RT20 medio del área del público que del teatro al completo, pues la densidad de posiciones de micrófono en el escenario es considerablemente menor, lo que se traduce en una menor ponderación espacial del área STG. A su vez, esto resulta en una mayor incertidumbre a la hora de determinar la absorción de los materiales del escenario, obteniendo unos resultados en la simulación que difieren más de los obtenidos con las medidas reales.

Por otro lado, en los niveles de claridad se observa unos resultados bastante suavizados comparados a los reales, lo cual podría pulirse editando los coeficientes de *scattering* de los materiales con mayor detalle. Además, el área STG (escenario) obtiene un nivel de claridad considerablemente mayor que el de las demás áreas del público. Esto se debe a que los puntos de medición se encontraban bastante más cerca de la fuente y, al no tratarse de un campo difuso ideal, esto se traducía en una mayor relación sonido directo-reverberante.

Sin embargo, curiosamente, a pesar de que el área RX A de la platea se sitúa más próxima al escenario, se observa una claridad ligeramente menor respecto a las áreas RX C y RX D del anfiteatro, e incluso respecto a RX B. La explicación a este fenómeno podría ser la siguiente:

La platea se sitúa en una zona más abierta del teatro, en la que las paredes y superficies reflectantes se encuentran más separadas entre sí. En consecuencia, las reflexiones tardan más en llegar al receptor y se concentra mayor energía después del límite temporal precoz de 80 milisegundos, lo que se traduce en una menor claridad. Por el contrario, el anfiteatro es más cerrado: el techo y el suelo están bastante más próximos entre sí y, además, está la presencia de la pared del fondo del teatro. Esto provoca que el sonido tarde menos en llegar de una superficie a otra, lo que se traduce en una mayor concentración de energía en las reflexiones tempranas (mayor claridad). Lo mismo ocurriría con RX B que, si bien se sitúa en la platea, se encuentra muy próxima a la pared lateral del teatro.

La conclusión es que, si bien la claridad depende en gran parte de la distancia entre el receptor y la fuente (siempre que el campo no sea idealmente difuso), también viene determinada por las dimensiones del recinto y la separación entre las superficies reflectantes.

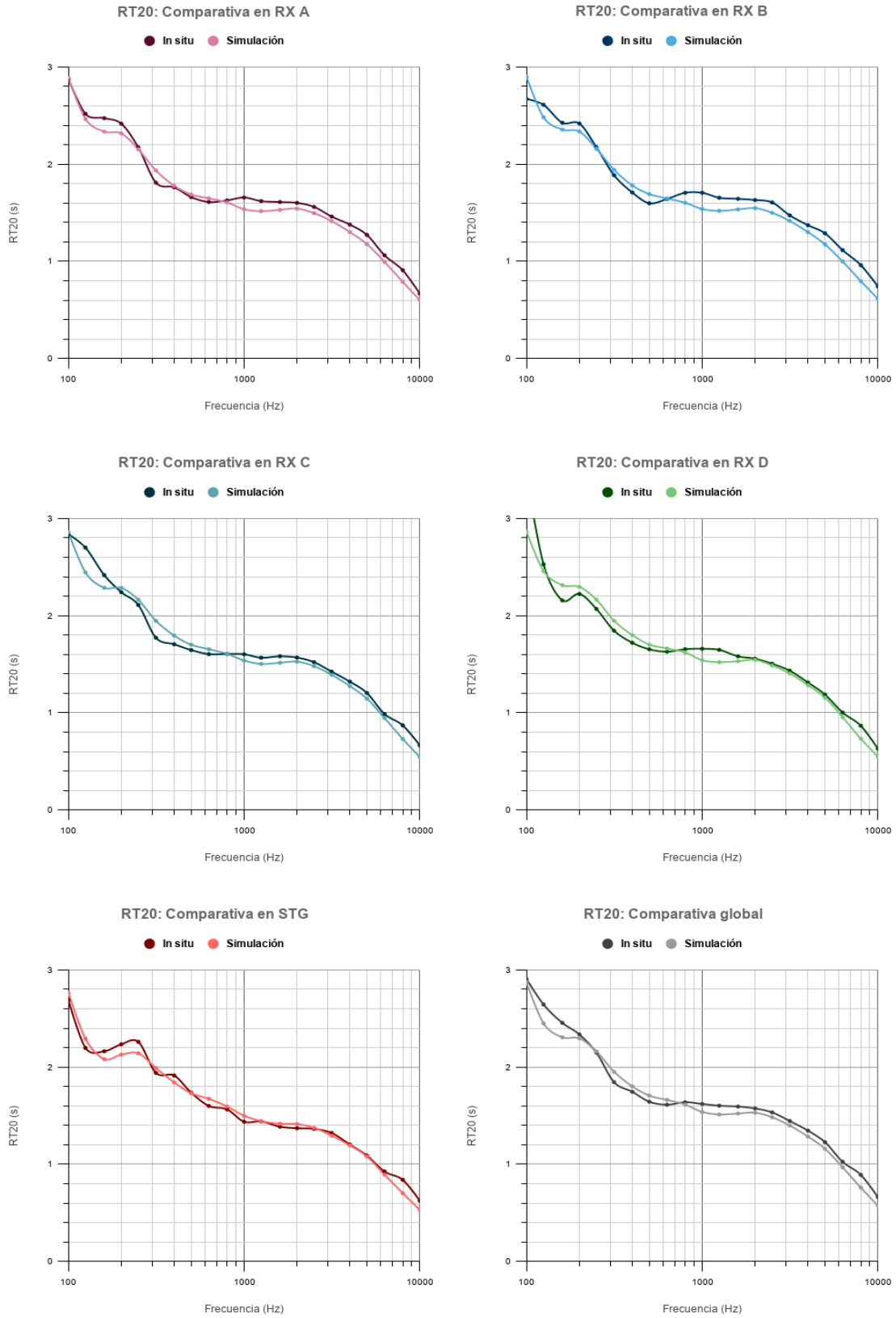


Fig. 6.6 Comparación gráfica del RT20 para cada área de audiencia

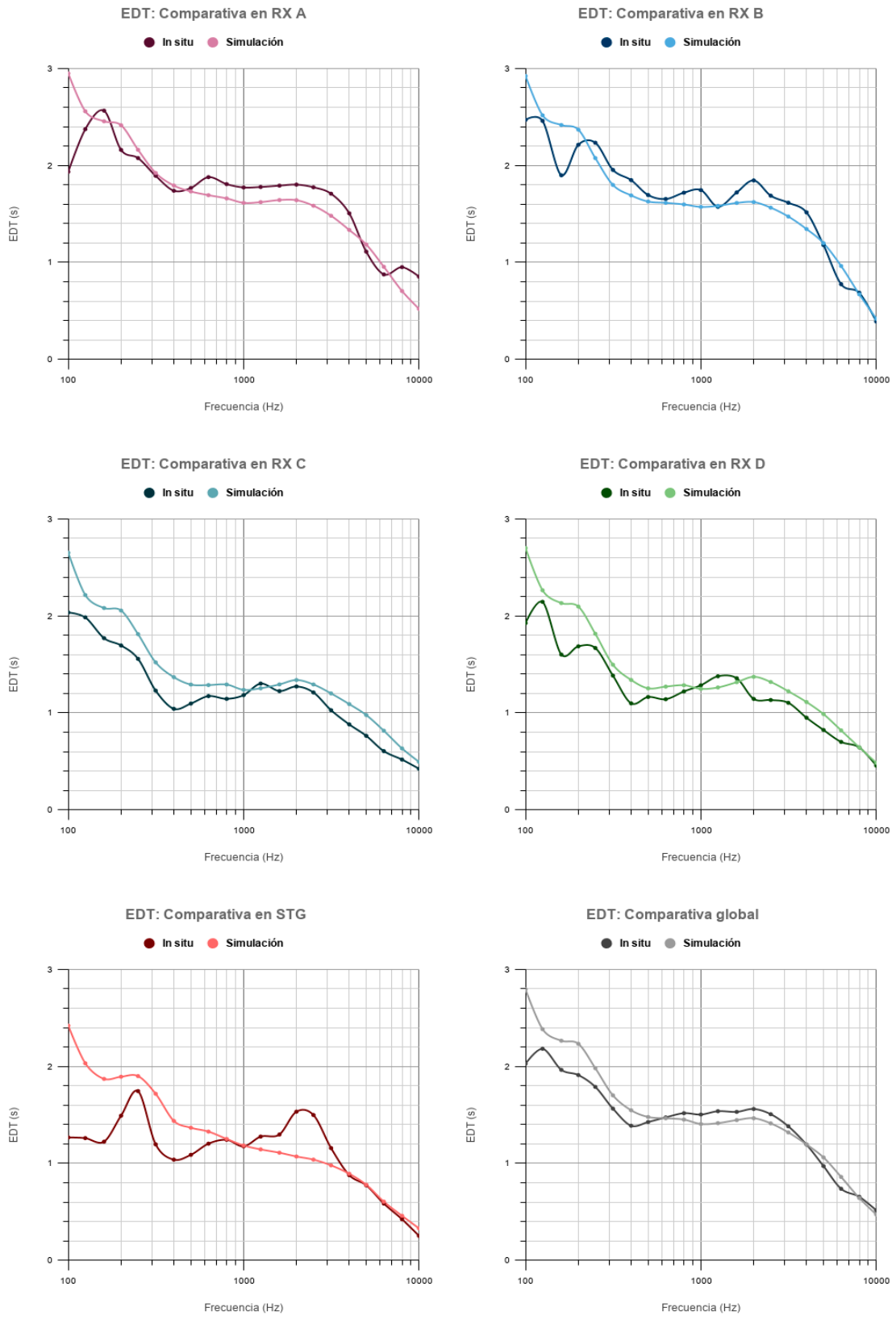


Fig. 6.7 Comparación gráfica del EDT para cada área de audiencia

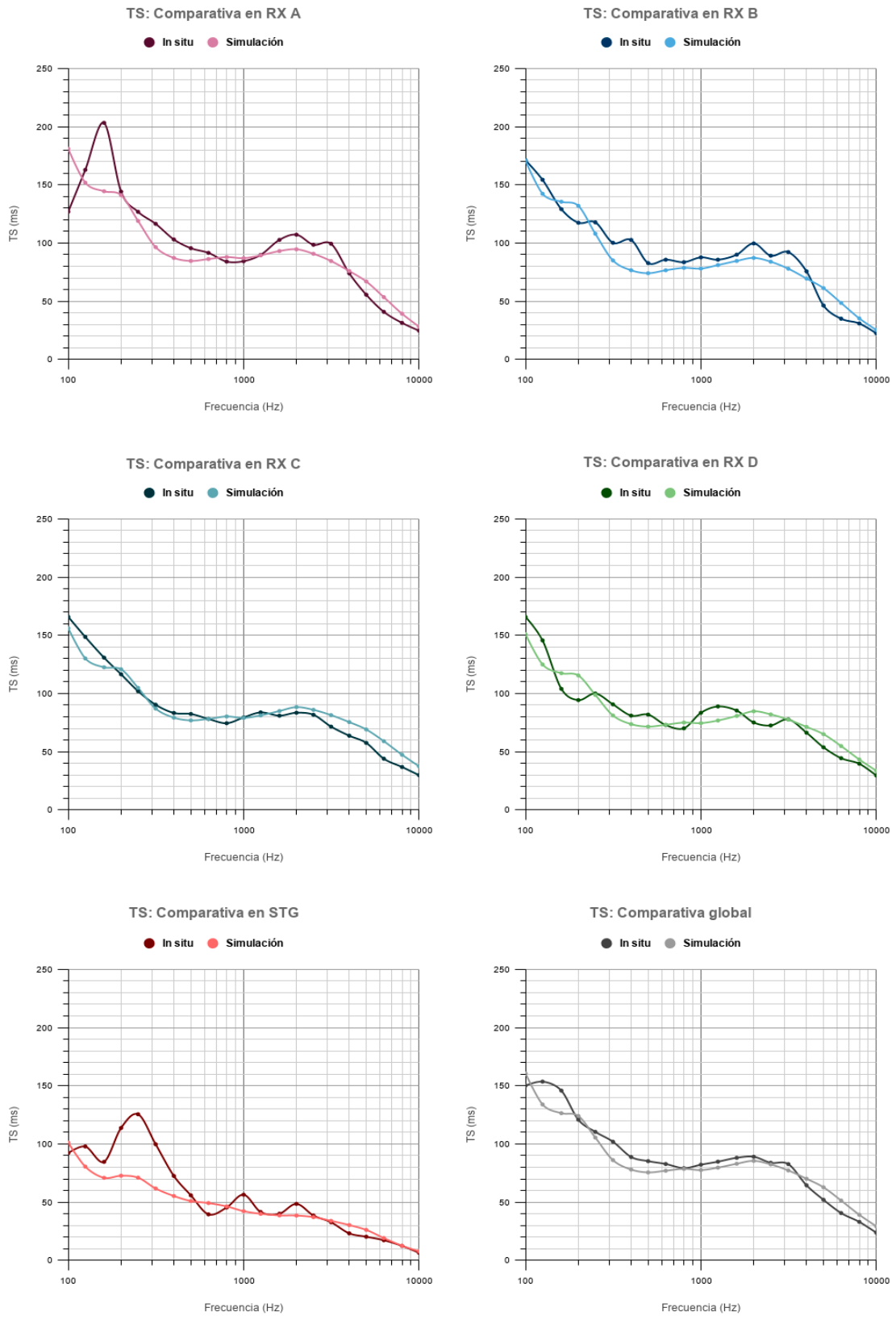


Fig. 6.8 Comparación gráfica del TS para cada área de audiencia

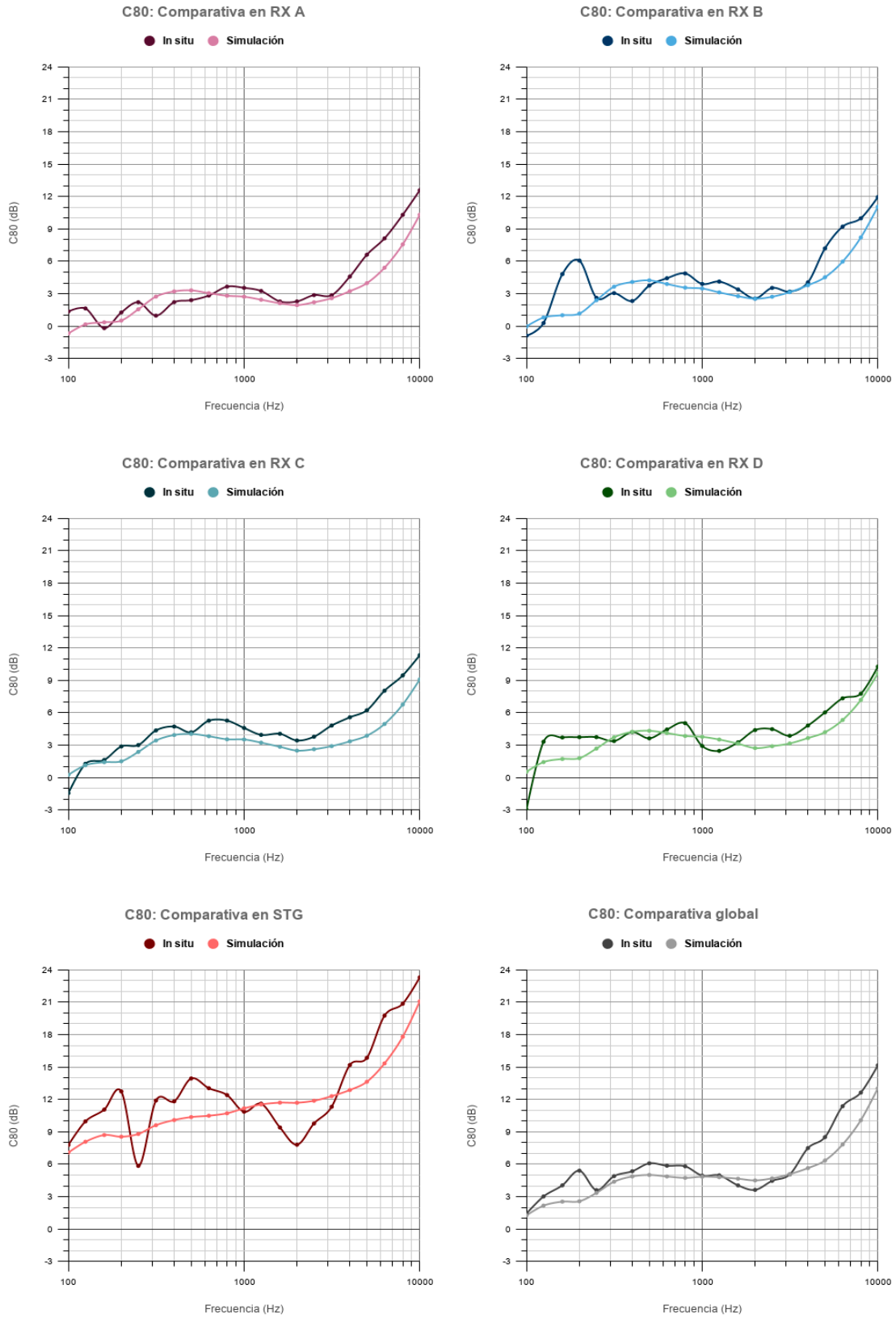


Fig. 6.9 Comparación gráfica del C80 para cada área de audiencia

Puesto que las anteriores gráficas comparan únicamente los parámetros promediados para las áreas de audiencia, adicionalmente se muestra una comparativa gráfica para un subconjunto de muestras. Concretamente, para el parámetro RT20, ya que fue el que más peso tuvo en el proceso de calibración del modelo.

Para ello, se ha escogido un subconjunto formado por un total de diez muestras, cinco por cada mitad del auditorio, dispersadas de manera homogénea, para distintas posiciones de fuente y teniendo una muestra por área de audiencia (dos en el caso del área STG).

Como el plan de muestreo dio lugar a un conjunto muy grande de muestras, se optó por usar los promedios como orientación, y no tanto los casos puntuales. Es por ello que, como se puede observar en las siguientes gráficas, los RT20 devueltos por la simulación difieren en algunos casos con las muestras reales, especialmente en media-baja frecuencia.

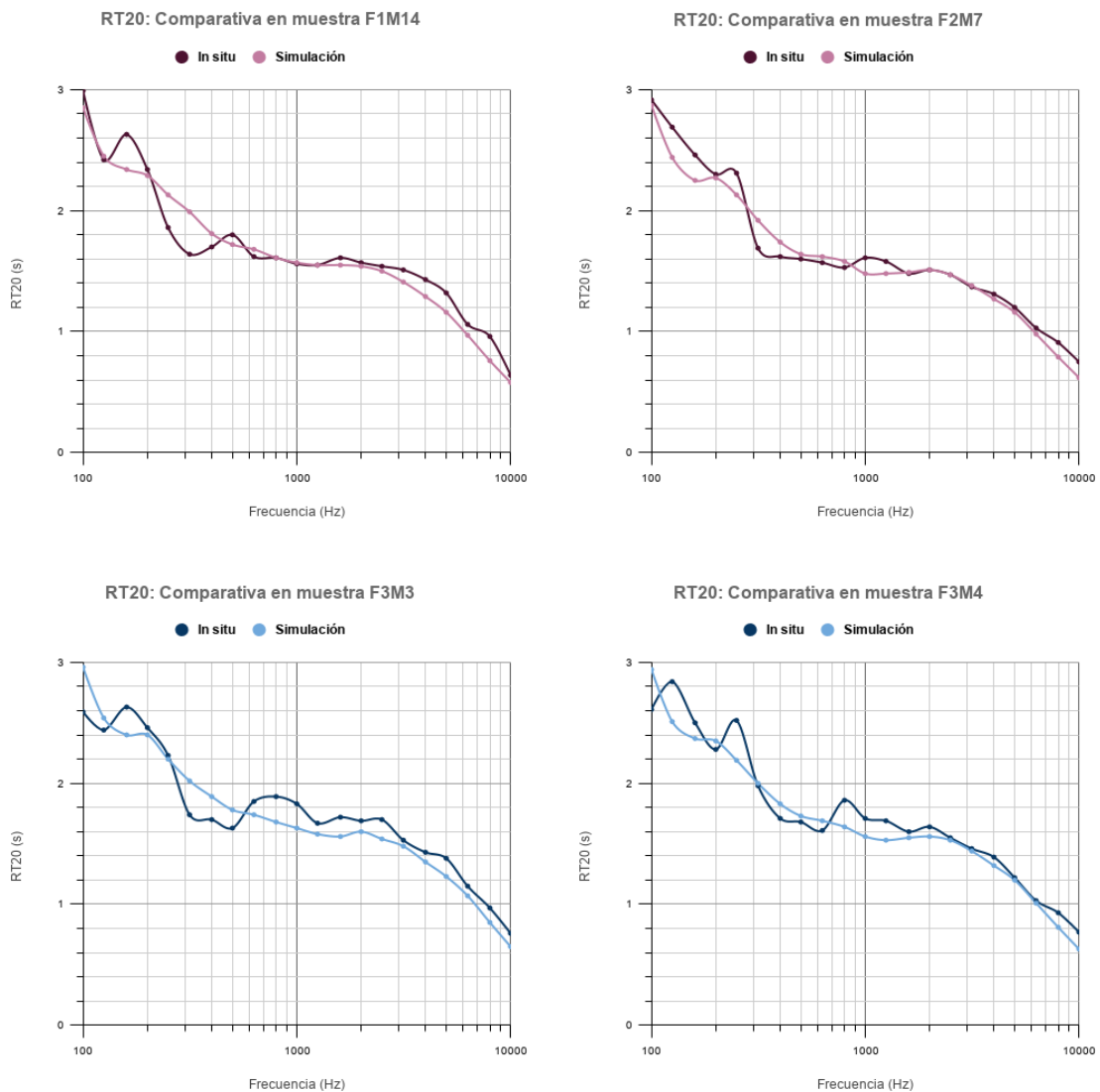


Fig. 6.10 Comparación gráfica del RT20 por muestras (1/2)

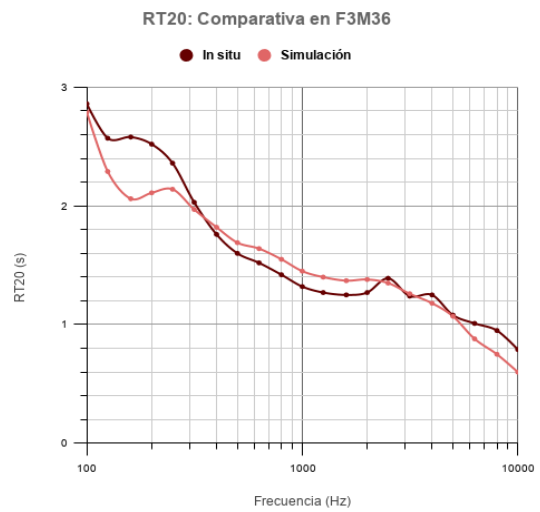
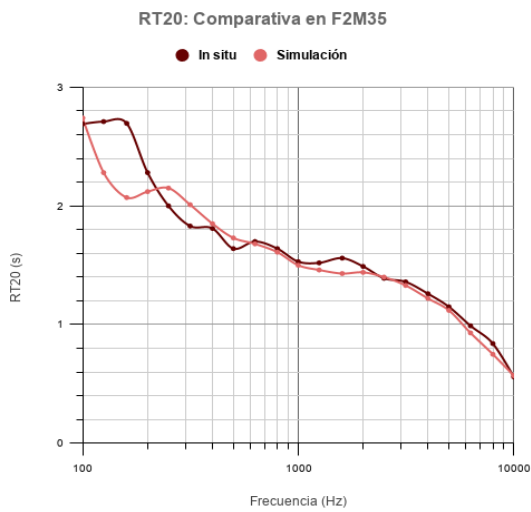
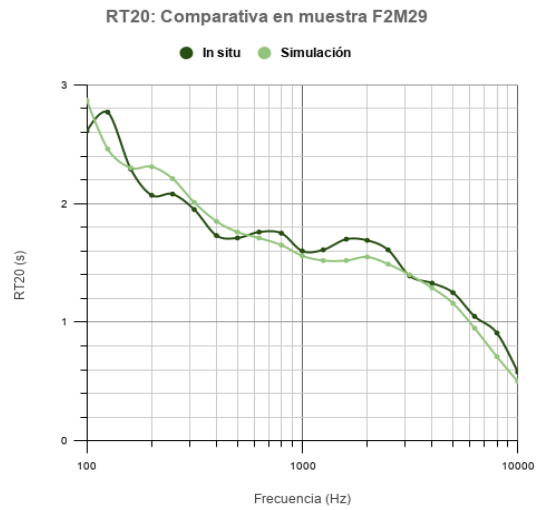
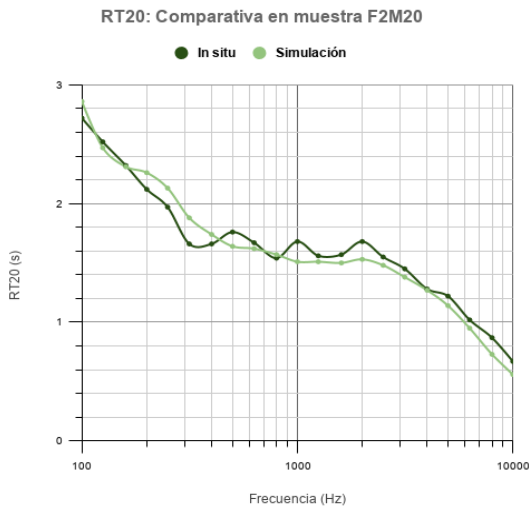
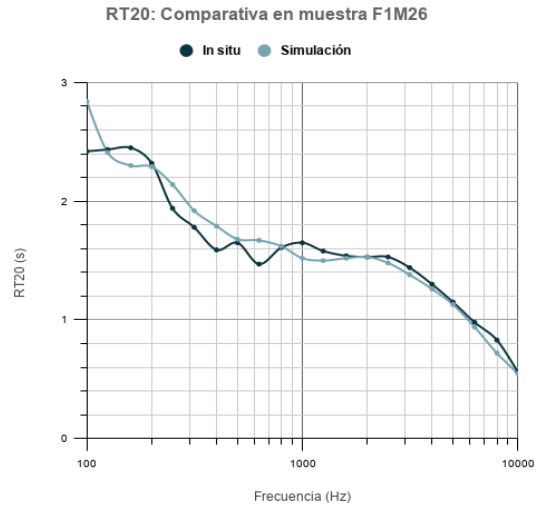
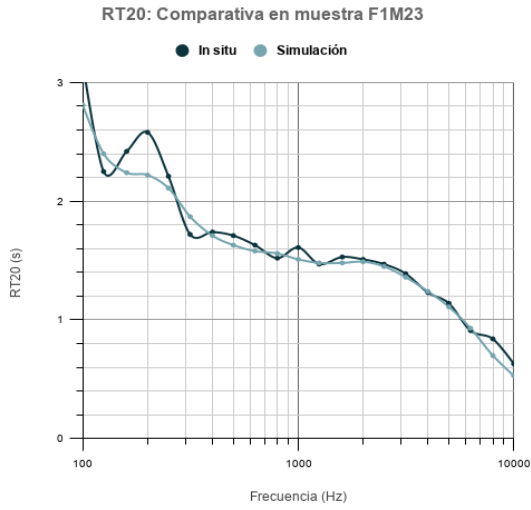


Fig. 6.11 Comparación gráfica del RT20 por muestras (2/2)

7. SISTEMA DE REPRODUCCIÓN *AMBISONICS*

7.1. Distribución de altavoces

Para reproducir completamente una señal ambisónica tridimensional de orden N , compuesta por $(N + 1)^2$ canales, se necesita un sistema de $L \geq (N + 1)^2$ altavoces, mientras que para una señal bidimensional compuesta por $2N + 1$ canales, se requiere $L \geq 2N + 2$.

Habitualmente, se emplea el mínimo número necesario de altavoces ya que, cuando se utiliza más de los necesarios, la interacción entre estos puede generar una degradación en alta frecuencia debido a un filtrado de tipo peine. Por tanto, si se desea sonorizar un área de audiencia de un tamaño considerable, no solo conviene aumentar L , sino también el orden N de la señal.

Por este motivo, y puesto que no se puede emplear altavoces de refuerzo, ya que la recreación espacial del campo sonoro podría verse comprometida; se optó por un sistema HOA de un orden que permitiese cubrir el área de audiencia del auditorio. Concretamente, se empleó un sistema de reproducción 3D de *Ambisonics* de orden $N = 3$, cuyo sistema óptimo de reproducción consta de 16 altavoces iguales equidistantes al *sweet spot* y orientados a él, distribuidos como en la Figura 7.1:

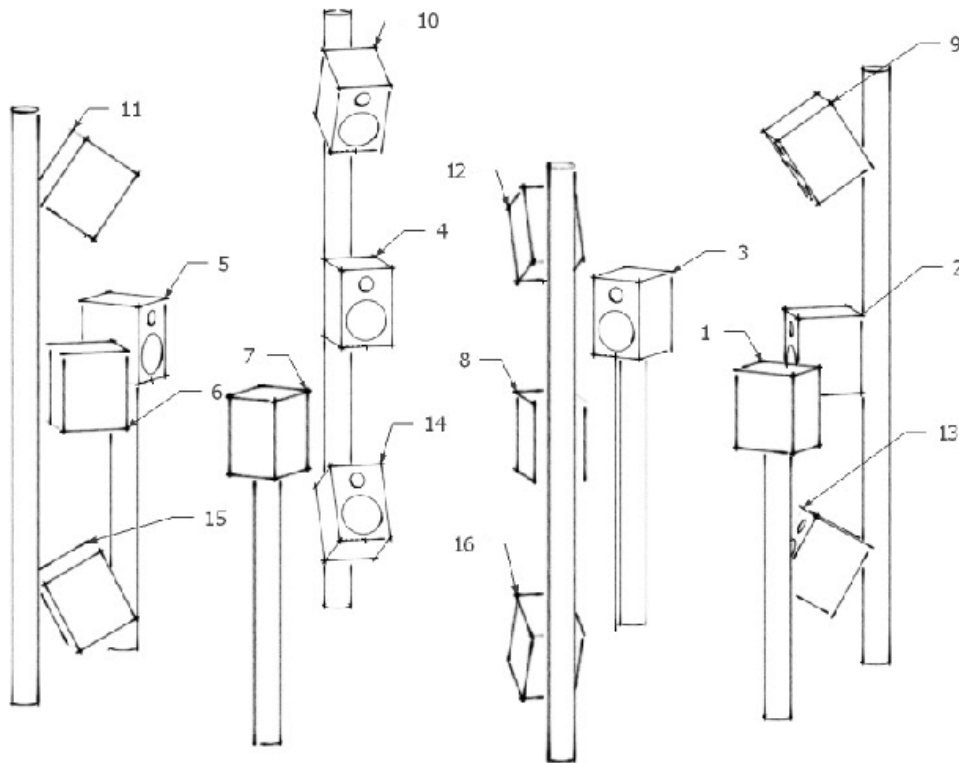


Fig. 7.1 Sistema de altavoces óptimo para *Ambisonics* de tercer orden

Los altavoces 1-8 cubren el plazo azimutal en saltos de 45° , mientras que los altavoces 9-16 cubren las esquinas superiores (para una elevación de $\vartheta = 45^\circ$) e inferiores ($\vartheta = 135^\circ$). Esta distribución ofrece, por tanto, mayor resolución espacial en el plano azimutal que en la componente de elevación.

Sin embargo, la geometría del teatro imposibilita una distribución óptima de los altavoces, por lo que es necesario situarlos de manera irregular y posteriormente compensar los desajustes temporales y de nivel producidos por las distintas distancias al *sweet spot* mediante ganancias y retardos a la señal que ha de emitir cada altavoz.

Tanto la disposición inclinada de las áreas de audiencia como la diferencia de altura entre la platea y el anfiteatro dificultan la distribución de los altavoces 1-8 en un plano horizontal, así como los altavoces inferiores 13-16. Como la platea tiene una menor inclinación y cuenta con un mayor número de butacas que el anfiteatro, se decidió focalizar la sonorización en ella, diferenciando un conjunto de áreas de audiencia con sonorización espacial (RX A/A* y RX B/B*) y otro conjunto convencional (RX C y RX D/D*).

Por tanto, el sistema propuesto consta de 12 altavoces distribuidos en forma de semiesfera irregular focalizada en un *sweet spot* situado en el centro de la platea, prescindiendo de los altavoces 13-16. Debido a la imposibilidad de instalar altavoces de refuerzo para el anfiteatro, este se verá cubierto acústicamente sin el beneficio de la percepción espacial, y con una menor claridad.

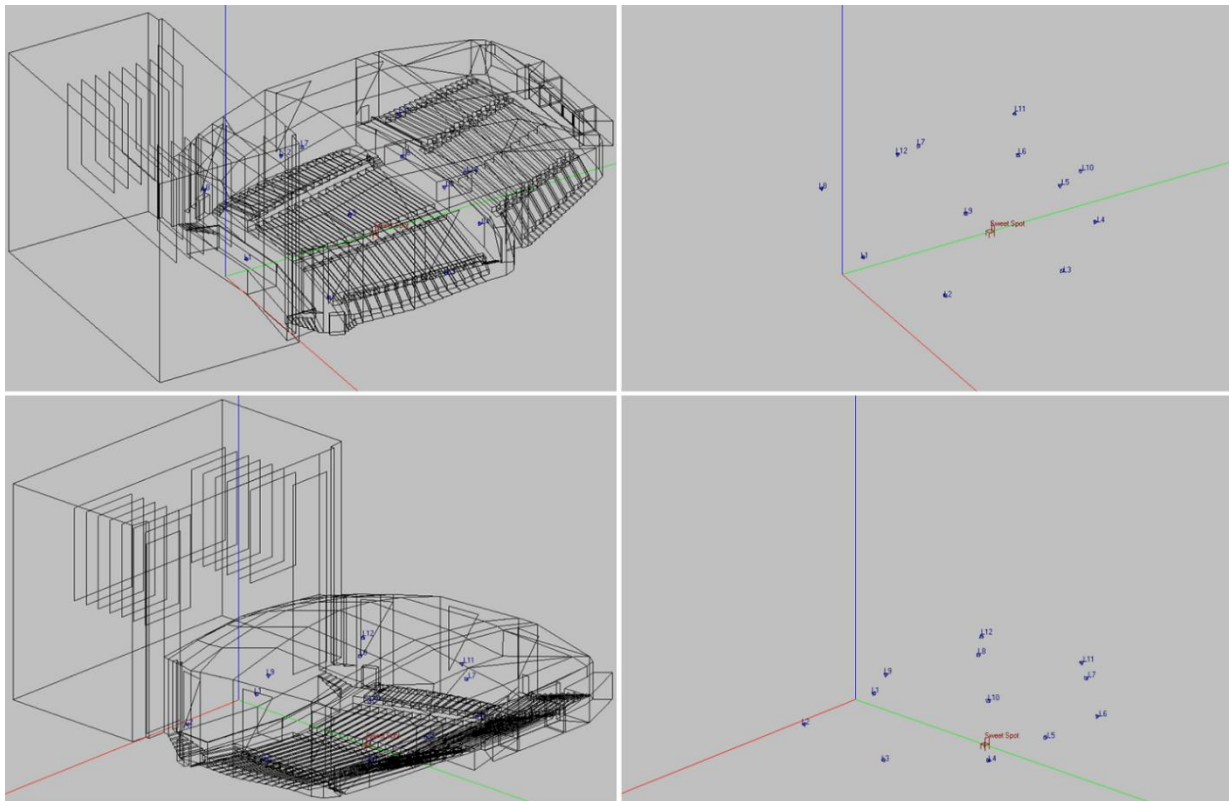


Fig. 7.2 Distribución del sistema de altavoces propuesto en EASE

El modelo de altavoz elegido fue el HM-1 de la compañía Meyer Sound. Se trata de un altavoz autoamplificado de rango completo (42 Hz – 20 kHz) con dos vías concéntricas: un *woofer* de 7" y un *tweeter* de 1". Cuenta además con dos cavidades de *bass-reflex* en la cara frontal para aprovechar la radiación trasera del *woofer* y encaminarla por un conducto de una longitud determinada para enfatizar los graves, logrando gracias a ello una frecuencia mínima de emisión algo más baja. Su cobertura a 1 kHz es de 100°, tanto en vertical como en horizontal (simetría de revolución).

7.2. Decodificador programado en Matlab

Como se explicó en el capítulo de *Ambisonics*, para reproducir estas señales en un sistema de altavoces previamente hay que decodificarlas para determinar la señal que ha de emitir cada altavoz. Por ello, para poder llevar a cabo una auralización en el modelo, se programó en Matlab un decodificador escalable en orden (desde $N = 0$ hasta $N = 3$) y en número de altavoces. Este código y las auralizaciones están accesibles en el enlace de descarga que aparece en el capítulo [9](#).

El decodificador propuesto permite además desplazar el *sweet spot* y/o las posiciones de los altavoces, adaptando de manera conveniente los pesos de cada canal para cada una de las señales decodificadas, y compensando mediante ganancias y retardos las diferencias en distancia que pudieran surgir entre los altavoces y el punto dulce de la sala.

7.2.1. Funcionamiento del decodificador

El decodificador, además de la señal en formato B, recibe las coordenadas del *sweet spot* y los altavoces en EASE. Como el criterio empleado para los ejes fue distinto al de AmbiX, fue necesaria una conversión:

Eje	EASE Room Editor	AmbiX
x^+	<i>Izquierda</i>	<i>Delante</i>
y^+	<i>Atrás</i>	<i>Izquierda</i>
z^+	<i>Arriba</i>	<i>Arriba</i>

Tabla 7.1 Criterios de coordenadas en EASE Room Editor vs AmbiX, ambos desde el punto de vista del público

Por tanto, siendo $\{x, y, z\}$ los ejes de EASE y $\{x', y', z'\}$ los de AmbiX, el decodificador lleva a cabo la siguiente operación para las coordenadas recibidas:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} -y \\ x \\ z \end{bmatrix} \quad (7.1)$$

Hecho esto, resta las coordenadas del *sweet spot* a las de los altavoces, obteniendo sus respectivos vectores de posición $\{\theta_1, \dots, \theta_L\}$, definidos en el apartado [3.6.2.1](#). Dados los vectores θ_l de cada altavoz l , realiza un cambio de coordenadas cartesianas a esféricas:

$$\theta_l = \begin{bmatrix} x'_l \\ y'_l \\ z'_l \end{bmatrix} \xrightarrow{\text{Conversión a esféricas}} \begin{bmatrix} \rho_l \\ \varphi_l \\ \vartheta_l \end{bmatrix} = \begin{bmatrix} \sqrt{x'_l{}^2 + y'_l{}^2 + z'_l{}^2} \\ \arctan \frac{y'_l}{x'_l} \\ \frac{\pi}{2} - \arctan \frac{z'_l}{\sqrt{x'_l{}^2 + y'_l{}^2}} \end{bmatrix} \quad (7.2)$$

A partir del azimut y la elevación (φ_l, ϑ_l) , y conociendo la función $Y_n^m(\varphi, \vartheta)$ que describe los armónicos esféricos definidos en el apartado 3.1, determina el peso $Y_n^m(\varphi_l, \vartheta_l)$ de cada canal en la señal que ha de emitir el altavoz l . Hecho esto, calcula los coeficientes a_n de acuerdo al modelo $max-r_E$ y normaliza el conjunto de señales para un número total de L de altavoces para mantener la sonoridad, logrando una decodificación óptima para un sistema de altavoces repartidos homogéneamente en forma de esfera y de manera orientada y equidistante al *sweet spot* (ver fórmula 3.32 y Figura 3.13).

Sin embargo, debido a las limitaciones físicas que impiden una distribución regular de los altavoces, se producen desfases y desajustes de nivel que comprometen la reconstrucción espacial del campo sonoro. Por ello, es necesario ajustar los altavoces con ganancias y retardos para compensar las diferentes distancias respecto al *sweet spot*. Para determinar las compensaciones necesarias para cada altavoz, basta con conocer su distancia ρ_l respecto al punto dulce de la sala.

El ajuste de ganancia G_l aplicado a cada señal decodificada se realiza tomando como referencia la distancia al *sweet spot* del altavoz más lejano. Representada en unidades naturales, la atenuación aplicada sobre el altavoz l es la siguiente:

$$G_l = \frac{\rho_l}{\max\{\rho_l\}} \leq 1 \quad (7.3)$$

De igual manera, el retardo k_l en muestras aplicado sobre la señal decodificada toma como referencia el altavoz más lejano.

$$k_l = f_s \cdot \frac{\max\{\rho_l\} - \rho_l}{c} \quad (7.4)$$

Puesto que este se implementó mediante la función *delayseq* de Matlab, permite aplicar retardos para un k_l con decimales, empleando la interpolación entre muestras. Esto permite un ajuste temporal más preciso e independiente de la frecuencia de muestreo f_s , ya que no es necesario truncar ni aproximar la cantidad de muestras a retardar.

7.2.2. Creación de señales de test en Nuendo

Para comprobar el funcionamiento del decodificador y poder realizar auralizaciones con la acústica del modelo, es necesario contar con señales de test de *Ambisonics* en formato B grabadas en cámara anecoica. Dado que, por limitaciones técnicas, no se pudo grabar señales HOA mediante sistemas como el expuesto en el apartado 3.3.3, se optó por grabarlas en mono en Nuendo 11 y posteriormente panoramizarlas para los órdenes $N = \{0, 1, 2, 3\}$ mediante el *plug-in* MultiPanner, empleando la convención *AmbiX*. En esencia, este proceso de codificación se basa en lo explicado en el apartado 3.4.

Nombre del archivo	Descripción
<i>audiotest_nth_axis.wav</i>	Grabación en la que se pronuncia las palabras “delante, detrás, izq., dcha., arriba, abajo”, con sus correspondientes panoramizaciones.
<i>audiotest_nth_corners.wav</i>	Grabación en la que se pronuncia las frases: “esquina frontal superior izquierda, esquina frontal superior derecha, esquina frontal inferior izquierda, esquina frontal inferior derecha, esquina trasera superior izquierda, esquina trasera superior derecha, esquina trasera inferior izquierda, esquina trasera inferior derecha”.
<i>audiotest_nth_clock_x=0.wav</i>	Grabación que recorre, en sentido antihorario y saltos de 45°, una vuelta completa de la elevación ϑ con un azimut fijado en $\varphi = 90^\circ$.
<i>audiotest_nth_clock_y=0.wav</i>	Grabación que recorre en saltos de 45°, una vuelta completa de la elevación ϑ con un azimut fijado en $\varphi = 0^\circ$.
<i>audiotest_nth_clock_z=0.wav</i>	Grabación que recorre, en sentido antihorario y saltos de 45°, una vuelta completa del azimut φ con una elevación fijada en $\vartheta = 0^\circ$.

Tabla 7.2 Descripción de las señales de test para el decodificador



Fig. 7.3 Plug-in MultiPanner de Nuendo 11

Para poder comprobar en primera persona que la panoramización realizada en MultiPanner era correcta, se empleó el decodificador biaural en tiempo real que proporciona Nuendo. Este permite monitorizar la señal mediante auriculares y percibir su contenido espacial a través de un formato estéreo.



Fig. 7.4 Decodificador Ambisonics/Biaural de Nuendo 11

Tal y como se explicó en el apartado [3.6.1.2](#), los decodificadores biaurales mapean la señal en un sistema virtual de altavoces y generan el estéreo mediante filtros HRTF. Si bien el decodificador de Nuendo permite insertar una HRTF personalizada, ofrece también la posibilidad de emplear dos modelos genéricos en caso de no disponer de los filtros personalizados.

A falta de una HRTF propia, la panoramización se comprobó mediante estos dos modelos genéricos: aunque en el plano azimutal la experiencia de espacialidad fue satisfactoria (salvo para $\varphi = 180^\circ$), la componente de elevación no se recreaba de manera fiable. Por tanto, fue necesario recurrir a un medidor de espacialidad para comprobar la panoramización de las señales de test de manera objetiva.

Fue entonces cuando se empleó el *plug-in* SuperVision que proporciona Nuendo. Este medidor representa, de manera gráfica y a tiempo real, la panoramización 3D de la señal, así como la forma de onda de las señales que portan cada uno de sus canales. Esto último permite observar, para una panoramización determinada, qué canales contienen información y cuáles son las amplitudes y polaridades de cada uno de ellos.

Teniendo en cuenta que la convención empleada en el codificador de Nuendo fue *AmbiX* (WYZX), en las siguientes capturas (Figuras 1, 2, 3 y 4) se puede observar que el comportamiento de los canales, tanto en contenido como en polaridad, es coherente con los armónicos esféricos descritos en la Figura 3.1 y la Tabla 3.2 del apartado [3.1](#).

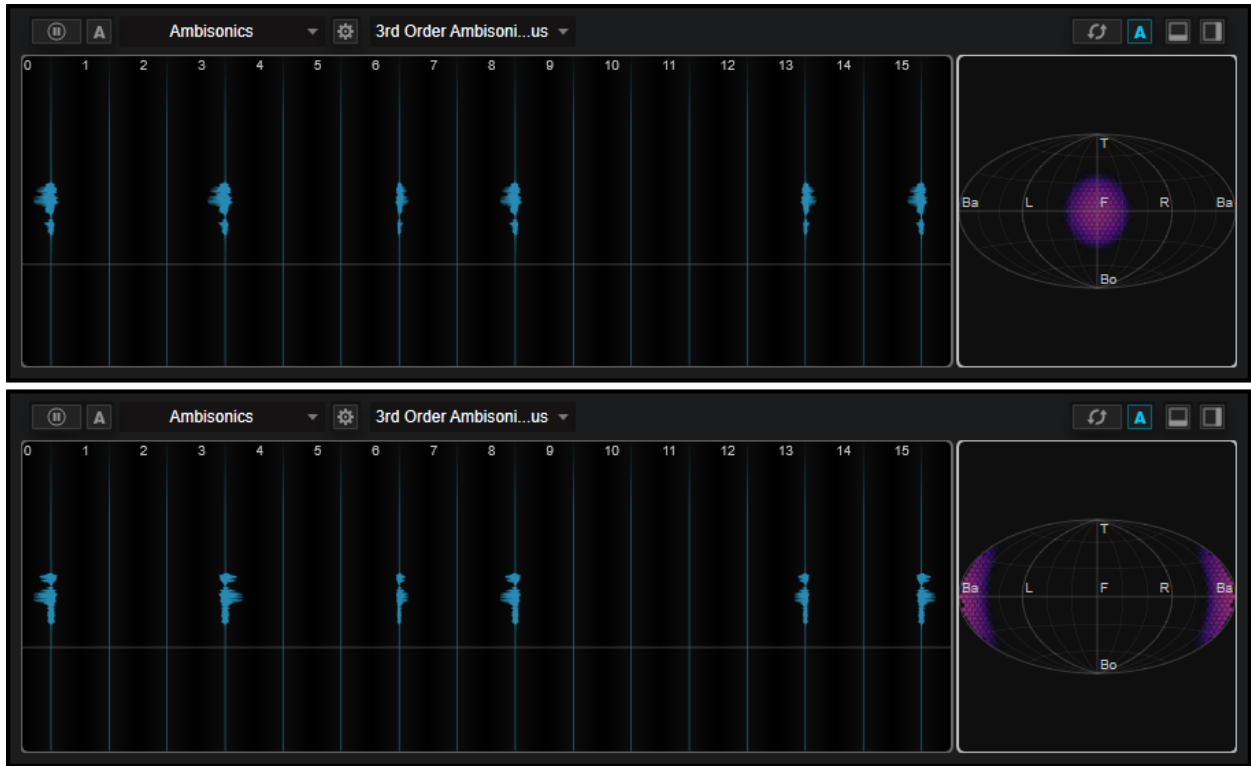


Fig. 7.5 Captura de la forma de onda y la panoramización en SuperVision (HOA orden 3): eje x (delante/detrás)

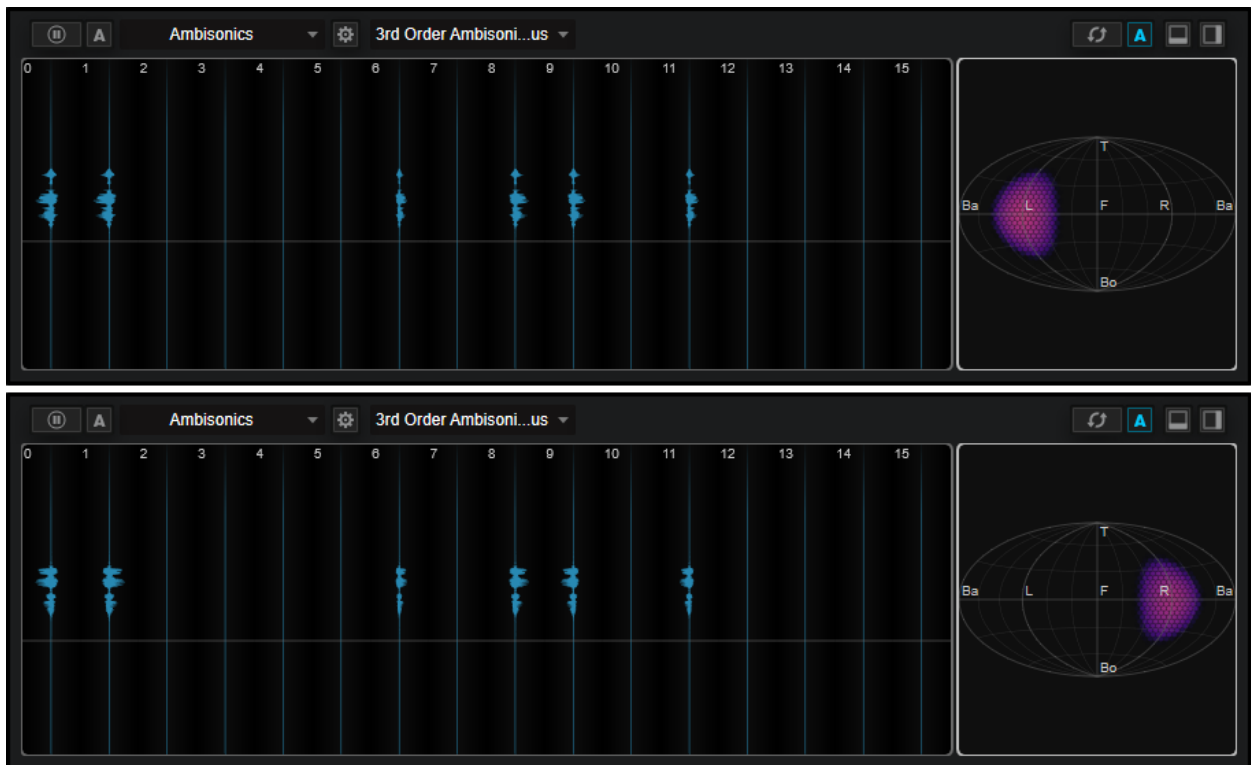


Fig. 7.6 Captura de la forma de onda y la panoramización en SuperVision (HOA orden 3): eje y (izquierda/derecha)

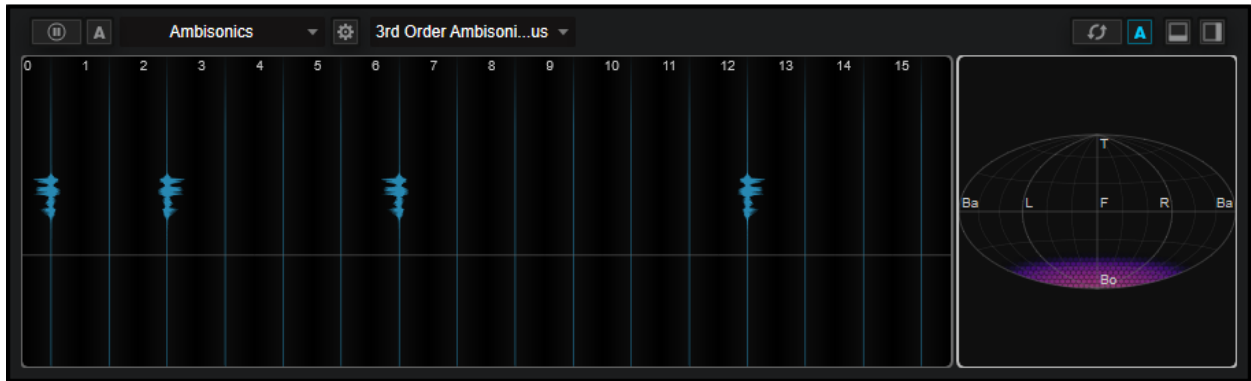
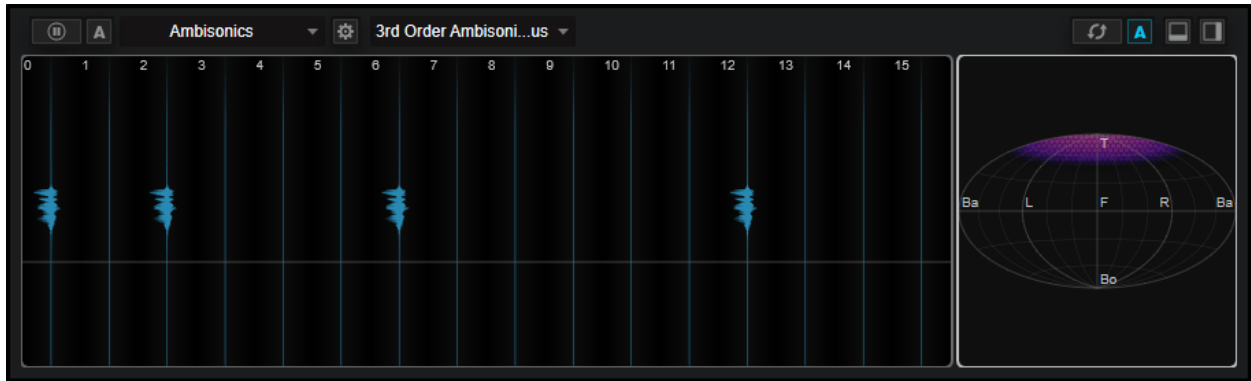


Fig. 7.7 Captura de la forma de onda y la panoramización en SuperVision (HOA orden 3): eje z (arriba/abajo)

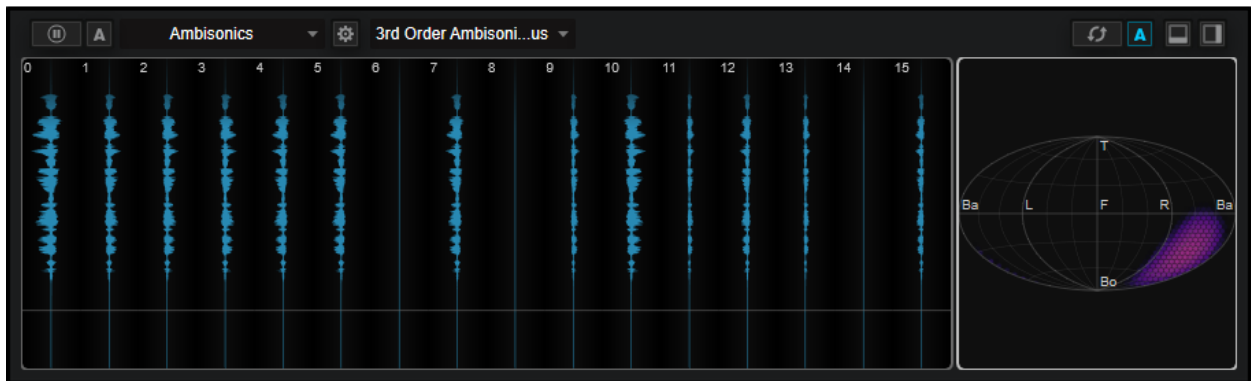
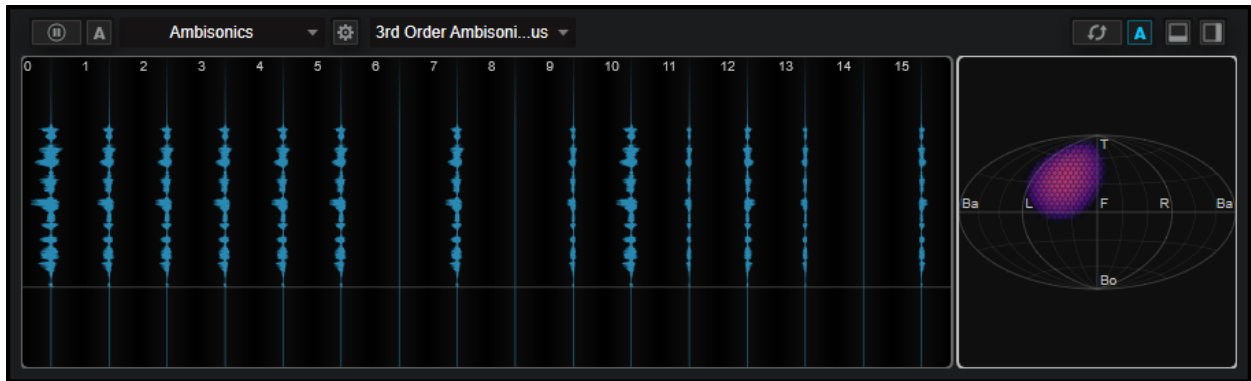


Fig. 7.8 Captura de la forma de onda y la panoramización en SuperVision (HOA orden 3): equinas front-sup-izq y tras-inf-dcha

7.3. Auralización en EASE Ears

Las auralizaciones se construyeron de acuerdo a la fórmula 3.30 del apartado 3.6.1.2. Mediante varias simulaciones realizadas en AURA sobre el modelo del auditorio, se generó en el *sweet spot* una respuesta al impulso biaural por cada altavoz del sistema. Estas doce respuestas emplean la HRTF por defecto de EASE, por lo que contienen la L-HRTF en su canal izquierdo y la R-HRTF en el derecho.

Partiendo del principio de linealidad y superposición, y empleando algunas de las señales de test creadas en Nuendo, se sumó todas las señales decodificadas de cada altavoz convolucionadas cada una de ellas con su respectiva respuesta al impulso biaural mediante el módulo *Ears* de EASE, obteniendo así la auralización: una aproximación de lo que un espectador situado en el *sweet spot* escucharía si se reprodujese la señal en cuestión por el sistema de altavoces propuesto.

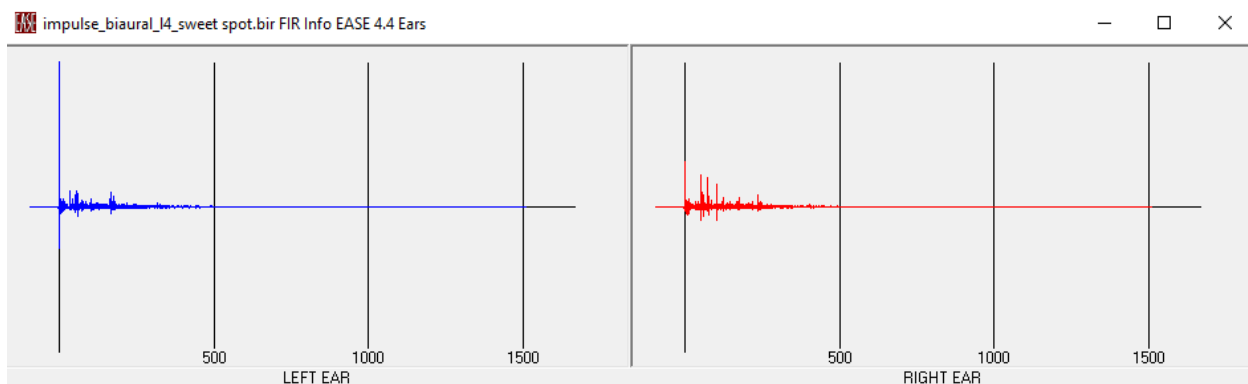


Fig. 7.9 L-HRIR y R-HRIR generados en AURA para el altavoz L4, teniendo en cuenta la acústica del recinto.

Para evaluar el impacto que tiene la acústica del teatro sobre la localización del sistema *Ambisonics*, se construyó las auralizaciones para dos situaciones, que fueron reproducidas en auriculares para ser evaluadas y comparadas entre sí:

- **Situación 1: Acústica del auditorio.**
Emplea los materiales con los coeficientes de absorción y *scattering* obtenidos tras la calibración del modelo, conteniendo la respuesta del altavoz HM-1 más la respuesta aproximada del recinto.
- **Situación 2: Campo libre ideal.**
Se estableció una absorción total en la sala, eliminando las reflexiones y convirtiendo su respuesta al impulso una delta de Dirac; esta auralización solo contiene la respuesta del altavoz HM-1.

Como era de esperar, la situación 1 ofrece una localización algo más confusa que la situación 2, aunque en ambas se percibe claramente la espacialidad, especialmente en el plano azimutal. Sin embargo, la percepción de altura no está tan lograda, sobre todo en panoramizaciones situadas debajo del usuario, debido a la distribución semiesférica de altavoces. En términos generales, la sala es lo suficientemente “seca” como para aceptar un sistema de reproducción *Ambisonics*, aunque podría mejorarse la localización reduciendo su tiempo de reverberación.

Hay que tener en cuenta que, dado que las respuestas biaurales generadas en AURA emplean una HRTF genérica, el realismo en la espacialidad percibida en las auralizaciones depende del usuario que las evalúe.

7.4. SPL directo y C80 en el área de audiencia

Adicionalmente, a continuación, se muestra el resultado obtenido en el módulo *Room Mapping* de EASE para la cobertura del SPL directo y la claridad (C80) en todas las áreas de audiencia, emitiendo cada altavoz 93 dB a un metro de distancia. Dado que el contenido más común a reproducir en un teatro es la voz humana, se evalúa la cobertura en bandas de octava con frecuencias centrales entre 500 Hz y 4 kHz.

Como era de esperar, la platea obtiene una mayor relación sonido directo sobre reverberante que el anfiteatro y el escenario. En términos globales, el SPL directo se sitúa en torno a 81 dB en la platea, 73 dB en el anfiteatro y 68 dB en el escenario, mientras que el C80 ronda los 10 dB en la platea, 1 dB en el anfiteatro y aproximadamente -7 dB en el escenario. Esto último puede resultar beneficioso de cara a prevenir acoples, ya que la coherencia entre las reflexiones se reduce y es más difícil que los micrófonos de los actores provoquen realimentaciones con el sistema de altavoces.

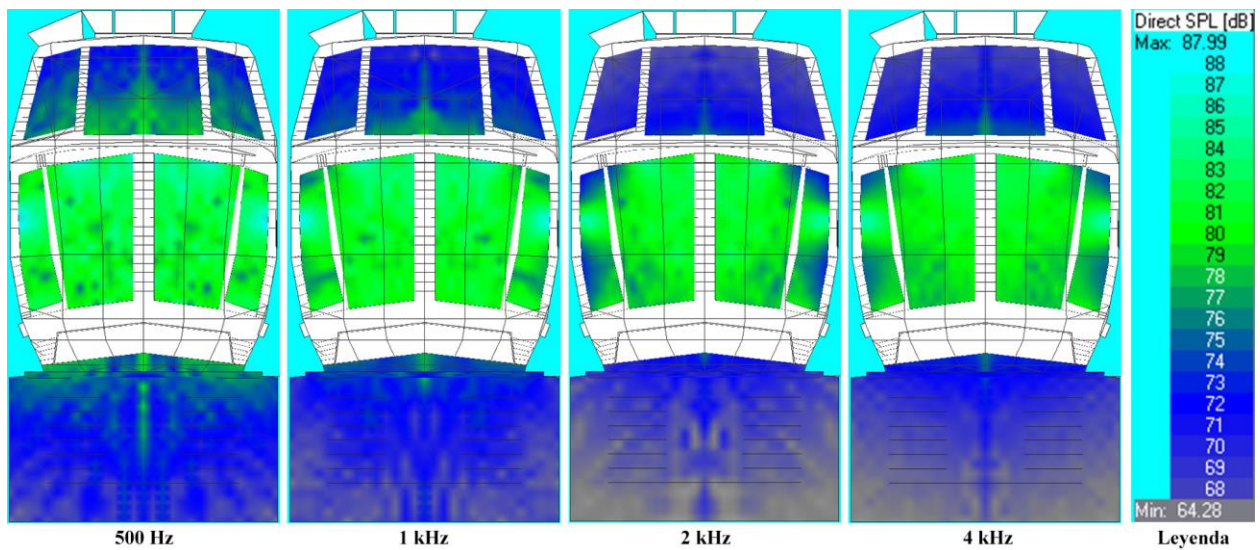


Fig. 7.10 Vista aérea del modelo con la representación de la cobertura del SPL directo en bandas de octava

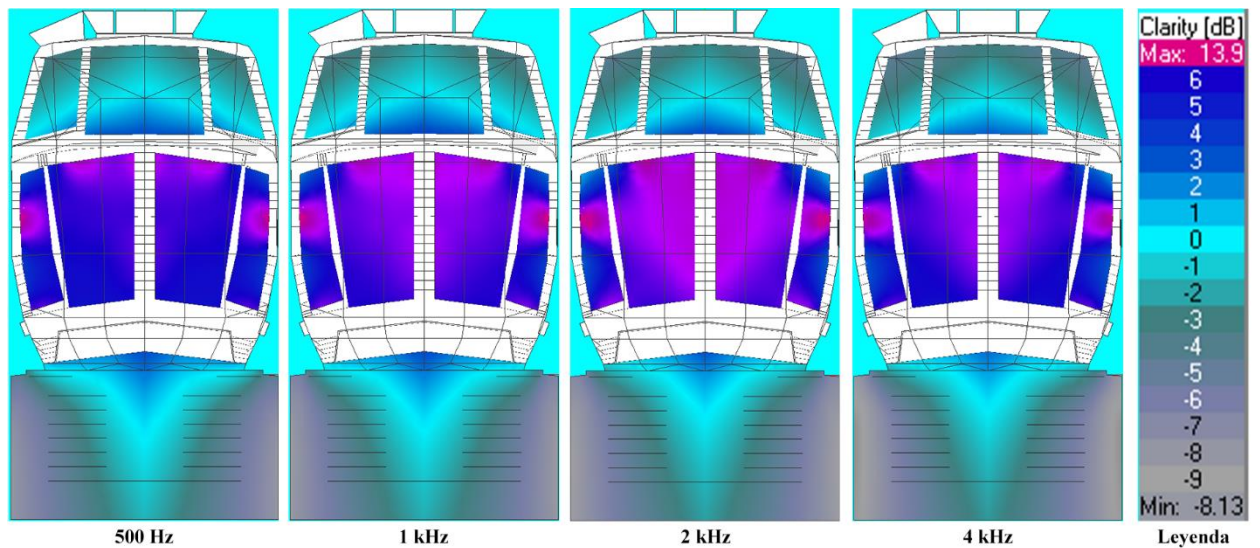


Fig. 7.11 Vista aérea del modelo con la representación de la cobertura del C80 en bandas de octava

8. MARCO REGULADOR E IMPACTO SOCIO-ECONÓMICO

8.1. Marco regulador

La normativa encargada de regular la medición de parámetros acústicos en recintos como teatros y salas de espectáculos es la ISO 3382-1:2009. Esta fue empleada como referencia en el proceso de caracterización del auditorio.

Dicha normativa define algunos de los parámetros acústicos más relevantes, regula la instrumentación a emplear y sus requisitos necesarios para obtener una medida fiable, y detalla los criterios a seguir en el plan de muestreo, además de ofrecer otras consideraciones y pautas a seguir en el proceso de medición. También expone distintos métodos para la obtención de parámetros como el tiempo de reverberación, que es el más importante para la caracterización de recintos. Finalmente, ofrece unas indicaciones para normalizar la presentación de los resultados.

También existen otras normativas relacionadas con espectáculos y la medición de altavoces. Sin embargo, dado que no están directamente relacionadas con el tema tratado en este proyecto, no se contemplan en este apartado, pues el responsable de la sonorización no tiene por qué emplearlas.

8.2. Presupuesto

El coste aproximado para la instalación del sistema de sonorización espacial propuesto en este trabajo es de 10.145 €. El presupuesto es el que se detalla a continuación:

Recurso	Descripción	Precio/Ud.	Uds.	Precio
Meyer Sound HM-1	Altavoz autoamplificado	325 €	12	3.900 €
DBX Driverack Venu360	Sistema digital de control de altavoces	819 €	2	1.638 €
Sistema de cableado	Carretes de cables XLR de larga longitud (coste aproximado)	116 €	12	1.392 €
Licencia Nuendo 11	DAW (<i>Digital Audio Workstation</i>). Contiene analizadores RTA/FFT y generadores de ruido.	610 €	1	610 €
Sistemas de suspensión	Sistemas de suspensión de altavoces (coste aproximado)	-	-	1.500 €
DPA 4006	Micrófono de medición para el proceso de ajuste <i>in situ</i> del sistema de altavoces (alquiler durante dos días)	185 €	1	185 €
Recursos humanos	Personal necesario para la instalación (coste aproximado)	230 € (3 días)	4	920 €
Coste total aproximado		10.145 €		

Tabla 8.1 Presupuesto aproximado de la instalación del sistema Ambisonics

8.3. Impacto socio-económico

La implementación de un sistema de reproducción de audio tridimensional en el Teatro Buero Vallejo de Guadalajara podría tener un impacto socio-económico muy positivo en la ciudad, ya que supondría un atractivo artístico que podría aumentar el tipo y número de eventos que componen la oferta cultural del teatro, lo que podría atraer a más espectadores por temporada, así como a más turistas a la ciudad, fomentando también los negocios locales.

Esta tecnología ofrece una herramienta muy potente a los ingenieros de sonido con vocación artística; abre las puertas a nuevas experiencias inmersivas, creativas e innovadoras, así como a eventos más llamativos susceptibles de atraer a más espectadores, lo que se traduciría en mayores ingresos, mayor renombre del teatro y en nuevos puestos de trabajo en el sector.

9. CONCLUSIÓN Y TRABAJOS FUTUROS

El sistema de reproducción basado en *Ambisonics* propuesto en este proyecto capacitaría al Teatro Buero Vallejo para albergar eventos que ofrezcan al espectador una experiencia inmersiva. Si bien este sistema ha sido evaluado de manera aproximada mediante simulaciones sobre una réplica virtual del auditorio, lo ideal sería instalarlo en la sala y estudiarlo *in situ*. Para cuantificar su calidad, podría emplearse como referencia el vector de localización r_V , tratando de maximizar su módulo empleando el modelo de suma coherente de ganancias (r_V) o el energético (r_E), en función de las observaciones.

Dado que el decodificador desarrollado en este proyecto es flexible en lo que respecta al número de altavoces y sus posiciones, así como a la posición del *sweet spot*, podría evaluarse configuraciones alternativas de altavoces que se considerasen más adecuadas. De igual manera, podría modificarse el decodificador para aumentar la escalabilidad del orden de *Ambisonics* más allá de $N = 3$, pudiendo así aumentar el número de altavoces y mejorar la cobertura de las áreas de audiencia, en especial las del anfiteatro, sin llegar a degradar la respuesta del sistema en alta frecuencia.

En caso de que la localización resultase confusa, podría adecuarse la acústica del recinto, instalando materiales absorbentes que lograsen reducir el tiempo de reverberación de la sala y aproximarla más al caso de un campo libre.

Por otro lado, también podría emplearse el decodificador propuesto en situaciones ajenas al Teatro Buero Vallejo, como podrían ser sistemas de *Ambisonics* en otros recintos, al aire libre o en cámara anecoica.

Tanto el modelo construido en EASE, como las Hojas de Cálculo de Google empleadas para su calibración, así como las auralizaciones y el código programado en Matlab para el decodificador de *Ambisonics*, están accesibles en el siguiente enlace: https://www.tsc.uc3m.es/~azpicueta/TFG_OCD.rar

REFERENCIAS

Ambisonics

- [1] Zotter, Franz and Frank, Matthias, *Ambisonics. A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement and Virtual Reality*, vol. 19. Cham: Springer International Publishing AG, 2019.
- [2] Martínez Téllez, Armando. *Polinomios de Legendre: aspectos matemáticos*. La Mecánica Cuántica. <https://bit.ly/2RWx9BJ> (acceso: 3 mayo 2021).
- [3] IFC 2018. *Aaron Heller: Ambisonics Decoder*. (21 julio 2018). Acceso: 5 febrero 2021. [Video en línea]. Disponible en: https://www.youtube.com/watch?v=E11g_UIDXPY
- [4] De la Prida, Daniel. (2019). *Formatos de reproducción de audio 3D*. [Presentación de PowerPoint]
- [5] De la Prida, Daniel. (2019). *Sistemas de reproducción de audio 3D: ejemplos de sistemas y su ecualización*. [Presentación de PowerPoint]
- [6] Simon Hutchinson. *Ambisonic Audio Fundamentals*. (13 julio 2020). Acceso: 5 febrero 2021. [Video en línea]. Disponible en: <https://www.youtube.com/watch?v=9m4VxFSYSEA>
- [7] Nuendo. *Ambisonics Setup Guide*. (20 agosto 2020). Acceso: 23 mayo 2021. [Video en línea]. Disponible en: <https://www.youtube.com/watch?v=-PRjLar7lyw>
- [8] J.J.G. Roy. *Técnicas de microfonía estéreo*. Sonidoyaudio. (acceso: 14 abril 2021).

EASE

- [9] Jason Romney's Classroom. *EASE Training Part 1*. (6 abril 2020). Acceso: 1 febrero 2021. [Video en línea]. Disponible en: <https://www.youtube.com/watch?v=gzLC2xs1ysU&t=2686s>
- [10] Jason Romney's Classroom. *EASE Training Part 2*. (10 abril 2020). Acceso: 1 febrero 2021. [Video en línea]. Disponible en: https://www.youtube.com/watch?v=l_oGyE6TxGA&t=1525s
- [11] Jason Romney's Classroom. *EASE Training Part 3*. (13 abril 2020). Acceso: 2 febrero 2021. [Video en línea]. Disponible en: <https://www.youtube.com/watch?v=BdREgzTXUU8&t=1171s>

Acústica

- [12] *Acústica. Medición de parámetros acústicos en recintos. Parte 1: Salas de espectáculos*, AEN/CTN 74 (AECOR) ISO 3382-1:2009.
- [13] *Acústica. Medición de parámetros acústicos en recintos. Parte 2: Tiempo de reverberación en recintos ordinarios*, AEN/CTN 74 (AECOR) ISO 3382-2:2008.
- [14] De los Reyes Sánchez, Pedro y Azpicueta Ruiz, Luis Antonio, *Estudio y Simulación Acústica del Auditorio de la Universidad Carlos III de Madrid*, Trabajo de fin de grado, Dpto. Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, Leganés, España, 2016.
[En línea]. Disponible en: <https://e-archivo.uc3m.es/handle/10016/27345>
- [15] Palafox Pascual, Lorena y Sancho Gil, Juan, *Caracterización y Optimización acústica del Teatro Auditorio Buero Vallejo*, Trabajo de fin de carrera, Dpto. Ingeniería Audiovisual y Comunicaciones, Universidad Politécnica de Madrid, España, 2013.
[En línea]. Disponible en: <http://oa.upm.es/21910/>
- [16] Azpicueta Ruiz, Luis A. *Sistemas Electroacústicos y Sonorización. Tema 3: Micrófonos*.
[Presentación de PowerPoint]
- [17] Azpicueta Ruiz, Luis A. *Sistemas Electroacústicos y Sonorización. Último tema: Sonorización*.
[Presentación de PowerPoint]
- [18] Jens Hee. (2003). *Impulse response measurements using MLS*.

Otros apartados

- [19] *History of Sound Recording*. Wikipedia. <https://bit.ly/34vGVgP> (acceso 15 mayo 2021).
- [20] *Surround Sound*. Wikipedia. <https://bit.ly/2SEznG2> (acceso 15 mayo 2021).

Figuras

Fig. 2.1. Esquema de un sistema óptimo de reproducción estéreo mediante altavoces

Fuente: <https://cuzgeek.com/best-budget-studio-monitors/>

Fig. 2.2. Configuración de un array Blumlein XY con dos cápsulas bidireccionales

Fuente: [1] (*Chapter 1*)

Fig. 2.3. Configuración microfónica ORTF

Fuente: <https://www.macprovideo.com/article/audio-hardware/review-m-audio-pulsar-ii>)

Fig. 2.4. Configuración microfónica AB

Fuente: [8]

Fig. 2.5. Configuración microfónica Deca Tree

Fuente: [8]

Fig. 2.6. Configuración de un array MS con una cápsula omnidireccional y otra bidireccional

Fuente: [1] (*Chapter 1*)

Fig. 2.7. Configuración de un array MS con dos cápsulas cardioides enfrentadas

Fuente: [1] (*Chapter 1*)

Fig. 2.8. Configuración de un array MS con una cápsula bidireccional y otra cardioide

Fuente: [8]

Fig. 2.9. HRTF para distintas direcciones de llegada del sonido

Fuente: <https://www.hispasonic.com/tutoriales/sonidos-espacio-hrtf-otros-indicios/43298>

Fig. 2.10. HRIR y HRTF para cada hemisferio, para una fuente en una localización determinada

Fuente: <https://slidetodoc.com/on-the-manifolds-of-spatial-hearing-vikas-c/>

Fig. 2.11. Neumann KU-100 Dummy-Head

Fuente: <https://es-es.neumann.com/ku-100>

Fig. 2.12. Roland CS-10EM

Fuente: <https://www.roland.com/es-es/products/cs-10em/>

Fig. 2.13. Crosstalk producido en la reproducción de señales binaurales en un sistema de altavoces Fuente:

Fuente: <https://www.ece.ucdavis.edu/cipic/spatial-sound/tutorial/hrtfsys/>

Fig. 2.14. Configuración óptima de un sistema 5.1

Fuente: [4]

Fig. 2.15. Configuración óptima de un sistema 7.1

Fuente: [4]

Fig. 2.16. Esquema del funcionamiento de un sistema VBAP de tres canales

Fuente (modificado): <https://bit.ly/3wV94dt>

Fig. 2.17. Configuración de un sistema VBAP

Fuente: http://impala.utopia.free.fr/pd/patches/externals_libs/vbap/vbap.html

Fig. 3.1. Armónicos esféricos en *Ambisonics*

Fuente (modificado): [1] (*Chapter 4*)

Fig. 3.2. Representación de los armónicos esféricos 2D en escala logarítmica, para $m = -3, \dots, 3$

Fuente: [1] (*Chapter 4*)

Fig. 3.3. Sistema FOA para la captación del campo sonoro en formato A

Fuente (modificado): [6]

Fig. 3.4. Configuración de un array FOA 2D para la captación de señales ambisónicas en formato B

Fuente: [1] (*Chapter 1*)

Fig. 3.5. Configuración de un array FOA 2D formado por tres micrófonos cardioides equiangulados

Fuente: [1] (*Chapter 1*)

Fig. 3.6. Configuración de un array FOA 2D formado por cuatro micrófonos cardioides equiangulados

Fuente: [1] (*Chapter 1*)

Fig. 3.7. Configuración de array FOA 3D para captación de señales ambisónicas en formato B

Fuente: [1] (*Chapter 1*)

Fig. 3.8. Configuración de array FOA 3D para captación de señales ambisónicas en formato A, y su conversión a formato B

Fuente: [1] (*Chapter 1*)

Fig. 3.9. Micrófono para la captación de *Ambisonics* de tercer orden en Formato A

Fuente: <https://endabates.wordpress.com/2017/06/19/comparing-ambisonic-microphones/>

Fig. 3.10. Diagrama de bloques de la codificación y decodificación en *Ambisonics*

Fuente (modificado): http://gyronymo.free.fr/audio3D/the_experimenter_corner.html

Fig. 3.11. Configuración de un sistema FLBR

Fuente: [1] (*Chapter 1*)

Fig. 3.12. Configuración de conversor FOA 2D/Biaural con rastreo de la orientación de la cabeza del usuario

Fuente: [1] (*Chapter 1*)

Fig. 3.13. Sistema de altavoces para la reproducción de HOA 3D

Fuente: <https://new.steinberg.net/es/nuendo/virtual-reality/>

Fig. 4.1. Foto exterior del acceso principal al edificio (izq.) y foto interior del hall (dcha.)

Fuente (modificado): <https://www.redescena.net/escenario/136/teatro-buero-vallejo/>

Fig. 4.2. Evento teatral con música en vivo en el auditorio

Fuente: https://movil.lacronica.net/articulo_lc.asp?idarticulo=63613

Fig. 4.3. Asociación “Libros y Más” en sala de exposiciones

Fuente: <https://www.libroymas.es/es/>

Fig. 4.13. Comparativa entre el teatro y el modelo en Vision, desde el anfiteatro [...]

Fuente: <https://www.jacksonlive.es/sala/teatro-buero-vallejo-de-guadalajara>

Fig. 5.3. Forma de onda de un sweep sine

Fuente: <https://theaudioprogrammer.com/signal-analysis-ii-linear-vs-logarithmic-sine-sweep/>

Fig. 5.4. Evolución temporal de la energía al interrumpir el ruido de banda ancha

Fuente: <https://www.nti-audio.com/es/aplicaciones/acustica-de-salas-y-edificios/medicion-del-tiempo-de-reverberacion-rt60>

Fig. 5.5. Integral de Schröder aplicada sobre una respuesta al impulso cuadrática

Fuente: <https://microflown.com>

Fig. 5.6. Representación gráfica de la evaluación del EDT sobre una curva de energía sobre tiempo

Fuente: <https://www.acousticbulletin.com/room-acoustic-descriptors-rt-c50-and-gain>

Fig. 5.10. Amplificador de potencia 2716C

Fuente: <https://bksv.com>

Fig. 7.1. Sistema de altavoces óptimo para *Ambisonics* de tercer orden

Fuente: https://www.researchgate.net/figure/The-loudspeaker-setup-contains-16-loudspeakers-Eight-loudspeakers-are-arranged-on-a_fig1_263007887