



2024

XIII CONGRESO/CONGRESO IBEROAMERICANO
DE ACÚSTICA

Santiago de Chile, 2–4 diciembre 2024

Implementación y evaluación de un diseño sonoro en formato 3D-Ambisonics

Aguilar, M.¹; Otondo, F.¹

¹ Laboratorio de Artes y Tecnología, Instituto de Acústica, Universidad Austral de Chile, Valdivia, Chile, martin.agui.del@gmail.com,
felipe.otondo@uach.cl.

Resumen

El presente estudio explora como las características acústicas del recinto afectan la reproducción de audio espacial respecto a la percepción de un oyente, con la finalidad de comprender de mejor manera este tipo de sistemas para un mejor desarrollo creativo y artístico. Para lograr esto, se generó y optimizó un algoritmo capaz de controlar un sistema de múltiples canales mediante la metodología Ambisonics, para luego comprobar su rendimiento a través de pruebas auditivas de localización de fuentes puntuales en dos recintos acústicamente distintos. Tanto las pruebas piloto como las definitivas arrojaron resultados que indicaban una clara dificultad para diferenciar las alturas para los oyentes, pero con un desempeño favorable en el plano horizontal, probablemente debido a las limitaciones propias de la escucha humana. Los resultados también mostraron una menor influencia de las reflexiones tardías en el rendimiento, resaltando la importancia de mantener controladas las reflexiones tempranas para una óptima reproducción multicanal.

Palabras clave: Audio espacial, Test auditivo, Instalación sonora, Ambisonics.

1. Introducción

La creciente demanda de experiencias auditivas inmersivas y realistas ha impulsado el avance de las tecnologías de audio espacial en las últimas décadas. Estas tecnologías permiten la recreación de entornos sonoros tridimensionales que sumergen al oyente en una experiencia acústica envolvente y auténtica. Entre las diversas opciones de audio disponibles, Ambisonics se ha destacado como un método versátil y eficaz para capturar, procesar y reproducir audio espacial [1].

El método Ambisonics, una técnica para sintetizar y reproducir campos sonoros, emplea una matriz de micrófonos para capturar el sonido desde todas las direcciones alrededor de una fuente acústica [2]. Esto permite una recreación precisa de la dirección y ubicación de las fuentes de sonido en el espacio tridimensional [3]. Su flexibilidad es evidente en su compatibilidad con varias configuraciones de altavoces, lo que lo hace adecuado para su aplicación en una amplia gama de entornos y sistemas de reproducción de audio.

Este estudio busca ofrecer información pertinente sobre cómo los factores ambientales y las características acústicas influyen en la reproducción de dichos sistemas multicanal. Esta información resultará valiosa en la creación de obras artísticas de audio espacial, simplificando el proceso creativo al mejorar nuestra comprensión de los diversos impactos de la acústica de la sala. En última instancia, esto permitirá una difusión más amplia de estas obras y llegar a un público más amplio.

2. Metodología

En esta sección se presentan los componentes principales para el experimento, explicando su desarrollo e implementación.

2.1 Diseño de algoritmo

Debido a que este estudio fue realizado en el marco de un proyecto más longevo se gestó en base a los avances realizados por anteriores miembros del mismo [4]. De esta manera, el algoritmo no fue diseñado en su totalidad, sino más bien se rediseñó para que fuera compatible con reproducción multicanal.

Para la confección del algoritmo se utilizó el software de programación modular Max/MSP [5] en conjunto con la librería de audio espacial SPAT 5 de IRCAM [6], esto debido a las posibilidad que entrega de utilizar la metodología Ambisonics para la codificación, decodificación y reproducción de audio espacial. A continuación, se presenta una simplificación del algoritmo final utilizado para las pruebas auditivas.

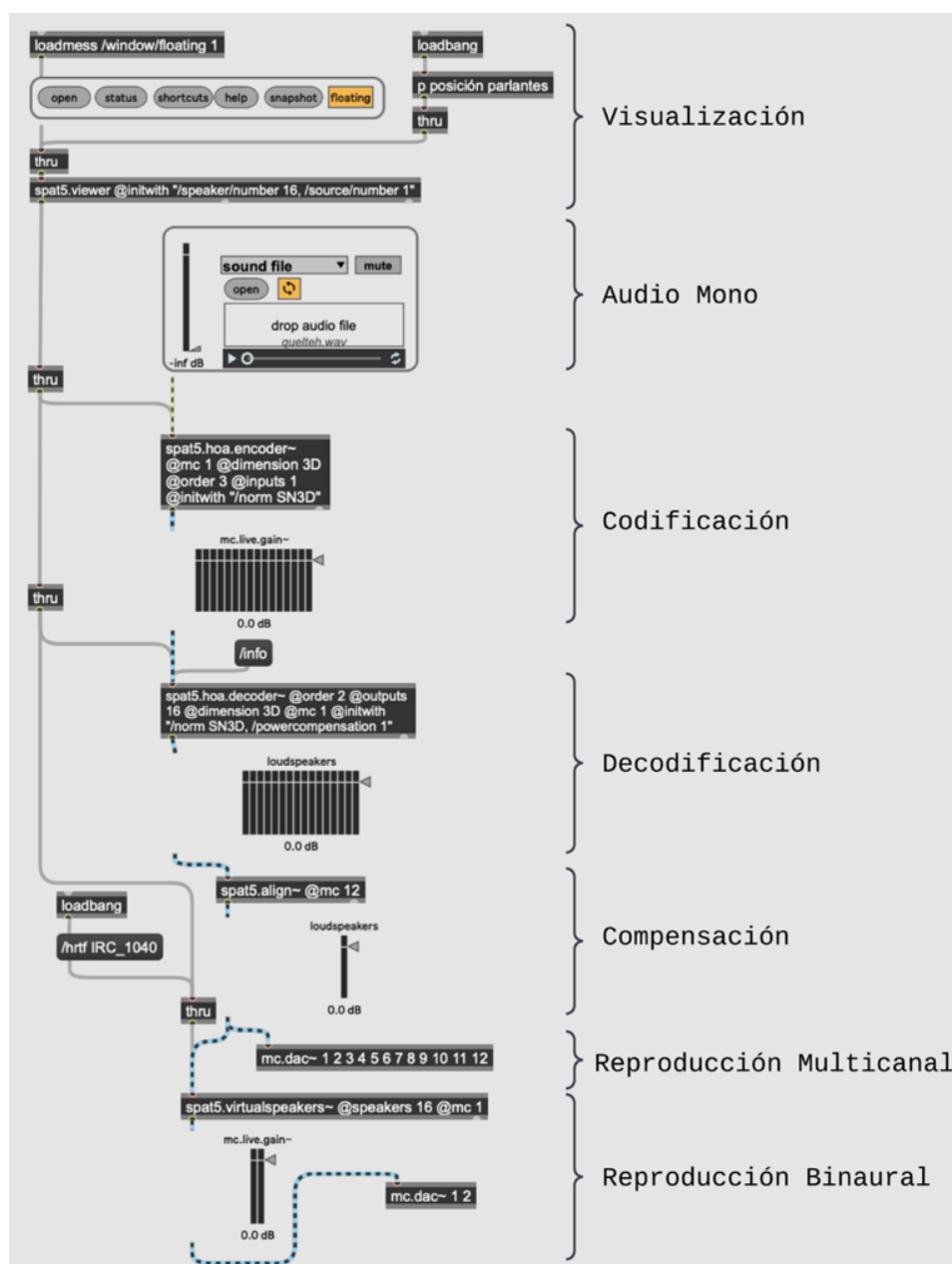


Figura 1: Patch del algoritmo generado en Max/MSP que muestra las etapas de espacialización. (Autor)

El algoritmo mostrado en la figura 1 presenta los procesos necesarios para la reproducción del sistema multicanal utilizando formatos de audio Monofónicos y formato A de Ambisonics, con la capacidad de reproducir en formato B Ambisonics en la cantidad de canales que se le indique según las necesidades del sistema. Esta es una de las grandes ventajas de la metodología [1].

2.2 Sistemas multicanal

Para nuestro estudio, empleamos dos configuraciones de sistemas multicanal para comparar su rendimiento en entornos acústicamente diversos. Por un lado, utilizamos el sistema multicanal del Laboratorio de Artes y Tecnología (LATE) de la Universidad Austral [7], que consta de 12 altavoces dispuestos de manera semiesférica alrededor del oyente. En contraste, en el Museo de Sitio Castillo de Niebla [8], utilizamos un sistema con 16 altavoces dispuestos en forma de ortoedro. Este último fue diseñado para replicar lo más fielmente posible la instalación de una exhibición, considerando todos los detalles asociados.



Figura 2: Sistema de 12 canales en recinto acústicamente controlado utilizado para las pruebas auditivas, Laboratorio de Artes y tecnología.

En esta configuración inicial, el sistema consta de dos niveles: uno a la altura del oyente y otro por encima de su cabeza, con 8 y 4 altavoces respectivamente, dispuestos en círculo a la misma distancia del oyente. Debido al espacio limitado en el laboratorio, el sistema funciona óptimamente para una sola persona adentro del área de escucha, por lo que las pruebas auditivas se realizaron de manera individual en este recinto. Es importante destacar que inicialmente se utilizaron 4 altavoces adicionales a nivel del suelo para las pruebas piloto, pero posteriormente fueron descartados.



Figura 3: Sistema de 16 canales utilizado para las pruebas auditivas, Museo de Sitio Castillo de Niebla.

En este caso, se realizaron modificaciones en la arquitectura del sistema para satisfacer mejor las necesidades de una instalación sonora, especialmente la ampliación del área de escucha, lo cual era crucial para obtener resultados precisos en el experimento. A diferencia del sistema anterior, en este caso fue necesario aplicar un ligero retraso y ajustar los niveles de los altavoces para compensar las diferencias de distancia entre cada altavoz y el oyente.

2.3 Pruebas auditivas

Con el objetivo de optimizar el algoritmo y evaluar el rendimiento del sistema multicanal, se llevaron a cabo diversas pruebas auditivas en diferentes etapas del proyecto. En primer lugar, se realizaron pruebas piloto para refinar las preguntas, el algoritmo y otros aspectos que fueron comentados por los participantes. Es importante destacar que aunque no todos los participantes eran expertos en el área, tenían al menos conocimientos educativos básicos al respecto. Este proceso proporcionó información valiosa, como la dificultad para identificar sonidos cercanos al suelo, mejoras necesarias para la accesibilidad de las pruebas auditivas y la selección del material sonoro adecuado.

Las pruebas auditivas finales se enfocaron en la localización de fuentes puntuales, ya que este aspecto es fundamental para una composición espacial y para evaluar las capacidades del sistema multicanal propuesto.

Dado el contexto del proyecto, se optó por utilizar cantos de aves como fuentes sonoras puntuales, ya que eran los sonidos más adecuados disponibles en las grabaciones de paisajes sonoros de los humedales cercanos a Valdivia. Se utilizó el canto de un Queltehue, con una duración aproximada de 5 segundos, como material sonoro para las pruebas. A continuación, se presenta el análisis espectral del material sonoro.

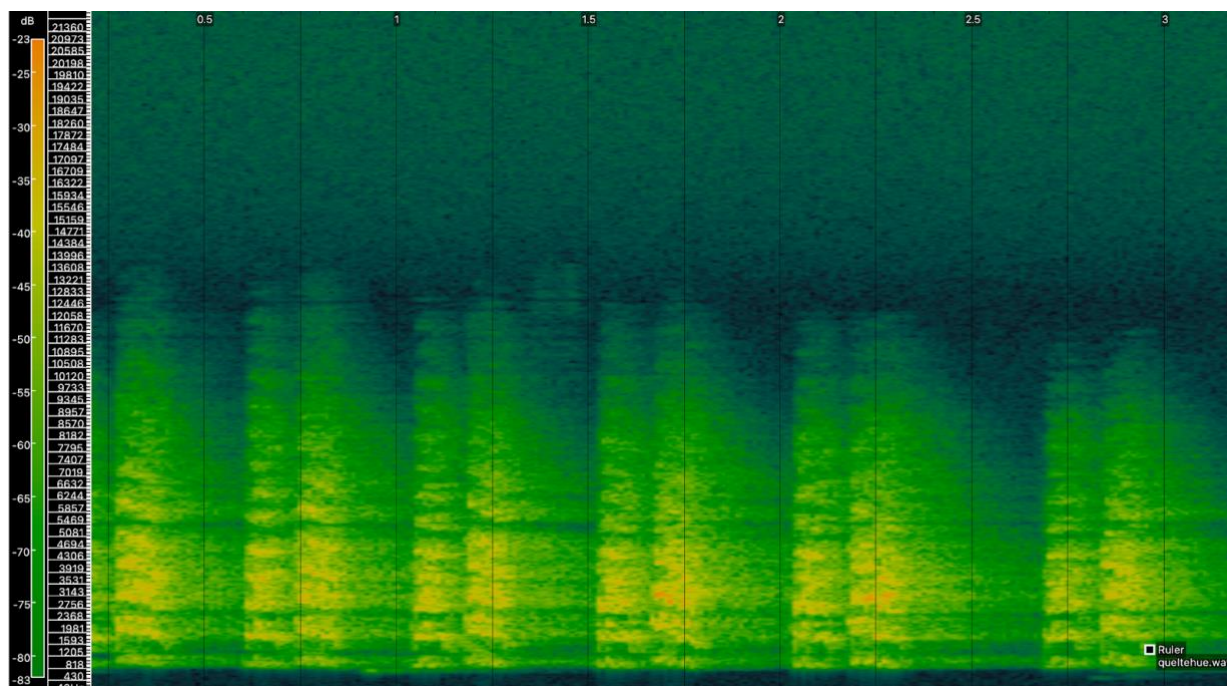


Figura 4: Componentes espectrales de la muestra del canto de ave seleccionada para las pruebas auditivas.

La composición de las pruebas consistió en dos partes, cada una compuesta por una serie de ejercicios de localización de fuentes alrededor del oyente. La principal diferencia entre ambos segmentos radicaba en la altura de las fuentes sonoras. Mientras que en la primera parte se buscaba que el oyente identificara fuentes en un plano horizontal justo a la altura de sus oídos, en la segunda parte se añadió la componente de la altura, requiriendo que los participantes indicaran tanto la dirección como la altura de cada sonido emitido.

A continuación, se pueden apreciar un esquema de ambos sistemas multicanal expuestos con anterioridad.

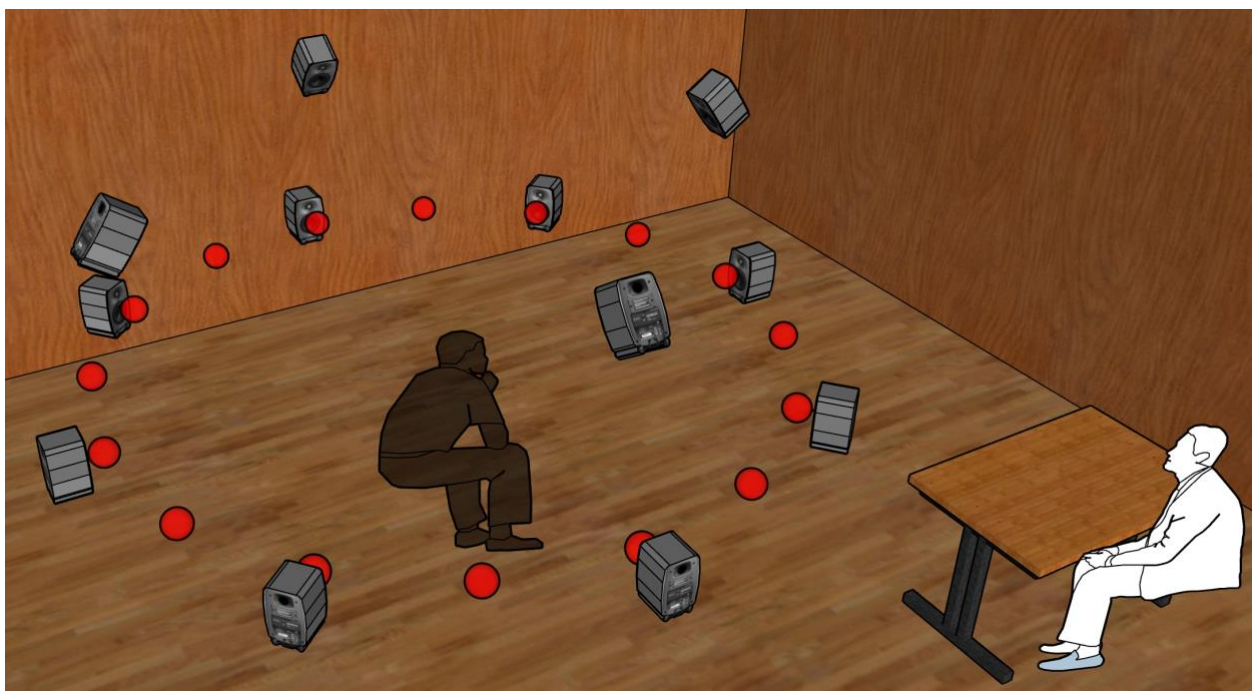


Figura 5: Diagrama de la configuración para la primera prueba auditiva en el Laboratorio de Artes y Tecnología.



Figura 6: Diagrama de la configuración para la segunda prueba auditiva en Museo de Sitio Catillo de Niebla.

En las figuras 5 y 6 además de exponer la arquitectura de los sistemas de reproducción espacial, se presentan las posibles posiciones de las distintas pruebas. Por un lado, la figura 5 muestra el sistema del laboratorio y las posiciones la prueba en el plano horizontal, mientras que la figura 6 presenta el sistema instalado en el museo junto con las posiciones de la prueba con alturas. Mencionar que las posiciones de ambas pruebas fueron las mismas en distancia respecto al punto central de ambos sistemas, con la finalidad de mantener algunos de los parámetros de las pruebas intacto entre los distintos recintos.

3. Resultados

La muestra para este experimento constó de 29 personas diferentes, de las cuales 23 participaron tanto en las pruebas realizadas en el laboratorio como en el museo. Así, se recolectaron 528 resultados por prueba en el laboratorio y 552 en el museo. Esto se traduce en 88 y 92 datos, respectivamente, por posición estudiada en cada prueba. Estos datos se procesaron y organizaron en diversos tipos de gráficos para visualizar los resultados de manera más efectiva.

La primera gráfica, conocida como de violín, proporciona información sobre la densidad de respuestas en un rango de valores de manera concisa, permitiendo la comparación entre dos situaciones diferentes, lo cual es ideal para representar esta comparación. Además, se superpuso una gráfica de cajas y bigotes, que ofrece información sobre la dispersión de los datos y la presencia de valores atípicos.

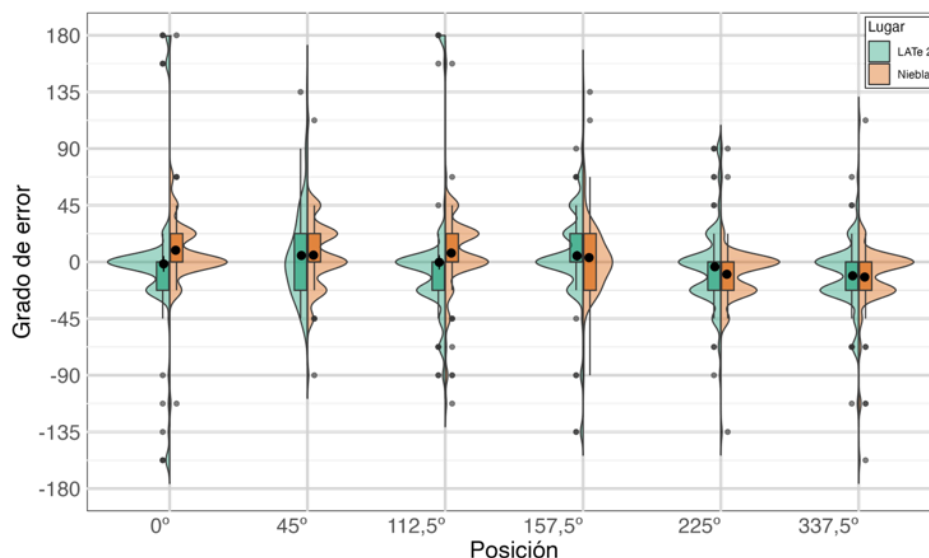


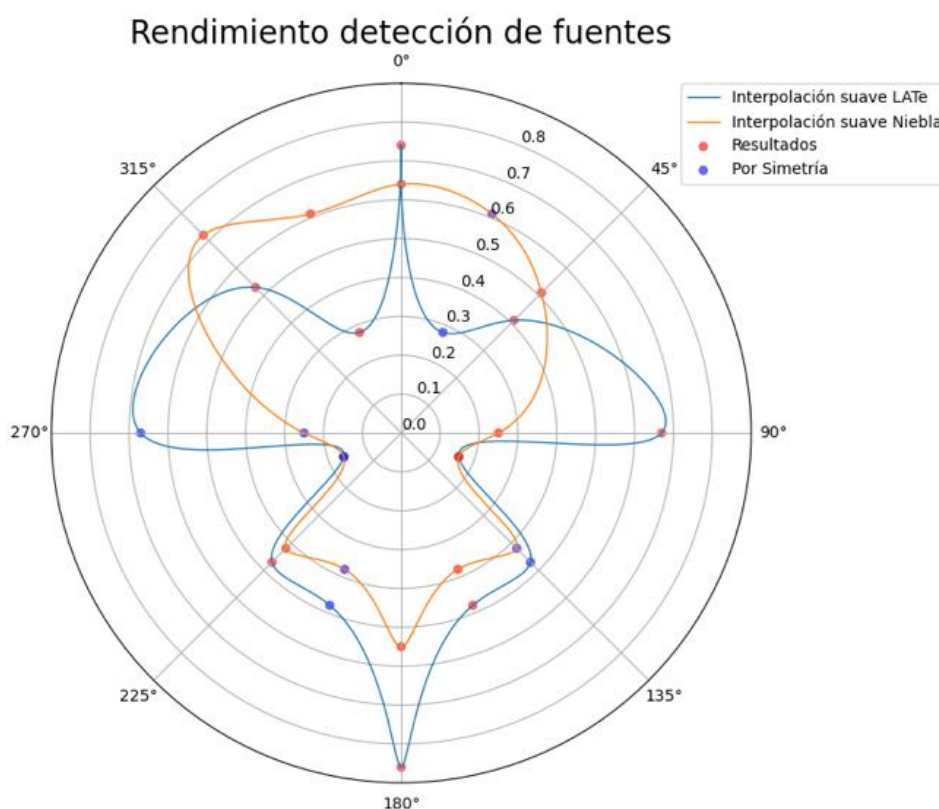
Figura 7: Gráficas de Violín con los resultados del experimento en el laboratorio y en el museo.

La Figura 7 muestra la comparación de la prueba en el eje horizontal sin diferencias de altura. En el eje horizontal se representa la ubicación de la fuente sonora puntual reproducida, donde la angulación se asigna en relación al punto frontal del oyente en sentido del reloj. Mientras tanto, en el eje vertical se muestran las respuestas de los participantes. Para esto, se parametrizaron las respuestas de cada posición de modo que el valor cero represente una respuesta correcta, y el resto del eje represente los niveles de errores, con valores positivos hacia la derecha de la posición correcta y negativos hacia la izquierda.

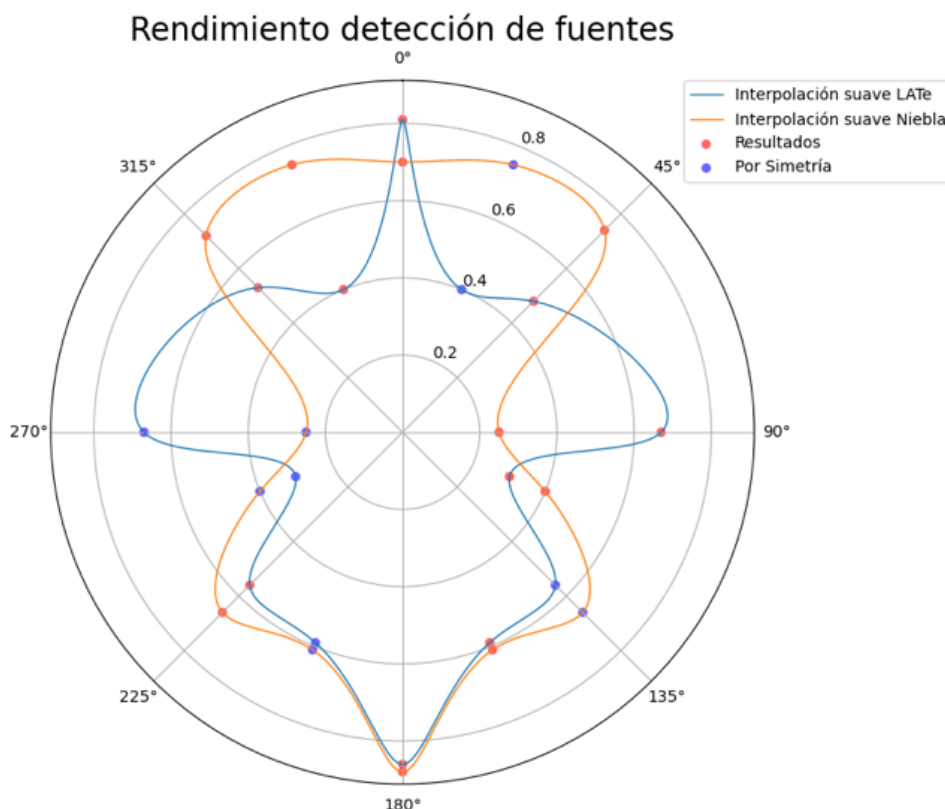
En general, las curvas se concentran alrededor de 0° , lo que indica un buen desempeño por parte de los participantes en lo que respecta al posicionamiento en un plano horizontal. En la mayoría de los casos, se observa una ligera tendencia hacia el lado derecho o izquierdo de la posición correcta, y solo en dos ocasiones se aprecia una tendencia homogénea en ambas direcciones. Este patrón se destaca aún más en la gráfica de cajas y bigotes, debido a la naturaleza de la prueba y sus opciones de respuesta.

En cuanto a los valores atípicos, la mayoría de las posiciones presentan varios de estos, con la excepción de la posición a 45° . Lo interesante es que esta tendencia se mantiene consistente en ambos lugares donde se llevó a cabo el experimento.

Por otro lado, se crearon patrones polares para representar visualmente el rendimiento del sistema. En un principio, se consideró procesar los datos en dos categorías, como es común: respuestas correctas o incorrectas. Sin embargo, después de un análisis detallado y una discusión sobre el tema, se decidió agregar una tercera categoría llamada "error leve" para brindar información más detallada y realista del experimento. Esta categorización tuvo como objetivo distinguir respuestas casi correctas de errores mayores, que la primera opción agrupa en un mismo grupo. Se propusieron dos gráficas polares para comprender mejor el rendimiento de los sistemas.



(a) Solo respuestas correctas.



(b) Respuestas correctas con un margen de $\pm 22.5^\circ$.

Figura 8: Gráficas Polares con los resultados del experimento en el laboratorio y en el museo.

En la Figura 8 (a) se muestra la comparación de respuestas totalmente acertadas entre ambos recintos. En la parte trasera del oyente, el espectro generado es bastante similar en ambos casos, con algunas diferencias de magnitud. Sin embargo, en la parte frontal hay diferencias significativas entre los recintos. Lo más notable es la mejora en el rendimiento a $\pm 45^\circ$.

El patrón polar de la Figura 8 (b) muestra no solo las respuestas correctas, sino también los errores dentro de $\pm 22.5^\circ$ de diferencia con respecto a la posición correcta. De esta manera, se observa una mejora considerable en comparación con la gráfica anterior, manteniendo en su mayoría el patrón original. Se puede apreciar que los lados traseros al oyente tienen un rendimiento más bajo en ambos sistemas, lo cual coincide con los resultados de trabajos anteriores.

4. Discusión

A continuación se expondrán las reflexiones más relevantes del proyecto, para esto se presentan ciertos apartados que engloban las características más relevantes del experimento comparando ambos recintos de manera de visualizar las diferencias principales de los sistemas y cuales podrían ser las causas de estas.

4.1 Influencia de la arquitectura del sistema de altavoces

Es evidente la diferencia entre las arquitecturas de ambos sistemas estudiados, esto debido a las condiciones físicas de los emplazamientos y la finalidad que cumpliría el sistema. En consecuencia, el sistema del museo es más amplio y con más altavoces que el del laboratorio. La diferencia de altavoces no afectó de gran manera debido a que el sistema fue calibrado para reproducir el mismo nivel en el área de escucha de ambos sistemas. Por otro lado, el posicionamiento si pudo afectar a la percepción de los participantes.

Las diferencias notorias de las gráficas de la figura 8, a los costados del sistema y en la parte frontal del mismo, pudieron surgir debido a la diferencia en la densidad de altavoces. Por un lado, en la parte delantera del sistema del museo hay mayor cantidad de altavoces lo que podría generar una imagen sonora más detallada y por ende mejores respuestas. En contraste, en los costados del sistema los altavoces estaban más espaciados y podría haber generado el impacto contrario.

Otra característica asociada al posicionamiento de los altavoces es el área de escucha, esta se ve afectada principalmente por el área resultante dentro del arreglo de altavoces, mientras mayor distancia existe entre el oyente y los altavoces mayor será el área de escucha [8]. Este fue uno de los puntos que se tuvieron en cuenta al proponer un sistema diferente para las pruebas en el recinto para exhibiciones, con un sistema con estos fines es más lógico conseguir un área de escucha lo más amplia posible.

4.2 Localización de fuentes

Este punto es la variable estudiada más relevante de la investigación, una buena localización de fuentes puntuales da paso a experimentar de mejor manera al crear cualquier tipo de ambiente sonoro, además de entregar más seguridad al momento de realizar escuchas críticas.

Respecto al desempeño en las pruebas en el plano horizontal, existe una tendencia hacia la respuesta correcta en todas las posiciones (figura 7). De igual modo, se pueden apreciar algunos valores atípicos en todas las posiciones, algunos bastante lejanos a la posición correcta, pero gran parte de estas respuestas tienen relación a la posición original, esto podría deberse a confusiones que se generan entre puntos simétricos del sistema debido a los conos de confusión [9].

Por otro lado, existe también el fenómeno de la confusión frontal-trasera, que sucede cuando la localización de la fuente según el oyente se ve distorsionada posicionando la fuente simétricamente detrás o delante de la original dependiendo del caso. En base a los resultados se puede afirmar que dicho fenómeno no solo se presenta en las posiciones de 0° y 180° , sino que también en $157,5^\circ$ o incluso en los $112,5^\circ$ junto a las posiciones simétricas del sistema.

Algunas de las falencias del sistema se podría decir que son sencillas de solucionar, teniendo en consideración la teoría relacionada al ITD e ILD que explica que la diferencia de tiempo e intensidad de los sonidos que llegan ambos oídos nos ayudan a posicionar los sonidos en el espacio [8], de esta manera el movimiento de la cabeza agrega más referencias al oyente. Esta es una acción que todas las personas realizamos instintivamente para localizar sonidos a nuestro alrededor.

La segunda prueba incluyó la altura como variable, pero los resultados fueron casi aleatorios, lo que dificulta un análisis detallado. Esto se debe en parte a que solo hay dos opciones de respuesta (altura de la cabeza y sobre la cabeza) y a la dificultad para identificar fuentes en diferentes alturas debido a la fisiología del aparato auditivo humano. Se consideró agregar una tercera altura (nivel del suelo), pero generó confusión en los participantes de la prueba piloto, esto podría deberse a la naturalidad de escuchar aves en alturas elevadas y a las sutiles diferencias en la percepción del sonido en distintas alturas.

4.3 Impacto de las características acústicas de las salas

Para este análisis se hizo énfasis en la comparación de ciertos parámetros acústicos debido a su importancia en la reproducción multicanal, estos son las reflexiones tempranas y tardías de los recintos. En este caso, los recintos tienen niveles de reflexión tempranas, pero en lo que respecta a las tardías hay una diferencia de 0,4 segundos de tiempo de reverberación.

Basado en los resultados del experimento y las recomendaciones de la experta en el área, Natasha Barrett [9], se concluye que las reflexiones tardías no tienen un gran impacto en el rendimiento del sistema utilizando la metodología Ambisonics. A pesar de las diferencias en el tiempo de reverberación entre los dos lugares evaluados, los resultados en terreno fueron mayormente positivos.

Por otro lado, las reflexiones tempranas fueron un punto crítico, la presencia de reflexiones dentro de los primeros 10 milisegundos puede confundir al oyente significativamente [11]. Esto puede atribuirse al hecho de que estas reflexiones proporcionan información "falsa", lo que puede desviar la percepción auditiva original esperada al reproducir fuentes sonoras en el sistema.

Con lo propuesto, se podría decir que los recintos pueden influir en gran medida en la reproducción de sistemas multicanal, pero si se logran controlar algunas características acústicas en particular dentro del lugar, se vería positivamente afectada la reproducción multicanal dentro del mismo.

5. Conclusiones

Basándonos en las pruebas realizadas en el laboratorio LATE y en el Museo de Sitio Castillo de Niebla, parece que el rendimiento de un sistema multicanal utilizando Ambisonics no se ve gravemente afectado al cambiar de un entorno controlado acústicamente a uno no controlado. Sin embargo, la localización precisa en las alturas no es consistente en ninguno de los entornos evaluados.

Los resultados de las pruebas de localización en ambos sitios mostraron un buen desempeño del sistema multicanal y del algoritmo. Sin embargo, se detectaron algunas deficiencias durante el análisis, como zonas de confusión y el fenómeno de confusión frontal-trasera. Estos aspectos podrían estar relacionadas con la percepción auditiva humana o la arquitectura de los sistemas y la posición del oyente.

Los resultados de las pruebas de localización a diferentes alturas no fueron concluyentes. Esta dificultad puede atribuirse a la percepción humana del sonido y a nuestra limitación para ubicar fuentes en altura, a menudo compensada por movimientos involuntarios.

En cuanto a las características acústicas de los recintos, es destacable la mínima influencia del tiempo de reverberación, a pesar de las claras diferencias entre ambos lugares. Esto subraya la importancia de controlar las reflexiones tempranas para la reproducción multicanal, aspecto en el que ambos sitios mostraron buen desempeño.

6. Agradecimientos

El presente trabajo se realizó gracias al Laboratorio de Artes y Tecnología [7] y los recursos entregados por el proyecto FONDECYT regular "Herramienta espacio-temporal para aplicaciones creativas y educativas de grabaciones de paisaje sonoro de humedales" (Folio 1220320), proyectos liderados por Felipe Otondo.

Mi sincero agradecimiento al profesor Felipe Otondo tanto por su apoyo académico como por el acceso a los recursos para este proyecto, su ayuda fue vital para completar este trabajo.

Referencias

1. Wiggins, B. (2021). Ambisonics comes of age. *Audio Engineering Society*, 69(9), 696–700.
2. Bates, E. B., Dooney, S. D., Gorzel, M. G., O'Dwyer, H. O., Ferguson, L. F., & Boland, F. M. B. (s. f.). Comparing Ambisonic Microphones. *Convention Of Audio Engineering Society*, 142.
3. Frank, M. & Zotter, F. (2017) Exploring the perceptual sweet area in Ambisonics. *AES (Audio Engineering Society)*. <http://www.aes.org/e-lib/browse.cfm?elib=18604>
4. Fernandez, J. (2023) Diseño y evaluación perceptual de un algoritmo de espacialización sonora en formato Ambisonics.
5. *What is Max? | Cycling '74*. (s. f.). <https://cycling74.com/products/max>
6. *Spat | IRCAM Forum*. (s. f.). <https://forum.ircam.fr/projects/detail/spat/>
7. *LATE | Laboratorio de Arte y Tecnología*. (s. f.). <https://soundlapse.net/latelab/>
8. Museo de Sitio Castillo de Niebla. <https://www.museodeniebla.gob.cl/>
9. Strong, J. T. Braasch, J. & Xiang, N. (2010) A Binaural Model with Head Motion That Resolves Front-Back Confusions for Analysis of Room Impulse Responses. *AES (Audio Engineering Society)*. <http://www.aes.org/e-lib/browse.cfm?elib=15701>

10. *Natasha Barrett's website*. (s. f.). <https://www.natashabarrett.org/>
11. ITU (2015). *Métodos para la evaluación subjetiva de pequeñas degradaciones en los sistemas de audio*.