

SIMULACIÓN Y MEDICIÓN DE UN SISTEMA DE CANCELACIÓN DE CROSSTALK POR REGULARIZACIÓN DE PARÁMETROS

**FRANCO JOFRÉ¹ GEORGINA LIZASO²
Y JORGE PETROSINO³**

¹ Universidad Nacional de Tres de Febrero, Caseros, Argentina.

francoj.jofre@gmail.com¹

² Universidad Nacional de Lanús - CONICET, Buenos Aires, Argentina

glizaso@unla.edu.ar²

³ Universidad Nacional de Tres de Febrero - Universidad Nacional de Lanús, Buenos Aires, Argentina

jpetrosino@untref.edu.ar³

Resumen

El audio binaural es una técnica utilizada para crear una percepción auditiva de un campo de sonido tridimensional. La idea central es reproducir los niveles de presión sonora en la entrada de los canales auditivos como si se tratara de una escena acústica real. Cuando la intención es recrear el campo sonoro con altavoces las señales de cada uno de ellos llegan a ambos oídos interfiriendo entre sí (generando crosstalk), lo que representa un problema ya que se busca que a cada oído llegue información diferenciada. Los sistemas de cancelación de crosstalk aéreo modifican la señal enviada a los parlantes a partir de la aplicación de filtros para que el audio de cada canal sea percibido sólo por el oído ipsilateral cancelando la contribución en el contralateral. En el presente trabajo se aborda e implementa analíticamente la técnica de regularización de matrices descrita por Choueiri (2018) y se comparan los resultados obtenidos por medio de simulaciones y mediciones realizadas en condiciones no anecoicas. Esta investigación fue realizada en el marco de una beca EVC-CIN.

Abstract

Binaural audio is a technique that creates the sensation of a three-dimensional sound field. The central idea is to reproduce the sound pressure levels at the input of the auditory channels so that they mimic those of a real acoustic scene. When the intention is to recreate the sound field using loudspeakers, each speaker's signal will reach both ears and interfere with each other (causing crosstalk), which is problematic because it is desirable that each ear only receives information from its channel. Crosstalk cancellation systems modify the signal sent to the loudspeakers by applying filters so that each ear signal is only perceived by the ipsilateral ear, canceling the contribution in the contralateral. In this work, the technique of regularization of matrices, described by Choueiri (2018), is implemented analytically. This paper compares the results of simulations and measurements made in non-anechoic conditions. This research was conducted under an EVC-CIN fellowship.

Palabras claves: Cancelación de crosstalk, Método pseudoespectral, Medición de crosstalk.

SIMULACIÓN Y MEDICIÓN DE UN SISTEMA DE CANCELACIÓN DE CROSSTALK POR REGULARIZACIÓN DE PARÁMETROS

1. Introducción

Cuando se utilizan altavoces para emitir eventos sonoros en un espacio físico ya sea en condiciones de campo libre o dentro de un recinto, se generan variaciones de presión de aire que crean un campo sonoro. El cómo este campo sonoro sea percibido por quienes estén en situación de escucha depende de múltiples factores, entre ellos las características acústicas del ambiente y las características físicas del cuerpo humano, como la forma de la cabeza, torso y pabellón auricular del oyente. Estas características físicas modifican el espectro de la señal sonora en función del ángulo de incidencia en relación con el eje medio de la cabeza y es lo que se conoce como función de transferencia de la cabeza (HRTF por Head Related Transfer Function). Si bien en cuanto a la reproducción esto podría resolverse con el uso de auriculares, en donde a cada oído llegue una señal distinta que no esté alterada por el HRTF ni por el ambiente, para ciertas aplicaciones esto no representa una solución. A la hora de reproducir campos sonoros que incluyan espacialización o características tridimensionales las alteraciones producidas por el ambiente y el HRTF se vuelven determinantes. Es por ello que para este caso las señales enviadas a cada oído deben incluir procesos que imiten las condiciones físicas del HRTF y el ambiente, o bien simulen indicios que promuevan la localización (señales binaurales). Sin embargo, para lograr una correcta binauralidad es necesario compensar la función de transferencia entre la señal emitida por el auricular y la detectada por el tímpano. Para lograrlo resulta necesario aplicar una ecualización precisa y diferente para cada individuo [1]. En la mayoría de los casos esta compensación no se realiza por la dificultad de obtener la función de transferencia específica para cada oyente y además existen aplicaciones donde el uso de auriculares puede resultar indeseable.

Las técnicas de audio transaural intentan reproducir señales binaurales por medio de altavoces en donde cada oído recibe una señal distinta. Como se dijo anteriormente, el campo sonoro percibido por los oídos del oyente contendrá las modificaciones por el HRTF por lo que las señales a emitir deben incluir procesos que cancelen la función de transferencia de la cabeza debido a la propagación al momento de la escucha, es decir que la señal binaural deberá incluir una HRTF que podrá ser la del oyente o una genérica.

Los desafíos técnicos principales que implica el audio transaural se presentan ya en el caso más básico que corresponde a la utilización de dos altavoces. En la Figura 1 la señal del canal derecho llega en forma más directa al oído derecho, pero parte de su emisión también alcanza al oído izquierdo luego de rodear la cabeza del oyente. Lo mismo ocurre con la señal emitida por el canal izquierdo. Esto implica que lo detectado por cada oído será una mezcla de lo emitido por cada altavoz. Los sistemas de audio transaural tienen que procesar la señal de información de ambos canales de modo que al emitirlos la contribución cruzada tienda a cancelarse para que cada oído reciba la información original de cada canal. Por este motivo también se los denomina sistemas de cancelación de crosstalk aéreo (sistemas XTC).

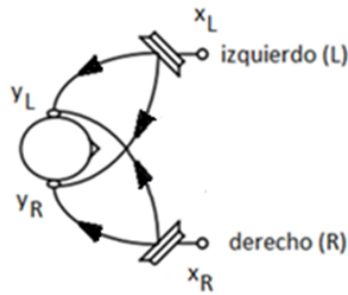


Figura 1: Recorrido de las señales emitidas por un par de altavoces.

La primera patente de un sistema XTC fue presentada en 1966 por Atal y Schroeder [2] con el fin de obtener grabaciones binaurales para su auralización por medio de altavoces. Donde el desafío central a resolver fue poder transmitir dos señales independientes desde un conjunto de altavoces a los dos oídos de un oyente mediante la aplicación de filtros analógicos que compensen el crosstalk producido cuando un oído recibe información del altavoz ubicado del lado opuesto. Sin embargo los sistemas XTC a priori funcionan para un único oyente en una posición y orientación predeterminadas (poseen un área de funcionamiento denominado sweet spot muy limitado) en donde un sistema razonablemente bien diseñado tolerará pequeñas variaciones en la posición de la cabeza, pero los grandes cambios hacen que el filtrado no entregue la señal binaural adecuada al oyente [3]. Por otra parte, pueden mencionarse tres problemas asociados a estos sistemas: lograr una separación de canales para poder conseguir el nivel XTC deseado, mantener dentro de niveles aceptables la exigencia de rango dinámico en los altavoces y la coloración tonal en relación al tamaño del sweet spot. Para lograr sistemas XTC perfectos deberían darse las condiciones de que los altavoces y toda la cadena electroacústica se encuentren apareados, un recinto anecoico y que la posición de la cabeza del oyente no varíe. A lo largo de los años se han propuesto variadas técnicas que intentan dar solución a las dificultades mencionadas.

En este trabajo se presenta la técnica de regulación de matrices de crosstalk propuesta por Choueiri [4], técnica que no requiere la compensación por HRTF ya que calcula los filtros a utilizar suponiendo dos puntos de escucha que representan la separación entre oídos y sin presencia de obstáculos (en este caso la representación de la cabeza).

El objetivo de esta investigación es implementar y evaluar la técnica mediante una medición acústica en un espacio no anecoico y comparar los resultados obtenidos con una simulación utilizando el toolbox k-Wave para Matlab que implementa la resolución de ecuaciones diferenciales de propagación acústica en base al método pseudo espectral del espacio k en el dominio del tiempo (PSTD).

2. Cancelación de crosstalk

En la Figura 2 se muestra el diagrama en bloques de un sistema de cancelación de crosstalk, donde se busca que al oído Y_L llegue la señal S_L y al Y_R llegue S_R . Lo que se emite por el altavoz X_L va a llegar tanto al oído Y_L como al Y_R . Por lo cual el altavoz X_R deberá emitir una señal que logre cancelar esta contribución. Esta nueva señal también llegará al oído Y_L , por lo cual, se debe diseñar un conjunto de filtros C que logre cancelar ambas contribuciones simultáneamente.

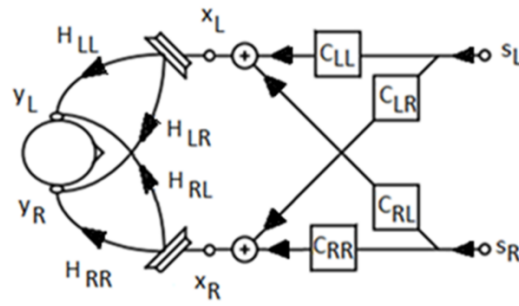


Figura 2: Diagrama en bloques de un sistema de cancelación de crosstalk.

Esto implica resolver un sistema de ecuaciones que puede representarse en forma matricial. Este sistema de ecuaciones tiene la desventaja de que frente a pequeñas variaciones en las condiciones iniciales del sistema, se producen grandes variaciones en el resultado. Pequeños movimientos de la cabeza generan una caída del nivel de cancelación y grandes coloraciones. Por otro lado, las condiciones no ideales del ambiente y los altavoces utilizados también afectan negativamente. Para reducir este problema se aplica la técnica conocida como regularización de matrices (1) mediante la cual los resultados teóricos de cancelación de crosstalk desmejoran pero el conjunto de filtros se vuelve más estable ante las variaciones del sistema.

$$C_{\beta} = [H^{TC} \cdot H + \beta \cdot I]^{-1} \cdot H^{TC} \cdot K \cdot e^{-j\omega\Delta t} \quad (1)$$

En la ecuación (1) C es la matriz que contiene los cuatro filtros de la Figura 2, H representa los caminos de propagación de sonido al aire, β es el parámetro de regularización (una constante mayor que cero), I es la matriz identidad y el superíndice TC significa transpuesto conjugado.

3. Mediciones acústicas

Para evaluar la técnica implementada se deben realizar mediciones del nivel de cancelación de crosstalk conseguido. Las mismas fueron realizadas en la sala del estudio de grabación Enrique Santos Discépolo de la Universidad Nacional de Lanús. El filtro inverso fue calculado para ser utilizado con un parlante bluetooth (JBL Flip 4) a una distancia aproximada de un brazo semiextendido del oyente. Esto se realizó en base a un trabajo de Choueiri [5], disponible en la página web de la Princeton University que contiene audios binaurales procesados como ejemplo para su descarga. Estos fueron calculados para ser utilizados con un dispositivo portátil similar al mencionado. Para la medición se utilizó un micrófono Earthworks M-30 y una placa de audio externa Komplete Audio 6.

El micrófono se colocó en la posición donde estaría el oído derecho de un potencial oyente con una distancia entre el plano del parlante y el del micrófono de 35 cm. Se consideró que el oído se encuentra a una distancia de 6 cm del plano de simetría de los parlantes. No se utilizaron obstáculos físicos que representen la cabeza del oyente (propagación libre), siguiendo los lineamientos descritos por Choueiri [4]. Posteriormente, se repitió la medición colocando el micrófono en la posición del oído izquierdo. En la Figura 3 se puede observar la disposición de los instrumentos en la medición.



Figura 3: Distribución de los instrumentos de medición.

Como señal para la medición se utilizó una secuencia de dos barridos exponenciales (sine sweep). El primero fue localizado totalmente a la derecha y el segundo a la izquierda. Al procesar los audios utilizando el filtro diseñado se obtiene una señal que pretende reproducir el barrido en primer lugar en el oído izquierdo, cancelándolo en el derecho y posteriormente en el oído derecho, cancelando la señal en el oído izquierdo. En la Figura 4 se puede ver un esquema de las señales enviadas al altavoz y el registro en cada oído.

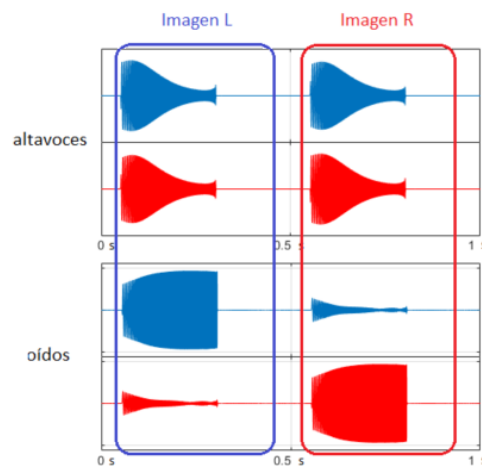


Figura 4: Esquema de señales. a) Señales procesadas para los altavoces; b) Señales registradas en los oídos.

Con esta metodología puede obtenerse el nivel de cancelación de crosstalk siguiendo dos métodos. La primera posibilidad es comparar los registros entre los oídos, es decir, comparar el nivel del oído derecho con el del izquierdo (las señales registradas en el oído como imagen L o R en la Figura 4). La segunda posibilidad es analizar los canales por separado, donde uno de los barridos debe estar presente y el otro tiende a cancelarse. Adicionalmente, se determinó el nivel de ruido de fondo del estudio para asegurar que sea lo suficientemente bajo para que no afecte las mediciones.

4. Simulaciones numéricas en el dominio del tiempo

Las simulaciones fueron realizadas utilizando el toolbox kWave [6] para Matlab que implementa el método pseudo espectral del espacio k de propagación de ondas acústicas en el dominio del tiempo (PSTD). Este método no posee dispersión numérica espacial. Si bien, presenta dispersión numérica temporal, esta puede ser corregida en forma exacta cuando la

velocidad de propagación es la misma en todos los puntos del espacio de simulación, como es el caso que corresponde a la situación en estudio [7].

Se utilizó un modelo 3D, aún cuando casi todos los elementos podrían situarse en un mismo plano, para que las características de propagación puedan compararse con las mediciones. En un modelo 2D, los elementos se comportan como fuentes cilíndricas y no como fuentes puntuales.

El espacio de simulación utilizado fue una cuadrícula tridimensional de 64 cm x 64 cm x 32 cm con resolución de 1 cm. La elección de dimensiones con forma 2^N es por motivos de eficiencia informática en el cómputo de la transformada rápida de Fourier (FFT). Todas las paredes de la superficie externa de la cuadrícula incluyen una capa de absorción para evitar reflexiones conocida como PML, por *Perfectly Matched Layer* [8] de un espesor de 10 cm por lado. Sin esta capa, el método pseudo espectral genera propagaciones que una vez que alcanzan un extremo continúan su propagación por el extremo opuesto. Esta capa modeliza un entorno anecoico (suprimiendo tanto la propagación mencionada como las reflexiones). Una capa de 10 puntos en un modelo 3D provoca una disminución de la propagación indeseada de -64 dB y de las reflexiones de -80 dB [6]. El agregado de esta capa de PML implica una disminución de 20 puntos en el espacio útil de cada una de las dimensiones, dejando un volumen efectivo de 44 cm por 44 cm por 12 cm. Suficiente para el tipo de estudio realizado.

Para la simulación se utilizó un valor de CFL (condición de Courant-Friedrich-Lewis) de 0,5. Este coeficiente, que siempre debe ser menor que uno para lograr estabilidad de las soluciones [9], se relaciona con la resolución espacial y la velocidad de propagación ($c = 340$ m/s) para determinar la frecuencia de muestreo adecuada para la simulación. En las simulaciones desarrolladas esta frecuencia fue de 68 kHz. Este valor permite obtener simulaciones exactas (sin dispersión numérica para el caso de c constante) para señales limitadas en banda en 17 kHz, que surge de multiplicar el CFL por la frecuencia de Nyquist. El método permite trabajar con varias fuentes con señales independientes. Se plantearon dos escenarios de simulación. En el primero se utilizan dos fuentes puntuales, emitiendo cada una de las señales de los altavoces. En el segundo, se utilizan dos fuentes extensas de 2 cm de radio para representar el tamaño de los altavoces, como se muestra en la Figura 3.

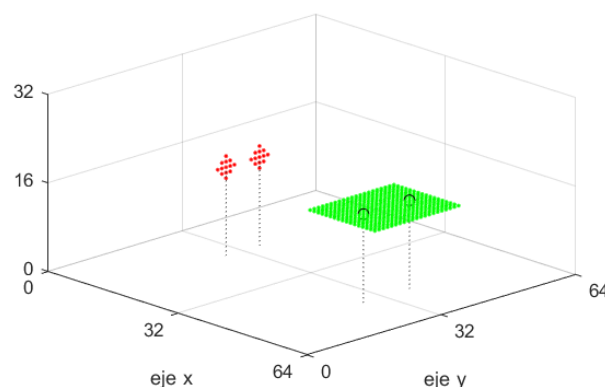


Figura 5: Fuentes (rojo) y sensores (verde) en la simulación con fuentes extensas. Los semicírculos negros representan la posición de los oídos.

Las fuentes de la simulación emiten las mismas señales que se envían a los altavoces en la medición, con la única salvedad de un remuestreo previo para convertir los 44.1 kHz, utilizados en el audio para la medición, en una versión con muestreo de 68 kHz. La simulación utilizada establece la señal de las fuentes en modo "dirichlet" [6].

El tiempo de la simulación utilizando el método pseudo espectral es prácticamente independiente de la cantidad de fuentes y sensores utilizados, por lo cual el uso de fuentes extensas y de un arreglo de sensores no implica un incremento significativo del tiempo de procesamiento. En general se teme que las simulaciones en el dominio del tiempo de resolución de las ecuaciones de propagación de onda acústica exijan tiempos de simulación y un nivel de recursos inalcanzables. Las simulaciones fueron realizadas en Matlab en una laptop con procesador Ryzen 5 5600H de 3.3 GHz con 32 GB de RAM. Cada simulación, correspondiente a la propagación de la secuencia de señales durante 1 s, tomó 13 minutos de tiempo de cómputo.

5. Resultados

5.1 Medición acústica

En primer lugar se muestran los resultados de la medición de ruido de fondo en la Figura 6. Dado que el parlante utilizado no es capaz de reproducir correctamente frecuencias bajas, se toma como frecuencia de corte inferior 200 Hz. Como se puede ver, el ruido de fondo es aceptable para este rango de frecuencias.

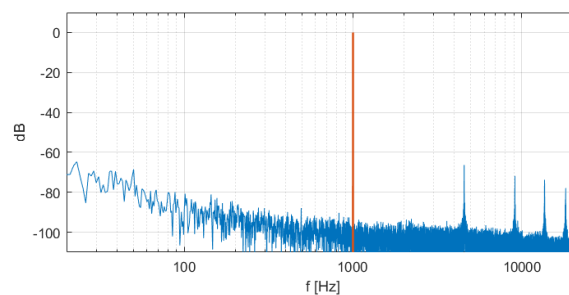


Figura 6: Medición de ruido de fondo.

Los picos presentes después de los 4000 Hz son provocados por la placa de audio utilizada, los de frecuencia inferior a 10000 Hz fueron eliminados utilizando filtros notch con un $Q = 20$ centrado en cada una de las frecuencias de interés, el resto fueron eliminados con un filtro pasabajos en 10 Khz y esto representa el límite superior de la medición.

Para determinar el crosstalk se registraron tres secuencias correspondientes a la señal imagen R y tres para la señal imagen L. Se calculó la potencia por tercios de octava de cada registro y se promediaron los resultados. En la Figura 7 se pueden observar los niveles de señal para el oído ipsilateral y para el contralateral.

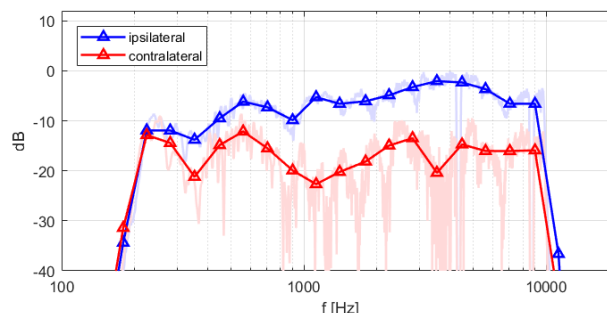


Figura 7: Promedio en tercios de octava de la potencia obtenida en el oído ipsilateral y contralateral.

Al restar estos valores se obtiene la cancelación de crosstalk para el sistema en la configuración utilizada para la medición. En la Figura 8 se muestra el valor de cancelación por frecuencia obtenido.

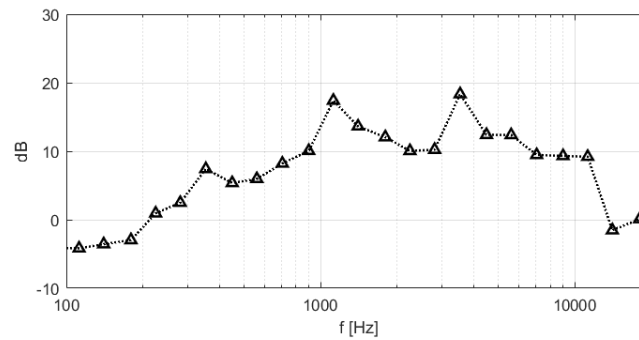


Figura 8: Resultados de crosstalk para la medición realizada.

5.2 Simulación numérica

Para la simulación fueron utilizados los mismos audios que en la medición con altavoces. La Figura 9 muestra el registro en la posición del oído L para una secuencia de barrido L y barrido R, utilizando fuentes puntuales.

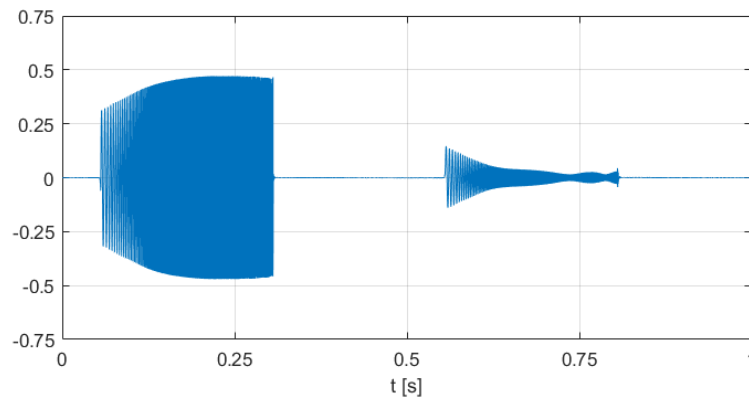


Figura 9: Secuencia temporal recibida en el sensor del oído izquierdo utilizando fuente puntual.

En la Figura 10 se observa el mismo caso pero para fuentes extensas.

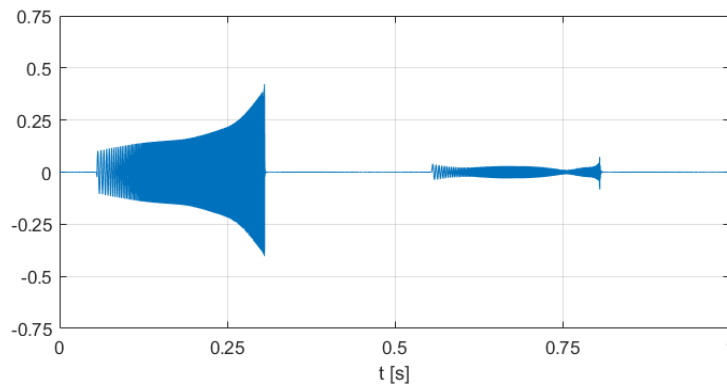


Figura 10: Secuencia temporal recibida en el sensor del oído izquierdo utilizando fuente extensa.

Considerando que los tiempos de procesamiento eran bajos, se decidió simular las dos secuencias, de 1 segundo de duración cada una, para mantener condiciones estrictamente

similares a las de la medición. Los registros obtenidos al completarse la propagación de ondas hasta los sensores ubicados en los oídos fueron divididos en dos partes de 0,5 s cada una. Se procedió a analizar cada parte con el mismo procedimiento utilizado en la medición. Se obtuvieron los registros de nivel ipsilateral (azul) y contralateral (rojo) correspondientes a los casos de fuentes puntuales (Figura 11) y de fuente extensa (Figura 12).

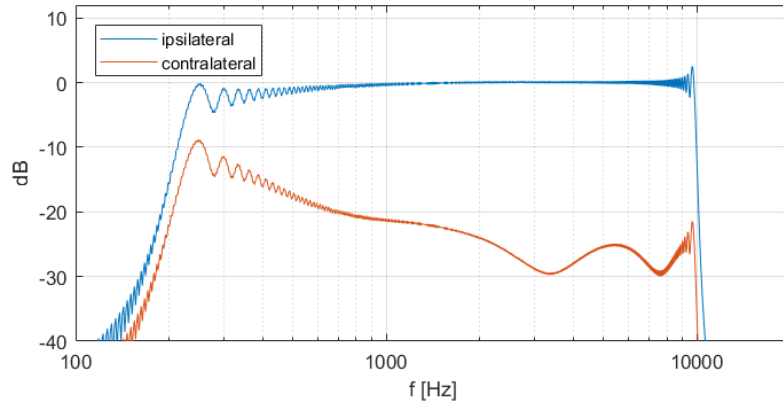


Figura 11: Niveles ipsilateral y contralateral para fuentes puntuales.

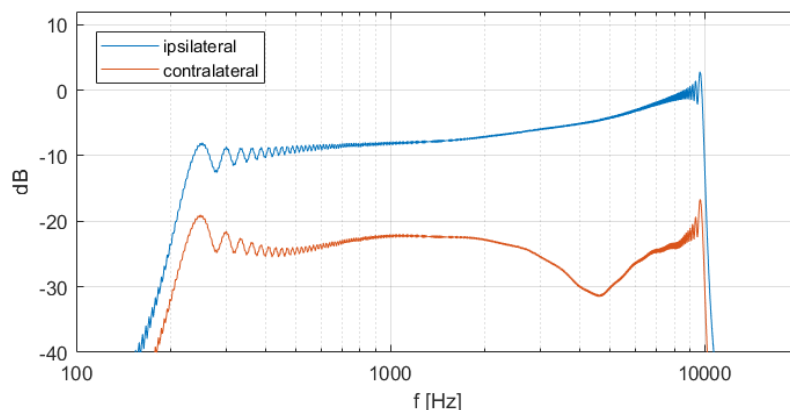


Figura 12: Niveles ipsilateral y contralateral para fuentes extensas.

6. Comparación y análisis de los resultados

La Figura 13, muestra una comparación de los resultados de la medición y de la simulación con fuentes extensas. Se muestra solamente la comparación del caso de fuentes extensas ya que es el que más relación guarda con la realidad de la medición. Además se agrega el resultado analítico que surge de las ecuaciones propuestas por Choueiri. El análisis debe limitarse al rango de frecuencias entre 200 Hz y 10 kHz, dado que son las frecuencias de corte utilizadas en el barrido exponencial. Como se puede observar en la Figura 13, existe un amplio rango de frecuencias para el cual las mediciones acústicas coinciden con la simulación realizada utilizando el método pseudoespectral para el oído ipsilateral. En el caso de la cancelación en el oído contralateral la simulación se aleja de la medición en mayor porcentaje. Esto es posiblemente debido a que, si bien se consideran fuentes no puntuales, hay otros aspectos de la situación real de medición que no son tenidos en cuenta en la

simulación, como pueden ser reflexiones, tiempo de reverberación, y la cadena electroacústica no está perfectamente apareada.

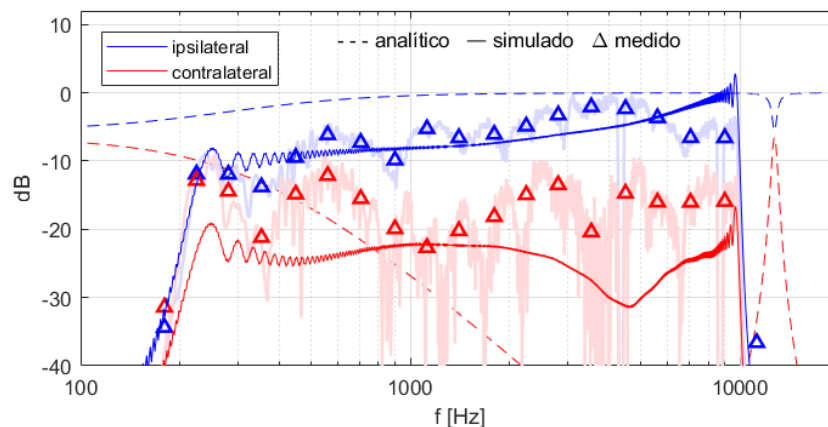


Figura 13: Comparación entre los resultados de la medición acústica y la simulación numérica.

En cuanto a la solución analítica, los resultados no son similares a lo obtenido mediante simulación y medición. Esto es probablemente ocasionado porque el cálculo de Choueiri no contempla la dimensión de los transductores utilizados, así como tampoco los efectos del espacio acústico no ideal donde fue realizada la medición.

El crosstalk obtenido en la medición no llega a los 20 dB que normalmente se mencionan como necesarios para lograr binauralidad [10]. Sin embargo, según afirma Choueiri, este nivel de cancelación no es alcanzable en la mayoría de las situaciones prácticas ni tampoco necesario. Las pruebas de escucha parciales realizadas parecen apoyar esta afirmación ya que la percepción de binauralidad es buena aún no llegando a los 20 dB de cancelación.

7. Conclusiones

Las simulaciones numéricas basadas en el modelo ondulatorio permiten obtener resultados más precisos si se las compara con las que utilizan el método geométrico. Sin embargo, estas suelen implicar muchos recursos computacionales y por lo tanto mucho tiempo de cómputo ya que deben resolver las ecuaciones de onda en el dominio temporal para cada punto espacial que se simule. Este estudio sugiere que el método pseudoespectral del espacio k en el dominio del tiempo (PSTD) es una herramienta eficiente para la resolución de ecuaciones como las planteadas. A diferencia del método de diferencias finitas, que implicaría tiempos de simulación similares a los de este estudio, el método aquí utilizado tiene la ventaja de no presentar dispersión numérica espacial y que la dispersión temporal puede corregirse de forma exacta cuando se considera una única velocidad de propagación. Al mismo tiempo, permite utilizar una gran cantidad de sensores y fuentes sin que esto incremente significativamente el tiempo de cómputo. A futuro se propone la posibilidad de analizar la información en todo un plano de sensores a la altura de los oídos del oyente con la intención de mapear las variaciones de presión en distintos puntos del espacio. A partir de este análisis es posible obtener diferentes sweet spot para distintos valores del parámetro de regularización. Esta herramienta puede ser útil para determinar valores óptimos de acuerdo a las distintas aplicaciones de la técnica.

Por otro lado, los resultados obtenidos en las mediciones acústicas indican que las ecuaciones planteadas por Choueiri no logran una buena predicción de la cancelación de

crosstalk que se obtiene en situaciones reales. Sin embargo, el resultado perceptivo de la aplicación de este método parece ser bueno. Esto debe investigarse en el futuro mediante pruebas estadísticamente significativas que permitan afirmar que sin llegar a los 20 dB propuestos en la bibliografía se logren alcanzar resultados satisfactorios.

Referencias

[1] Vorländer, Michael. Acoustic load on the ear caused by headphones. The Journal of the Acoustical Society of America, 107(4):2082–2088, 2000.

[2] Atal, BS; Hill, M e Schroeder, MR. Apparent sound source translator. us patent no. 3,236,949. US Patent and Trademark Office, 1966.

[3] Gardner, William Grant. *Transaural 3-D audio*. Perceptual Computing Section, Media Laboratory, Massachusetts Institute of Technology, 1995.

[4] Choueiri, Edgar. Binaural audio through loudspeakers. En *Immersive Sound*, págs. 124–179. Routledge, 2018.

[5] Choueiri, Edgar. Bacch demos for the jambox.
<http://www.princeton.edu/3D3A/Jambox+BACCHDemos.html>, Visitado por última vez el 17 de agosto de 2022.

[6] Treeby, Bradley; Cox, Ben e Jaros, Jiri. k-wave a matlab toolbox for the time domain simulation of acoustic wave fields user manual. Manual Version 1. 0. 1, 2012.

[7] Robertson, James L; Cox, Ben T e Treeby, Bradley E. Quantifying numerical errors in the simulation of transcranial ultrasound using pseudospectral methods. En 2014 IEEE International Ultrasonics Symposium, págs. 2000–2003. IEEE, 2014.

[8] Liu, Qing-Huo. The PSTD algorithm: A timedomain method combining the pseudospectral technique and perfectly matched layers. Tesis doctoral, Acoustical Society of America, 1997.

[9] Igel, Heiner. Computational seismology: a practical introduction. Oxford University Press, 2017.

[10] Parodi, Yesenia Lacouture e Rubak, Per. Analysis of design parameters for crosstalk cancellation filters applied to different loudspeaker configurations. Journal of the Audio Engineering Society, 59(5):304–320, 2011.