



SEÑALES Y SISTEMAS

(66.74 - 85.05)

TRABAJO PRÁCTICO ESPECIAL

Procesamiento de señales de un arreglo de
micrófonos

Primer Cuatrimestre 2021

Objetivo

Aplicar técnicas de procesamiento digital de señales estudiadas en la materia en audios obtenidos con un arreglo de micrófonos. Estimar los retardos temporales de la señal sonora en entre cada micrófono del arreglo. Estimar la posición de las fuentes sonoras.

Para la realización de este trabajo práctico el alumno deberá aplicar los conocimientos aprendidos en la materia acerca de:

- Relación entre señales continuas y discretas. Muestreo, sobremuestreo e interpolación.
- Transformada discreta de Fourier (DFT).
- Transformada de Fourier de corto tiempo (STFT).
- Diseño de filtros digitales.

Requisitos de aprobación

El proyecto especial es de carácter individual, tendrá una fecha límite de entrega y posterior defensa. Habrá un rango de días anterior a la fecha definitiva de entrega en el cual el alumno podrá hacer consultas a los docentes. Si el docente lo indica deberá realizar una pre-entrega del proyecto especial en versión electrónica, PDF solamente, en <http://campus.fi.uba.ar/>. Durante ese rango de fechas el docente puede aconsejar al alumno la revisión de ciertos puntos en el proyecto. Luego del cierre del período de pre-entrega, se habilitará un período de entrega definitiva, donde el alumno debe depositar su versión en formato PDF del informe y los algoritmos correspondientes. Luego del vencimiento del período de entrega definitivo, no se admitirán más entregas y el alumno que no cumpla este requisito quedará libre. Luego el docente de cada curso evaluará el mismo en tiempo y forma.

La modalidad de la evaluación se realizará según el docente lo crea conveniente, de modo de asegurar el conocimiento del tema desarrollado y la realización individual del trabajo por parte del alumno. El trabajo sólo podrá ser presentado y evaluado en el curso en el cual el alumno se halla inscripto. La evaluación final puede incluir preguntas sobre:

- Ítems particulares sobre los ejercicios de esta guía y su implementación.
- Conceptos teóricos necesarios para realizar los ejercicios.
- Puede requerirse también al alumno que implemente alguno de los ejercicios similares en la computadora en el momento de la evaluación virtual.

Nota del trabajo práctico especial

El proyecto especial se aprobará con 60 puntos sobre 100. La nota final de cursada se determina ponderando la nota del parcial N_{par} con la nota del proyecto especial N_{pe} del siguiente modo:

$$N_{cur} = 0,6N_{par} + 0,4N_{pe}$$

Como tanto la nota del parcial N_{par} como la nota del proyecto especial N_{pe} deben ser mayores a 60, para la aprobación de la cursada la nota de cursada N_{cur} debe ser mayor que 60 puntos, y se recuerda que la aprobación del proyecto especial también es obligatoria para la aprobación de la cursada.

Material

Los archivos necesarios para realizar el trabajo práctico se encuentran en:

<https://drive.google.com/drive/folders/1A8NmCySLFsTjzyBluyNZ5HuV3BoviZQu?usp=sharing>

Desarrollo

La localización de una fuente sonora es de interés en muchas aplicaciones, por ejemplo: un robot puede descubrir la posición de la persona que le esté hablando; detectar fuentes de ruido para su posterior filtrado; sistemas de navegación; sonares; vigilancia; entre otros. En la Figura 1 se pueden observar algunos ejemplos de aplicación.

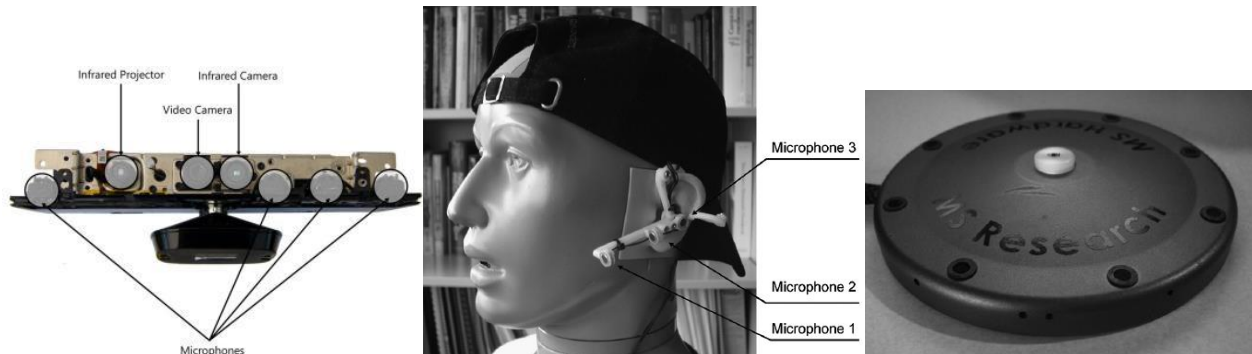


Figura 1: Ejemplos de arreglos de micrófonos: a) Sensor Kinetics para la consola Xbox; b) 3 micrófonos para un headset; y c) circular de 8 micrófonos. Extraído de [1].

Una alternativa para estimar la posición de una fuente sonora es utilizar un arreglo de micrófonos. Si consideramos una fuente de sonido puntual, la difusión del sonido se realiza en forma esférica y a distancia lo suficientemente lejos de la fuente el frente de onda se puede considerar plano (Figura 2). La fuente puede encontrarse por ejemplo en una habitación y el sonido será reflejado y absorbido por las paredes y techo, produciéndose efectos de reverberación. Por otro lado, se puede deducir en forma intuitiva que si aumenta la distancia entre el micrófono y la fuente de sonido la amplitud de la presión sonora disminuye.

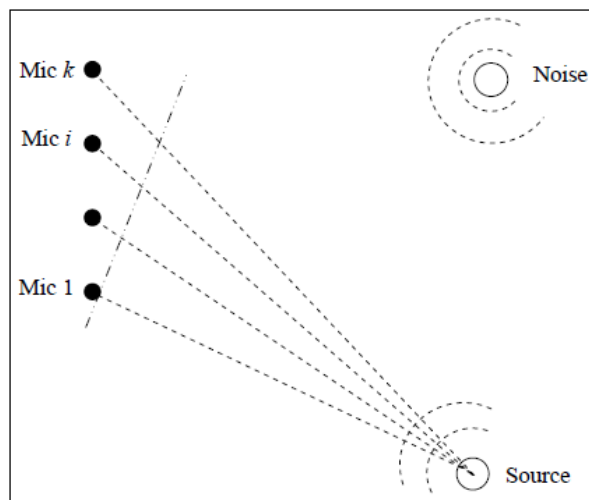


Figura 2: Esquema donde se puede observar dos fuentes de sonido y un arreglo lineal de micrófonos.

En un arreglo de micrófonos las señales registradas por cada micrófono serán iguales salvo por algunas pequeñas diferencias, que nos pueden ayudar a estimar la posición de la fuente sonora. La onda de sonido de la fuente sonora viaja a una velocidad determinada que podemos considerar constante, y dependiendo de la ubicación de los micrófonos el sonido registrado por cada uno de ellos tendrá un desfase temporal (TDOA, del inglés *Time Difference of Arrival*) dado que los frentes de onda llegan a tiempos diferentes a cada micrófono del arreglo. Si la fuente de sonido se encuentra en movimiento, la estimación de los TDOA se debe realizar en un esquema de ventanas temporales de análisis.

En lo siguiente realizaremos un análisis geométrico para describir como llega el frente de onda sonora a los micrófonos del arreglo. Se puede suponer que la onda de sonora se desplaza en el aire a una velocidad

constante, de aproximadamente $c = 340$ m/s. También se puede considerar que es un frente de onda plano, lo cual se cumple considerando que la fuente está lo suficientemente lejos de los micrófonos.

En la Figura 3 se observa un esquema de dos micrófonos, M_i y M_d , separados una distancia d , a los cuales llega un frente de onda plano con un ángulo de incidencia θ , registrando las señales $s_i(t)$ y $s_d(t)$ respectivamente. El micrófono más lejano M_i a la fuente tiene un retardo τ en la llegada del frente de onda con respecto al micrófono más cercano M_d .

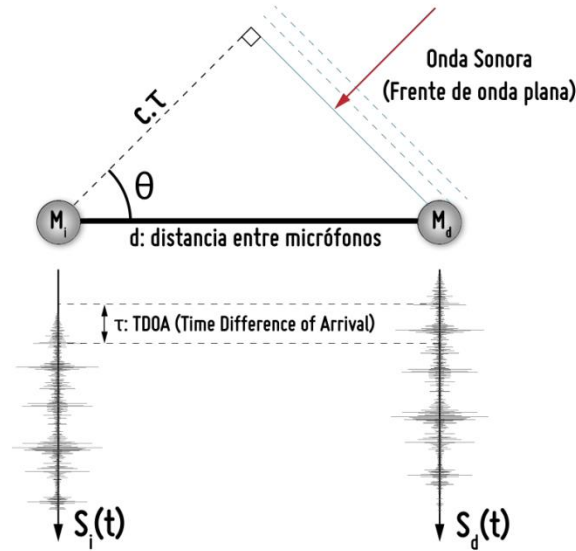


Figura 3: Esquema del arribo de un frente de onda plano a dos micrófonos.

Si la fuente emite una señal $s(t)$ y demora t_0 segundos en llegar al micrófono más cercano M_d , podemos escribir:

$$s_d(t) = s(t - t_0) + r_d(t) \quad (\text{Ec. 1})$$

$$s_i(t) = s(t - t_0 - \tau) + r_i(t) \quad (\text{Ec. 2})$$

$$s_i(t) = s_d(t - \tau) + r_i(t) - r_d(t) \quad (\text{Ec. 3})$$

Donde las señales $r(t)$ son fuentes de ruido aditivas. Aquí hemos omitido los efectos de la habitación.

Ejercicio 1

En el archivo *audios1.mat* se guardaron las señales registradas por un arreglo lineal de 5 micrófonos separados 5 cm ubicados en una habitación de 3x4 metros, según el esquema de la Figura 4. Se utilizó una frecuencia de muestreo de 48 kHz.

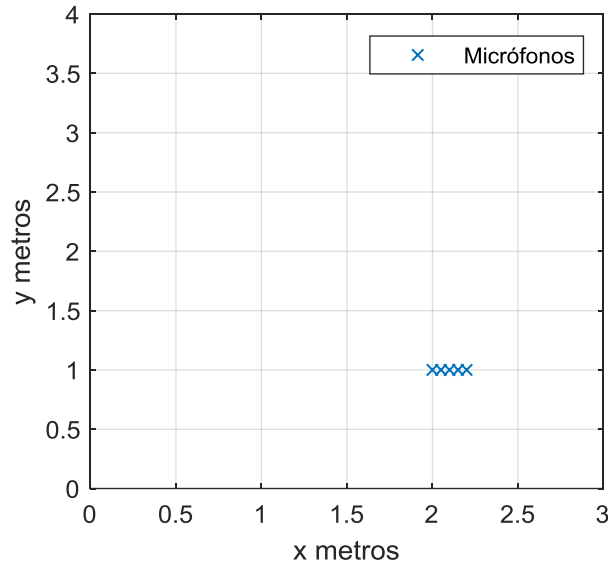


Figura 4: Esquema de la habitación donde se realizó la grabación de las señales guardadas en el archivo audios1.wav.

Graficar en forma superpuesta las señales registradas por todos los micrófonos. Etiquetar ejes y sumar leyenda. Realizar una segunda figura de un segmento donde se puedan identificar los desplazamientos de las señales. Analizar los espectros de las señales mediante transformada de Fourier y espectrogramas.

Ejercicio 2

Utilizando la figura del ejercicio anterior, estimar los retardos de llegada a todos los micrófonos. Suponer una posible posición de la fuente de sonido con respecto a los micrófonos a partir de los valores estimados de los retardos.

Una forma de estimar en forma automática el retardo de llegada de una señal sonora a un micrófono respecto de otro es utilizando la función de correlación cruzada, que permite medir el grado de similitud entre dos señales. En tiempo discreto la correlación cruzada entre dos señales $x[n]$ e $y[n]$ se define como [6][7]:

$$r_{xy}[l] = \sum_{n=-\infty}^{\infty} x[n]y[n-l] \quad (\text{Ec. 4})$$

El índice l es un parámetro de desplazamiento temporal, que determina el tiempo relativo entre ambas señales. El cálculo de esta secuencia resulta impreciso en ambientes sumidos en ruido. Como alternativa, se puede recurrir a una propiedad que vincula a una secuencia de correlación cruzada entre dos señales con sus respectivas transformadas de Fourier en Tiempo Discreto:

$$\mathcal{F}(r_{xy}[l]) = X(\Omega)\overline{Y(\Omega)} \quad (\text{Ec. 5})$$

Donde $\mathcal{F}(\cdot)$ representa la transformada de Fourier, $X(\Omega)$ es la transformada de Fourier de la señal $x[n]$ e $\overline{Y(\Omega)}$ es el complejo conjugado de la transformada de Fourier de la señal $y[n]$.

Una alternativa es una técnica basada en correlación de señales en el dominio frecuencial llamada GCC-PHAT (del inglés, *Generalized Cross Correlation with Phase Transform*) [3], que presenta mayor robustez en ambientes ruidosos. Esta se define de la siguiente forma:

$$G_{PH}[k] = \frac{X[k]\overline{Y[k]}}{|X[k]||Y[k]|} \quad (\text{Ec. 6})$$

Donde $X[k]$ representa la Transformada de Fourier Discreta (DFT) de la señal $x[n]$ y $\overline{Y[k]}$ es el complejo conjugado de la DFT de la señal $y[n]$. El denominador de la ecuación actúa como normalizador a la vez que se preserva la información de fase. La Ec. 6 se puede reescribir considerando la Ec. 3, como:

$$G_{PH}[k] = e^{j\frac{2\pi}{N}k\tau} \quad (\text{Ec. 7})$$

Calculando la transformada inversa obtenemos:

$$g_{PH}[l] = \delta(l - \tau) \quad (\text{Ec. 8})$$

En la práctica no obtenemos un $\delta(l - \tau)$ que nos marcaría la posición del retardo, pero sí una función que presenta un pico máximo en \mathcal{T} . En la Figura 5 se muestra un ejemplo de aplicación.

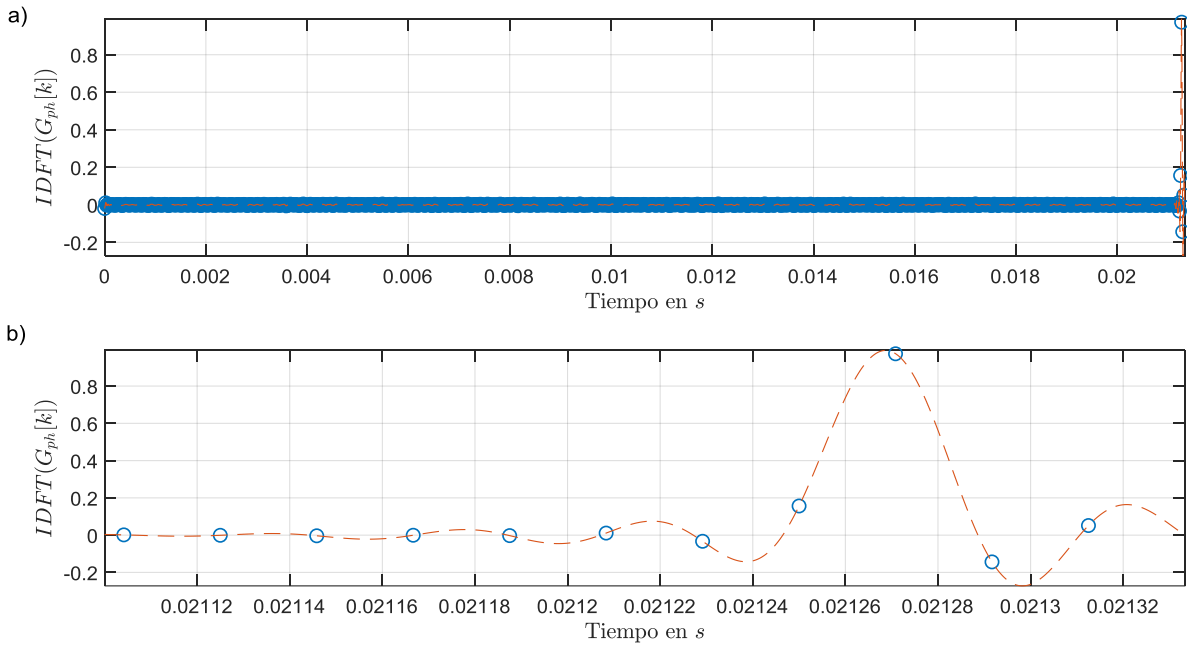


Figura 5: a) Transformada de Fourier Discreta Inversa de la función $G_{PH}[k]$; y b) un zoom en la zona de su máximo.

Ejercicio 3

Calcule las correlaciones entre cada par de señales de micrófonos consecutivos en el dominio temporal (Ec. 4) y mediante la IDFT de GCC-PHAT (Ec. 8). Obtenga los retardos a partir de ambos métodos.

El algoritmo de estimación de los retardos, estimados por ventanas y par de micrófonos, se puede escribir cómo:

$w(n) \leftarrow$ Ventana de N muestras

$\Delta n \leftarrow$ Nro. de muestras del desplazamiento de la ventana

$x(n) \leftarrow$ señal de audio registrada por el micrófono i – ésimo

$y(n) \leftarrow$ señal de audio registrada por el micrófono $(i + 1)$ – ésimo

Para cada par de señales (x, y) , hacer:

$$n_0 \leftarrow \frac{N}{2}$$

Mientras $n_0 + \frac{N}{2} < \text{tamaño de } x$, hacer

$$DFT_x \leftarrow DFT\{x(n) w(n - n_0)\}$$

$$DFT_y \leftarrow DFT\{y(n) w(n - n_0)\}$$

$$G_{ph} \leftarrow \frac{DFT_x DFT_y^*}{|DFT_x| |DFT_y|}$$

$$m \leftarrow \max_k (IDFT\{G_{ph}\}(k))$$

$$\text{Si } m > \frac{N}{2} \Rightarrow m \leftarrow m - N$$

$$\tau_{xy}(n_0) \leftarrow \frac{m}{fs}$$

$$n_0 \leftarrow n_0 + \Delta n$$

Si \mathcal{T} es el retardo entre dos micrófonos separados una distancia d , el ángulo de llegada del frente de onda plano (Figura 3) se puede estimar como:

$$\cos(\theta) = \frac{c \cdot \tau}{d} \quad (\text{Ec. 9})$$

$$\theta = \cos^{-1} \left(\frac{c \cdot \tau}{d} \right) \quad (\text{Ec. 10})$$

Donde c es la velocidad del sonido. En un arreglo de más de dos micrófonos, podemos estimar la posición de la fuente encontrando el punto de intersección de las rectas que pasan por cada uno de los micrófonos y de pendientes:

$$m_i = \tan(\theta) = \tan \left(\cos^{-1} \left(\frac{c \cdot \tau_i}{d} \right) \right) \quad (\text{Ec. 11})$$

En la Figura 6 se muestra un esquema de la estimación de la posición de la fuente a partir de los retardos, utilizando la Ec 11.

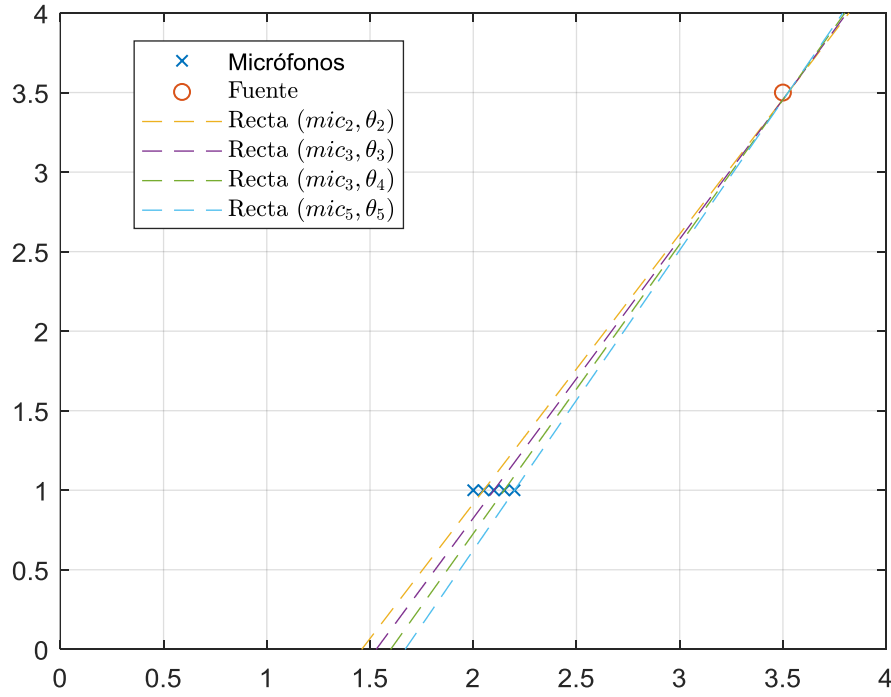


Figura 6: Esquema del trazado de las rectas que pasan por los micrófonos, con la pendiente estimada con la Ec. 6. La intersección determina la posición de la fuente.

Ejercicio 4

Estimar los retardos de llegada y graficar su evolución temporal utilizando el algoritmo propuesto. Realizar los histogramas de los retardos calculados para cada micrófono. Obtener un valor representativo de retardo para cada señal. Estimar y graficar la posición de la fuente.

Las señales que hemos utilizado están libres de ruido. Si suponemos que la señal $s[n]$ responde la Ec. 1, entonces el nivel de ruido puede medirse mediante la Relación Señal Ruido (SNR), como:

$$SNR = 10 \log_{10} \left(\frac{\sum_{n=0}^{N-1} |s(n)|^2}{\sum_{n=0}^{N-1} |r(n)|^2} \right) \quad (\text{Ec. 12})$$

Ejercicio 5

Agregar, en forma independiente, un ruido blanco con una SNR de 20 dB a las señales capturadas por los micrófonos. Escuchar ambas versiones y comparar sus espectros. Repetir los ejercicios 1 a 4 con las nuevas señales.

Una de las limitaciones del algoritmo propuesto es su carácter discreto, acotando la resolución temporal/espacial, como se puede observar en la Figura 5.b) y se ve reflejado en la Figura 6. Una alternativa para mitigar esta limitación es realizar un sobremuestreo de $IDFT(G_{PH}[k])$ antes de estimar su máximo.

Ejercicio 6

Modificar el algoritmo de estimación de los retardos realizando un sobremuestreo de las señales $IDFT(G_{PH}[k])$. Repetir los ejercicios 2 y 4.

Si tenemos más de una fuente, por ejemplo un locutor y ruido de fondo, con diferentes distribuciones de energía en su espectro de frecuencias, podemos reducir el ancho de banda a analizar al de la fuente de interés, reduciendo los errores de cálculo.

Ejercicio 7

Utilizando los espectros obtenidos en el ejercicio 5, diseñar un filtro pasabanda con el objetivo de reducir el ruido y por lo tanto mejorar la estimación de los retardos. Graficar la respuesta en frecuencia y diagrama de polos y ceros del filtro obtenido. Repetir los ejercicios 2 y 4.

Una aplicación de los arreglos de micrófonos es mejorar la relación señal ruido de una fuente en particular, eliminando o atenuando las fuentes secundarias. Una técnica clásica para mejorar el SNR es mediante una combinación lineal de las señales registradas por cada micrófono, realizando una conformación de haz o *beamforming*, donde las perturbaciones indeseadas se atenúan por el fenómeno de selectividad espacial del arreglo. Para un arreglo de N micrófonos, la señal resultante $\hat{s}[n]$ se puede expresar como:

$$\hat{s}[n] = \frac{1}{N} \sum_{i=1}^N s_i[n - d_i] \quad (\text{Ec. 13})$$

donde $s_i[n]$ es la señal registrada por el micrófono i -ésimo, con retardo de llegada d_i .

Ejercicio 8

Alinear las señales utilizando los retardos estimados en el ejercicio anterior y obtener una nueva señal promediada (Ec. 13). Escuchar ambas versiones. Tener en cuenta que es necesario sobremuestrear las señales con la misma tasa del ejercicio 6 de modo de obtener una mejor alineación.

Ejercicio 9

Graficar los espectrogramas de la señal del micrófono 1, antes y después del filtrado del ejercicio 8. Identificar los efectos del filtrado.

Ejercicio 10

Repetir los ejercicios 8 y 9 realizando una interpolación para mejorar el alineado de las señales.

Caso particular de localización de fuentes sonoras: el oído humano

Escuche el audio <https://www.youtube.com/watch?v=XgcUyCKurHA> con auriculares y los ojos cerrados. Se puede apreciar como nuestro oído actúa como un sistema de localización de fuentes sonoras, es decir, un arreglo de dos micrófonos. Si bien la diferencia de fase no es la única información que utilizamos para localizar la posición de la fuente sonora, esta juega un papel muy importante en dicho proceso. Y lo que es interesante también es que jugando con la fase de una señal de audio mono podemos producir señales estéreo que den la sensación de “espacialidad” o “sonido 3D”.

Ejercicio 11 (opcional)

Escuche la señal *audio2.wav* donde se simula el comportamiento del oído humano para localizar fuentes de sonido. Determine perceptualmente de qué lado provienen los distintos segmentos de la señal. Utilice el algoritmo de estimación de retardos para confirmar las posiciones.

Ejercicio 12 (opcional)

Escuche la señal *audio3.wav* donde se simula una fuente sonora en movimiento. ¿Que percibe que está sucediendo? Utilice el algoritmo de estimación de retardos para analizar la evolución temporal de los mismos.

Bibliografía

- [1] Tashev, I. Sound Capture and Processing: Practical Approaches. John Wiley & Sons, Ltd. ISBN: 978-0-470-31983-3. 2009.
- [2] Benesty, J.; Chen, J. & Huang, Y. Microphone Array Signal Processing. Springer Topics in Signal Processing. Ed. by J. Benesty and W. Kellermann. Springer-Verlag Berlin Heidelberg. 2008. ISBN: 978-3-540-78611-5]
- [3] Liou, J.; Joshi, J. & Vador, G. Real-Time Sound Source Localization. Project Report for EE 586. 2009.
- [4] Lee, R.; Kang, M.; Kim, B.; Park, K; Lee, S. & Park, H. Sound Source Localization Based on GCC-PHAT With Diffuseness Mask in Noisy and Reverberant Environments. *IEEE Access*, vol. 8, pp. 7373-7382, 2020, doi: 10.1109/ACCESS.2019.2963768.
- [5] Knapp, C. & Carter, G. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320-327, August 1976, doi: 10.1109/TASSP.1976.1162830.
- [6] Soria, E.; Martínez, M.; Francés, J.V.; Camps, G. Tratamiento digital de señales. Problemas y ejercicios resueltos. ISBN: 84-205-3559-1. 2003.
- [7] Howell, K. Principles of fourier analysis. Chapman & Hall/CRC. ISBN 0-8493-8275-0. 2001.