

# Estimación de dirección de arribo de fuentes sonoras

**Carmelo Cappello, Pablo Muia, Ivan Sánchez, Joaquín Varela**

*carmelcv.cv@gmail.com, p.muia94@gmail.com*  
*joaquo9909@gmail.com, sanchez61783@estudiantes.untref.edu.ar*

Ingeniería en Sonido, Universidad Nacional de Tres de Febrero.

*Resumen - En este trabajo se estudia la estimación de dirección de arribo (DOA) de una fuente sonora mediante la estimación de diferencias temporales de arribo (TDOA) entre múltiples micrófonos. Se emplea inicialmente un arreglo lineal de cuatro sensores con separación de 10 cm, y se implementan técnicas de correlación cruzada para estimar los TDOA entre pares de señales. A partir de esta base, se extiende el análisis a arreglos con un número arbitrario de micrófonos, evaluando cómo impacta la cantidad de sensores en la precisión de la estimación. Se utilizan tanto la correlación cruzada clásica como su versión generalizada (GCC), aplicando distintas funciones de ponderación. Las pruebas se realizan en condiciones de campo libre y campo reverberante, con el objetivo de comparar el desempeño de cada método frente a diferentes escenarios acústicos. El trabajo busca analizar la precisión de estimación angular obtenida por cada técnica, identificando sus ventajas y limitaciones según el contexto de aplicación.*

## 1. INTRODUCCIÓN

La estimación de la dirección de arribo (DOA) de señales acústicas se basa en el cálculo de las diferencias temporales de arribo (TDOA) entre pares de sensores. Dado un arreglo de micrófonos con geometría conocida y bajo el supuesto de incidencia de frentes de onda planos, estas diferencias temporales pueden vincularse directamente con el ángulo de arribo mediante relaciones trigonométricas.

Una de las técnicas más empleadas para la estimación de TDOA es la correlación cruzada, tanto en su versión clásica como en su forma generalizada (GCC), que incorpora funciones de ponderación para mejorar el desempeño en presencia de ruido o reverberación.

En este trabajo se implementan y comparan distintas variantes del algoritmo de correlación cruzada aplicadas a un arreglo lineal de cuatro micrófonos separados 10 cm entre sí. Se evalúa el desempeño de cada técnica en condiciones de campo libre y campo reverberante. Finalmente, se extiende el análisis a configuraciones con un número mayor de sensores, a fin de analizar cómo influye la cantidad de micrófonos en la precisión

de la estimación de DOA.

## 2. MARCO TEÓRICO

### 2.1. CORRELACIÓN CRUZADA

La estimación de la dirección de arribo (DOA, por sus siglas en inglés) de una señal acústica puede abordarse a partir del cálculo de las diferencias temporales de arribo (TDOA) entre múltiples sensores. En configuraciones donde la fuente sonora se encuentra en el campo lejano, el frente de onda que incide sobre el arreglo puede considerarse plano, y el retardo temporal entre dos micrófonos separados una distancia  $d$  se expresa como:

$$\tau_{1,2} = \frac{d \cos(\theta)}{c} \quad (1)$$

donde  $\theta$  es el ángulo de incidencia de la señal respecto a la normal del arreglo, y  $c$  es la velocidad del sonido en el medio.

Una técnica ampliamente utilizada para estimar TDOA es la correlación cruzada, que permite detectar el desplazamiento temporal entre dos señales similares [1]. Dado un par de señales discretas  $x_1(t)$  y  $x_2(t)$ , la correlación cruzada clásica se define como:

$$R_{1,2}[k] = \sum_{n=0}^{N-1} x_1[n] \cdot x_2[n-k] \quad (2)$$

El valor de  $k$  que maximiza  $R_{1,2}[k]$  se interpreta como el retardo estimado entre ambas señales.

Sin embargo, en entornos ruidosos o reverberantes, la versión clásica puede volverse poco robusta. En estos casos, se utiliza la correlación cruzada generalizada (GCC), que opera en el dominio de la frecuencia aplicando una función de ponderación  $\Psi(f)$  sobre la densidad espectral cruzada:

$$R_{1,2}^{GCC}[\tau] = \int_{-\infty}^{\infty} \Psi(f) S_{1,2}(f) e^{j2\pi f\tau} df \quad (3)$$

donde  $S_{1,2}(f)$  es la densidad espectral cruzada entre las señales  $x_1(t)$  y  $x_2(t)$ .

## 2.2. PONDERACIONES DE LA GCC

Las ponderaciones son funciones de frecuencia  $\Psi(f)$  que se aplican sobre la densidad espectral cruzada  $S_{1,2}(f)$  antes de realizar la transformada inversa para obtener la función de correlación en el dominio del tiempo [3]. Distintas elecciones de  $\Psi(f)$ , como PHAT, SCOT, ROTH, dan lugar a diferentes variantes del algoritmo GCC, cada una con propiedades particulares de robustez y precisión según el entorno acústico.

### 2.2.1. ROTH

La ponderación ROTH busca mejorar la estimación del retardo temporal al suprimir las frecuencias dominadas por el ruido en el canal de referencia [2].

$$\Psi_R(f) = \frac{1}{G_{x1,x1}(f)} \quad (4)$$

donde  $G_{x1,x1}(f)$  es la densidad espectral de potencia de la señal  $x_1(t)$ .

### 2.2.2. SCOT

La ponderación SCOT, o Smoothed Coherence Transform, busca equilibrar la influencia de ambos canales al normalizar la densidad espectral cruzada por la raíz cuadrada del producto de

las densidades espectrales individuales. Esto ayuda a mitigar los efectos de diferencias de potencia entre los canales [2].

$$\Psi_S(f) = \frac{1}{\sqrt{G_{x1,x1}(f) \cdot G_{x2,x2}(f)}} \quad (5)$$

donde  $G_{xi,xi}(f)$  es la densidad espectral de potencia de la señal  $x_i(t)$ .

### 2.2.3. PHAT

La ponderación PHAT, o Phase Transform, busca desestimar la magnitud del espectro cruzado para estudiar la fase. Esto es útil ya que el retardo temporal de una señal se manifiesta en el dominio de las frecuencias como un desfase [2]. De esta manera, la función de ponderación  $\Psi(f)$  consiste en:

$$\Psi_P(f) = \frac{1}{|G_{x1,x2}(f)|} \quad (6)$$

donde  $G_{x1,x2}(f)$  es la densidad espectral de potencia entre las señales  $x_1(t)$  y  $x_2(t)$ .

### 2.2.4. ECKART

El filtro Eckart tiene como objetivo maximizar la relación señal-ruido de la salida del operador de correlación, es decir, busca la ponderación  $\Psi(f)$  que maximiza la relación entre el espectro cruzado de  $s_1$  y  $s_2$  y el espectro cruzado del ruido de cada micrófono. Como indica Knapp[2], la función  $\Psi_E(f)$  queda de la siguiente manera:

$$\Psi_E(f) = \frac{G_{s1,s1}(f)}{G_{n1,n1}(f) \cdot G_{n2,n2}(f)} \quad (7)$$

donde  $G_{s1,s1}(f)$  es la auto-correlación de la señal real  $s_1(t)$ , y  $G_{nx,nx}(f)$  es la auto-correlación de los ruidos  $n_1(t)$  y  $n_2(t)$  correspondientemente. Es pertinente aclarar que para poder usar este filtro en la práctica, es necesario tener conocimiento tanto de la señal de referencia como también del ruido presente en los micrófonos.

### 2.2.5. HT (ML)

Al filtro HT o ML se lo puede interpretar como una extensión de la ponderación PHAT. Este filtro se encarga de obtener un pico mejor definido a pesar de la presencia de ruido. La ponderación

posee un término con el cuál se deshace de la amplitud (como en la ponderación de PHAT) y otro término que depende de la coherencia en frecuencia entre ambas señales. De esta manera, el filtro desestima los desfases causados por ruido y define mejor el pico por desplazamiento en la correlación cruzada resultante. [2]. La función  $\Psi_{HT}(f)$  queda definida de la siguiente manera:

$$\Psi_{HT}(f) = \frac{1}{G_{x1,x2}(f)} \cdot \frac{|\gamma(f)|^2}{1 - |\gamma_{1,2}(f)|^2} \quad (8)$$

Aquí, nuevamente  $G_{x1,x2}(f)$  es correlación cruzada entre  $x_1(t)$  y  $x_2(t)$  y  $\gamma_{1,2}(f)$  es la coherencia en frecuencia de las señales  $x_1(t)$  y  $x_2(t)$ :

$$\gamma_{1,2}(f) = \frac{G_{x1,x2}(f)}{\sqrt{G_{x1,x1}(f) \cdot G_{x2,x2}(f)}} \quad (9)$$

### 3. DESARROLLO EXPERIMENTAL

Con el objetivo de evaluar el rendimiento de los distintos métodos de estimación de TDOA (Time Difference of Arrival), se diseñó una serie de pruebas en entorno simulado, implementadas en el entorno Python. En principio, las pruebas se realizaron utilizando un arreglo lineal de cuatro micrófonos equiespaciados a 10 cm.

#### 3.1. PRUEBAS TEORICAS

Inicialmente, se emplearon señales de prueba con el objetivo de verificar el funcionamiento del algoritmo. En primer lugar, se hizo uso de una señal compuesta por un pulso unitario limpio como fuente sonora. Esta señal fue retrasada artificialmente para simular el tiempo de arribo a cada micrófono, permitiendo establecer una referencia clara para la estimación del retardo entre pares de señales. A partir de dichas señales, se aplicó la correlación cruzada clásica (CC) para determinar los retardos entre micrófonos a partir del TDOA, y de esa forma obtener el ángulo estimado promedio de la dirección de ubicación de la fuente.

Posteriormente, a las mismas señales anteriores se les incorporó ruido rosa no correlacionado en cada canal, a fin de simular condiciones no

ideales. Se repitió la estimación mediante correlación cruzada clásica, obteniendo así el retraso en tiempo buscado entre cada micrófono y obteniendo el ángulo estimado de dirección de la fuente. Además, sobre estas señales con ruido, se implementaron las distintas variantes del método de correlación cruzada generalizada (GCC), aplicando el algoritmo base en conjunto con cada una de las funciones de ponderación correspondientes: Roth, SCOT, PHAT, HT y Eckart. De esta forma se obtuvieron diferentes resultados dependiendo de que algoritmo se haya utilizado.

#### 3.2. SIMULACIONES

En una segunda etapa, se procedió a simular condiciones de propagación más realistas. Utilizando la librería pyroomacoustics, se representó el caso de campo libre usando la función de cámara anecoica que simula un recinto sin reflexiones alguna y sin dimensiones específicas, solamente limitada por las ubicaciones de las fuentes y micrófonos, delimitando que se considere la absorción acústica del aire y valores de temperatura y humedad de 20 °C y 40 % respectivamente, para considerar condiciones de medición típicas. A esto, se aplicó una fuente sonora reproduciendo un audio anecoico de una voz femenina [3]. En cuanto a los micrófonos, se utilizaron de patrón polar omnidireccional para partir de una situación general antes de analizar aplicaciones más prácticas. La simulación de este caso se basa en suponer que la fuente se encuentra en campo libre y en el campo lejano, lo que permite modelar el frente de onda como plano. Bajo estas condiciones, se asume que todos los micrófonos reciben una misma señal desplazada en el tiempo, lo que habilita la estimación del TDOA de forma análoga a los casos anteriores. Se aplicaron los algoritmos de CC y GCC al conjunto de señales obtenidas, manteniendo fija la configuración del arreglo de micrófonos. Posteriormente, se repitió el proceso incorporando ruido blanco gaussiano a la simulación, a través de parámetros propios de la función de simulación estableciendo una relación SNR de 15 dB, con el fin de analizar las diferencias en el desempeño de cada ponderación bajo condiciones no ideales.

Luego, se llevaron a cabo pruebas bajo condiciones de campo reverberante con el objetivo

de evaluar el desempeño de los algoritmos en un entorno desfavorable con exceso de reflexiones. Para ello, se simuló una sala cerrada de  $10 \times 10 \times 10$  m, con un coeficiente de absorción promedio de  $\alpha = 0,01$ . Se añadió ruido al entorno y se emplearon micrófonos con patrón polar omnidireccional, nuevamente, a fin de captar las reflexiones provenientes de todas las direcciones. En estas condiciones, se repitieron las estimaciones de TDOA utilizando las técnicas de correlación cruzada clásica y generalizada con sus respectivas ponderaciones, con el fin de comparar su rendimiento en escenarios complejos.

Para un caso más complejo, se procedió con una simulación en donde se buscó recrear la situación de operación en un ambiente abierto considerando solamente reflexiones provenientes del suelo y micrófonos de patrón polar cardioide, para estudiar la cancelación de las mismas en la práctica. El arreglo de micrófonos se mantuvo igual y se realizaron pruebas con diferentes ubicaciones de fuentes. En este caso se tuvieron que especificar dimensiones del recinto, al no poder usar la función de habitación anecoica, por lo cuál se determinó un área de  $1 \text{ km}^2$  y  $100 \text{ m}$  de alto para que los límites del espacio no representaran un problema en la simulación. Al suelo se le asignaron coeficientes de absorción por banda de octava respectivas a la grava [4], mientras que para el resto de superficies se les asignaron coeficientes de absorción de  $\alpha = 1$  para representar la no reflexión del campo libre.

### 3.3. SIMULACIONES CON DIFERENTES CANTIDADES DE RECEPTORES

En última instancia se repitieron las diferentes simulaciones con pyroomacoustics pero variando

la cantidad de micrófonos dispuestos en el arreglo lineal para comprobar si una mayor o menor cantidad de receptores en el arreglo lineal implica una relación con mayor precisión en la estimación del TDOA y, respectivamente, en el DOA.

## 4. RESULTADOS

## 5. DISCUSIONES

## 6. CONCLUSIONES

## REFERENCES

- [1] Benesty, J., Chen, J., Huang, Y. (2008). Microphone Array Signal Processing. Springer. Capítulo 9: Time Delay Estimation.
- [2] Knapp, C. H. & Carter, G. C. (1976). The generalized correlation method for estimation of time delay. IEEE Transactions on Acoustics, Speech, and Signal Processing, 24(4), 320–327.
- [3] Quinton, P. (2018). HRIR measurements of a KEMAR dummy-head microphone [Dataset]. University of Edinburgh. <https://datashare.ed.ac.uk/handle/10283/2651>
- [4] Furet, R. (2012). Bunker Audio. Coeficientes de absorción acústica de materiales. <https://www.bunker-audio.com/bunker-audio-portal-sonido-documentos.php?id=3>