

# Estimación de dirección de arribo de fuentes sonoras

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

*carmelcv.cv@gmail.com, p.muia94@gmail.com*  
*joa909909@gmail.com, sánchez61783@estudiantes.untref.edu.ar*

Ingeniería en Sonido, Universidad Nacional de Tres de Febrero.

*Resumen - En este trabajo se estudia la estimación del ángulo de incidencia de un frente de onda sobre un arreglo de micrófonos, mediante la estimación de diferencias temporales de arribo (TDOA) entre múltiples sensores. 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, variación del tiempo de reverberación, distancia entre micrófonos, distancia entre fuente y el arreglo lineal, ángulo de elevación de la fuente y frecuencia de muestreo. 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. Además, se analiza el error de estimación angular en función de distintos parámetros del sistema. El trabajo busca analizar la precisión de estimación angular obtenida por cada técnica, identificando sus ventajas y limitaciones.*

## 1. INTRODUCCIÓN

La estimación de la dirección de arribo (DOA, por sus siglas en inglés) de señales acústicas se basa en el cálculo de las diferencias temporales de arribo (TDOA, por sus siglas en inglés) 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 error de estimación para cada método, analizando el impacto de diversos parámetros en condiciones de campo libre y campo reverberante.

La evaluación se extiende a configuraciones con distinta cantidad de sensores, con el objetivo de estudiar cómo influye este factor en la precisión de la DOA. También se analiza la influencia de otras variables, como el tiempo de reverberación, la distancia entre la fuente y el arreglo, la separación entre micrófonos, el ángulo de elevación de la fuente respecto del plano del arreglo lineal y la frecuencia de muestreo.

## 2. MARCO TEÓRICO

### 2.1. CORRELACIÓN CRUZADA

La estimación de la dirección de arribo (DOA) de una señal acústica puede abordarse a partir del cálculo de las diferencias temporales de arribo (TDOA) entre múltiples sensores. Teniendo en cuenta que el frente de onda que incide sobre el arreglo se considera 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 al eje axial del arreglo lineal, y  $c$  es la velocidad del sonido en el medio. Dado que  $\tau_{1,2}$  tiene su máximo en  $0^\circ$  y  $180^\circ$ , y físicamente el retardo es máximo cuando los micrófonos se encuentran alineados con la dirección de propagación de la onda sonora, se establece como referencia de  $0^\circ$  al eje de alineación de micrófonos.

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) y DOA (Direction of Arrival), el desarrollo experimental se dividió en dos etapas:

- Validación del algoritmo en campo libre.
- Simulaciones con Pyroomacoustics [5].

En todos los casos, las simulaciones se realizaron aplicando los algoritmos de correlación cruzada entre todos los pares posibles de micrófonos, y no únicamente respecto de un micrófono de referencia.

Para evaluar el desempeño de cada método, se implementó una medida de error basada en el error cuadrático medio angular, tal como se describe en la ecuación (10). Esta métrica calcula el cuadrado de la distancia angular entre el ángulo estimado ( $\hat{\theta}$ ) y el ángulo real ( $\theta$ ), considerando la naturaleza circular de los ángulos. Un valor de  $\varepsilon$  cercano a 0 corresponde a una estimación precisa, mientras que errores mas cercanos a 1 reflejan estimaciones altamente inexactas, correspondientes a direcciones opuestas de incidencia del frente de onda. Por ejemplo, una diferencia angular de 90° de diferencia se corresponde con un  $\varepsilon = 0,25$ .

$$\varepsilon = \frac{((\theta - \hat{\theta} + \pi) \text{Mod}(2\pi) - \pi)^2}{\pi^2} \quad (10)$$

### 3.1. DISEÑO DE SISTEMA

Se implementó el algoritmo para estimar TDOA y DOA a partir de un arreglo lineal a partir de los distintos métodos de correlación cruzada generalizada.

El sistema cuenta inicialmente con cuatro micrófonos omnidireccionales separados por una distancia de 10 cm, cuyas señales serán correlacionadas entre sí para la obtención de un  $\bar{\theta}$  promedio. Previo a la ejecución de la correlación cruzada, se colocó un filtro butterworth de orden 1 con frecuencia de corte de ecuación:

$$f_c = \frac{2 \cdot d \cdot c}{f_s \cdot 0,5} \quad (11)$$

para evitar aliasing en la correlación cruzada por la distancia de separación entre micrófonos [1], con  $f_s$  representando a la frecuencia de muestreo. Este fenómeno puede generar picos en la correlación cruzada, aumentando la probabilidad de la selección de un  $\tau$  erróneo.

Dado que la diferencia máxima de tiempo de arribo entre 2 puntos para un mismo frente de onda plano se produce cuando estos se encuentran alineados con la dirección de propagación de la misma, se ventanearon las señales de correlación cruzada para evitar retrasos superiores al máximo para la onda. Dicho retraso corresponde a  $\tau_{max} = d/c$ , por lo que se colocó una ventana de cantidad de muestras  $N$  a partir de la ecuación (12):

$$N = \frac{A \cdot d \cdot f_s}{c} \quad (12)$$

en donde  $A$  es una relación entre el tamaño de muestras de la ventana y el  $\tau_{max}$ .

Finalmente se realiza la correlación cruzada clásica y generalizada entre todas las señales de pares de micrófono y se obtiene un  $\bar{\theta}$  promedio.

### 3.2. SIMULACIÓN EN CAMPO LIBRE

En primer lugar, se evaluó el sistema en condiciones de campo libre para verificar su correcto funcionamiento bajo condiciones ideales.

En un recinto simulado sin reflexiones, considerando absorción del aire a 20° C y 40 % de humedad, se reprodujo una señal anecoica (voz femenina) donde se recorrieron todas las variantes de correlación cruzada generalizada (GCC), concretamente las funciones de ponderación Roth, SCOT, PHAT, HT y Eckart, además de la correlación cruzada clásica (CC).

La simulación se realizó variando la posición relativa entre la fuente y el arreglo de micrófonos, con el objetivo de analizar cómo se modifican los retardos temporales de arribo (TDOA) y su impacto en los ángulos estimados. Durante este proceso, se mantuvo constante la separación entre micrófonos en 10 cm, así como la cantidad de sensores (4 micrófonos) y las dimensiones de la sala. Además, se analizaron dos casos donde se varió el nivel de la relación señal ruido (SNR), en un caso 30 dB y en otro caso 5 dB, para poder analizar visualmente que métodos funcionan de una manera mas eficiente.

### 3.3. SIMULACIONES EN RECINTO CERRADO

Se realizaron simulaciones a partir de la librería Pyroomacoustics, generando un recinto cerrado de  $10m \times 10m \times 3m$  con un SNR de 10 dB para poder verificar el comportamiento del algoritmo frente a un ruido excesivo y en presencia de las reflexiones del recinto.

A continuación, se realizaron simulaciones en el mismo recinto pero analizando y variando sistemáticamente los siguientes parámetros:

- Tiempo de reverberación ( $RT_{60}$ ).

- Cantidad de micrófonos y distancia de separación entre los mismos.
- Distancia entre la fuente y el centro del arreglo.
- Elevación de la fuente respecto al arreglo lineal.
- Posición relativa de la fuente y del arreglo dentro de la sala.
- Frecuencia de muestreo.

Las simulaciones se llevaron a cabo en una sala de dimensiones constantes ( $10m \times 10m \times 3m$ ), modelando un entorno reverberante. Además, estas dimensiones son representativas de una sala real de tamaño medio, lo que aporta realismo a la simulación sin introducir complejidades propias de espacios muy grandes o irregulares. A su vez, permiten mantener controlado el tiempo de reverberación ( $RT_{60}$ ) de forma independiente al volumen de la sala. Por último, mantener constante el tamaño del recinto a lo largo de todas las simulaciones asegura la comparación entre casos, permitiendo aislar el efecto de otros parámetros como la separación entre micrófonos, la distancia fuente-array o la posición relativa de los elementos. La elección de estas dimensiones responde a variaciones de diferentes parámetros.

#### 3.3.1. VARIACIÓN DE $RT_{60}$

Con la sala fija, se recorrió un rango de tiempos de reverberación desde 0,2s a 1,1s para estudiar la influencia del tiempo de reverberación ( $RT_{60}$ ) sobre el sistema. Para cada valor de  $RT_{60}$  se generaron 100 simulaciones, variando la posición horizontal de la fuente y del arreglo lineal  $x, y$ , manteniendo la cota  $z$  constante. En cada iteración se calculó el ángulo estimado ( $\hat{\theta}$ ) mediante los algoritmos de correlación cruzada clásica (CC) y generalizada (GCC), y se comparó con el ángulo calculado  $\theta$  para extraer el valor medio del error angular y su desviación estándar.

#### 3.3.2. DISTANCIA ENTRE FUENTE Y ARREGLO LINEAL

A continuación, se evaluó el efecto de la distancia entre la fuente y el arreglo. Con la geometría del arreglo inmóvil, se desplazó la fuente en el plano horizontal  $x, y$  para modificar su distancia euclídea al centro del arreglo lineal. Para cada nueva distancia, se repitió el barrido de

$RT_{60}$ , realizando 100 simulaciones por cada variación de posición de fuente, en combinación con los otros parámetros, con el fin de obtener estimaciones estadísticamente representativas del valor medio del error angular y su desviación estándar.

### 3.3.3. CANTIDAD DE MICRÓFONOS Y SEPARACIÓN ENTRE SENSORES DEL ARREGLO

Seguidamente, se examinó cómo la separación entre los micrófonos afecta la precisión de la estimación. Manteniéndose constantes la sala y la posición relativa de la fuente, se fue modificando el  $RT_{60}$ ,  $d_{mic}$ , y número de micrófonos. Por cada parámetro a analizar, cantidad de micrófonos y separación de los mismos, se realizaron 125 simulaciones. Luego de realizar las simulaciones se agruparon los datos necesarios a partir del DataFrame generado para poder obtener reportar en cada combinación del parámetro analizado, el valor medio del error angular y la desviación estándar.

Con este análisis se buscó analizar cuál configuración puede llegar a ser la más óptima para estimación de dirección de arribo.

### 3.3.4. ANGULO DE INCIDENCIA DE ELECCIÓN DE LA FUENTE

Con el arreglo en posición fija, se modificaron las coordenadas tridimensionales  $x, y, z$  de la fuente para cubrir distintos ángulos de elevación. De manera que en cada posición de  $x, y$  se modifique la variable de la altura  $z$ , realizando 25 movimientos de altura de la fuente, en conjunto con los demás parámetros. Al modificar la altura de la fuente, se permitió calcular las distancias euclidianas entre la fuente y el centro del array para obtener el ángulo real de incidencia. De este modo se cuantificó el impacto conjunto de la reverberación y la geometría espacial en el error de estimación de DOA.

### 3.3.5. FRECUENCIA DE MUESTREO

Con el objetivo de evaluar cómo influye la frecuencia de muestreo en la precisión de la estimación del ángulo de arribo, se realizaron 2250 simulaciones por cada variación de frecuencia de

muestro manteniendo constantes varios parámetros: la cantidad de micrófonos del arreglo, la separación entre ellos, y variando cinco configuraciones distintas de tiempo de reverberación  $RT_{60}$ . En cada caso, se variaron las posiciones relativas de la fuente y del arreglo dentro de la sala, a fin de representar distintas geometrías de incidencia.

Sobre esta base, se iteraron cinco valores de frecuencia de muestreo: 48 kHz, 44.1 kHz, 16 kHz, 8 kHz y 4 kHz. Para cada combinación de parámetros se calcularon los errores en la estimación angular, utilizando las métricas ya mencionadas con el fin de analizar cómo impacta la resolución temporal del sistema sobre el desempeño de los algoritmos de correlación.

## 4. RESULTADOS

### 4.1. RESULTADOS EN CAMPO LIBRE

A continuación se presentan los resultados obtenidos bajo condiciones ideales de campo libre, es decir, sin reflexiones ni ruido ambiental. Se analizaron tanto la correlación cruzada clásica (CC) como las cinco variantes del método de correlación cruzada generalizada (GCC): PHAT, Roth, SCOT, Eckart y HT.

Para ilustrar el comportamiento de cada método y ponderación, se muestra en la Figura 1 las funciones de correlación obtenidas para un par específico de micrófonos del arreglo con un SNR de 30 dB. Se utilizó el par de micrófonos 1-2, siendo el micrófono 1 el de referencia.

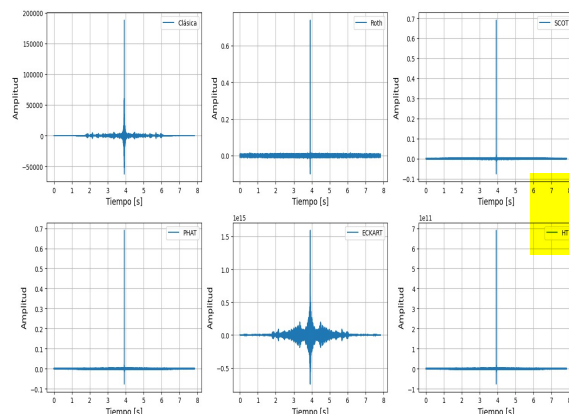


Figura 1: Correlación cruzada clásica y generalizada entre pares de micrófonos 1-2 en caso de campo libre, con SNR de 30 dB.

Por otro lado, en la Figura 2 se analiza el mismo caso para un SNR de 5 dB.

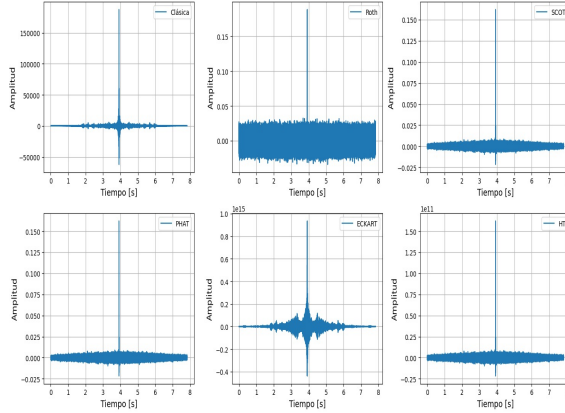


Figura 2: Correlación cruzada clásica y generalizada entre pares de micrófonos 1-2 en caso de campo libre, con SNR de 5 dB.

Adicionalmente, se llevó a cabo un conjunto de aproximadamente 100 simulaciones independientes para cada método, a fin de evaluar su desempeño estadístico. En la Figura 3 se presenta el valor medio del error angular de estimación del ángulo de arribo ( $\theta$ ) junto con su desviación estándar, permitiendo comparar la precisión relativa entre métodos bajo condiciones ideales.

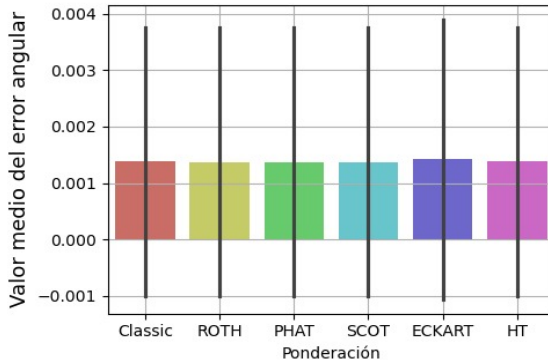


Figura 3: Valor medio del error angular y desviación estándar de ponderaciones de la correlación cruzada generalizada.

## 4.2. RESULTADOS EN CAMPO REVERBERANTE

A continuación en la Figura 4 se presentan los resultados de la simulación en la condición de campo reverberante, es decir, con reflexiones, además de tener ruido blanco agregado con un

SNR de 10 dB. Se analizaron tanto la correlación cruzada clásica (CC) como las cinco variantes del método de correlación cruzada generalizada (GCC): PHAT, Roth, SCOT, Eckart y HT.

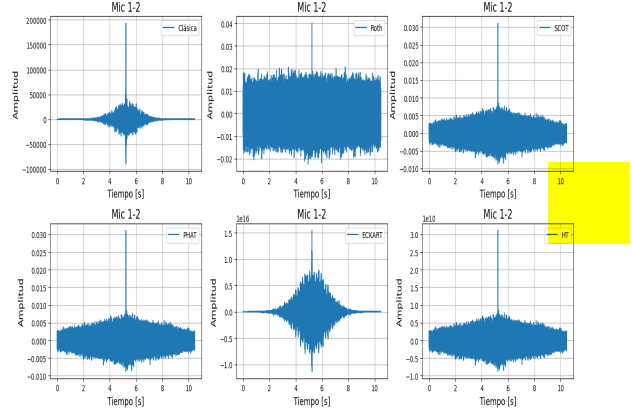


Figura 4: Correlación cruzada clásica y generalizada entre pares de micrófonos 1-2 en caso de campo reverberante, con SNR de 10 dB.

En este caso, se presentan los resultados obtenidos a partir de simulaciones realizadas en salas cerradas, modelando condiciones de campo reverberante. Estos escenarios permiten analizar el desempeño de los distintos métodos de estimación en presencia de reflexiones múltiples, típicas de entornos reales no controlados.

### 4.2.1. TIEMPO DE REVERBERACIÓN

En la Figura 5 se presenta un gráfico de barras que muestra el valor medio del error angular medio en la estimación del ángulo de arribo ( $\theta$ ) en función del tiempo de reverberación  $RT_{60}$ , variando desde 0,2 s hasta 1,1 s. Para cada valor de  $RT_{60}$ , se realizaron múltiples simulaciones, y se representó la media del error junto con su desviación estándar mediante barras de error.



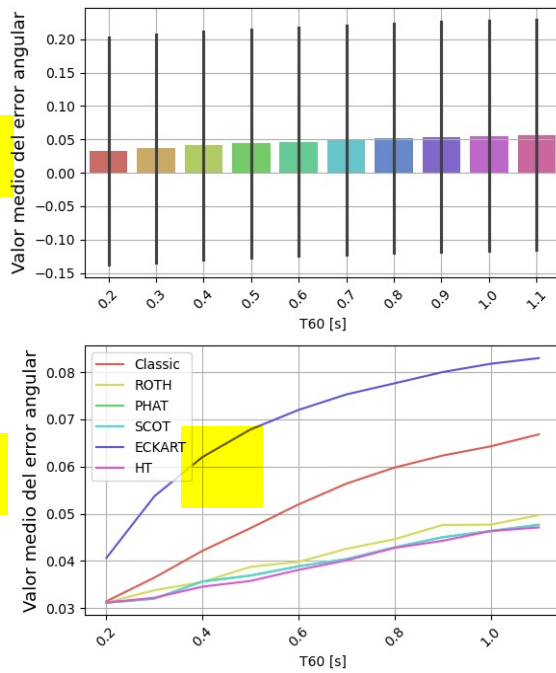


Figura 5:

- a) Valor medio del error angular y desviación estándar del ángulo estimado a partir del  $RT_{60}$ .  
 (b) Valor medio del error angular por ponderación de GCC.

#### 4.2.2. DISTANCIA ENTRE ARREGLO Y FUENTE

La Figura 6 muestra la evolución del error angular promedio en función de la distancia entre la fuente emisora y el arreglo de micrófonos, evaluado bajo múltiples esquemas de ponderación aplicados al algoritmo GCC (Classic, PHAT, ROTH, SCOT, ECKART y HT) en función a la separación fuente-array bajo distintas condiciones simuladas.

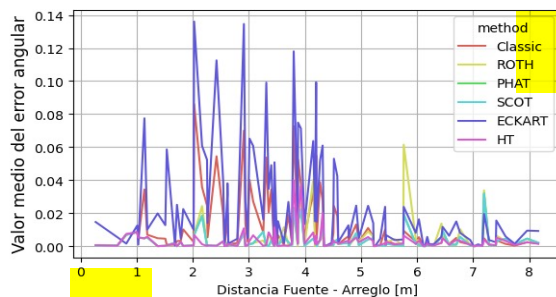


Figura 6: Valor medio del error angular del ángulo estimado en función a la separación fuente-array, con métodos de ponderación.

#### 4.2.3. CANTIDAD Y DISTANCIA ENTRE MICRÓFONOS

En la Figura 7 se presentan los errores angulares medios con sus respectivas desviaciones estándar en función de la cantidad de micrófonos utilizados en el arreglo. De forma análoga, la Figura 8 muestra los resultados obtenidos al variar la separación entre micrófonos, manteniendo constantes las demás condiciones de simulación.

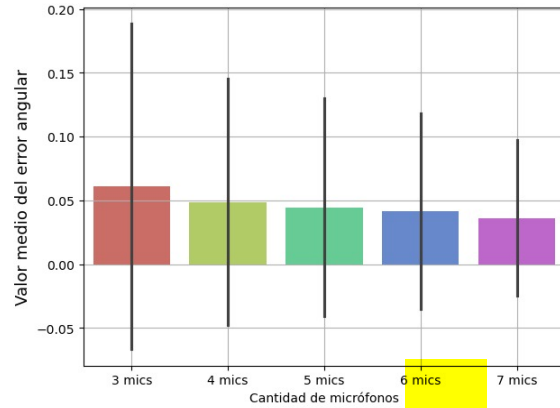


Figura 7: Valor medio del error angular y desviación estándar del ángulo estimado a partir de la cantidad de micrófonos del arreglo lineal.

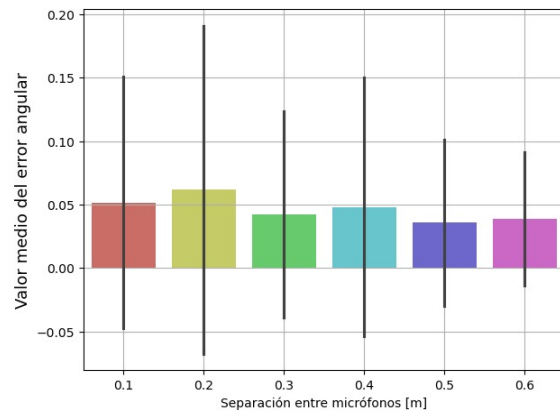


Figura 8: Valor medio del error angular y desviación estándar del ángulo estimado a partir de la distancia de separación entre micrófonos del arreglo lineal.

Adicionalmente, se presentan los resultados del error angular medio en función de la separación entre micrófonos con curvas que representan distintas cantidades de micrófonos en el arreglo.

#### 4.2.4. ESTIMACIÓN DEL ÁNGULO DE INCIDENCIA DE LA FUENTE

En la Figura 10 se muestran los resultados del valor medio del error angular continuos y su desviación estándar en la estimación del ángulo de incidencia de la fuente, obtenidos a partir de variaciones en la posición vertical de la fuente (eje  $z$ ), el tiempo de reverberación ( $RT_{60}$ ) y la separación entre micrófonos. Además, en la Figura 11 se muestra el valor medio del error angular a partir de las diferentes ponderaciones del GCC. El arreglo permaneció en una posición fija en todas las simulaciones.

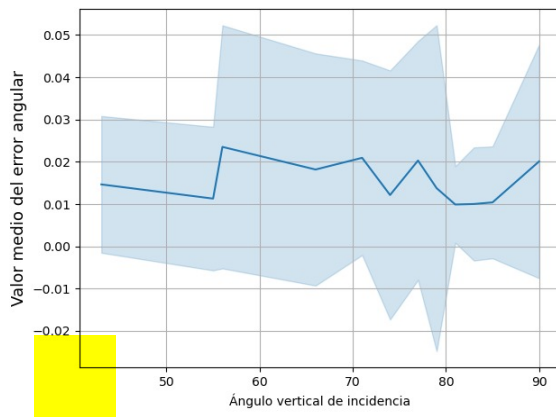


Figura 10: Valor medio del error angular y desviación estándar del angulo estimado a partir del angulo de incidencia de la fuente.

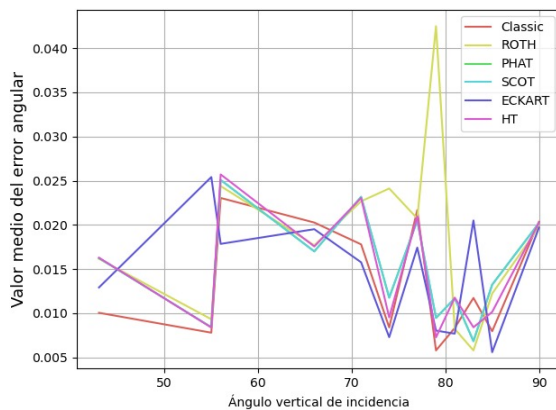


Figura 11: Valor medio del error angular por ponderación de GCC.

El gráfico superior muestra errores medios generalmente bajos ( $< 0.03$ ), con mayores desviaciones en ángulos intermedios ( $66^\circ$ – $77^\circ$ ), lo que sugiere una mayor sensibilidad en ese rango.

#### 4.2.5. FRECUENCIA DE MUESTREO

En la Figura 12 se presentan los resultados del valor medio del error angular y su desviación estándar en la estimación del ángulo de arribo para distintos valores de frecuencia de muestreo. Cada punto del gráfico corresponde al promedio del error angular obtenido a partir de múltiples simulaciones con posiciones variadas de la fuente y el arreglo. Las barras de error representan la desviación estándar de dichas estimaciones.

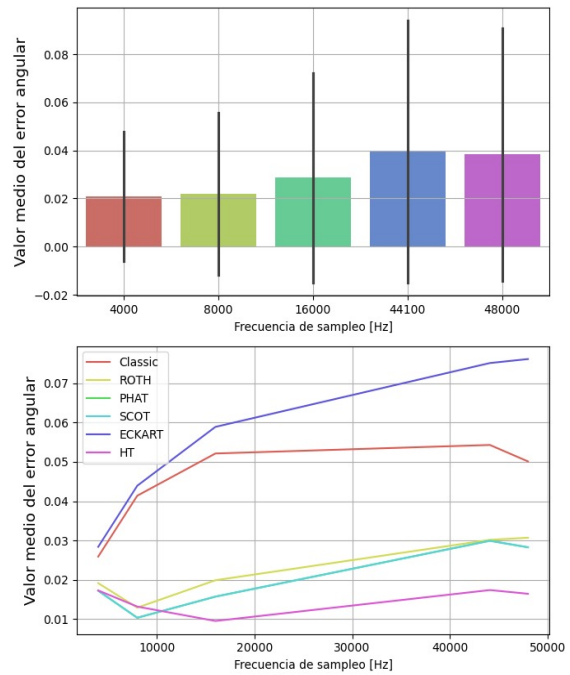


Figura 12:

- (a) Valor medio del error angular y desviación estándar del angulo estimado a partir de la frecuencia de muestreo.
- (b) Valor medio del error angular por ponderación de GCC en funcion de la frecuencia de muestreo.

Los resultados muestran que, al analizar el error angular medio en función de la frecuencia de muestreo, nuevamente las ponderaciones Eckart y clásica son las que presentan mayores errores. En contraste, SCOT, PHAT y HT mantienen un desempeño más preciso, con menor error angular a lo largo de las distintas frecuencias evaluadas.



## 5. DISCUSIONES

### 5.1. ANÁLISIS DE CC Y GCC

Al comparar los métodos de correlación cruzada (CC y GCC) en distintos escenarios acústicos, se observa una diferencia clara en su desempeño dependiendo de las condiciones del entorno. Los resultados muestran que, en condiciones de campo libre, las variantes de ponderación PHAT, SCOT y HT ofrecen los mejores resultados, con picos de correlación claros y menor sensibilidad al ruido, lo que mejora la estimación del TDOA.

En campo reverberante, con un SNR de 10 dB, se observa que la eficacia del algoritmo posee cierta dependencia ante el efecto de las reflexiones, las cuales introducen ruido no correlacionado, empeorando la estimación del TDOA y, en consecuencia, del DOA.

A partir del análisis de los distintos casos, se observa que la ponderación de Eckart es, en general, la menos recomendable. Los datos obtenidos muestran que este método presenta un mayor error promedio en la estimación del ángulo, y su rendimiento se ve especialmente afectado en condiciones de bajo SNR.

Por el contrario, las ponderaciones SCOT, PHAT y HT demostraron poseer mayor precisión, respecto a los demás algoritmos, en condiciones de ruido elevado. Ante esto, se observa que al aplicar la GCC es necesario aplicar filtrado o preprocesamiento a la señal captada en entornos no ideales.

Por otro lado, se aprecia que los métodos PHAT y SCOT **presentan resultados muy similares tanto en las simulaciones como en su formulación matemática**. En casos donde se obtienen resultados idénticos puede llegar a deberse por el redondeo numérico en la implementación y por su estructura matemática comparable. Esta similitud también ha sido reportada en la literatura especializada [2].

### 5.2. ANÁLISIS DE PARÁMETROS EN SIMULACIONES

Los resultados presentados en las simulaciones en campo reverberante de la estimación del ángulo  $\theta$  mediante CC y GCC muestran un aumento del error con mayores niveles de reverberación, debido a la ambigüedad temporal en la de-

tección del retardo. Si bien este efecto no resulta significativamente perjudicial a los valores simulados, se debe posiblemente a que el tiempo de reverberación no es lo suficientemente elevado como para afectar de forma crítica el desempeño.

En relación con la distancia entre la fuente y el arreglo, se observa que la precisión de la mayoría de las ponderaciones no se ven afectadas en su error angular y, además, presentan relativa estabilidad para los diferentes valores analizados. Las ponderaciones que menos afección de este tipo presentan son las de HT, PHAT y SCOT, siendo la ponderación de ECKART la que otorga mayor error y con mayor variabilidad de valores, seguido por la correlación clásica.

A partir de los resultados analizados, se observa que el aumento en la cantidad de micrófonos del arreglo tiende a reducir el error en la estimación angular y disminuyendo la dispersión de la misma, mejorando la precisión del sistema. En contraposición, a mayor cantidad de micrófonos, también el tiempo de computo resulta mayor, por lo que colocar un número de sensores de manera excesiva puede traer desventajas en cuanto a eficiencia. Por otro lado no se encontró una tendencia entre la separación entre micrófonos y error de estimación, por lo que estadísticamente se puede inferir que **los 0.5 m** de separación brindan mejores resultados.

Al considerar ambas variables en conjunto, los resultados indican que la configuración más precisa, dentro de los límites de simulación planteados, corresponde a una combinación de 5 a 7 micrófonos con una separación de 0,5 m.

En el caso del análisis de error a partir de las técnicas de ponderación, al evaluar el ángulo estimado a partir del ángulo de incidencia de la fuente, se aprecia que los métodos de ponderación muestran un error promedio similar en la mayoría de los ángulos, y quizás se podría destacar la ponderación PHAT que se mantiene en todos los casos un valor de error más bajo. En contraste, ROTH presenta un pico de error notorio en valores cercanos a los  $80^\circ$ , lo cual podría deberse a su mayor sensibilidad a pequeñas variaciones de fase o condiciones específicas de propagación.

En los resultados correspondientes a la frecuencia de muestreo, se observa un comportamiento, contrario a lo esperado, donde el error angular es menor para frecuencias de muestreo más

bajas, teniendo en cuenta que en todos los casos se partió de un audio originalmente grabado en 48 kHz y se realizó el proceso de downsampling. Se nota que el caso de peor precisión es al utilizar una frecuencia de 44.1 kHz.

## 6. CONCLUSIONES

Del análisis realizado sobre los métodos de correlación cruzada, se concluye que la ponderación Eckart fue la menos eficiente en la mayoría de los escenarios evaluados. En cambio, las variantes SCOT, PHAT y HT demostraron un mejor desempeño, especialmente en condiciones de bajo SNR, mostrando mayor precisión general frente al ruido. Estas características las presentan como opciones más recomendables para contextos acústicamente adversos. Asimismo, los resultados obtenidos en las simulaciones del ángulo de elevación —y en otros escenarios evaluados— muestran que la correlación cruzada clásica presenta un error considerable en la estimación del DOA. Esto refuerza la conveniencia de utilizar variantes ponderadas al aplicar la GCC.

Respecto a los parámetros analizados, se observó que el  $RT_{60}$  no produce una variación sustancial en el error de estimación angular. En cambio, la relación señal-ruido (SNR) sí tiene un impacto significativo en la precisión.

Además, cabe destacar que al utilizar la librería `pyroomacoustics`, como herramienta nos otorgó la posibilidad de estudio y desarrollo de simulaciones, pero tiene ciertas limitaciones y dificultades intrínsecas que complicaron el desarrollo de este trabajo, como el excesivo tiempo de simulación o alinealidades en las mismas que no corresponden con lo esperable en la realidad.

Finalmente, se concluye que el sistema de estimación de dirección de arribo debería estar compuesto por 5 micrófonos, una separación de 50 cm, ponderación HT y frecuencia de muestreo de 4 kHz. Se recomienda cómo estudios futuros analizar el desempeño de esta configuración en pruebas similares a las realizadas a lo largo de todo el estudio.

## 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>
- [5] Scheibler, R., Bezzam, E., & Dokmanić, I. (2018). *Pyroomacoustics: A Python package for audio room simulations and array processing algorithms*. <https://github.com/LCAV/pyroomacoustics>