Localización y separación de múltiples fuentes de voz usando un arreglo de micrófonos

Luis M. Gato Díaz
lmiguelgato@comunidad.unam.mx

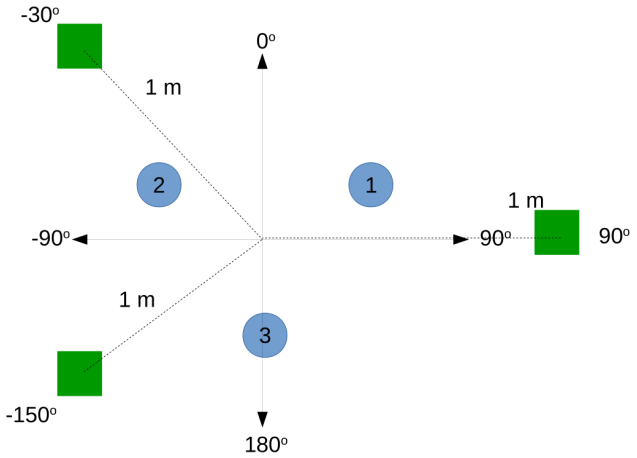
Maestría en Ingeniería Eléctrica, UNAM
Posgrado de Procesamiento Digital de Señales

Proyecto Final - Procesamiento Digital de Audio
Profesor: Dr. Caleb Rascón Estebané

Sumario

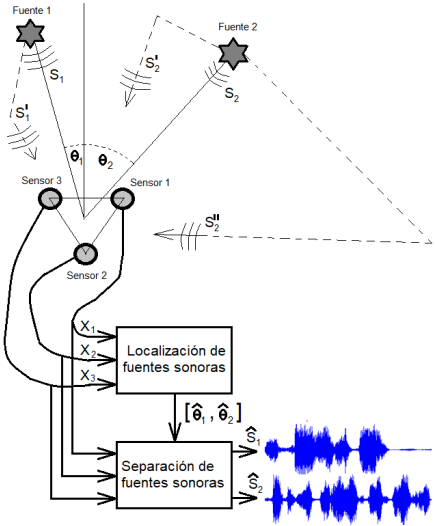
Sumario

Descripción del problema



Tarea No. 1: Estimar la dirección en donde se localizan las fuentes de voz.

Descripción del problema

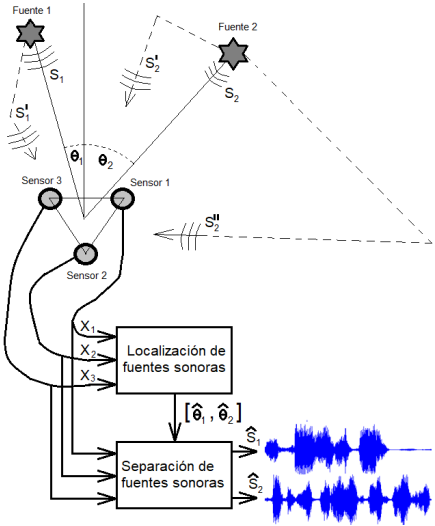


Requerimientos:

- Separación a ciegas: tanto las fuentes de voz como el proceso de mezclado son desconocidos.
- Únicamente se dispone de las grabaciones asociadas a cada elemento del arreglo.
- Maximizar la relación señal a interferencia.
- Presencia de niveles de ruido y de reverberación moderados.
- Reducido costo computacional.

Tarea No. 2: Separar las distintas fuentes de voz presentes en la mezcla.

Descripción del problema

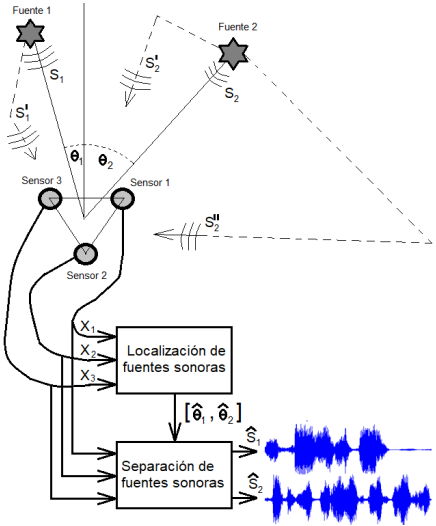


Requerimientos:

- Separación a ciegas: tanto las fuentes de voz como el proceso de mezclado son desconocidos.
- Únicamente se dispone de las grabaciones asociadas a cada elemento del arreglo.
- Maximizar la relación señal a interferencia.
- Presencia de niveles de ruido y de reverberación moderados.
- Reducido costo computacional.

Tarea No. 2: Separar las distintas fuentes de voz presentes en la mezcla.

Descripción del problema

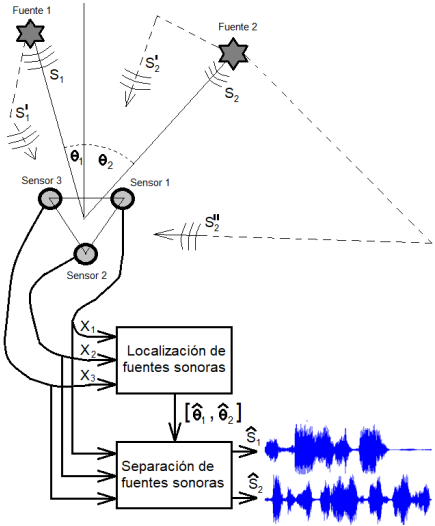


Requerimientos:

- Separación a ciegas: tanto las fuentes de voz como el proceso de mezclado son desconocidos.
- Únicamente se dispone de las grabaciones asociadas a cada elemento del arreglo.
- Maximizar la relación señal a interferencia.
- Presencia de niveles de ruido y de reverberación moderados.
- Reducido costo computacional.

Tarea No. 2: Separar las distintas fuentes de voz presentes en la mezcla.

Descripción del problema

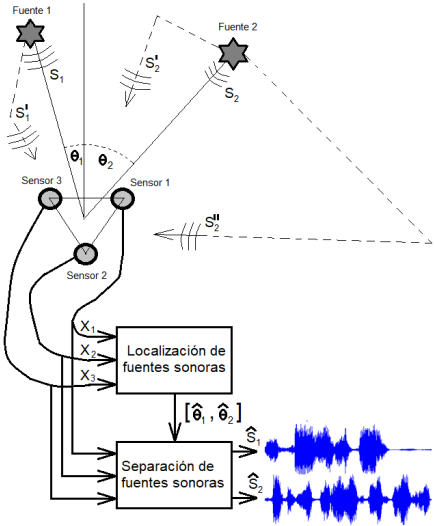


Requerimientos:

- Separación a ciegas: tanto las fuentes de voz como el proceso de mezclado son desconocidos.
- Únicamente se dispone de las grabaciones asociadas a cada elemento del arreglo.
- Maximizar la relación señal a interferencia.
- Presencia de niveles de ruido y de reverberación moderados.
- Reducido costo computacional.

Tarea No. 2: Separar las distintas fuentes de voz presentes en la mezcla.

Descripción del problema



Requerimientos:

- Separación a ciegas: tanto las fuentes de voz como el proceso de mezclado son desconocidos.
- Únicamente se dispone de las grabaciones asociadas a cada elemento del arreglo.
- Maximizar la relación señal a interferencia.
- Presencia de niveles de ruido y de reverberación moderados.
- Reducido costo computacional.

Tarea No. 2: Separar las distintas fuentes de voz presentes en la mezcla.

Sumario

Modelo geométrico de propagación

Ecuación de onda en medios homogéneos y no dispersivos:

$$\frac{\partial^2 E(t,\mathbf{r})}{\partial x^2}+\frac{\partial^2 E(t,\mathbf{r})}{\partial y^2}+\frac{\partial^2 E(t,\mathbf{r})}{\partial z^2}=\frac{1}{c^2}\frac{\partial^2 E(t,\mathbf{r})}{\partial t^2} \tag{1}$$

Si la fuente es un emisor puntual:

$$\frac{\partial^2 \{rE(t,\mathbf{r})\}}{\partial r^2}=\frac{1}{c^2}\frac{\partial^2 \{rE(t,\mathbf{r})\}}{\partial t^2} \tag{2}$$

Solución de la ecuación de onda:

$$E(t,\mathbf{r})=s(t-r/c) \tag{3}$$

Ejemplo: señales complejas de banda estrecha $E(t,\mathbf{r})=Ae^{j(\omega t-\mathbf{k}\cdot\mathbf{r})}=Ae^{j\omega t}e^{-j\mathbf{k}\cdot\mathbf{r}}$

Modelo geométrico de propagación

Detección tridimensional de una fuente:

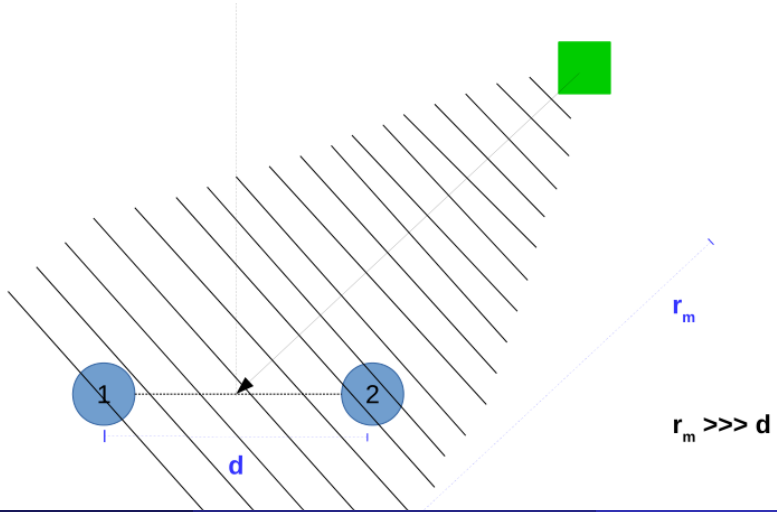
$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \begin{bmatrix} e^{-j\mathbf{k}(\theta,\phi)\cdot\mathbf{r}_1} \\ e^{-j\mathbf{k}(\theta,\phi)\cdot\mathbf{r}_2} \\ \vdots \\ e^{-j\mathbf{k}(\theta,\phi)\cdot\mathbf{r}_M} \end{bmatrix} s(t) = \mathbf{a}(\theta,\phi)s(t) \tag{4}$$

Detección bidimensional de varias fuentes contaminadas con ruido:

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{w}(t) \tag{5}$$

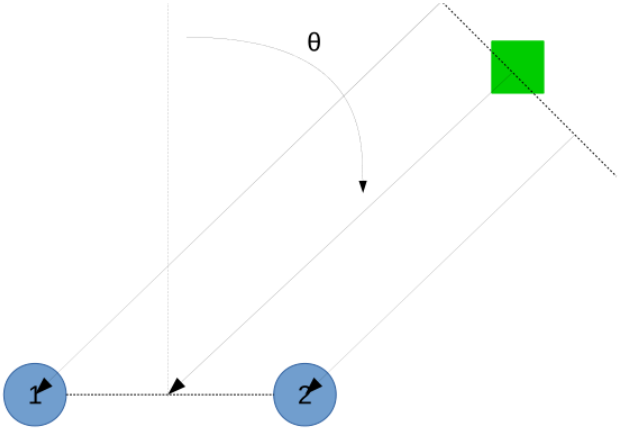
Modelo geométrico de propagación

Modelo de campo lejano:



Modelo geométrico de propagación

Modelo de campo lejano:

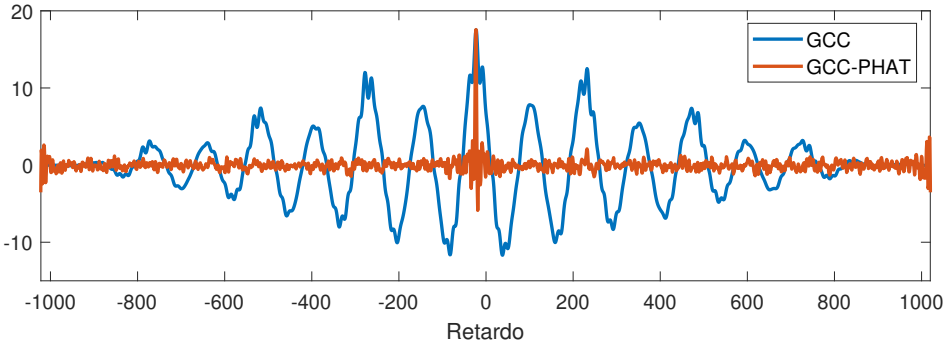


Sumario

Estimación de las direcciones de arribo

Vector de correlación cruzada con transformada de fase:

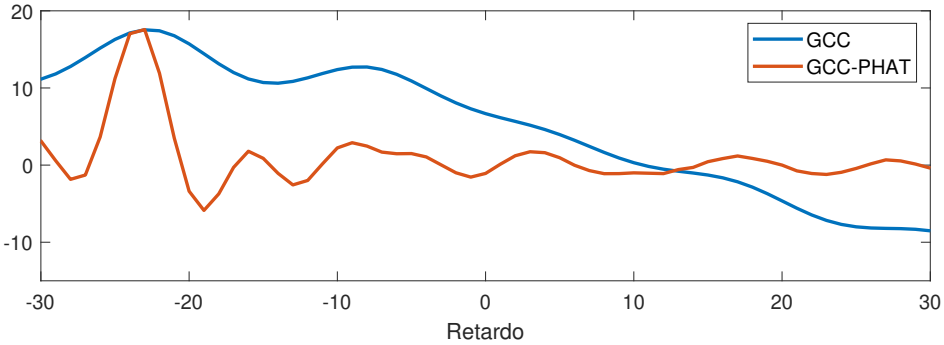
$$r_{pq}[k] = \frac{1}{N}\sum_{m=0}^{N-1}\frac{X_p[m]X_q^*[m]}{|X_p[m]||X_q[m]|}e^{j2\pi km/N} \qquad \text{para } k_{\min} \leq k \leq k_{\max} \tag{6}$$



Estimación de las direcciones de arribo

Vector de correlación cruzada con transformada de fase:

$$r_{pq}[k] = \frac{1}{N}\sum_{m=0}^{N-1}\frac{X_p[m]X_q^*[m]}{|X_p[m]||X_q[m]|}e^{j2\pi km/N} \qquad \text{para } k_{\min} \leq k \leq k_{\max} \tag{7}$$

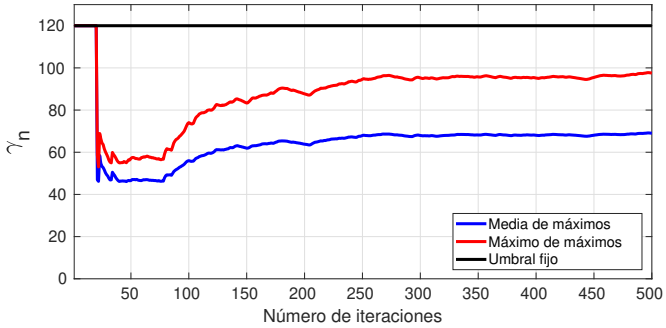


Estimación de las direcciones de arribo

Ajustes al método de correlación:

- Se usó un umbral para descartar niveles bajos de correlación: $r_{pq}[k]_{max} > \gamma_0$
- Luego se incluyó un umbral adaptativo:

a) $\gamma_n = 0.9\frac{(n-1)\gamma_{n-1}+max\{r_{pq}[k]_{max}\}}{n}$ b) $\gamma_n = 0.9\frac{(n-1)\gamma_{n-1}+\overline{r_{pq}[k]_{max}}}{n}$

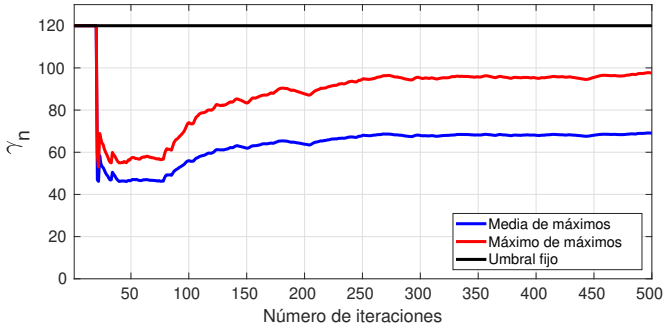


Estimación de las direcciones de arribo

Ajustes al método de correlación:

- Se usó un umbral para descartar niveles bajos de correlación: $r_{pq}[k]_{max} > \gamma_0$
- Luego se incluyó un umbral adaptativo:

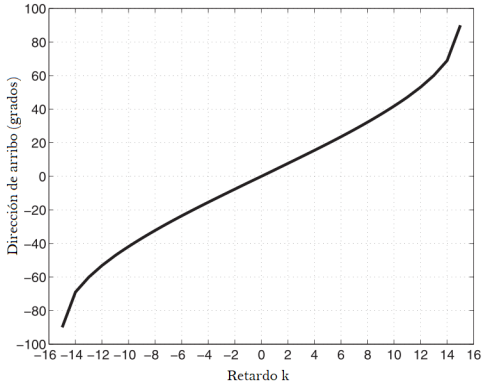
  a) $\gamma_n = 0.9\frac{(n-1)\gamma_{n-1}+max\{r_{pq}[k]_{max}\}}{n}$ b) $\gamma_n = 0.9\frac{(n-1)\gamma_{n-1}+\overline{r_{pq}[k]_{max}}}{n}$

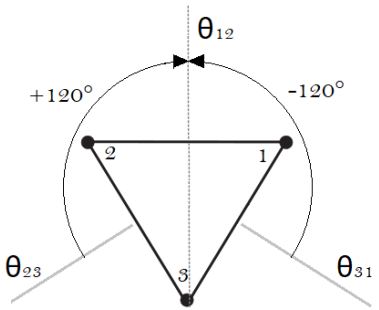


Estimación de las direcciones de arribo

Se estima, para cada par de micrófonos la dirección de arribo:

$$\theta_{12} = \frac{\text{sen}^{-1}(c\Delta t_{12})}{d} \qquad \theta_{23} = \frac{\text{sen}^{-1}(c\Delta t_{23})}{d} \qquad \theta_{31} = \frac{\text{sen}^{-1}(c\Delta t_{31})}{d} \tag{8}$$

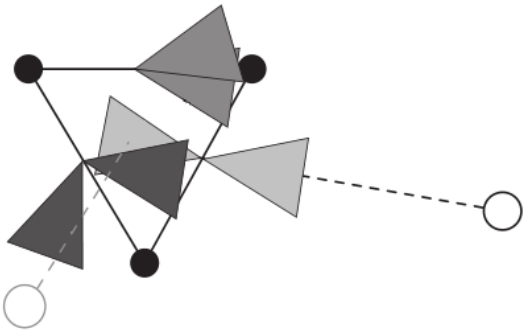




Estimación de las direcciones de arribo

Se determina si existe redundancia en las direcciones de arribo:

$$[\theta_{12}; \theta'_{12}] \qquad [\theta_{23}; \theta'_{23}] + 120^{\circ} \qquad [\theta_{31}; \theta'_{31}] - 120^{\circ} \tag{9}$$



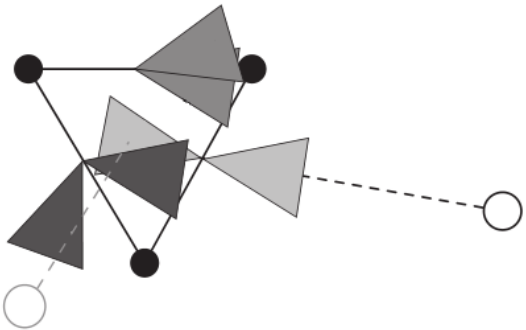
Existen 2^3 pares distintos de ángulos, de los cuales se pueden obtener 2 o hasta 3 pares que *apuntan* aproximadamente a la misma dirección.

Se estableció un umbral de coherencia: $\Delta\theta < 15^{\circ}$

Estimación de las direcciones de arribo

Se determina si existe redundancia en las direcciones de arribo:

$$[\theta_{12}; \theta'_{12}] \qquad [\theta_{23}; \theta'_{23}] + 120^{\circ} \qquad [\theta_{31}; \theta'_{31}] - 120^{\circ} \tag{9}$$

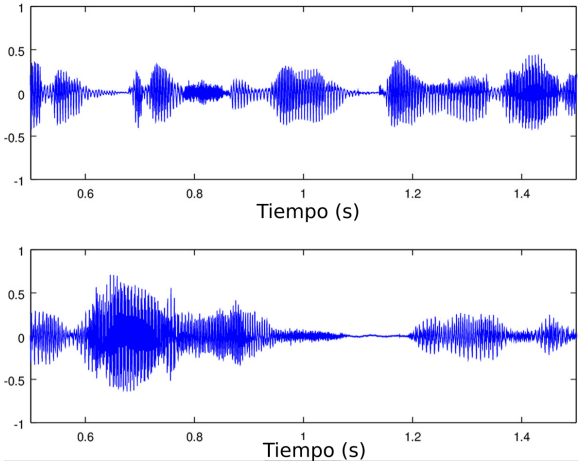


Existen 2^3 pares distintos de ángulos, de los cuales se pueden obtener 2 o hasta 3 pares que *apuntan* aproximadamente a la misma dirección.

Se estableció un umbral de coherencia: $\Delta\theta < 15^{\circ}$

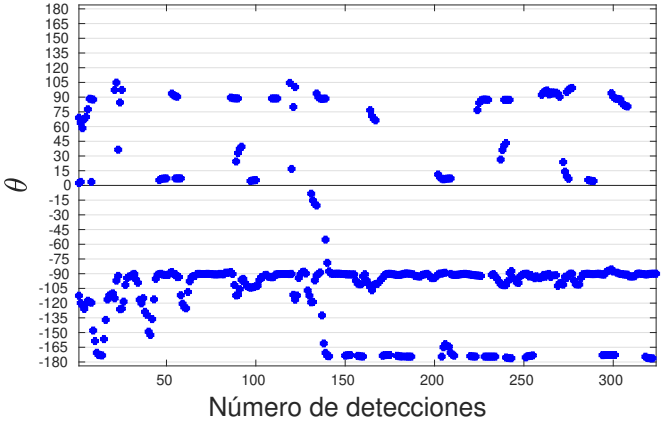
Estimación de las direcciones de arribo

Fuentes de voz que no coinciden por breves segmentos de tiempo ...



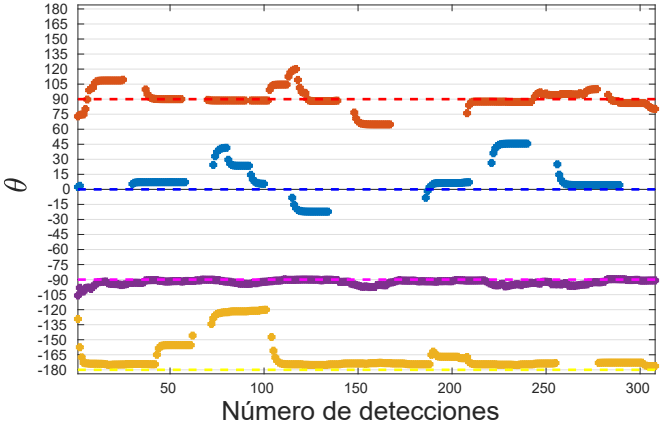
Estimación de las direcciones de arribo

... proporcionan conjuntos distinguibles de direcciones de arribo, asociados a las distintas fuentes presentes:



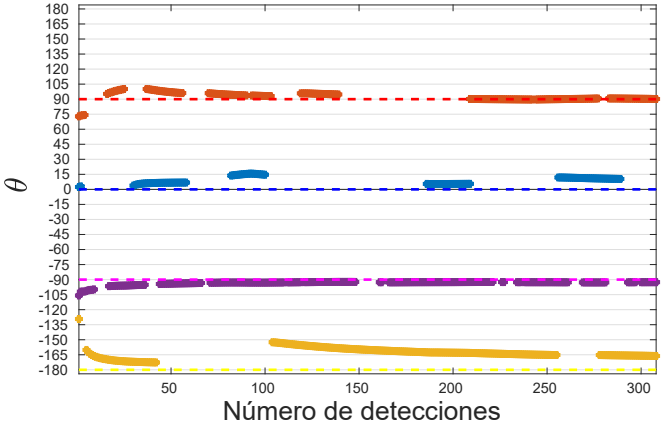
Estimación de las direcciones de arribo

Se usó el algoritmo de clasificación k-means para distinguir las distintas fuentes, suponiendo que no hay dos fuentes en una misma dirección.



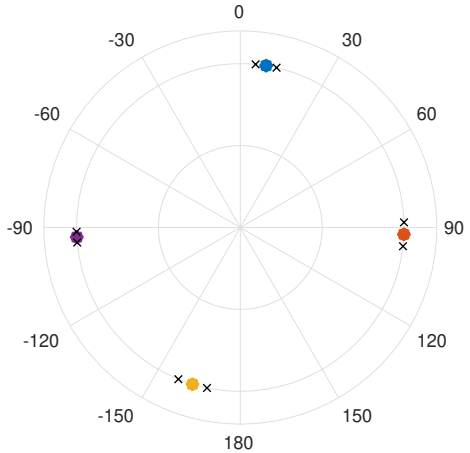
Estimación de las direcciones de arribo

Se aplicó a los centroides de k-means un filtro de media móvil y un umbral de desviación respecto a la media.



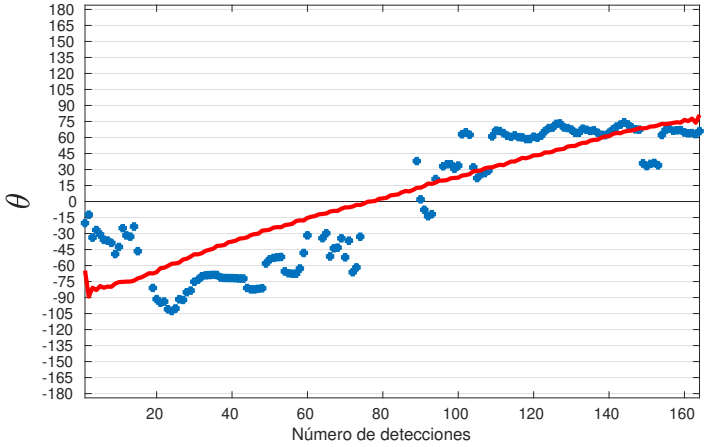
Estimación de las direcciones de arribo

Se aplicó a los centroides de k-means un filtro de media móvil y un umbral de desviación respecto a la media.



Estimación de las direcciones de arribo

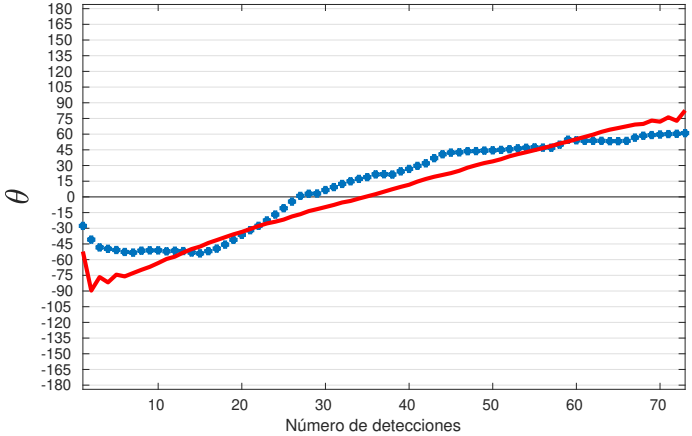
Estas modificaciones permiten también extender el método a fuentes móviles:



antes ...

Estimación de las direcciones de arribo

Estas modificaciones permiten también extender el método a fuentes móviles:

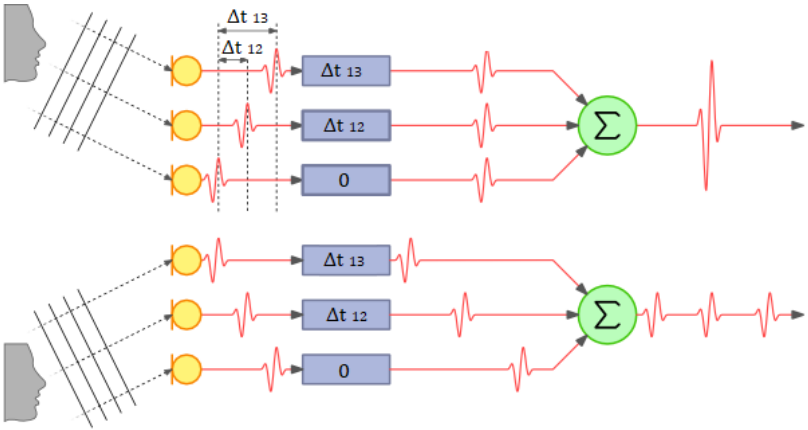


... después.

Sumario

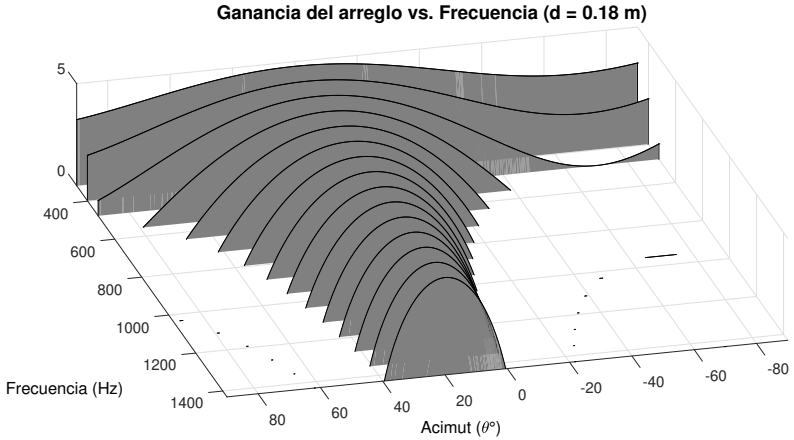
Separación de las fuentes de voz

Formador de haz de retardos y sumas:



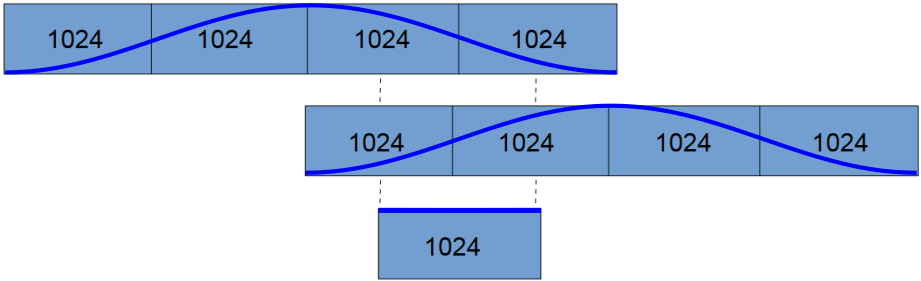
Separación de las fuentes de voz

Formador de haz de retardos y sumas:



Separación de las fuentes de voz

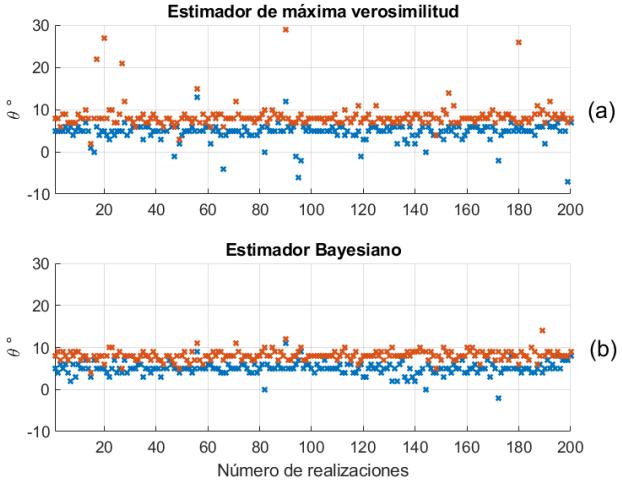
Se aplican los retardos operando en el dominio de la frecuencia, sobre dos buffers solapados de 4 ventanas de datos (overlap-add):



Sumario

Resultados

Estimación de las direcciones de arribo de dos fuentes localizadas en $\boldsymbol{\theta} = [5^{\circ} \;\; 8^{\circ}]^T$.



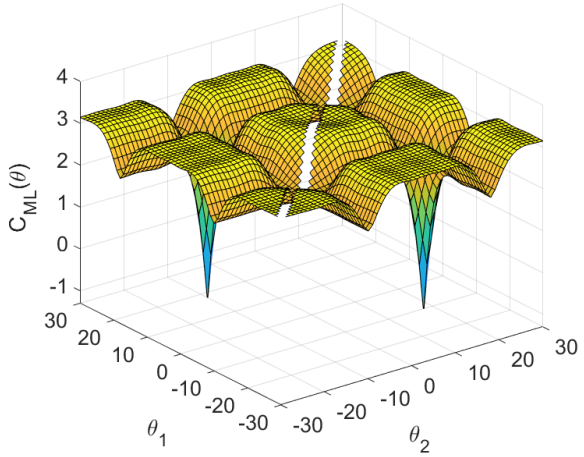
Resultados

Error cuadrático medio de los estimadores.

σ^2	SNR (dB)	$\theta = 5^o$		$\theta = 8^o$	
		MV	**Bayes**	**MV**	**Bayes**
0.0001	40.0	0.2236^o	0.2236^o	0.3391^o	0.1871^o
0.001	30.0	0.4743^o	0.4301^o	0.4416^o	0.3808^o
0.01	20.0	2.4135^o	1.1136^o	2.1107^o	1.7176^o
0.02	16.9	5.0453^o	3.0668^o	1.4265^o	1.1832^o
0.06	12.2	9.9088^o	5.7615^o	4.2988^o	3.1177^o
0.1	10.0	10.064^o	4.9487^o	5.0813^o	3.5043^o

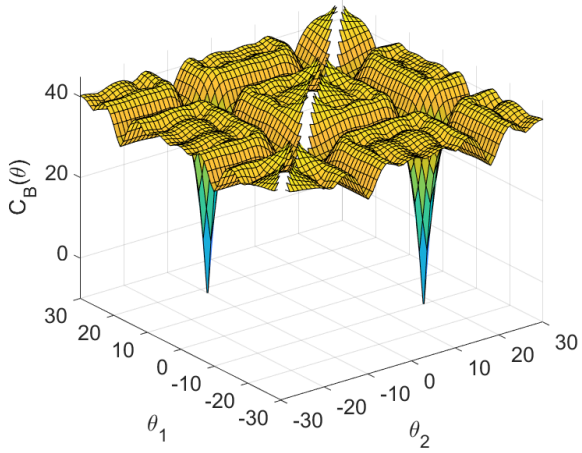
Resultados

Función de costo C_{ML} evaluada sobre la superficie $\{\theta_1, \theta_2\} \in \{-30^o, 30^o\} \times \{-30^o, 30^o\}$ para un escenario con dos fuentes localizadas en $\boldsymbol{\theta} = [-10^o \;\; 18^o]^T$ y $\sigma^2 = 0.02$.



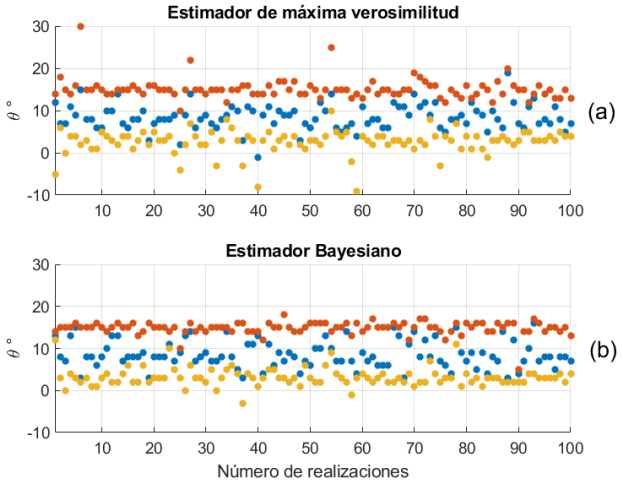
Resultados

Función de costo C_B evaluada sobre la superficie $\{\theta_1, \theta_2\} \in \{-30^o, 30^o\} \times \{-30^o, 30^o\}$ para un escenario con dos fuentes localizadas en $\boldsymbol{\theta} = [-10^o \quad 18^o]^T$ y $\sigma^2 = 0.02$.



Resultados

Estimación de las direcciones de arribo de tres fuentes localizadas en $\boldsymbol{\theta} = [3^{\text{o}} \quad 8^{\text{o}} \quad 15^{\text{o}}]^T$.



Sumario

Conclusiones

- Se obtuvo un método que permite estimar simultáneamente el número de fuentes y sus direcciones con un **reducido número de observaciones**.
- Presenta un **menor error cuadrático medio** que el estimador de máxima verosimilitud.
- Relativamente **elevado costo computacional**.
- No es aplicable a:
  - señales de banda ancha,
  - campo no lejano,
  - ruido correlacionado.

Conclusiones

- Se obtuvo un método que permite estimar simultáneamente el número de fuentes y sus direcciones con un **reducido número de observaciones**.
- Presenta un **menor error cuadrático medio** que el estimador de máxima verosimilitud.
- Relativamente **elevado costo computacional**.
- No es aplicable a:
  - señales de banda ancha,
  - campo no lejano,
  - ruido correlacionado.

Conclusiones

- Se obtuvo un método que permite estimar simultáneamente el número de fuentes y sus direcciones con un **reducido número de observaciones**.
- Presenta un **menor error cuadrático medio** que el estimador de máxima verosimilitud.
- Relativamente **elevado costo computacional**.
- No es aplicable a:
  - señales de banda ancha,
  - campo no lejano,
  - ruido correlacionado.

Conclusiones

- Se obtuvo un método que permite estimar simultáneamente el número de fuentes y sus direcciones con un **reducido número de observaciones**.
- Presenta un **menor error cuadrático medio** que el estimador de máxima verosimilitud.
- Relativamente **elevado costo computacional**.
- No es aplicable a:
  - señales de banda ancha,
  - campo no lejano,
  - ruido correlacionado.

Localización y separación de múltiples fuentes de voz usando un arreglo de micrófonos

Luis M. Gato Díaz
lmiguelgato@comunidad.unam.mx

Maestría en Ingeniería Eléctrica, UNAM
Posgrado de Procesamiento Digital de Señales

Proyecto Final - Procesamiento Digital de Audio
Profesor: Dr. Caleb Rascón Estebané