

Separación de Fuentes

Modelo de Señales Capturadas

- Se asume que hay una o varias señales de interés (SOI), con direcciones de arribo conocidas, tal que:

$$\mathbf{A} = \begin{bmatrix} e^{-i2\pi f T_{1:1}} & e^{-i2\pi f T_{1:2}} & \dots & e^{-i2\pi f T_{1:D}} \\ e^{-i2\pi f T_{2:1}} & e^{-i2\pi f T_{2:2}} & \dots & e^{-i2\pi f T_{2:D}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-i2\pi f T_{M:1}} & e^{-i2\pi f T_{M:2}} & \dots & e^{-i2\pi f T_{M:D}} \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} s_1(1) & s_1(2) & \dots & s_1(N) \\ s_2(1) & s_2(2) & \dots & s_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ s_D(1) & s_D(2) & \dots & s_D(N) \end{bmatrix}$$

$$\mathbf{X} = \mathbf{S} \mathbf{A}$$

Donde:

s_d : es una señal de origen

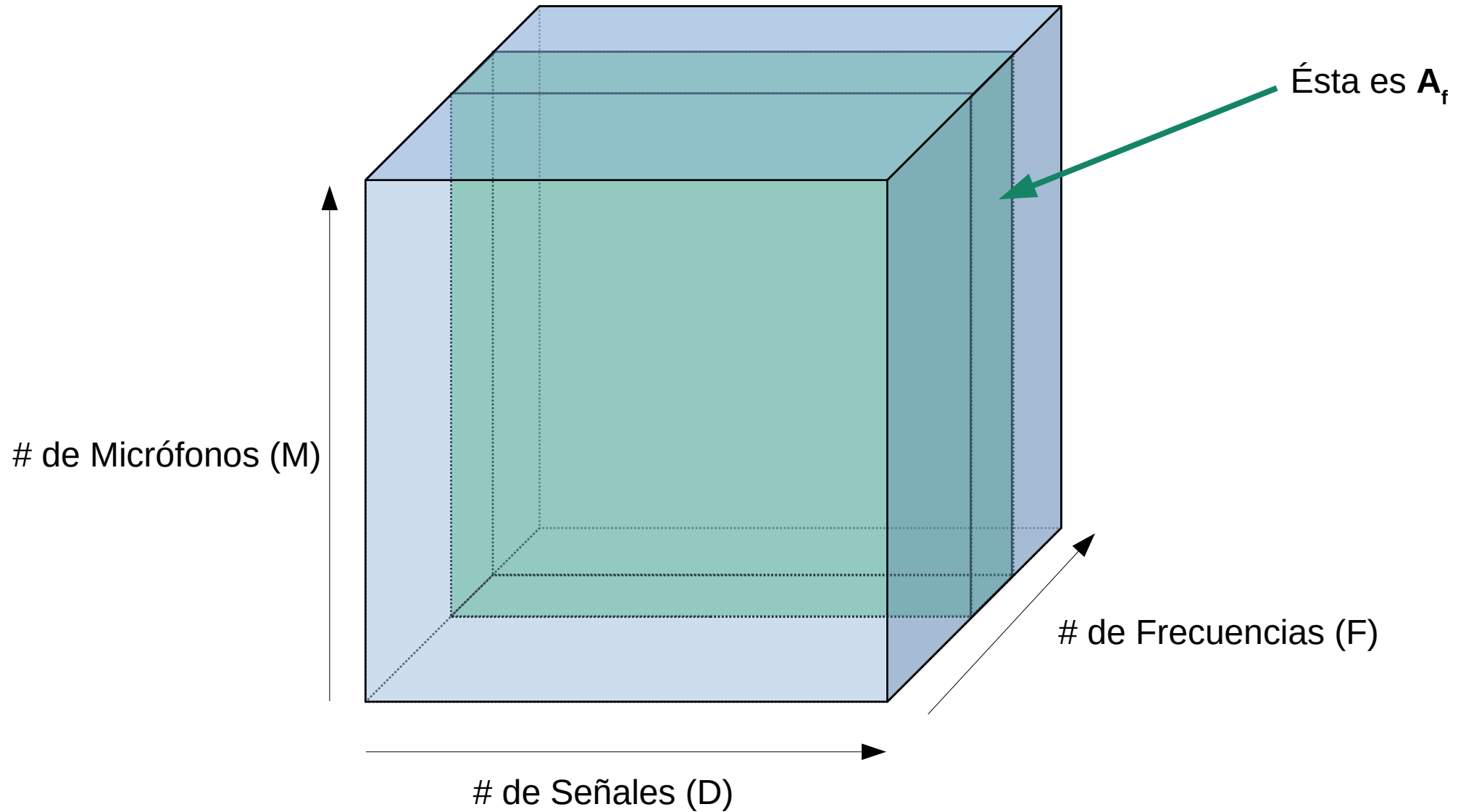
N : tamaño de la señal (o de la ventana de la señal)

$T_{m:d}$: es el retraso recibido de la señal s_d en el micrófono m

A: es la matriz que contiene los vectores de dirección (*direction vectors*)

X: son las señales capturadas (en los micrófonos); cada renglón representa un micrófono

RECORDATORIO



RECORDATORIO

- Por lo tanto, realmente tenemos que hacer esto para cada frecuencia:

$$\mathbf{A}_f = \begin{bmatrix} e^{-i2\pi f T_{1:1}} & e^{-i2\pi f T_{1:2}} & \dots & e^{-i2\pi f T_{1:D}} \\ e^{-i2\pi f T_{2:1}} & e^{-i2\pi f T_{2:2}} & \dots & e^{-i2\pi f T_{2:D}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-i2\pi f T_{M:1}} & e^{-i2\pi f T_{M:2}} & \dots & e^{-i2\pi f T_{M:D}} \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} s_1(1) & s_1(2) & \dots & s_1(N) \\ s_2(1) & s_2(2) & \dots & s_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ s_D(1) & s_D(2) & \dots & s_D(N) \end{bmatrix}$$

$$\mathbf{X}_f = \mathbf{S} \mathbf{A}_f$$

Donde:

s_d : es una señal de origen

N : tamaño de la señal (o de la ventana de la señal)

$T_{m:d}$: es el retraso recibido de la señal s_d en el micrófono m

\mathbf{A}_f : es la matriz que contiene los vectores de dirección para la frecuencia f

\mathbf{X}_f : son las señales capturadas en la frecuencia f ; cada renglón representa un micrófono

Objetivo

- Estimar las señales en **S** por medio de aplicar una matriz adicional **W** a **X**.

$$\hat{\mathbf{S}} = \mathbf{W} \mathbf{X}$$

Objetivo

- Estimar las señales en **S** por medio de aplicar una matriz adicional **W** a **X**.

$$\hat{\mathbf{S}} = \mathbf{W} \mathbf{X}$$

La matriz mágica que necesitamos calcular.

Escuelas de Pensamiento

- Beamforming.
- Análisis Estadístico.

3era Escuela de Pensamiento

Igual que Multi-DOA

- Aprendizaje Automático/Profundo***
 - No están diseñados para trabajar en línea.
 - Se “aprenden” la geometría del arreglo.
 - Si hay cambio en la geometría, se requiere re-entrenar.
 - Son “caja negra”, poco explicables.

Nos quedamos con éstas 2 Escuelas de Pensamiento

- Beamforming.
- Análisis Estadístico.

Escuelas de Pensamiento: Beamforming

- W es conocida como la matriz de *steering vectors*.
 - Está ligado a A , la matriz de *direction vectors*.
- Se pueden manipular las señales X , desfasándolas y sumándolas, para obtener una señal de S .
- Y luego repetir por cada dirección deseada.

Escuelas de Pensamiento: Análisis Estadístico

- W es una matriz “desmezcladora”, que puede separar de una sólo vez a todas las señales de X en S .
- Normalmente tiene la limitante de separar un máximo de señales como hay micrófonos.

Evaluación de Separación

- Para **localización**, es trivial llevar a cabo una evaluación, conociendo la localización verdadera:
 - Distancia entre la verdadera y la estimada.
- Pero evaluar separación es un poco más complicado:

Evaluación de Separación

- La métrica más utilizada es la Razón Señal-Interferencia (SIR):

$$SIR = \frac{P_{SOI}}{\sum P_{Interf}}$$

- Entre más alta, mayor la presencia de la señal de interés relativo al cúmulo de las intensidades de las interferencias.

Evaluación de Separación

- Hay otras métricas como:
 - Razón de Señal-Distorsión (SDR): *distorsión* siendo la diferencia entre la señal deseada y la capturada.
 - Razón de Señal-Artefactos (SAR): *artefactos* siendo sonidos insertados por el proceso de separación. También conocido como “ruido musical”.
 - Ejemplo: los sonidos insertados al introducir discontinuidades en el dominio de la frecuencia y regresar al dominio del tiempo.

Evaluación de Separación

- Al separar las señales (\hat{S}) de las mezclas (X), se utilizan las señales originales (S) para calcular estas métricas.
 - Si no, sería un cálculo “psíquico”.
- Pero, implementar este cálculo va más allá de la intención de este curso.

Evaluación de Separación

- Una implementación conocida es la documentada en:
 - E. Vincent, R. Gribonval and C. Fevotte, "Performance measurement in blind audio source separation," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 4, pp. 1462-1469, July 2006.
- Conocido como *bss_eval_sources*.
 - Está en Matlab/Octave
 - También hay una implementación en Python

BSS_EVAL_SOURCES

- Pueden descargar el artículo original y el código de la página del curso.
- También pueden descargar un script para probar `bss_eval_sources`.
 - Se llama *test_bss_eval_sources*.
- Descarguen ambos el código `bss_eval_sources` y el script.

TEST_BSS_EVAL_SOURCES

- El script crea dos señales originales (s_1 y s_2) tipo senoidales.
- De estas dos señales, crea dos (\hat{s}_{at1} y \hat{s}_{at2}).
 - Estas señales son mezclas de s_1 y s_2 .
 - \hat{s}_{at1} contiene la suma de s_1 y una proporción de s_2 (definido por amp_2)
 - \hat{s}_{at2} contiene la suma de s_2 y una proporción de s_1 (definido por amp_1)
- La Figura 1 muestra a s_1 y \hat{s}_{at1} .
- La Figura 2 muestra a s_2 y \hat{s}_{at2} .
- Al final se estima el SDR, SIR y SAR de ambos \hat{s}_{at1} y \hat{s}_{at2}

BSS_EVAL_SOURCES

- La función de `bss_eval_sources` recibe dos argumentos:
 - Una matriz con cada renglón una señal **estimada**.
 - Una matriz con cada renglón una señal **original**.
 - No tienen que estar en orden.
- Regresa una serie de arreglos, uno por cada métrica.
 - Están en decibeles (dB).
 - Por ejemplo, en el arreglo del SIR, su primer valor es el SIR de la primera señal estimada; su segundo valor es el SIR de la segunda señal estimada; etc.

BSS_EVAL_SOURCES

- También regresa un arreglo adicional llamado “perm”, que presenta una asociación de cada señal estimada a una señal original.
 - Por ejemplo, si el primer valor de perm es 3, significa que la primera señal estimada se parece más a la tercera señal original.

Escuelas de Pensamiento

- Beamforming.
- Análisis Estadístico.

Siguiente Tema:
Beamforming