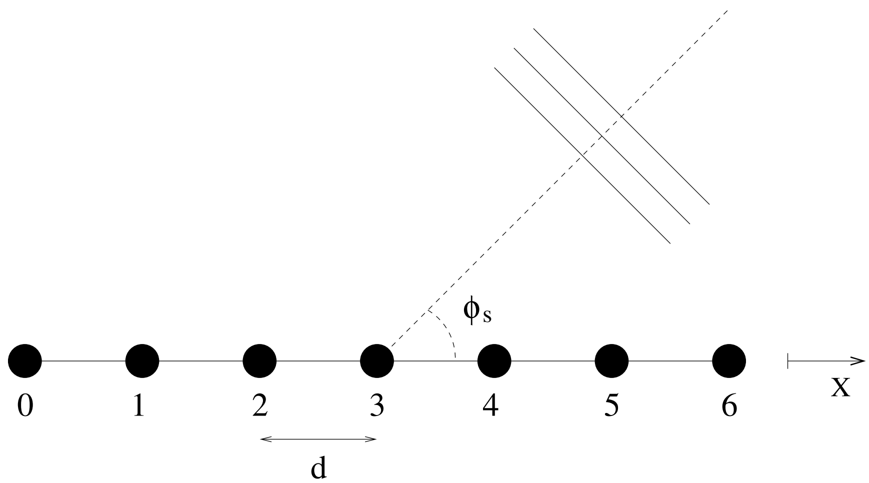
PRÁCTICA 5: BEAMFORMER DELAY-AND-SUM PARA REALCE DE VOZ EN AMBIENTE RUIDOSO

1. Objetivo y características del sistema

Se pretende desarrollar un beamformer delay-and-sum con el objetivo de realzar una señal de voz capturada por un array lineal uniforme y procedente de una dirección espacial dada por el ángulo de incidencia ϕ_s según se indica en la figura. La señal de voz está contaminada por una interferencia procedente de otra dirección espacial (desconocida). La implementación se realizará en el dominio de la frecuencia mediante la transformada STFT.



El sistema está caracterizado por los siguientes parámetros:

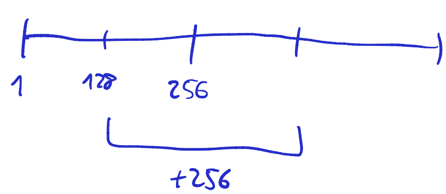
1. Frecuencia de muestreo: $F_s = 16000$ Hz.
2. Dirección de procedencia de la voz: $\phi_s = \pi/4$.
3. Velocidad de propagación del sonido: $c = 340$ m/s.
4. Número de elementos del array: $N = 7$.
5. Separación entre elementos del array: $d = 0{,}04$ m.

Señales recibidas por el array: cargarlas desde el fichero `signals_array.mat`. Las 7 señales están disponibles en variables del tipo `xc{n}` (n=1,...7). Además, la fuente original está en la variable `xorg16`.

Ejercicios:

1. Obtener la señal de voz realzada por beamforming y obtener los valores de SNR antes y después del beamforming. NOTA: para el cómputo de la SNR, puede considerarse que las 3000 primeras muestras de señal (voz en silencio) corresponden únicamente a la interferencia.
2. Dibujar el patrón de directividad polar del beamformer para frecuencias $f = 100, 400, 700, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000$ Hz. Explicar el comportamiento del array en función de la frecuencia.

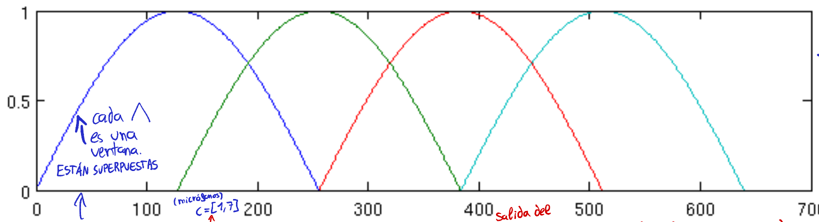
Dist $\geq \lambda/2$
para que el
beamformer
funcione bien.
(justifica con la teoría de eso)



2. Implementación en el dominio de la frecuencia mediante STFT

La implementación de algoritmos de beamforming sobre señales de banda ancha requiere el uso de una transformación al dominio de la frecuencia que normalmente se realizará mediante la transformada de Fourier de tiempo corto (short-time Fourier transform, STFT) sobre las señales discretas $x_n(m)$ de cada uno de los canales ($n = 1, \dots, N = 7$). La STFT consiste en ir tomando tramas de señal de $L = 256$ muestras, con un desplazamiento entre tramas de $L_{shift} = L/2$, aplicar una ventana de análisis $w(m)$ a cada una de estas tramas, y obtener finalmente la transformada discreta de Fourier (FFT) de las tramas enventanadas. El proceso de segmentación en tramas y enventanado se ilustra en la siguiente figura:

La ventana se reparte entre una transformada y su inversa.
Así, al reconstruir, la señal se reconstruye bien.
Media ventana es $\sqrt{\text{ventana Hanning}}$
La ventana cumple → $w^2(n) + w^2(n-L) = 1$



cada ∧ es una ventana.
ESTÁN SUPERPUESTAS

→ tramas de $L = 256$ // Ventana de Hanning

Le haces FFT $[\sqrt{w_{HAN}} * x_n^{(c)}]$ (micrófonos) $c = [1,7]$; 256 muestras cada → PROCESADO → IFFT $[X_{out}(k)] \cdot \sqrt{w_{HANN}}$ (salida del beamformer) ⇒ Esto constituye un 'sistema de reconstrucción perfecta' (A la salida se suman todas)

El resultado de este proceso es que cada trama queda representada en el dominio de la frecuencia mediante su FFT que notaremos como $X_n(k)$, donde el índice k representa a la frecuencia $\omega_k = 2\pi(k-1)/L$ ($k = 1, \dots, L$). Siguiendo una aproximación de *banda estrecha*, el beamforming se aplicará a cada frecuencia con significado físico ($k = 1, \dots, L/2+1$) de forma independiente, de acuerdo con las expresiones vistas en teoría (NOTA: obsérvese que la aplicación de la ventana implica la aparición de leakage, por lo que la FFT no puede mostrar frecuencias totalmente independientes).

Una vez realizado el beamforming, la señal debe devolverse al dominio temporal mediante la IFFT. Sin embargo el procesado realizado no garantiza que la señal resultante en el dominio del tiempo sea real. Para garantizar esto debemos tener en cuenta lo siguiente:

ojo comprobar esto!

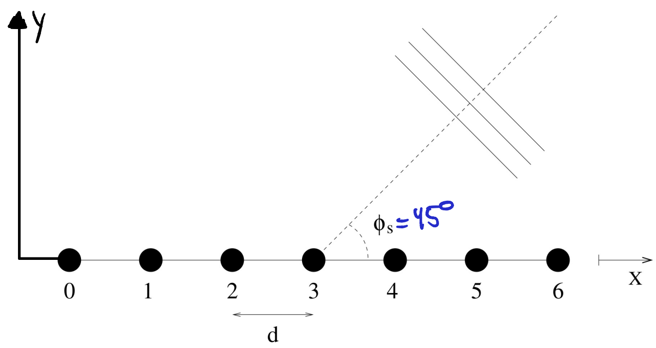
1. El espectro de una señal real es conjugadamente simétrico. Por tanto, la señal realzada en el dominio frecuencial debe simetrizarse antes de su paso al dominio del tiempo para garantizar que la señal resultante sea real. Para ello, debemos forzar que el espectro realzado $X_{out}(k)$ cumple que $X_{out}(L/2+2:L) = X^*_{out}(L/2:-1:2)$. Esta simetrización también justifica que el beamforming solo se aplique a la primera mitad del espectro ($k = 1, \dots, L/2+1$) (lo que adicionalmente redunda en un ahorro computacional).

2. El espectro de una señal real debe ser real en la componente $\omega = \pi$ ($k = L/2+1$). Para asegurar ésto, debemos quedarnos únicamente con la parte real de la señal de salida $x_{out}(n)$. Este procesado solo modifica la componente $\omega = \pi$ (que suele ser casi nula o haber sido eliminada por el filtrado anti-aliasing).

Para la síntesis de la señal completa se sigue un procedimiento *overlap-add* mediante el que a cada trama t enventanada $w(m)x_{out,t}(m)$ de la señal de salida se le suma la trama siguiente $t+1$desplazada $w(m+L/2)x_{out,t+1}(m+L/2)$, siguiendo un esquema similar al mostrado en la figura anterior. El uso de una ventana acampanada asegura una transición suave de una trama a otra.

Para que el resultado final sea coherente (reconstrucción perfecta) la ventana de análisis/síntesis debe cumplir que $w^2(m) + w^2(m+L/2) = 1$ en la zona de solapamiento. Para asegurar reconstrucción perfecta se usará una ventana de Hanning $w_H(m)$ (comando `hanning`, con la opcion `'periodic'`) de forma que $w(m) = \sqrt{w_H(m)}$, ya que $w_H(m) + w_H(m+L/2) = 1$.

NOTAS:



Tenemos las 7 señales de 7 micrófonos

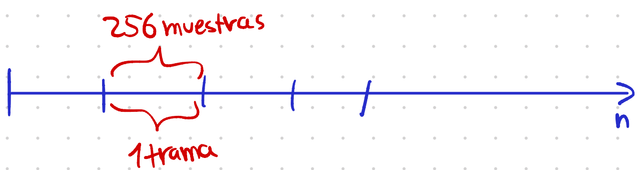
Hay un pitido que viene de una dirección desconocida y distinta que el target.

$f_s = 16 kHz$

dist = 4cm

$V_{prop} = 340 m/s$

Vamos a dividir la señal en tramas.



Pseudocódigo:

$N_{tramas} = L/128$ → for ntram = 1 : N_{Tramas}

for c = 1:7

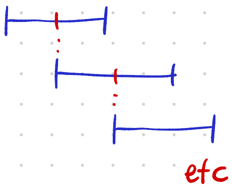
FFT[· ·]

Aplicamos el peso del B.F : $X_{out}^{(c)}(f)$

$$X'_{out} = X_{out} + X_{out}^{(c)}$$

$IFFT \longrightarrow X_{out}(n)$ ⇒ Esto te da una trama

Así montamos la señal: (overlap add se llama lo que estamos haciendo)

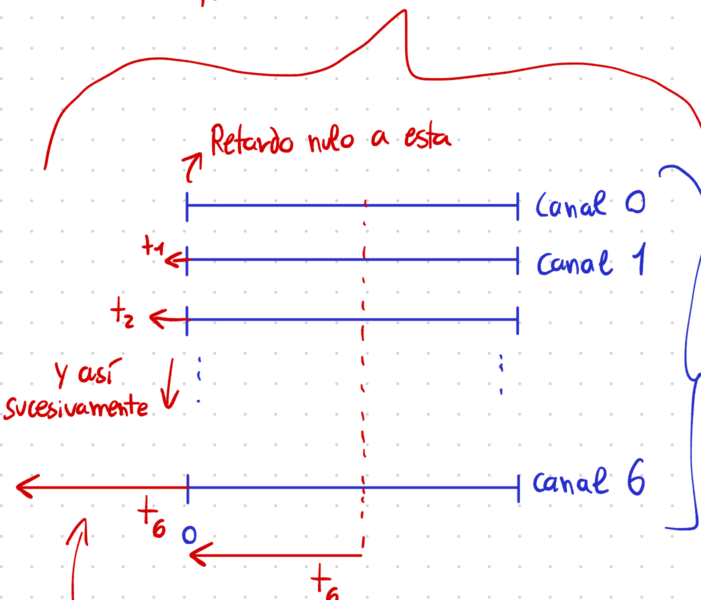


EN CUANTO AL BEAMFORMER:

$$W_c(f) = \frac{1}{N} e^{j 2\pi f t_n}$$

el retardo de cada canal

MI PUTO CASO



ojo con este retardo estarías antes de 0, por eso tienes que poner el 0 en el instante t_6, para que cuadre.

fórmula de retardos

$$t_n = \frac{n d \cos\phi_s}{c}$$

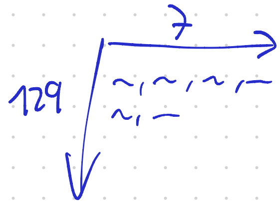
Hacer matriz de frecuencias y retardos

p.ej. pruebas con muchas frecuencias.

Poner el beamformer en una matriz, para $\forall$ de f y n

$f = 0, \ldots, 128$

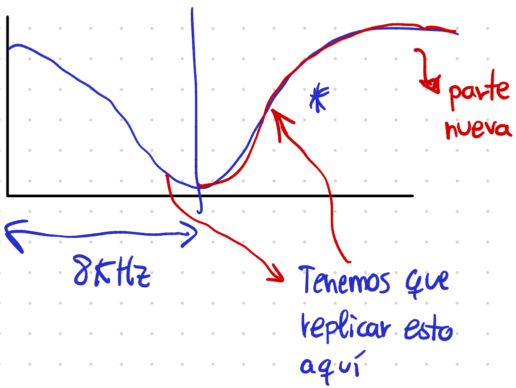
$\hookrightarrow F_s/2 \quad (L/2 + 1)$



$x_n(n)$

$x_{n}(n)$

Para la IFFT (mirar guión)



¡OJO!

Comprobar que la mejor directividad se da para $\lambda/2$ (mira la teoría)

Cuando hagos la IFFT, metes un Re[ ],

y quitas la parte imaginaria

Notas propias

Para cada iteración