Práctica 3 - TDS

Más representaciones tiempo-frecuencia: el espectrograma y el espectrograma Mel

Voice Activity Detector (VAD)

Práctica 3

OBJETIVOS

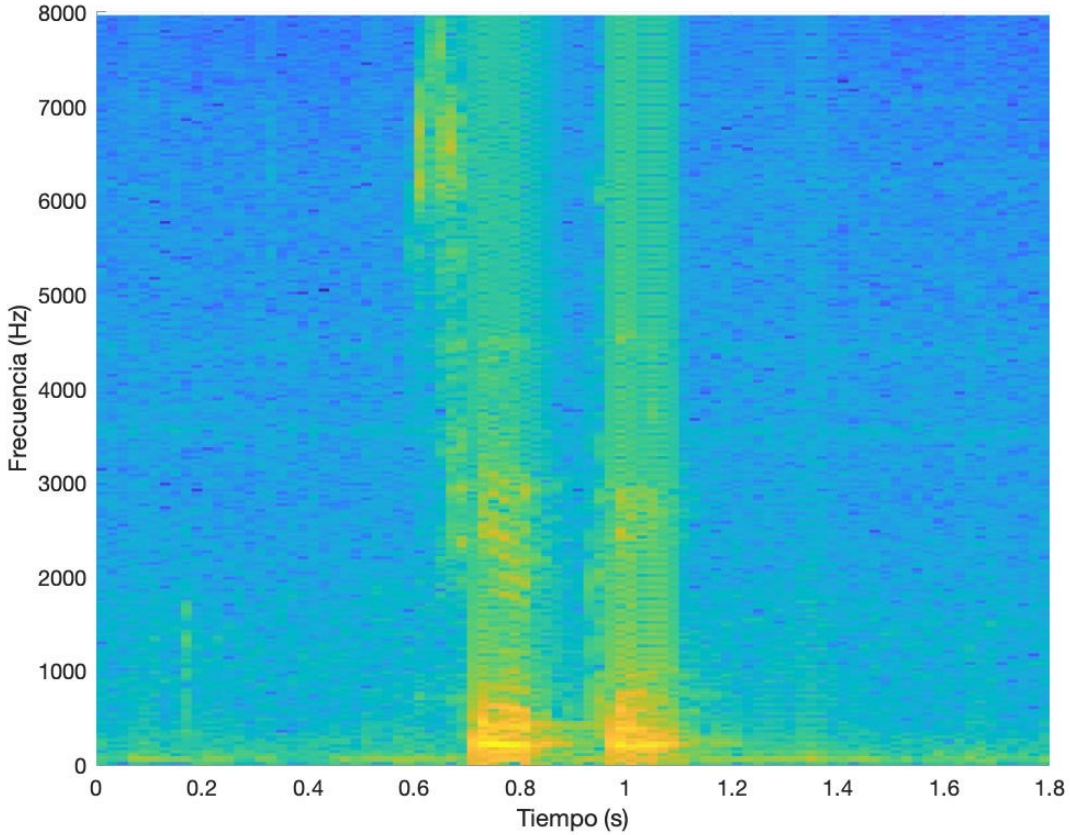
1. Veremos otras formas de visualizar el espectro tiempo-frecuencia (TF) muy habituales: el espectrograma y el espectrograma de Mel.
2. Funcionamiento de un detector de actividad vocal (*voice activity detector*, VAD)
3. Para casa: Acabar las grabaciones de los dígitos.

Importante:

- Se proporciona el fichero mi_practica3.mlx donde escribir el código de Matlab utilizado para cada ejercicio.
- Este **fichero se debe entregar antes de la siguiente sesión junto con su impresión digital en pdf** (se puede hacer desde Matlab). Se valorará que el fichero tenga comentarios explicativos.
- En las siguientes transparencias, el texto de color marrón indica que son variables o instrucciones para usar en Matlab

Ejercicio 3.1

- Carga el fichero siete.wav asignándole a la variable el nombre siete.
- Mediante instrucciones similares a las utilizadas en el Ejercicio 2.5 de la Práctica 2, usa la función surf para visualizar el espectro tiempo-frecuencia usando tramas de 20ms y un tamaño de FFT de 512 muestras. Al usar la función surf, el valor de la propiedad 'EdgeColor' debe ser 'none'
- Ejemplo de uso: surf(ejet,ejef,20*log10(??),**'EdgeColor','none'**)



Ejercicio 3.2 – el espectrograma

- En este ejercicio vamos a aprender a utilizar la función propia de Matlab para visualizar el espectro clásico tiempo-frecuencia de una señal, lo que se conoce como **espectrograma**. Se calcula mediante la función:

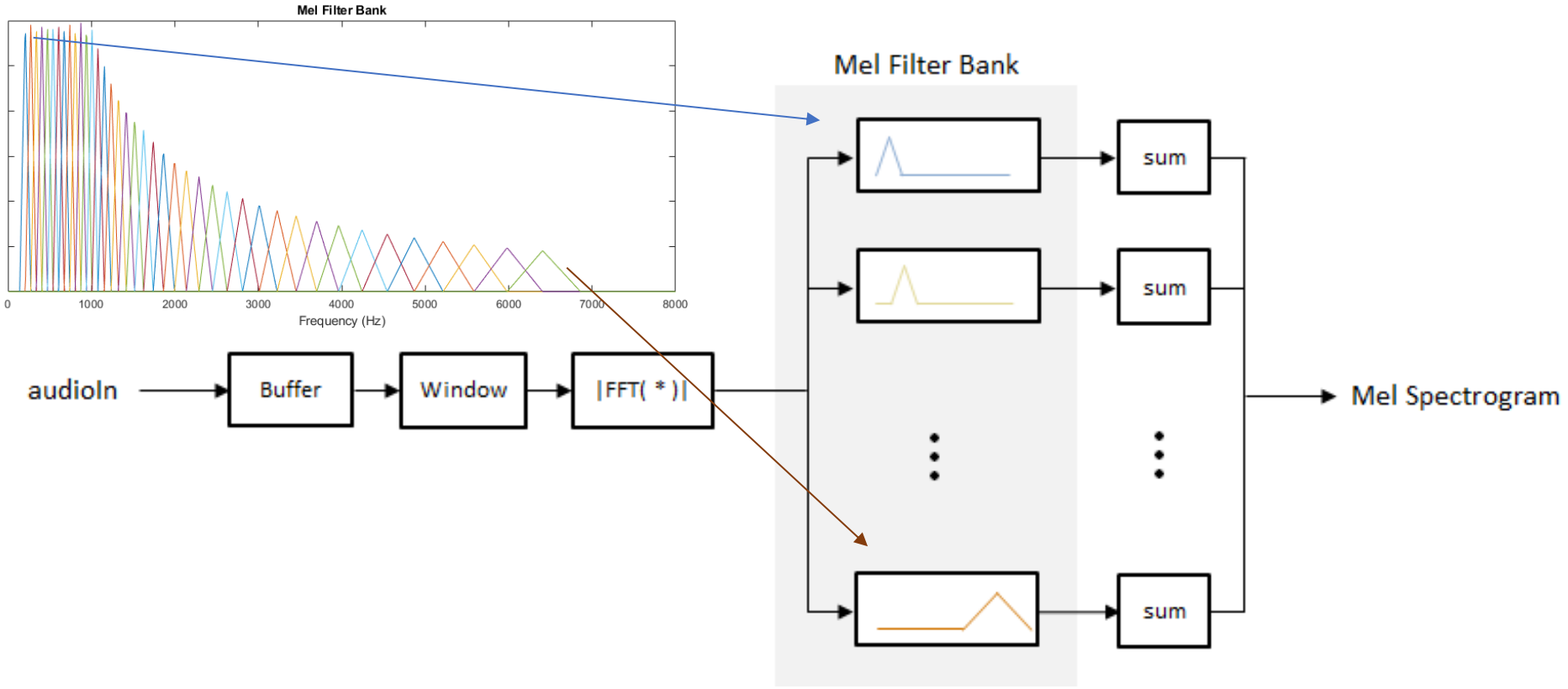
`spectrogram(x,window,noverlap,nfft,fs)`

- `x` es la señal (vector) a analizar.
- `window` es una función que permite visualizar mejor el espectro. Veremos las distintas ventanas en el Tema 2. Por ahora, usaremos la opción de introducir aquí un número. En este caso el valor de `window` es el número de muestras de la trama, lo que denominamos L_frame.
- `noverlap` es el número de muestras de solape entre tramas. Si se deja vacío (usando [] en la tercera posición de los argumentos de entrada) por defecto coge el 50%.
- `nfft` es el tamaño de la FFT y debe ser mayor o igual que el valor usado en window. Para que la computación sea eficiente, nfft debe ser una potencia de 2.
- `fs` es la frecuencia de muestreo.
- Mira los **ejemplos de uso la función spectrogram** antes de utilizarla tecleando "doc spectrogram" en la Command Window de Matlab. La función usada sin argumentos de salida da como resultado una nueva figura. Mira la documentación para ver ejemplos de espectrogramas de distintas señales cuyo contenido frecuencial varía en el tiempo como le pasa a la voz.
- Establece los parámetros del espectrograma para que use tramas de 20ms con un solape del 50% y con un valor de nfft de 512. Dibuja el **espectrograma** usando la opción 'yaxis':

spectrogram(x,window,noverlap,nfft,fs,'yaxis')

El espectrograma Mel

- Por último, vamos a visualizar el espectro tiempo-frecuencia de una señal mediante el **espectrograma Mel**.
- La diferencia con el **espectrograma** clásico es que agrupa la energía por bandas de frecuencias mediante un banco de filtros. Lo curioso es que ese banco de filtros actúa de una forma muy parecida a como procesa el sonido nuestro oído y la distribución de energía que obtiene es más "perceptual".



Ejercicio 3.3

- La función propia de Matlab para visualizar el **espectrograma Mel** de una señal es:

```
melSpectrogram(x,fs)
```

- `x` es la señal (vector) a analizar.
- `fs` es la frecuencia de muestreo.
- Esta función tiene además los mismos parámetros que la función spectrogram en cuanto a tamaño de la ventana, etc., pero se introducen como propiedades en lugar de definirlos como argumentos de entrada. Por ejemplo:

```
melSpectrogram(audioIn,fs,'Window',512,'OverlapLength',256,'NumBands',40);
```

- significa que la duración de la trama (`Window`) es 512, que el solape (`OverlapLength`) es de 256 muestras y por tanto abarca el 50% de la trama, y que el número de bandas o el número de filtros (`NumBands`) que usa el espectrograma Mel es 40.
- Establece los parámetros del espectrograma Mel para que use tramas de 20ms con un solape del 50%, un tamaño de la FFT de 512 y usando 40 filtros. Dibújalo.

Ejercicio 3.4 - Detector de actividad vocal (VAD)

- El *voice activity detector* (VAD) es un algoritmo que decide si la señal que está grabando el micrófono es la señal de voz del locutor o es otra cosa (habitualmente, el ruido de fondo)
- El VAD es un sistema que llevan todos los teléfonos móviles, pero que también utilizan los sistemas de videoconferencia, así como los reconocedores de voz tipo Alexa. Esto permite:
  - Ahorrar el envío de datos en un sistema de comunicaciones móviles o de video-conferencia cuando el locutor no está hablando
  - Ahorrar el esfuerzo de poner en marcha el reconocedor en los asistentes de voz (Alexa, Google nest home, Cortana) cuando lo que capta el micrófono no es voz.
- Un VAD básico es un sistema que va analizando la energía de la señal que graba, y cuando la energía supera un cierto umbral, entonces decide que lo que está captando el micrófono es voz.
- **En este ejercicio, hay que ejecutar lo que se va pidiendo en el Ejercicio 3.4 del MLX**

MUY IMPORTANTE

- Recuerda que debes realizar **20 grabaciones (a 4-5 personas distintas, no grabes más de 5 veces a la misma persona)** de cada dígito. No grabes a compañeros de clase para asegurar que no repetimos personas.
- Intenta realizar las grabaciones de los dígitos a alguna persona de tu entorno. Recuerda:
  - Procura hacer las grabaciones en un **entorno lo más silencioso posible**.
  - La **frecuencia de muestreo** debe ser $f_s = 16$ kHz. Deben ser monocanal, no sirven señales estéreo.
  - El locutor debe dejar una pausa de al menos 1 segundo entre número y número para poder segmentarlos bien.
  - Comprueba que el rango de valores de las muestras grabadas está comprendido en el rango (-1,1) y que no ha saturado dichos valores en ningún momento. Si no es así, aleja al locutor del micrófono y vuelve a grabar.
  - Por el contrario, si ves que los valores de las muestras no pasan de 0.1-0.2 en el eje de ordenadas, acerca el micrófono al locutor y vuelve a grabar hasta que alcancen al menos 0.5 (ó -0.5) en su máxima amplitud.
- Guárdalas en ficheros distintos de la siguiente forma: Asigna un número a cada persona y otro número a cada una de sus grabaciones. Utiliza el siguiente formato:

  etiqueta_#locutor_#grabacion.wav

- Por ejemplo, el fichero **cero_3_2.wav** correspondería al dígito ‘cero’ que ha dicho el locutor número 3 en la 2ª grabación de dicho locutor.

Entrega de las grabaciones

- **Recuerda que antes del 14 de noviembre**, debes haber realizado **20 grabaciones (a 4-5 personas distintas, no grabes más de 5 veces a la misma persona)** de cada dígito y haberlos subido a la carpeta compartida de OneDrive:

DATABASE_2022_23

(Si no funciona el enlace, la url es: https://upvedues-my.sharepoint.com/:f:/g/personal/gpinyero_upv_edu_es/EhzANwsjYplGnYPlvUXerGwBxAWk41OKPm7lWVKKKgSCaA?e=4cmZdx)

- Hay un vídeo en PoliformaT -> Recursos -> Prácticas que muestra cómo debes subir los ficheros a la carpeta