

PRÁCTICA 05: Captura de audio y detección del inicio-fin de una palabra.			
Apellidos, nombre		Fecha	

Implementar las siguientes funciones para:

1. Capturar la señal de audio:

```
señal = grabacion (t, Fs, Ch)
```

donde:

- *t* es el tiempo (en segundos) de grabación.
- *Fs* corresponde a la frecuencia de muestreo.
- *Ch* toma el valor 1 o 2 si trata de una grabación mono o estéreo.

Usa la función `wavrecord()` y como tipo de dato usa 'double'.

2. Implementar el filtro de preénfasis:

```
señal = preenfasis (a)
```

donde:

- *a* es el parámetro que refleja el grado de preénfasis (suele estar en el rango 0.9 a 1.0).

3. Realizar la segmentación y el enventanado de la señal de audio:

```
segmentos = segmentacion (señal,num_muestras,despl)
segmentos_enventanados = enventanado(segmentos,ventana)
```

donde:

- *num_muestras* corresponde al número de muestras por segmento.
- *despl* es el número de muestras desplazadas entre segmentos.
- *ventana* es una cadena de caracteres que indica el tipo de ventana a aplicar. Puede tomar alguno de los siguientes valores: rectangular, haming o hanning.

4. Calcular los valores de energía, magnitud y tasa de cruces por cero de los segmentos:

```
salida = energia (señal,num_muestras,despl,ventana)
salida = magnitud (señal,num_muestras,despl,ventana)
salida = cruces_por_cero (señal,num_muestras,despl,ventana)
```

5. Detectar el inicio-final de una palabra (basado en el algoritmo de Rabiner-Sambur):

Una forma sencilla de eliminar de la señal los tramos sordos y quedarnos sólo con los sonoros es calcular la energía promedio de cada segmento y la de la señal completa. Si la energía promedio del segmento es mayor que la energía de la señal completa ponderada por un valor umbral, se dice que el segmento es sonoro.

En este caso se va a implementar un método más elaborado para detectar el inicio y fin de una palabra basado en el algoritmo propuesto por Rabiner-Sambur. Este método considera las características de los sonidos:

- Sonidos sonoros: Tienen un alto contenido en energía y ocupan las frecuencias bajas del espectro de la voz.
- Sonidos no sonoros: Con un bajo contenido en energía y ocupan las frecuencias superiores del espectro de la voz.

Para la detección del inicio de la palabra realizamos la siguiente secuencia de pasos:

1. Por cada trama de 128 muestras, calcular las funciones: cruce por ceros $\{ Z[n] \}$ y la magnitud promedio de la señal $\{ M[n] \}$, estas funciones se definen a continuación:

$$M_n = \sum_{m=0}^{N-1} |x[m]|$$

$$Z_n = \frac{\sum_{m=0}^{N-2} |\text{sign}(x[m+1]) - \text{sign}(x[m])|}{2N}$$

2. Para obtener las estadísticas del ruido ambiental, se considera que las primeras diez ventanas son ruido, con lo cual se tiene:

$$Ms_n = \{M_1, M_2, \dots, M_{10}\}$$

$$Zs_n = \{Z_1, Z_2, \dots, Z_{10}\}$$

3. Calcular la media y la desviación estándar para las características del ruido y obtener los siguientes umbrales:

Umbral	Nombre del umbral	Valor
UmbSupEnrg	Umbral Superior de Energía	$0.5 * \max\{M_n\}$
UmbInfEnrg	Umbral Inferior de Energía	$\mu_{Ms} + 2 * \sigma_{Ms}$
UmbCruCero	Umbral de cruces por cero	$\mu_{Zs} + 2 * \sigma_{Zs}$

4. Recorrer la función M_n incrementando en una unidad a n de 11 hasta que $M_n > \text{UmbSupEnrg}$. En este punto estamos garantizando presencia de señal. A este punto lo marcaremos como ln .
5. Resulta lógico pensar que el inicio de la señal se encuentra en algún punto anterior a ln , por lo que ahora recorreremos la función M_n desde $n = ln$ hasta que $M_n < \text{UmbInfEnrg}$. Este punto lo marcaremos como le y lo reconocemos tentativamente como el inicio de la señal, determinado por la función de magnitud.
6. Ahora decrementamos n desde $n = le$ hasta $n = le - 25$ o en su defecto $n = 11$, verificando si sucede alguna de las siguientes condiciones en la función de cruces por cero, ya que lo que ahora buscamos es la posibilidad de que un sonido *no sonoro* preceda a un sonido *sonoro*.
 - Si $\{ Z_n < \text{UmbCruCero} \}$ significa que no encontramos alguna porción de la señal con aumento importante de frecuencia en 25 ventanas anteriores, por lo tanto el inicio es le .
 - Si encontramos que $\{ Z_n > \text{UmbCruCero} \}$ menos de tres veces seguidas significa que solo fue una espiga de ruido, el punto de inicio sigue siendo le .
 - Si encontramos que $\{ Z_n > \text{UmbCruCero} \}$ al menos tres veces seguidas hemos encontrado un sonido *no sonoro*, entonces buscamos el punto n para el cual $\{ Z_n > \text{UmbCruCero} \}$ la primera de las más de tres veces, es decir, el punto para el cual la función Z_n sobrepasa el umbral, indicando el comienzo del sonido *no sonoro* y desplazamos el inicio de la palabra a lz .

Para la detección del fin de palabra, hacer la misma secuencia de pasos (a partir del punto 4), pero en sentido inverso.

Implementar una función en MATLAB para obtener la señal recortada con la siguiente sintaxis:

```
señal_recortada = inicio_fin (señal, num_muestras, despl)
```