

# TB065-86.05 - SEÑALES Y SISTEMAS

## TRABAJO PRÁCTICO ESPECIAL 1: ANÁLISIS DE LA SEÑAL DE HABLA

🐍 = Python.

La señal de habla está compuesta de secciones de propiedades cambiantes. Si pensamos a la señal de habla como la salida de un sistema, podemos atribuir dichas variaciones a dos causas: cambios en la excitación o cambios en la configuración del tracto vocal, es decir en el sistema. Si la entrada se comporta como un tren de impulsos cuasi-periódicos, la salida será uno de los posibles sonidos vocálicos (/a/, /e/, /i/, /o/, /u/, /m/, /n/, /l/). Si la entrada en cambio es un generador de ruido blanco, el sonido obtenido será un fonema fricativo (/s/, /f/, /sh/). La distinción entre los fonemas de la misma clase se produce por la forma que va tomando el tracto vocal para cada uno de ellos. La variación de la transferencia del sistema se supone que es suficientemente lenta como para considerar que la señal de habla es la concatenación de porciones de señales que se originan como salida de un sistema LTI. Por esto aparecerán bien representados en un espectrograma. Los sonidos explosivos (/p/, /k/, /t/) en cambio tienen una naturaleza distinta, y son más parecidos a un transitorio que a un sonido estacionario. Un esquema del modelo de producción de la voz se muestra en la Figura 1.

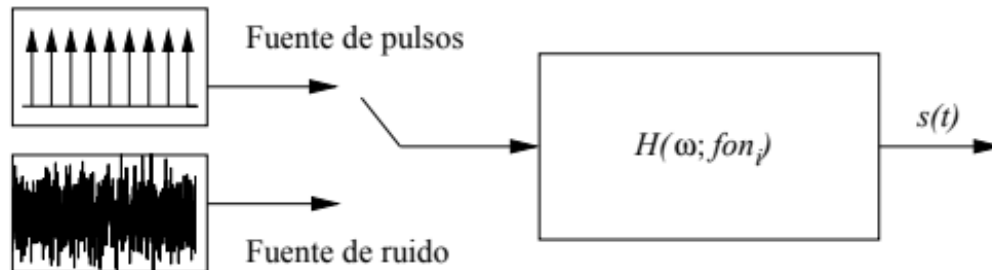


Figura 1: Modelo de producción de la voz

---

1. Grabar su propia voz diciendo la palabra “Mikasa”. Graficar en 🐍 la señal de voz ubicando en ella porciones de señales periódicas y no periódicas. Etiquetar los ejes adecuadamente.

---

2. Realizar una segmentación 🐍 de la señal localizando los segmentos de muestras donde aparece una [a] y una [s]. Describir que diferencias hay entre estos fonemas. Graficar en 🐍 el segmento [a] y estimar el periodo y la frecuencia de esta señal cuasiperiódica. Calcular lo mismo para la otra [a]

## TRABAJO PRÁCTICO ESPECIAL 2: SERIE DE FOURIER

Los sonidos sonoros son producidos forzando el aire a través de la glotis o a través las cuerdas vocales. La tensión de las cuerdas vocales se ajusta de manera tal que vibre en forma oscilatoria. La interrupción periódica del flujo de aire subglotal resulta en un soplido casi periódico de aire que excita el tracto vocal. El sonido producido por la laringe es llamado sonoro o con fonación. Este tipo de sonido consiste en una frecuencia fundamental ( $F_0$ ) y sus componentes armónicos producidos por las cuerdas vocales. El tracto vocal modifica esta señal de excitación que causa el formante. El término formante se utiliza para indicar el centro de estas frecuencias de resonancia, en la cual la concentración de energía es mayor. Los formantes son las frecuencias de resonancia del espectro, es decir, los picos de la envolvente del espectro de la señal de voz que representan las frecuencias de resonancia del tracto vocal. Cada formante tiene una amplitud y un ancho de banda. Las formantes en el espectro son usualmente llamadas  $F_1$ ,  $F_2$ ,  $F_3$ ,..., comenzando con la menor frecuencia. La frecuencia fundamental y frecuencias de formante son probablemente los conceptos más importantes en la producción y análisis del habla en general. Las frecuencias a las que se producen los primeros formantes son muy importantes para reconocer o sintetizar la voz. En la siguiente figura pueden verse representados los 3 primeros formantes de una señal de voz. 2.

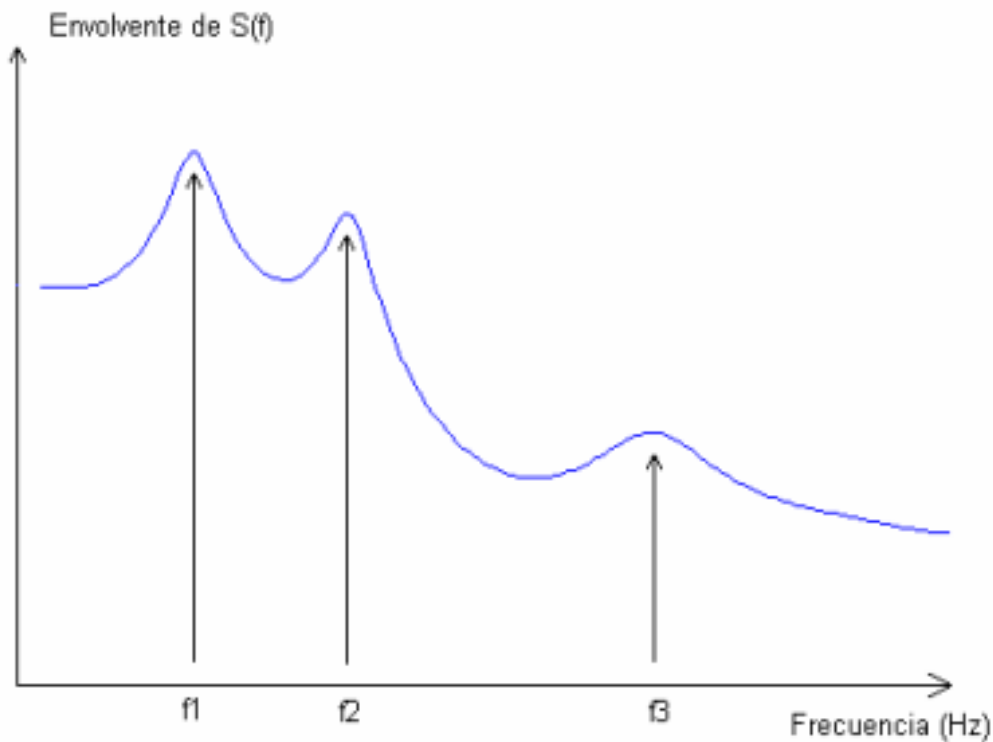



Figura 2: Envolvente del espectro de una vocal

- 
3. Estimar el periodo y la frecuencia fundamental de la otra vocal [i] presente en la señal.

---

4. Utilizando la FFT  grafique los coeficientes de Fourier de las porciones correspondientes a las vocales que hay en la señal. Hacer el cálculo tomando varios periodos de la vocal y tambien tomando un solo periodo. Identificar los primeros maximos en las envolventes de estos espectros y estimar los valores de frecuencia en que se producen.

# TRABAJO PRÁCTICO ESPECIAL 3: TRANSFORMADA DE CORTO TIEMPO

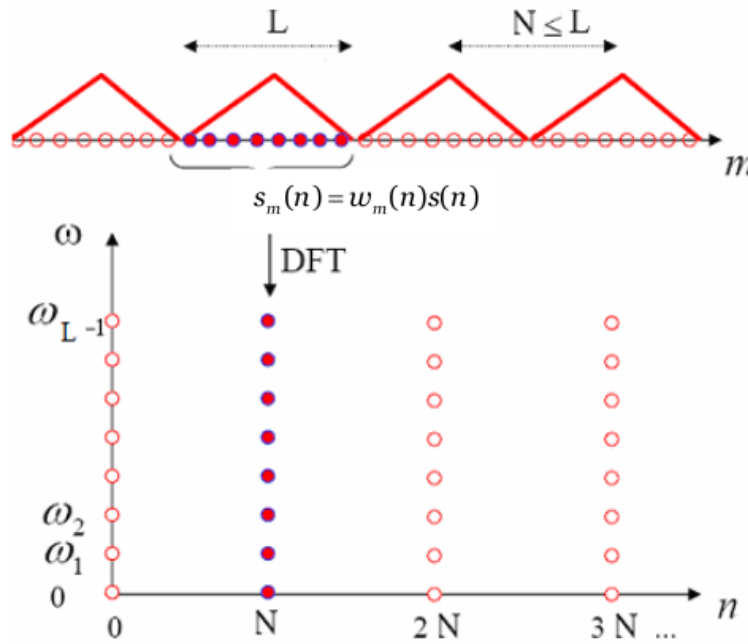
La STFT es una transformada de Fourier basada en la DFT. En la práctica, hay muchas aplicaciones en las que las propiedades de la señal que se trata, cambian muy rápidamente con el tiempo. Por ejemplo, esto sucede con señales no estacionarias tales como las de radar, sonar, voz y señales de comunicaciones. Pues bien, en estos casos calcular una única DFT para toda la señal no es suficiente, además de la dificultad añadida de que ésta podría ser larguísima siendo imposible de tratar en la práctica, ya que suelen usarse computadores digitales con una capacidad de cálculo y almacenamiento limitados. Todo ello nos guía hacia el concepto de transformada de Fourier de corta duración o STFT (Short-Time Fourier Transform). La STFT de una señal  $s(n)$  se define como:

$$S(n, w) = \sum_{m=-\infty}^{\infty} s(m)w(n-m)e^{-jwm}$$

Donde  $w(n)$  es la ventana. En la STFT, la secuencia unidimensional  $s(n)$ , función de una variable discreta, es transformada en una función bidimensional de la variable  $n$ , que es discreta, y de la frecuencia  $w$ , que es continua. Hay que ver de que la STFT es periódica en  $w$  con periodo  $2\pi$ , y por lo tanto sólo tendremos que considerar los valores incluidos en  $0 \leq w \leq 2\pi$ , o cualquier otro intervalo de longitud  $2\pi$ . Teniendo en cuenta la simetría de las ventanas, la ecuación anterior puede describirse como:

$$S(n, w) = \sum_{m=-\infty}^{\infty} s(m+n)w(m)e^{-jwm}$$

De esta forma, la STFT puede interpretarse como la transformada de Fourier de la señal desplazada  $s(m+n)$ , y vista a través de la ventana  $w(n)$ . La ventana tendría un origen fijo, y según  $n$  va cambiando, la señal se desliza pasando a través de la ventana de forma que para cada valor de  $n$  vemos una porción diferente de la señal.




El espectrograma es una herramienta muy útil para analizar los fonemas y sus transiciones. Un espectrograma de una señal en el tiempo es una representación especial en dos dimensiones, en el eje horizontal representa el tiempo y en el vertical representa la frecuencia. Normalmente se utiliza la escala de grises para indicar la energía en cada punto (t, f) representando con blanco las bajas energías y con negro las altas. El espectrograma se obtiene a partir de la STFT. El espectrograma solamente representa la energía y no la fase de la STFT. La energía la calculamos como:

$$\log |X(k)|^2 = \log (X_r^2(k) + X_i^2(k))$$


El valor de la ecuación anterior lo convertimos a escala de grises. Aquellos píxeles, cuyo valor no es calculado, se obtienen interpolando.

En Python podemos usar `scipy.signal.spectrogram` o `scipy.signal.ShortTimeFFT`

---

5. Graficar  el espectrograma de banda angosta de la palabra completa de forma que se pueda observar la frecuencia fundamental y sus armónicos.

---

6. Graficar  el espectrograma de banda ancha de la palabra completa de forma que se puedan observar los formantes de los sonidos vocálicos.



---

7. Graficar  el espectrograma de banda angosta y de banda ancha de las vocales presentes en la palabra.



## TRABAJO PRÁCTICO ESPECIAL 4: CAMBIOS DE VELOCIDAD DE LA SEÑAL DE HABLA

Hay muchas aplicaciones que requieren un cambio de la velocidad de la señal de habla. Un esquema simple que solo interpole o decime la señal de habla cambiaría la velocidad pero también alteraría las características fundamentales como la frecuencia glótica o la posición de los formantes. La señal alterada de esta manera podría incluso resultar ininteligible. Idealmente, lo que se requiere es variar la velocidad de la señal preservando sus características en frecuencia.

---

**8.** Decimar la señal Mikasa  de manera de aumentar al doble su velocidad quitando puntos en la señal temporal. Asegurarse de que no se produzca aliasing en el proceso (si es necesario utilizar un filtro antialiasing, que debe saber cómo construir con el método de ventaneo). Graficar  el espectrograma de forma que se puedan observar los cambios en sus características fundamentales.

---

**9.** Interpolarse la señal Mikasa  de manera de disminuir la mitad su velocidad con el método utilizado en la guía de interpolación. Graficar  el espectrograma de forma que se puedan observar los cambios en sus características fundamentales y analice.

El método anterior cambia la duración temporal de una señal de habla pero también desplaza los formantes y armónicos. Para reducir este efecto hay varios métodos. Uno de ellos (phase vocoder) trabaja con la TFCT (Transformada de Fourier de Corto Tiempo), que es la matriz de DFTs cuyo módulo se representa en un espectrograma:

1. **TFCT:** Se aplica la TFCT a la señal.
2. **Interpolación:** Se realiza una interpolación o remuestreo de las “filas” de la matriz de la TFCT, agregando o quitando columnas para ajustar a la duración deseada.
3. **Síntesis Temporal:** Se realiza la inversa de la TFCT (iTFCT). Esto puede consistir, por ejemplo, en hacer la iDFT por columnas, compensando el efecto de la ventana, y combinando los resultados.

Para más información sobre la transformada de corto tiempo, su inversa, el phase vocoder, y otros métodos para escalamiento temporal puede encontrar fácilmente múltiples referencias ().

Sólo se requerirá para este TP que lo apliquen usando una función de librería, como por ejemplo `librosa` (`librosa.phase_vocoder`) o alguna otra, que entienda la idea general del algoritmo, y que compare adecuadamente los resultados obtenidos y los efectos introducidos con respecto a los formantes, frecuencias fundamentales y armónicos, mediante espectrogramas,

FFT, y escuchando las señales.

---

**10.** Aumentar al doble la velocidad de la señal Mikasa  $\Rightarrow$  por modificación de su transformada de corto tiempo TFCT (phase vocoder). Eliminar una de cada 2 columnas y recuperar la señal temporal utilizando la iTFCT. Graficar  $\Rightarrow$  el espectrograma de forma que se puedan observar sus características fundamentales y analice.

---

**11.** Reducir a la mitad la velocidad de la señal Mikasa  $\Rightarrow$  de manera de disminuir a la mitad su velocidad por modificación de su TFCT intercalando columnas. Calcular los valores en las columnas intercaladas mediante interpolación de las columnas originales. Graficar  $\Rightarrow$  el espectrograma de forma que se puedan observar sus características fundamentales y analice.

---

**12.** (Optativo) Aplique algún método que modifique la señal en el tiempo para duplicar o reducir a la mitad la velocidad de la señal, como los metodos PSOLA para cambiar la duracion de la señal.