

Apellidos:	Nombre:
Apellidos:	Nombre:
Apellidos:	Nombre:
Apellidos:	Nombre:

PRÁCTICA 3

ANÁLISIS ESPECTRAL USANDO LA DFT

- Al finalizar la práctica debes entregar las respuestas de este cuestionario en el formulario web que se proporcionará al principio de la sesión.
- Justifica adecuadamente tus cálculos y razonamientos.
- Usa al menos cuatro cifras significativas en tus operaciones

Ejercicio 3.1 Características espectrales de una vocal

Cuando se analiza en la frecuencia una señal de voz real debe hacerse a tramos, puesto que los sonidos cambian rápidamente y cada uno de ellos puede tener un espectro muy diferente.

Inicia Matlab y abre el programa P3_1. Este programa importa el fichero *Entrada.wav*, creando en tu espacio de trabajo una variable *Fs* correspondiente a la frecuencia de muestreo en Hz y un vector *x* correspondiente a los datos de sonido. A continuación, el programa P3_1 implementa un análisis espectral básico: recorta la señal *x* seleccionando *M* muestras y toma la DFT de *N* puntos de las mismas. Inicialmente el programa toma *M* = 1000 muestras de la vocal “a” y *N* = 15000 puntos de su DFT.

- Q3.1** Ejecuta el programa P3_1. En la figura 1 aparece dibujada la señal recortada, que como observarás es casi periódica. A partir de la gráfica, obtén el período de esta señal. (Puede serte útil el botón *Data Cursor* de la barra de herramientas, el que es un rectángulo amarillo con un signo +). Promedia varios períodos para mejorar la precisión de tus cálculos. Indica los pasos realizados, no solo el resultado final.

Busco los picos que comenzarían cada repetición y busco el *x* de los 4 primeros picos (4, 116, 229, 340). Restándoles el del pico anterior obtenemos 3 periodos (112, 113, 111). Hacemos la media de los 3, da un periodo de 112.

- Q3.2** Obtén el período de la señal analógica correspondiente, en segundos. Para ello, usa la frecuencia de muestreo indicada por Matlab al importar el fichero *Entrada.wav*.

Siendo $F_s = 10000\text{Hz}$ $T = 112/10000 = 0.0112\text{s}$

- Q3.3** En la figura 2 se representa el espectro de magnitud de la señal. Verás que aparecen dos zonas relevantes al principio y al final de la figura, siendo el resto despreciable. Esas dos zonas son simétricas. ¿Por qué?

Son simétricas porque es una DFT de una señal que representa una señal de una magnitud real en el tiempo.

Debido a la simetría mencionada, podemos centrarnos únicamente en el principio de la figura, concretamente en las primeras 2000 muestras, aproximadamente, que es donde la magnitud adquiere valores más relevantes. Para ver mejor lo que sucede, puedes realizar un zoom de esa zona: pulsa la lupa con el signo + de la barra de herramientas de la figura (*Zoom In*) y dibuja un rectángulo que incluya la región de interés. Repite la operación varias veces si es necesario (si lo precisas también puedes usar la herramienta *Zoom Out*, la de la lupa con el signo -).

Comprobarás de este modo que aparecen varios picos, los cuales se corresponden con componentes aproximadamente sinusoidales de la voz denominadas *armónicos* o, en este contexto en particular, *formantes*. (Los formantes son picos que deben destacar claramente, no simplemente un máximo relativo cualquiera). El primer armónico (primer pico empezando por la izquierda), se denomina *armónico fundamental* o, en el ámbito concreto de la voz, *pitch*. Es el parámetro que controla principalmente lo que conocemos como el tono: a mayor frecuencia del *pitch*, más aguda será la voz. (Evidentemente, una persona puede modificar ese tono, como por ejemplo cuando canta, pero también incluso en el habla natural lo hace continuamente).

Q3.4 Encuentra el valor del índice k correspondiente al *pitch*. (Puedes ayudarte de nuevo del botón *Data Cursor*).
 $k=133$

Q3.5 Calcula la frecuencia analógica en Hz correspondiente a dicho *pitch*. Indica los pasos de tu cálculo, no solo el resultado final.

$$F_k = k \cdot F_s / N. \text{ Siendo } k=133, F_s=10000 \text{ y } N=15000. \text{ Por tanto } F_k=88.6667\text{Hz}$$


Q3.6 Determina exactamente los índices k del resto de formantes, hasta $k = 2000$.
 $\{267, 404, 535, 672, 802, 939, 1070, 1205, 1335, 1471, 1606, 1741, 1872\}$

Q3.7 Deberías haber obtenido que todos los formantes están aproximadamente en lugares múltiplos del fundamental. ¿Es así?
 Si lo es, ya que si dividimos esos números entre 133, nos da:
 $\{2.0075, 3.0376, 4.0226, 5.0526, 6.0301, 7.0602, 8.0451, 9.0602, 10.0376, 11.0602, 12.0752, 13.0902, 14.0752\}$.

Ejercicio 3.2 Efecto de la longitud de la señal en el tiempo

Q3.8 Toma ahora $M = 500$. No modifiques N . Observa que el aspecto de los formantes se ha ensanchado. Aún así, ¿se pueden seguir distinguiendo todos o casi todos ellos?
 Si, se pueden seguir distinguiendo todos casi todos.

Podemos reducir más el valor de M y aún seguir distinguiendo los formantes. Ello sucederá siempre y cuando el número de muestras M esté por encima de cierto valor, el cual depende de la separación $\Delta\omega$ entre los formantes.

Q3.9  Calcula dicha separación $\Delta\omega$. Promedia varios valores para mejorar la precisión de tus cálculos.

Para calcularlo, cogemos la n de cada formantes y restamos a cada uno la n del formante que esté inmediatamente a su izquierda. Quedando: $\{134, 137, 131, 137, 130, 137, 131, 135, 130, 136, 135, 135, 131\}$. Haciendo el promedio queda que la separación de los formantes es 133.7692. Esto serían muestras, para la separación de los formantes hay que pasarlo a rad/s, para ello $133.7692 \cdot 2 \cdot \pi / N$. Siendo $N=15000$. Por tanto queda 0.0560rad/s .

Q3.10 ¿A qué valor de frecuencia analógica en Hz se corresponde el valor anterior? Indica los pasos de tu cálculo, no solo el resultado final.

$$F_k = k \cdot F_s / N. \text{ Siendo } k=133.7692, F_s=10000 \text{ y } N=15000. \text{ Por tanto } F_k=89.1795\text{Hz}$$

- Q3.11** Teóricamente, el valor mínimo de M que permite distinguir los formantes en el caso de señales puramente sinusoidales es $M = 2\pi / \Delta\omega$. Calcúlalo.

$M = 2 * \pi / 0.056 = 112.1334$. Que acercandolo al entero más cercano que siga cumpliendo la condición de mínimo, $M = 113$.

- Q3.12** Ajusta el valor de M en el programa P3_1 para que sea una vez y media superior a lo obtenido en la cuestión Q3.11. (Por ejemplo, si has calculado $M = 120$ teóricamente, toma ahora $M = 180$. Ten en cuenta que estás trabajando con señales reales, no exactamente con componentes sinusoidales, con lo cual hay que dejar un cierto margen de error). ¿Qué valor de M obtienes? Observa la figura resultante. ¿Sigues pudiendo distinguir todos o casi todos los formantes?

$M = 169.5$. Se sigue pudiendo distinguir los formantes, aunque están menos definidos.

- Q3.13** Ajusta el valor de M para que sea como una vez y media inferior a lo obtenido en la cuestión Q3.11. (Por ejemplo, si has calculado $M = 120$ teóricamente, toma ahora $M = 80$). ¿Qué valor de M obtienes? Observa la figura resultante. ¿Sigues pudiendo distinguir todos o casi todos los formantes?

$M = 75.3333$. Ya no se pueden distinguir. Más de un formante han desaparecido.

- Q3.14** Compara los resultados de las cuestiones Q3.1 y Q3.11. ¿Encuentras alguna relación? Justifícalo.

Si, es casi el mismo valor el M mínimo y la N del pitch. Tiene sentido, porque si tenemos suficientes muestras para conseguir el pitch, podemos conseguir el resto de armónicos, cumpliendo la relación armónica que existe entre ellos.

Ejercicio 3.3 Efecto del número de puntos de la DFT

En este ejercicio toma $M = 180$. Veamos ahora qué sucede cuando modificamos el valor de N , es decir, el número de puntos de la DFT. Vamos a comprobar que al disminuir N la figura se hace más abrupta y por tanto menos clara.

- Q3.15** Según el resultado de la cuestión Q3.11, ¿este valor de M es adecuado para distinguir los formantes? Ejecuta el programa P3_1 y compruébalo.

Si lo es, visualmente se distinguen todos los formantes y teóricamente, es mayor que en M mínimo.

- Q3.16** Toma ahora $N = 10000$ y vuelve a ejecutar el programa. El espectro no debe haber cambiado significativamente respecto de la cuestión Q3.15, a excepción de la ubicación de los formantes respecto de los índices k . ¿Esa variación repercute en la frecuencia analógica en Hz de dichos formantes? Justifícalo.

Repercute, pero no porque los formantes hayan cambiado su frecuencia real, si no que la resolución espectral aumenta, por lo que va a haber mayor error; y por ende, mayor variación en los valores finales.

- Q3.17** Escoge ahora una N igual al doble de M . ¿Sigues pudiendo distinguir todos o casi todos los formantes?

Si, se pueden seguir distinguiendo.

- Q3.18** Finalmente, toma N igual a M . ¿Sigues pudiendo distinguir todos o casi todos los formantes? Si es que no, ¿crees que esto contradice el hecho de que tomando un número de puntos de la DFT al menos igual a la longitud de la secuencia, esta queda adecuadamente representada sin perder información?

No se pueden distinguir muy bien. Sin embargo, no lo contradice, porque aun podemos distinguir el pitch (por lo que tenemos el valor de todos los armónicos) y aún vemos el formante máximo absoluto.

Ejercicio 3.4 Reconocimiento automático de vocales

En una señal de voz real los fonemas cambian muy rápidamente por lo que es necesario tomar intervalos de muy corta duración, que en la práctica suelen estar en el orden de entre 5 y 10 ms.

Q3.19 Calcula el valor de M correspondiente a 8 ms. Según el resultado de la cuestión Q3.11, ¿es un valor suficiente de cara a distinguir todos los formantes?

$M=8\text{ms}\cdot F_s=80$ muestras. Según el apartado Q3.11 no es suficiente, para una $N = 15000$.

A pesar de lo anterior, es posible diseñar un software que permita distinguir automáticamente las vocales, puesto que no son necesarios todos los formantes para cumplir dicho objetivo. Veamos cómo.

El fichero *Entrada_2.wav* contiene un intervalo muy pequeño de una vocal (de hecho, escuchada así de forma aislada es casi ininteligible). Haz que el programa P3_1 importe dicho audio en lugar de *Entrada_2.wav*.

Q3.20 Inhabilita la línea $x = \text{data}(12551:12550+M)$; en el programa P3_1 añadiendo un % al inicio de la misma (ahora no vamos a recortar nada). Toma $N = 1000$, para poder ver con nitidez la figura. Ejecuta el programa y comprueba que la longitud de la nueva variable x es igual al valor de M obtenido en la cuestión Q3.19. Escribe el valor de k correspondiente a los dos formantes con mayor amplitud que se vean ahora.

Para el primer formante $k_1=744$. Y para el segundo $k_2=2764$

Q3.21 Calcula las frecuencias analógicas en Hz correspondientes a los formantes anteriores.

Utilizamos la fórmula $f_1=k_1\cdot F_s/N$ y $f_2=k_2\cdot F_s/N$. Siendo en este caso $F_s=10000$ y $N = 15000$. Dando $f_1=496$ Hz y $f_2=1842.7$ Hz

La posición de los dos primeros formantes de cada vocal (entendidos ahora como los dos de mayor amplitud), son aproximadamente los mismos independientemente de la persona que hable. (Varían de un idioma a otro, eso sí). En la figura siguiente, denominada *carta de formantes*, aparecen los valores de esos dos primeros formantes para el español.

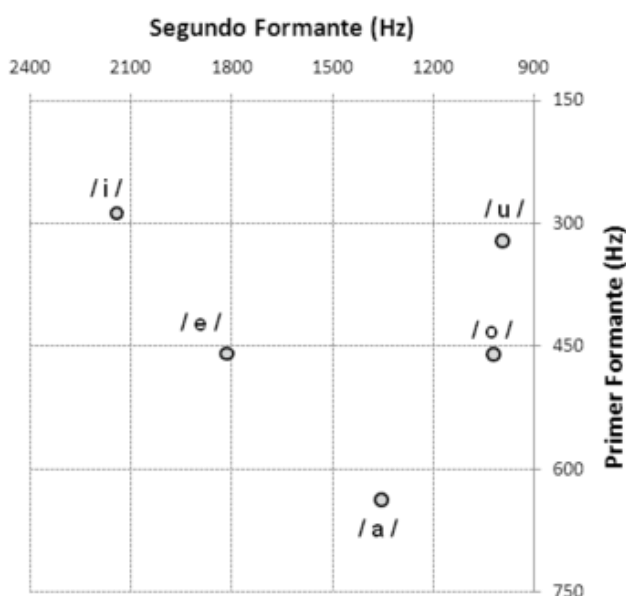


Figura 1: carta de formantes. (Extraída de es.wikipedia.org).

Q3.22 Observando la figura anterior, indica qué vocal está contenida en el archivo *Entrada_2.wav*.

Según los formantes obtenidos, sería la e.