

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO, ITAM
Laboratorio de Procesamiento Digital de Señales

Práctica No. 9
Espectrograma

Nombre del equipo: Chicken and Beer

Integrantes del equipo: Andrés Saucedo, Lorena Barrera.

Introducción

Durante la práctica haremos uso de dos pistas de audio, ambas mencionando la misma frase, “hola mundo”. Con ayuda de una herramienta de software, MatLab, crearemos un espectrograma de cada una de las grabaciones y con ayuda de los resultados obtenidos en la práctica 8, comparar los resultados de cada formante.

Un espectrograma es el resultado de calcular los espectros de una señal. Para lograr esto, el espectrograma divide la señal en ventanas temporales, es decir, divide la señal en N partes [1]. Cada uno de los segmentos se les analiza las frecuencias y se representan el resultado en un plano 3D (Figura 1), que después se cambia a un plano 2D (Figura 2).

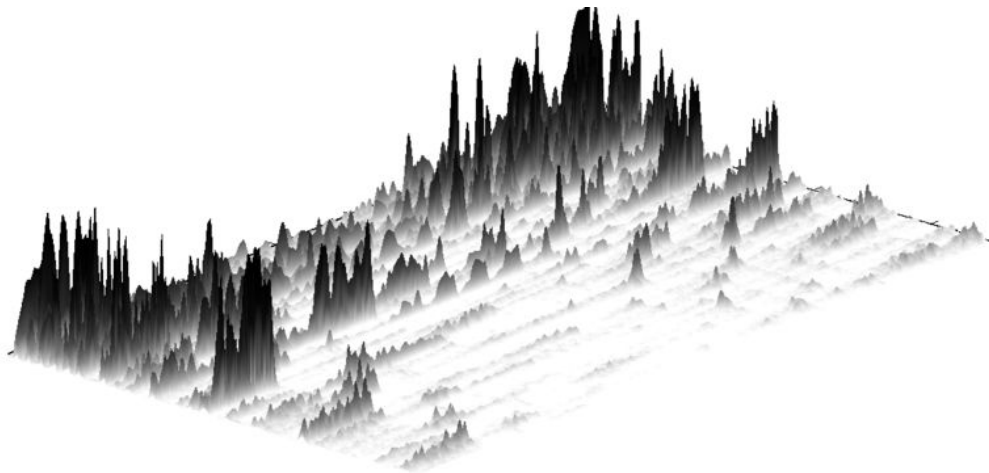


Figura 1. Espectrograma en 3D

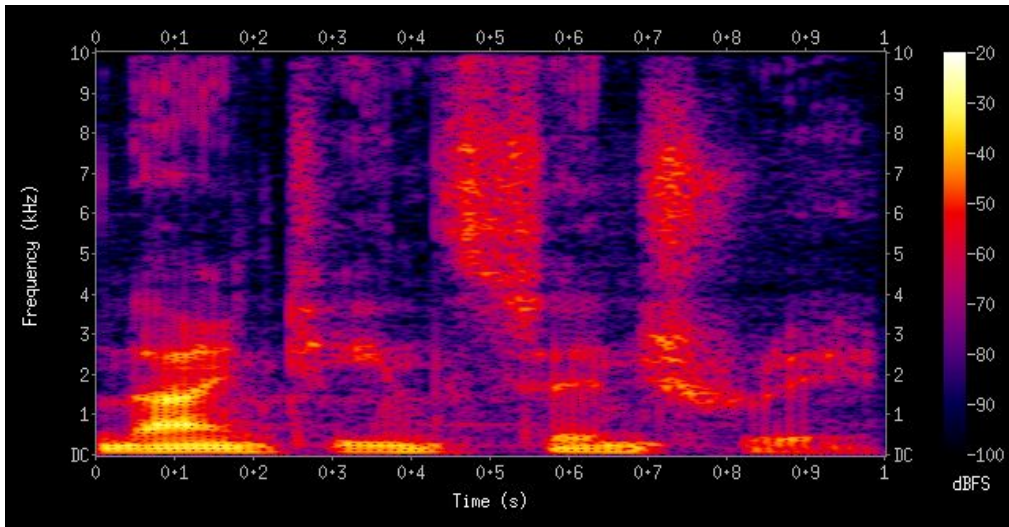


Figura 2. Espectrograma en 2D

El objetivo es encontrar las formantes que componen a la frase “Hola Mundo” y comparar los fonemas obtenidos con los valores obtenidos en las prácticas anteriores [2].

Desarrollo

Cada integrante del equipo grabó un audio con formato wav en el que se escuchara la frase “Hola Mundo”. Utilizamos la función *audioread()* para obtener la frecuencia de muestreo y la función *audioinfo()* para obtener la duración. Asignamos 128 puntos DFT (NFFT), 90 muestras superpuestas (noverlap) y una ventana (ventana) de 100 para generar el espectrograma (1).

```
spectrogram(y,segmentlen,noverlap,NFFT,Fs,'yaxis'); ... (1)
```

El espectrograma puede guardar en una matriz distintos valores: la transformación de Fourier de corta distancia (s) con una dimensión de 32769x8; las frecuencias normalizadas (w) que tienen la longitud igual al número de filas de .ws, en este caso 32768x1; finalmente, las instantáneas de tiempo (t) con una dimensión de 1x8.

```
[s, w, t] = spectrogram(y, 'yaxis'); ... (2)
```

Después, para cada vocal de la frase “Hola Mundo” se identificó la duración y en qué posición exacta de tiempo estaba para poder recortar ese segmento de audio.

```
inicio = round(segundoInicio/(1/Fs)); ... (3)
```

```
final = round(segundoFinal/(1/Fs)); ... (4)
```

```
y = y(inicio:final); ... (5)
```

Una vez identificado el segmento de cada vocal, se eligió una columna de las ocho de las transformaciones de Fourier que se adecúe con una posición de la vocal para obtener en vector del fonema.

```
vectorO = s(:, númeroColumna); ... (6)
```

Finalmente se utilizó la función desarrollada en la Práctica 8 para poder calcular los formantes de las vocales analizadas.

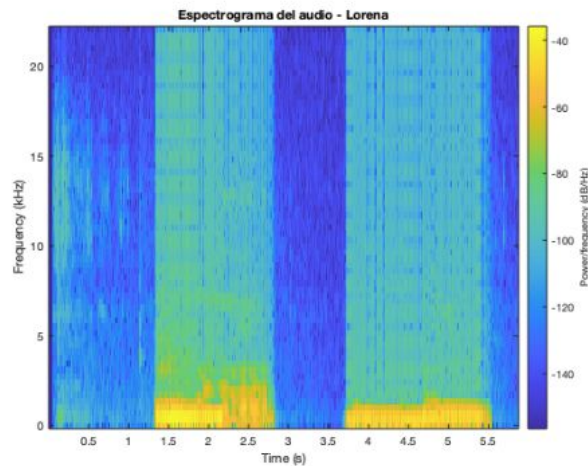
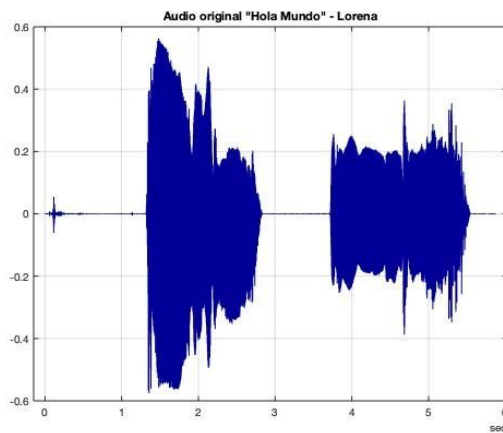
[Formantes] = Practica8 (y, Fs, 'Vocal') ... (7)

Resultados

Se analizaron las cuatro vocales que hay en “Hola Mundo” y se compararon con los datos de las formantes que ya se habían obtenido anteriormente.

Lorena

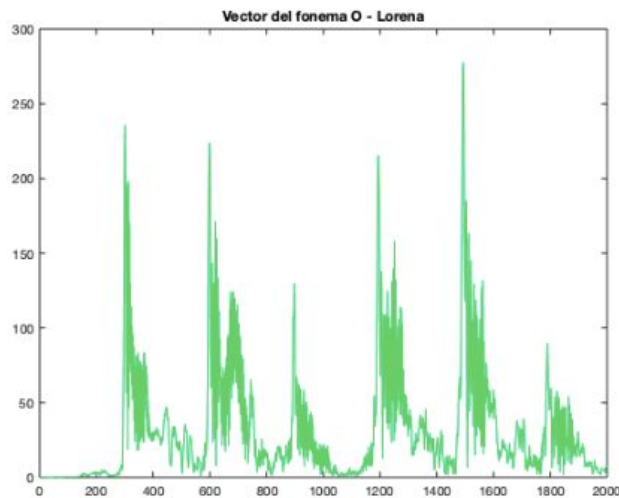
La señal resultante de la frase junto con su espectrograma son las siguientes:



Por lo tanto, se calculó que la duración aproximada de cada vocal fue:

Lorena	
Letra	Tiempo [s]
O	0.5
A	0.5
U	0.4
O	0.5

Para la primera *letra* O se obtuvo el siguiente vector de formante:

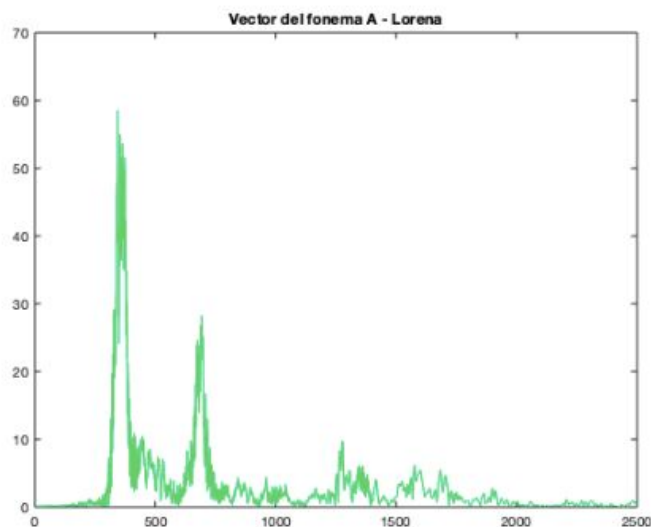


Los formantes de esta letra a comparación de los datos obtenidos en la Práctica 8 son:

Práctica 8	1.5813	1.0658	0.0808	0.0511
Práctica 9	1.6139	1.0860	0.7904	0.5713

En particular, los primeros dos formantes son muy parecidos entre ellos aunque los dos últimos son diferentes. Por lo tanto, la letra tuvo variaciones a comparación de los primeros datos obtenidos.

Para la *letra A* se obtuvo el siguiente vector de formante:



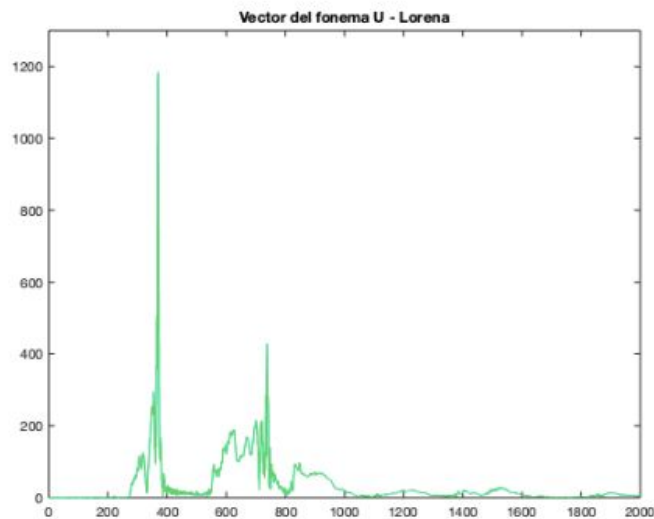
Los formantes de esta letra a comparación de los datos obtenidos en la Práctica 8 son:

Práctica 8	1.7069	1.3984	1.1924	0.9845
-------------------	--------	--------	--------	--------

Práctica 9	1.4869	1.2745	0.9588	0.6692
-------------------	--------	--------	--------	--------

En general, las formantes son muy parecidas a los datos de referencia que se tienen. Por lo tanto la vocal se pronunció muy similarmente.

Para la *letra U* se obtuvo el siguiente vector de formante:

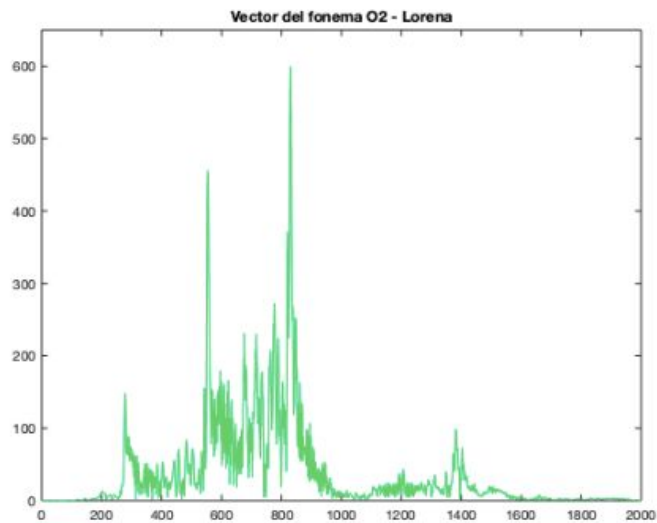


Los formantes de esta letra a comparación de los datos obtenidos en la Práctica 8 son:

Práctica 8	1.6207	0.9588	0.0457	-
Práctica 9	1.6047	0.9819	0.7592	0.0267

Notamos que los primeros dos formantes son casi iguales, pero el tercer formante de la Práctica 8 es más parecido al cuarto formante de la Práctica 9.

Para la segunda *letra O* se obtuvo el siguiente vector de formante:



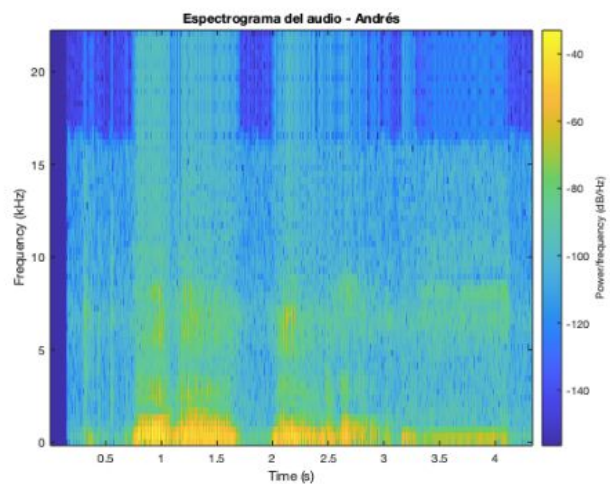
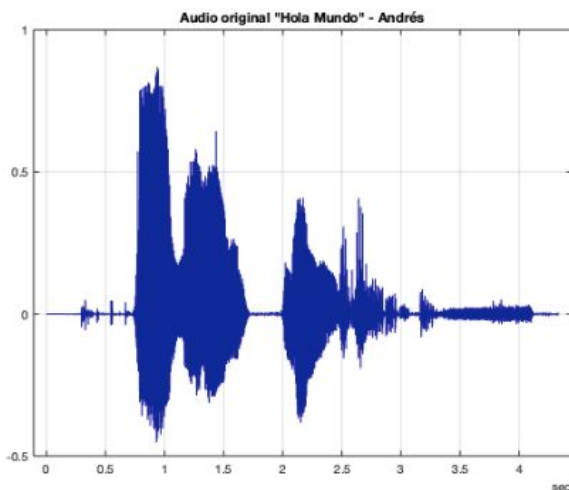
Los formantes de esta letra a comparación de los datos obtenidos en la Práctica 8 son:

Práctica 8	1.5813	1.0658	0.0808	0.0511
Práctica 9	1.6647	1.5512	1.0029	0.6980

En general, los formantes son muy cercanos entre ellos excepto el cuarto formante. Por lo tanto podemos decir que se dijo la letra casi de la misma forma que en los primeros datos obtenidos.

Andrés

La señal resultante de la frase junto con su espectrograma son las siguientes:

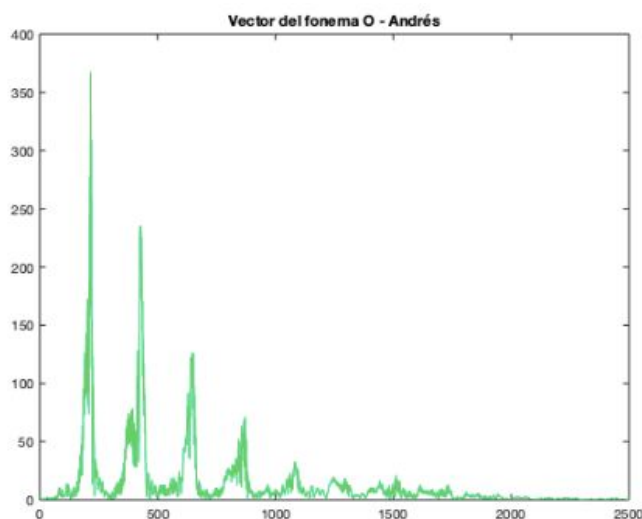


Por lo tanto, se calculó que la duración aproximada de cada vocal fue:

Andrés

Letra	Tiempo [s]
O	0.2
A	0.5
U	0.3
O	0.5

Para la primera *letra* O se obtuvo el siguiente vector de formante:

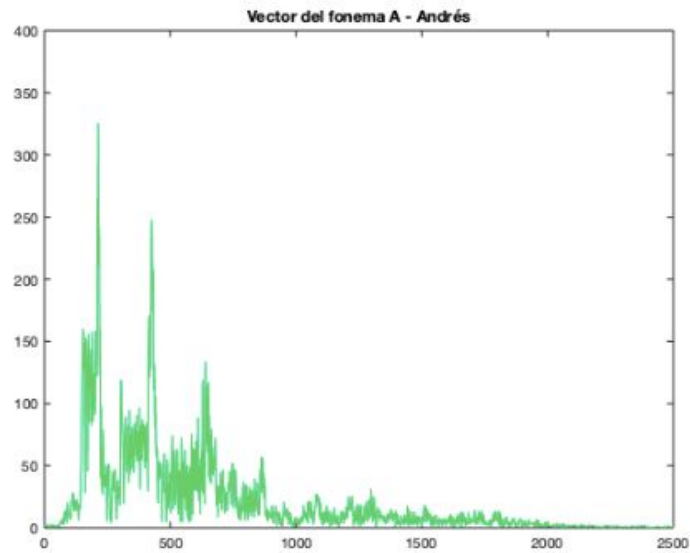


Los formantes de esta letra a comparación de los datos obtenidos en la Práctica 8 son:

Práctica 8	1.6339	1.4230	0.7503	0.0365
Práctica 9	1.3389	1.0178	0.7984	0.5835

En general, casi todos los formantes son parecidos entre ellos por lo que podemos decir que la letra se pronunció similarmente a la práctica pasada.

Para la *letra* A se obtuvo el siguiente vector de formante:

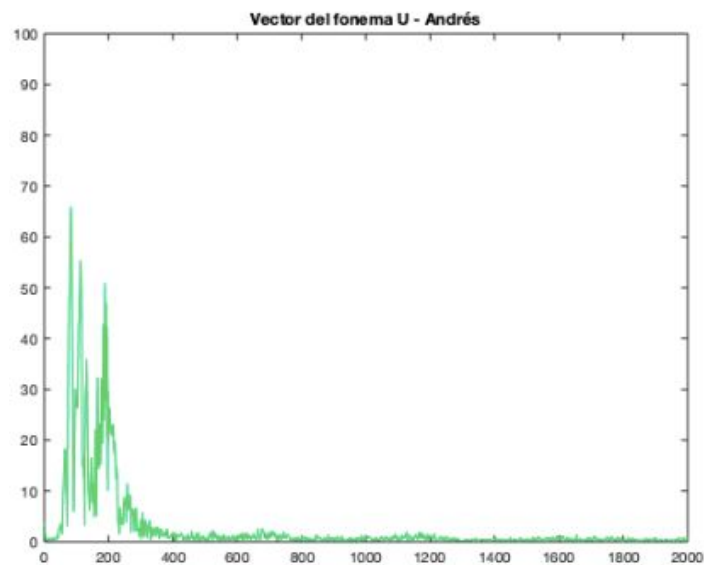


Los formantes de esta letra a comparación de los datos obtenidos en la Práctica 8 son:

Práctica 8	1.6301	0.7999	0.0551	-
Práctica 9	1.5533	1.4529	1.2884	1.0768

En particular, sólo el primer formante es un poco parecido, sin embargo los demás datos varían. Por lo tanto, la letra en Hola Mundo no fue pronunciada similarmente a la práctica pasada.

Para la *letra U* se obtuvo el siguiente vector de formante:

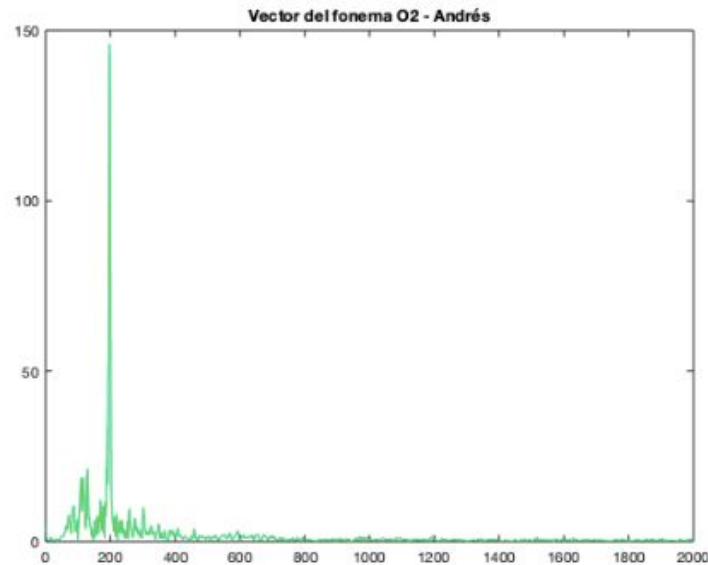


Los formantes de esta letra a comparación de los datos obtenidos en la Práctica 8 son:

Práctica 8	1.6442	1.4781	1.2855	0.5189
Práctica 9	1.5354	1.3207	0.8467	0.6737

En general, los formantes son muy parecidos entre ellos. Por lo tanto la letra se dijo muy similar a la práctica pasada.

Para la segunda *letra O* se obtuvo el siguiente vector de formante:



Los formantes de esta letra a comparación de los datos obtenidos en la Práctica 8 son:

Práctica 8	1.6339	1.4230	0.7503	0.0365
Práctica 9	1.5677	1.4720	1.3004	0.8426

En general, los formantes son muy cercanos entre ellos excepto el cuarto formante por lo que podemos decir que se dijo la letra casi de la misma forma que en los primeros datos obtenidos.

Conclusiones

Ya que la creación de ventanas para el espectrograma fue arbitrario, tomando en cuenta la duración en la pronunciación de una letra, la obtención del espectrograma, y por ende la de las formantes, puede mejorar si tomamos en cuenta parámetros en la industria. Además tendríamos que tomar en cuenta el tiempo de transición entre fonemas, ya que esto provoca diferencias entre los formantes tomados en prácticas anteriores y en esta práctica.

En conclusión, las formantes obtenidas en cada grabación son, en algún momento, similares a las anteriores, y los errores los podemos atribuir a la ventana seleccionada y las

transiciones entre fonemas. Aunque es posible reducir el error una vez implementadas técnicas aprobadas en la industria.

Referencias

- [1] Julius O. Mathematics of the Discrete Fourier Transform (DFT), with Audio Applications [Online]. Disponible en: <https://ccrma.stanford.edu/~jos/mdft/Spectrograms.html>
- [2] Olmo M. Formantes de vocales [Online]. Disponible en: <http://hyperphysics.phy-astr.gsu.edu/hbasees/Music/vowel.html#c3>