Universidad de la República

Facultad de Ingeniería

Procesamiento digital de señales de audio

Práctico 4

Autor:
Federico Bello

28 de septiembre de 2025

UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Índice

1. Ejercicio 1

En este ejercicio se estudia el cepstrum de señales de audio. Según el modelo del mecanismo de producción de la voz, la señal de voz se puede expresar como $s[n] = p[n] * h[n]$, donde $p[n]$ es la señal de excitación y $h[n]$ es la respuesta al impulso del tracto vocal. Mediante el cepstrum complejo se pretende deconvolucionar la señal de voz en la excitación y la respuesta al impulso.

Recordar que el cepstrum complejo de una secuencia es la transformada inversa de Fourier del logaritmo de la transformada de Fourier de la secuencia, es decir:

$$\hat{x}[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(X(e^{j\omega}))e^{j\omega n}\, d\omega \tag{1}$$

En el cepstrum, las características periódicas del espectro original (como armónicos o picos espectrales) aparecen como "picos.en el dominio quefreencial (dominio del cepstrum). Suele ser utilizado para analizar las propiedades de una señal que están relacionadas con sus componentes espectrales, como el pitch o las formantes en señales de habla.

1.1. Parte 1

Se comenzara calculando el cepstrum $\hat{p}[n]$ de $p[n]$, un tren de pulsos periódico:

$$p[n] = \beta^n \sum_{k=0}^{\infty} \delta[n - kP] \tag{2}$$

Se sabe que la transformada z del tren de pulsos se puede escribir como:

$$P(z) = \frac{1}{1 - (\beta z^{-1})^P} \tag{3}$$

Tomando el logaritmo, se llega a que:

$$\log P(z) = -log(1 - (\beta z^{-1})^P) \tag{4}$$

$$\overset{(1)}{=} P \sum_{n=1}^{+\infty} \frac{\beta^{nP}}{nP} (z)^{-nP} \tag{5}$$

$$\overset{(2)}{=} \sum_{n=1}^{+\infty} \left(P \sum_{k=1}^{+\infty} \delta[n - kP] \frac{\beta^n}{n} \right) z^{-n} \tag{6}$$

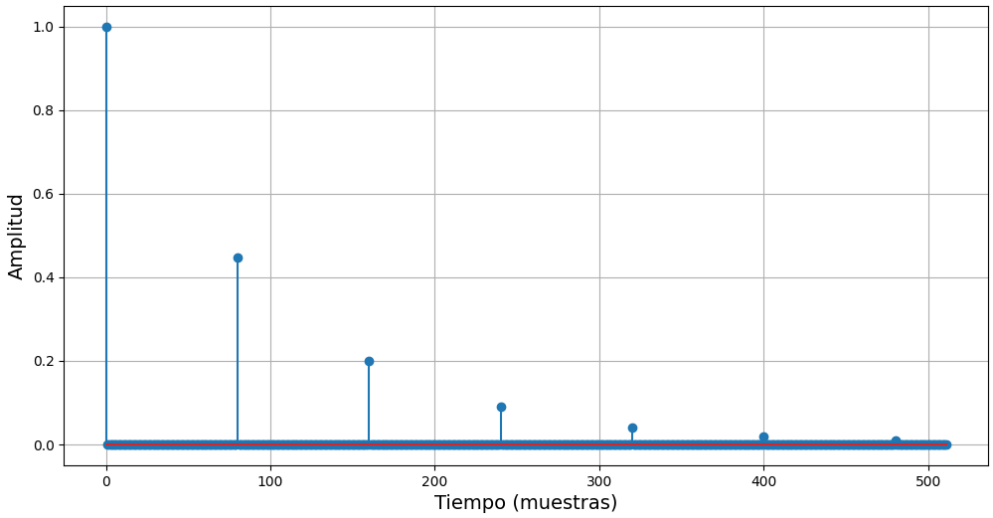
donde (1) surge de aplicar la serie de Taylor y multiplicar y dividir por P. Para el paso (2) se utilizo la descomposición discreta mediante deltas de Dirac. Observando que la ultima expresión es la transformada z de $P \sum_{k=1}^{+\infty} \delta[n - kP] \frac{\beta^n}{n}$ se llega a que el ceptrum del peine se puede escribir como:

$$\hat{p}[n] = P \frac{\beta^n}{n} \sum_{k=1}^{+\infty} \delta[n - kP] \tag{7}$$

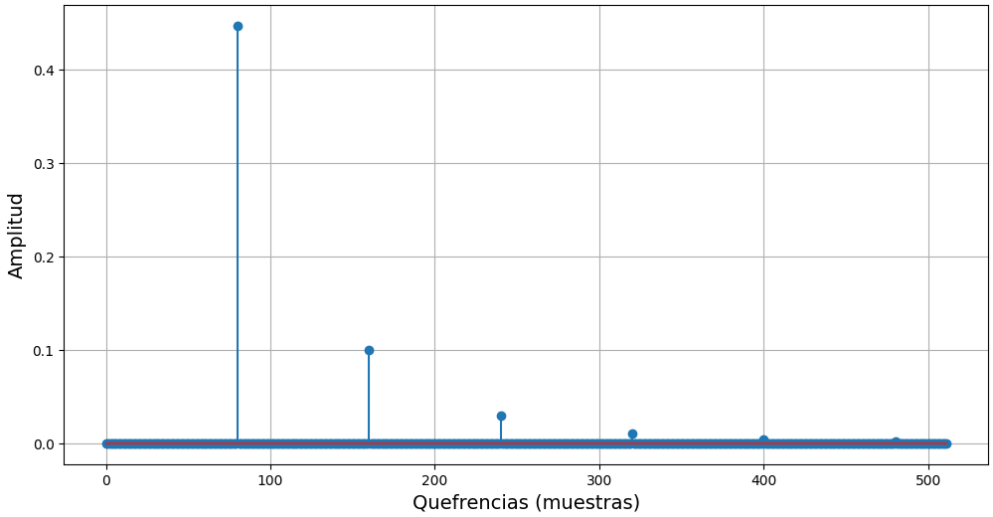
En la figura 1 se observa el tren periódico en cuestión y su correspondiente cepstrum. Se ve como en los picos del tren corresponden a picos también en el dominio del cepstrum.

Ahora, se calculara analíticamente el cepstrum complejo $\hat{h}[n]$ de la secuencia $h[n]$ cuya transformada $\mathcal{Z}$ es

$$H(z) = \frac{(1 - bz)(1 - b^* z)}{(1 - cz^{-1})(1 - c^* z^{-1})}, \quad \text{con } |b|, |c| < 1 \tag{8}$$



(a) Tren periódico



(b) Ceptrum

Figura 1: Tren periodico y su ceptrum

Tomando el logaritmo y utilizando series de potencia, se obtiene que:

$$\log H(z) = \log \frac{(1-bz)(1-b^*z)}{(1-cz^{-1})(1-c^*z^{-1})} \tag{9}$$

$$= \log(1-bz) + \log(1-b^*z) - \log(1-cz)\log(1-c^*z) \tag{10}$$

$$= -\sum_{n=1}^{+\infty} \frac{b^n}{n} z^n - \sum_{n=1}^{+\infty} \frac{(b^*)^n}{n} z^n + \sum_{n=1}^{+\infty} \frac{c^n}{n} z^{-n} + \sum_{n=1}^{+\infty} \frac{(c^*)^n}{n} z^{-n} \tag{11}$$

$$\tag{12}$$

Por lo tanto, se llega a que el cepstrum viene dado por:

$$\hat{h}[n] = \frac{b^{-n}}{n} u[-n-1] + \frac{(b^*)^{-n}}{n} u[-n-1] + \frac{c^n}{n} u[n+1] + \frac{(c^*)^n}{n} u[n+1] \tag{13}$$

donde b y c se pueden escribir como $b = |b|e^{j\theta_b}$ y $c = |c|e^{j\theta_c}$ respectivamente.

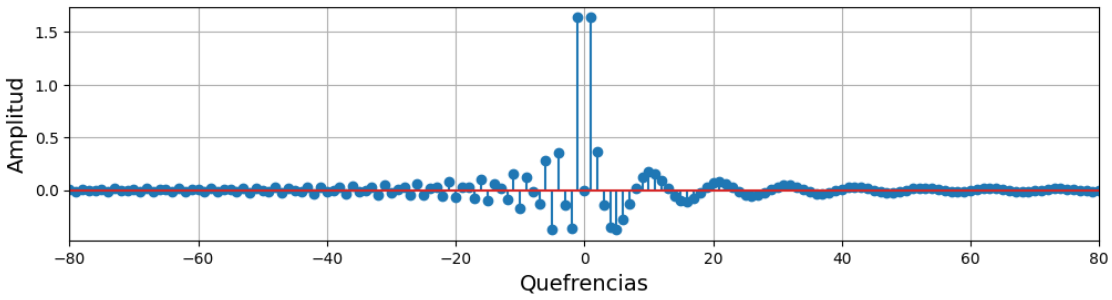


Figura 2: Cepstrum $\hat{h}[n]$

Sustituyendo estas expresiones en $\hat{h}[n]$, resulta:

$$\hat{h}[n] = \frac{|b|^{-n}e^{-jn\theta_b}}{n}u[-n-1] + \frac{|b|^{-n}e^{jn\theta_b}}{n}u[-n-1] + \frac{|c|^{n}e^{jn\theta_c}}{n}u[n+1] + \frac{|c|^{n}e^{-jn\theta_c}}{n}u[n+1] \tag{14}$$

$$= \frac{|b|^{-n}}{n}\left(e^{-jn\theta_b} + e^{jn\theta_b}\right)u[-n-1] + \frac{|c|^{n}}{n}\left(e^{jn\theta_c} + e^{-jn\theta_c}\right)u[n+1]. \tag{15}$$

Utilizando la identidad de Euler $e^{j\theta} + e^{-j\theta} = 2\cos\theta$, se llega al cepstrum complejo calculado analíticamente:

$$\hat{h}[n] = \frac{2}{n}|b|^{-n}\cos(n\theta_b)u[-n-1] + \frac{2}{n}|c|^{n}\cos(n\theta_c)u[n+1]. \tag{16}$$

El resultado se observa en la figura 2.

Ahora que se tienen ambos cepstrums, es posible calcular fácilmente el cepstrum de la señal resultante de la convolución $s[n] = h[n] * p[n]$. Aplicando la transformada z, se tiene que $S(z) = H(z)P(z)$. Luego, al aplicar el logaritmo se obtiene:

$$\hat{s}[n] = \log S(z) \tag{17}$$

$$= \log H(z) + \log P(z) \tag{18}$$

$$= \hat{h}[n] + \hat{p}[n] \tag{19}$$

Por lo que el cepstrum de la convolución no es mas que la suma de los cepstrums calculados anteriormente.

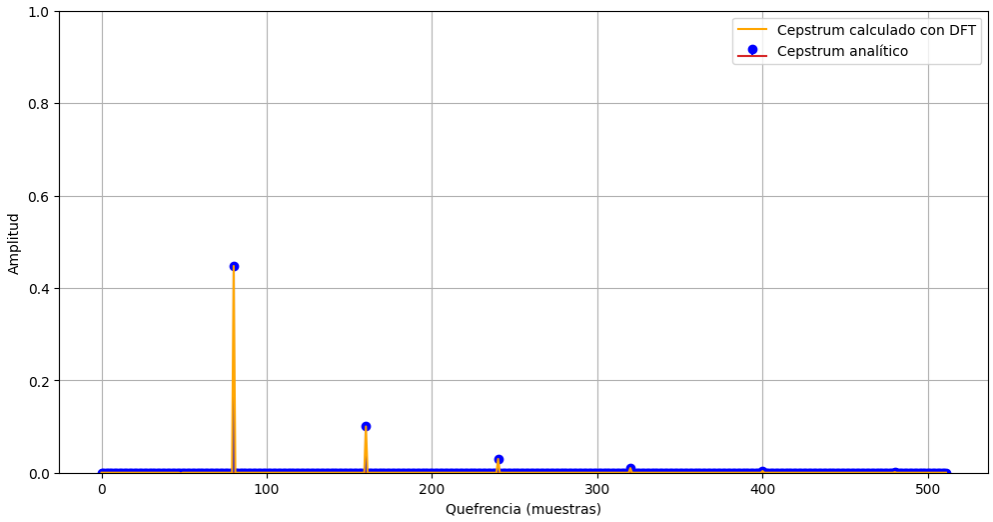
Luego, se procedió a realizar el calculo de los cepstrum utilizando la transformada de Fourier discreta (DFT). El resultado de la estimación se observa en la figura 3, donde se ve claramente la gran semejanza que tiene a los calculados analíticamente.

Finalmente, se aplicara un liftrado (filtrado en el dominio del ceptrum complejo para recuperar la respuesta al impulso $h[n]$ a partir de la señal $s[n]$. El liftrado en cuestión esta dado por la siguiente expresión:

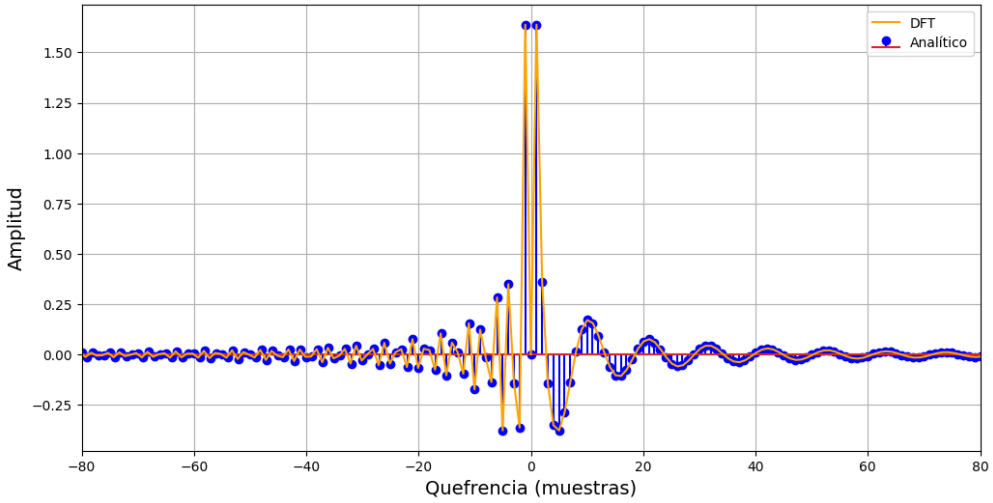
$$l[n] = \begin{cases} 1 & |n| \leq q_c \\ 0 & |n| > q_c \end{cases}$$

Para recuperar $h[n]$, se realiza el cálculo del cepstrum inverso. Este procedimiento consiste en aplicar la transformada inversa al resultado obtenido tras el proceso de liftrado, el cual previamente ha sido sometido a la operación exponencial. En la Figura 4 se ilustra el resultado de este proceso, destacándose una ligera discrepancia entre la señal original y la estimación reconstruida.

Dicha discrepancia puede atribuirse principalmente a dos factores. Primero, durante el cálculo, se llevó



(a) Tren periódico



(b) Ceptrum

Figura 3: Comparación entre cepstrum analíticos y calculados con la DFT

a cabo un truncamiento en el dominio de las quefrencias. Este paso, aunque necesario para limitar la complejidad computacional, conlleva una inevitable pérdida de información que afecta la precisión del resultado final. Segundo, se observa un desfasaje entre las señales, causado por el retardo introducido durante la linealización de la fase. Este ajuste introduce un desplazamiento temporal en la estimación, lo que explica el desfase evidente respecto a la señal original.

1.2. Parte 2

El análisis del cepstrum permite identificar picos en regiones específicas que corresponden a señales sonoras. Estos picos aparecen en las quefrencias medias y altas y su posición está directamente relacionada con el período de la señal, lo que facilita la identificación de la frecuencia fundamental. El algoritmo propuesto consiste en calcular el cepstrum para cada segmento de tiempo corto y buscar picos en las quefrencias mencionadas. Si el pico supera un umbral predefinido, se clasifica como una señal sonora, y la ubicación del pico determina la frecuencia fundamental. Por el contrario, si no se detecta un pico o este no supera el umbral, se clasifica como una señal sorda y la frecuencia fundamental se asigna a 0. Para enfocar la detección

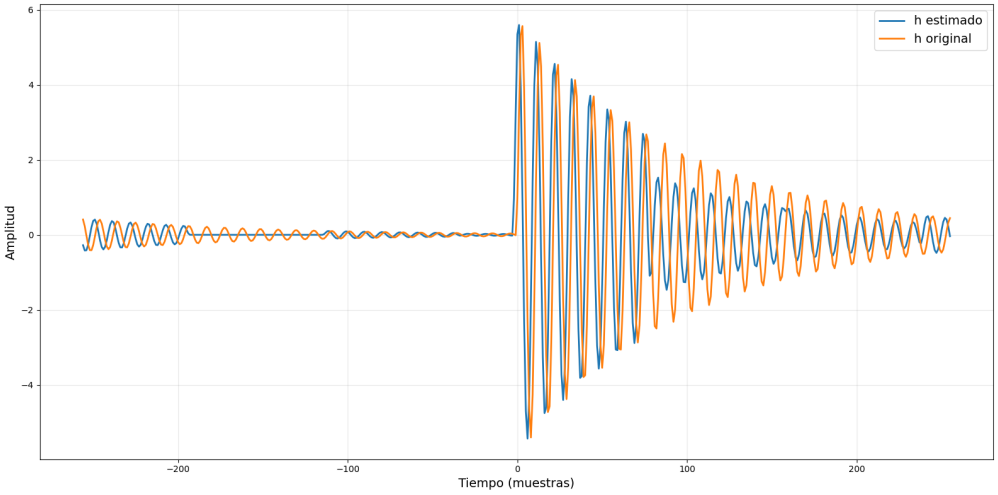


Figura 4: Comparación entre la estimación de $h[n]$ y $h[n]$

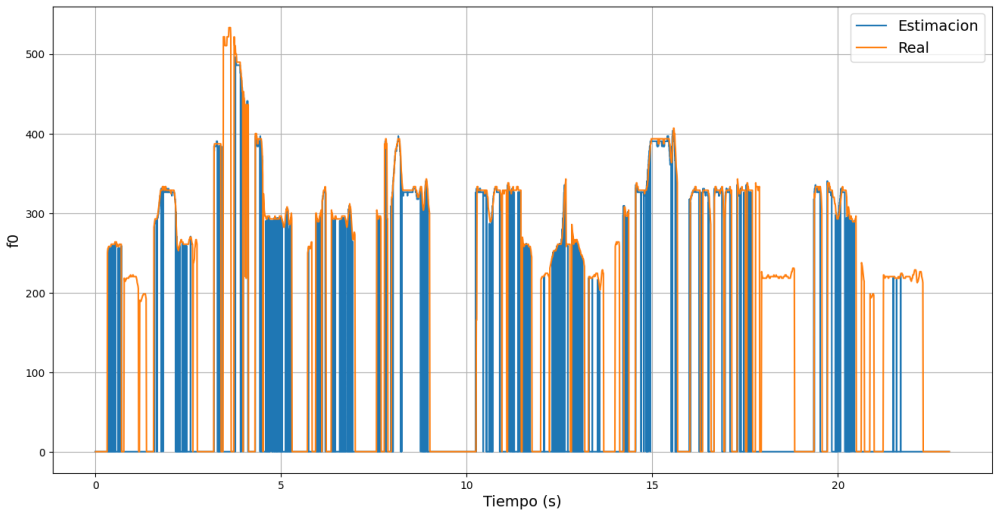


Figura 5: Resultados de la detección de frecuencias fundamentales utilizando el método basado en cepstrum, comparados con los valores de referencia

en las quefrencias medias y altas, cada segmento de tiempo se procesó mediante un filtro pasa altos.

Finalmente, el procedimiento descrito se implementó y sus resultados se compararon con un conjunto de datos de referencia. En la figura 5 se presentan los resultados obtenidos. Aunque se observaron algunos errores menores, estos podrían reducirse afinando la selección del umbral. Este método demostró ser altamente efectivo, produciendo los mejores resultados alcanzados en la resolución de este problema.

2. Ejercicio 2

2.1. Parte 1

En el modelo de predicción lineal se asume que la muestra actual de la señal de voz $s[n]$ es predecible a partir de una combinación lineal de p muestras previas,

$$\tilde{s}[n] = \sum_{k=1}^{p} \alpha_k s[n-k] \tag{20}$$

El problema consiste en encontrar los coeficientes α_k del predictor que mejor aproximan a la señal $s[n]$. Para eso se define el error de predicción como

$$e_n[m] = s_n[m] - \tilde{s}_n[m] \tag{21}$$

donde $s_n[m]$ es un fragmento de tiempo corto de la señal de voz elegido en torno a la muestra n.

Se define el error cuadrático medio de predicción como

$$E_n = \sum_m e_n^2[m] \tag{22}$$

para algún intervalo de muestras m. En el modelo de predicción lineal, el conjunto de coeficientes $\{\hat{\alpha}_k\}$ óptimo es el que minimiza el error cuadrático medio de predicción.

Se comenzara demostrando que los coeficientes que minimizan el error cuadrático medio obedecen el siguiente sistema lineal de ecuaciones (ecuaciones normales)

$$\sum_{k=1}^{p} \hat{\alpha}_k \sum_m s_n[m-i]s_n[m-k] = \sum_m s_n[m-i]s_n[m], \;\; 1 \leq i \leq p \tag{23}$$

Por definición, se tiene que:

$$E_n = \sum_m e_n^2[m] = \sum_m \left(s_n[m] - \sum_{k=1}^{P} \alpha_k s_n[m-k] \right)^2 \tag{24}$$

$$= \sum_m s_n^2[m] - 2\sum_m \left(\sum_{k=1}^{P} \hat{\alpha}_k s_n[m]s_n[m-k] \right) + \sum_m \left(\sum_{i=1}^{P} \hat{\alpha}_i s_n[m-i] \right)^2 \tag{25}$$

Se comienza por derivar E_{in} e igualar a 0 para encontrar su mínimo:

$$\frac{\partial E_n}{\partial \alpha_i} = \sum_m 2\left(s_n[m] - \sum_{k=1}^{P} \alpha_k s_n[m-k] \right) s_n[m-i] = 0 \tag{26}$$

Despejando, se obtiene que:

$$\sum_m s_n[m]s_n[m-i] = \sum_m \sum_{k=1}^{P} \hat{\alpha}_k s_n[m-k]s_n[m-i] \tag{27}$$

Reordenando términos, se llega a lo pedido::

$$\sum_{k=1}^{P} \hat{\alpha}_k \sum_m s_n[m-i]s_n[m-k] = \sum_m s_n[m-i]s_n[m] \tag{28}$$

Por otro lado, observar como multiplicando las ecuaciones normales de la parte anterior por α_i y sumando sobre i, se obtiene que:

$$\sum_{i=1}^{P} \hat{\alpha}_i \sum_{k=1}^{P} \hat{\alpha}_k \sum_m s_n[m-i]s_n[m-k] = \sum_{i=1}^{P} \hat{\alpha}_i \sum_m s_n[m-i]s_n[m] \tag{29}$$

Suplantando la expresión (29) en la ecuación (24), se tiene que:

$$E_n = \sum_m s_n^2[m] - 2\sum_m \sum_{k=1}^{P} \hat{\alpha}_k s_n[m] s_n[m-k] + \sum_{i=1}^{P} \sum_m \hat{\alpha}_i s_n[m-i] s_n[m] \tag{30}$$

$$= \sum_m s_n^2[m] - \sum_{k=1}^{P} \hat{\alpha}_k \sum_m s_n[m] s_n[m-k] \tag{31}$$

llegando a lo pedido.

2.2. Parte 2

El objetivo de esta parte es aplicar la técnica de LPC para la clasificación de vocales. El cuadro 1 muestra la frecuencia aproximada promedio de las dos primeras formates para cada vocal del idioma español.

Fonema	$\mathbf{F_1}$(Hz)	$\mathbf{F_2}$(Hz)
/a/	800	1170
/e/	480	2300
/i/	240	2800
/o/	510	960
/u/	250	630

Cuadro 1: Valores de F_1 y F_2 para diferentes fonemas.

Por lo tanto, un posible algoritmo para clasificar vocales consiste en:

1. Calcular las primeras dos formantes
2. Ver que fonema minimiza la distancia según la tabla 1

El algoritmo en cuestión se puede describir con el pseudocódigo **??**.

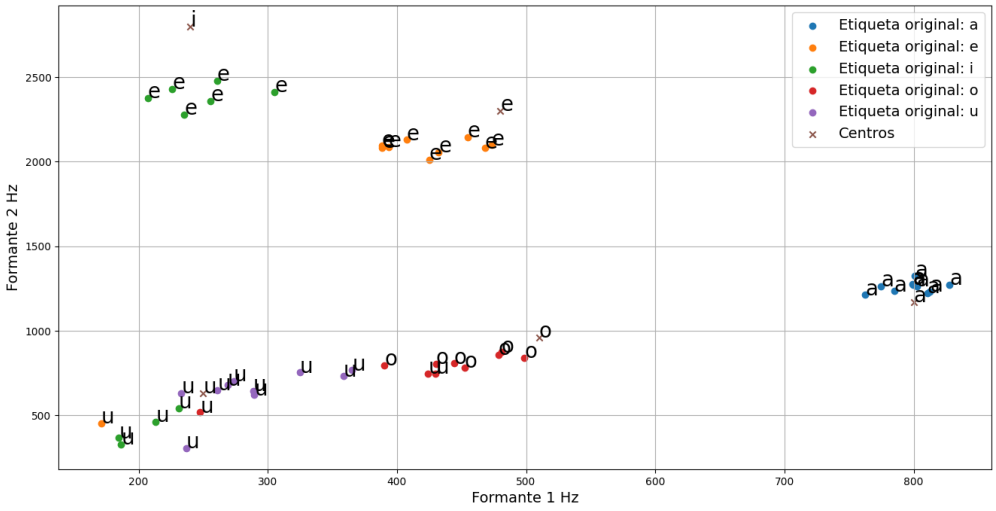


Figura 6: Clasificación de vocales para hablante masculino

Algorithm 1 Clasificación de vocales mediante análisis LPC.

Data: Señal de entrada $x(t)$, frecuencia de muestreo f_s, orden LPC p, valores de referencia de formantes.
Result: Clasificación de la vocal.
Preprocesamiento:

1. Centrar la señal en su nivel de energía promedio.
2. Aplicar ventana (por ejemplo, ventana de Hamming) para minimizar discontinuidades en los bordes.

Análisis LPC:

3. Realizar análisis LPC utilizando autocorrelación con el orden p.
4. Obtener los polos del modelo LPC.

Filtrado de polos:

5. Eliminar los polos con fase mayor a π (son redundantes).
6. Eliminar los polos que no cumplan con la condición:

$$\frac{f_s}{\pi} \log\left(\frac{1}{A_k}\right) < \delta$$

donde A_k es el módulo del polo y δ es un parámetro del algoritmo. En este caso se selecciono $\delta = 250Hz$

Extracción de formantes:

7. Seleccionar los dos primeros polos que cumplan las condiciones.
8. Calcular las dos primeras formantes a partir de los polos seleccionados.

Clasificación de la vocal:

9. Para cada par de formantes (F_1, F_2), calcular la distancia euclídea a los valores de referencia.
10. Clasificar la vocal según la menor distancia calculada.

Los resultados de la clasificación para un hablante masculino se observan en la figura 6. La matriz de confusión para el hablante masculino (figura 7) revela los siguientes patrones:

- La vocal */a/* presenta una alta precisión, con prácticamente ninguna confusión con otras vocales.
- La vocal */o/* muestra cierta tendencia a ser confundida con */u/*, lo que podría atribuirse a la proximidad de sus formantes.

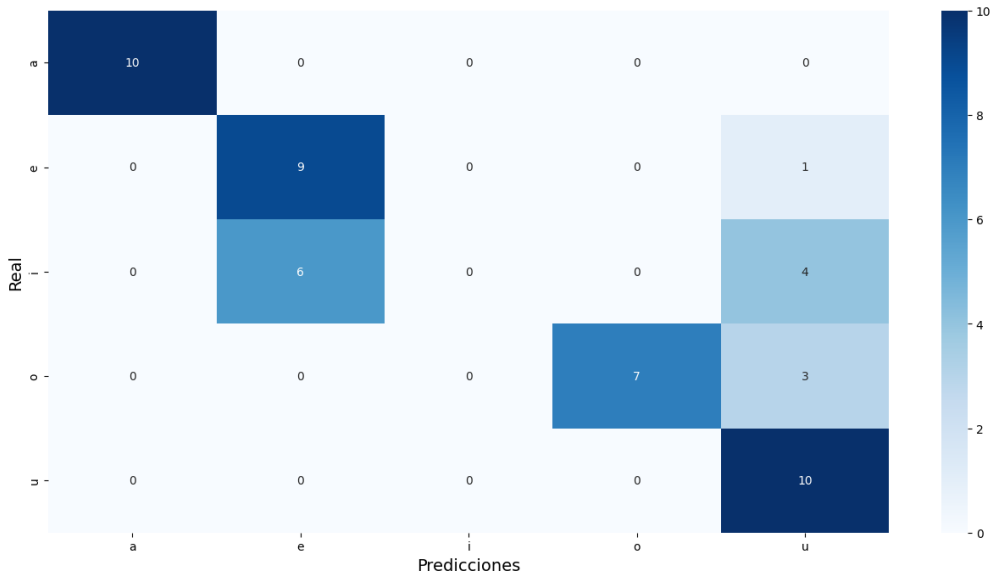


Figura 7: Matriz de confusión para hablante masculino

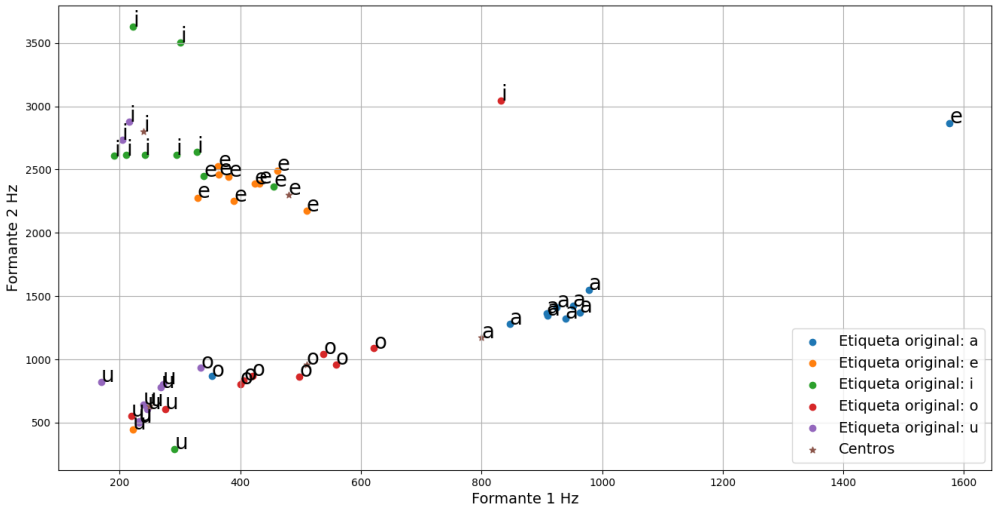


Figura 8: Clasificación de vocales para hablante femenino

- La vocal */e/* tiene una ligera confusión con */i/*, probablemente debido a la similitud en sus características espectrales.

Para el hablante femenino (figuras 8 y 9), los resultados muestran:

- Mayor dispersión en la clasificación comparado con el hablante masculino.
- Confusiones más frecuentes entre vocales con formantes cercanas.
- Posible influencia de diferencias en el tracto vocal entre géneros.

3. Conclusiones

El método de Predicción Lineal (LPC) para clasificación de vocales:

1. Demuestra alta efectividad en la discriminación de vocales.

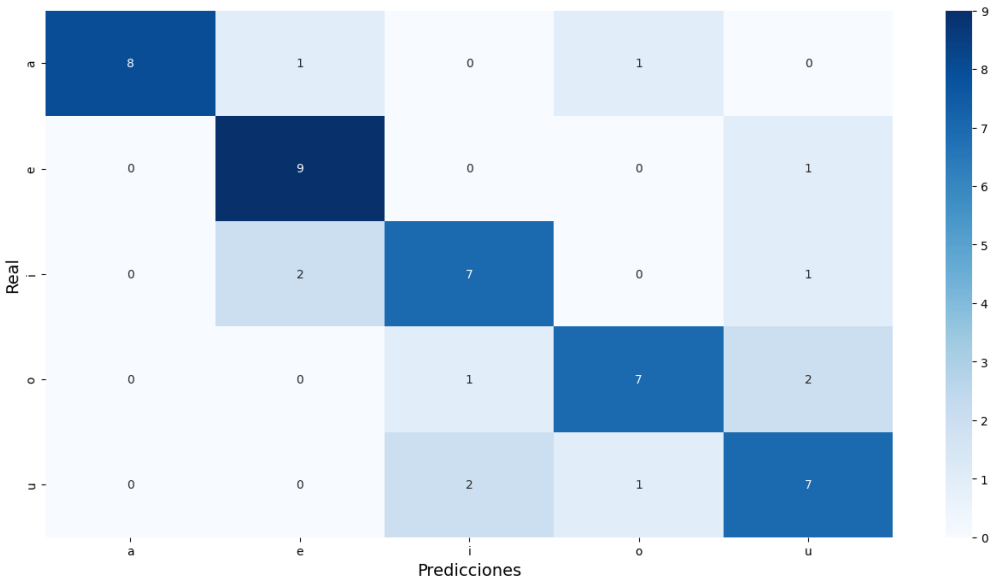


Figura 9: Clasificación de vocales para hablante femenino

2. Revela diferencias espectrales entre hablantes masculinos y femeninos.

3. Confirma la importancia de las dos primeras formantes en la caracterización de vocales.

Limitaciones: La precisión varía entre hablantes, sugiriendo la necesidad de modelos adaptativos.