



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE INGENIERÍA
Año 2018 - 1^{er} Cuatrimestre

PROCESAMIENTO DEL HABLA (86.53)

TRABAJO PRÁCTICO 1
TEMA: Linear Predictive Coding (LPC)
FECHA: 8 de abril de 2019

INTEGRANTES:

Manso, Juan
<juanmanso@gmail.com>

- #96133

Resumen

El objetivo del presente trabajo es el estudio del cálculo de coeficientes LPC para el análisis de señales de habla como también para la codificación de las mismas.

Índice

I	Caracterización de la señal	1
1.	Reconstrucción de una ventana	1
2.	Análisis de envolventes	2
3.	Espectograma	4
II	Codificación	5
4.	Reconstrucción con redondeo	5
5.	Compresión de la señal	6
III	Conclusiones	7

Parte I

Caracterización de la señal

A continuación se realiza el cálculo de los coeficientes del filtro del tracto vocal y a partir del mismo caracterizar la señal entrante.

1. Reconstrucción de una ventana

La señal a caracterizar es procesada bajo el algoritmo de LPC, iterando con muestras de 0,025 s para generar los coeficientes LPC que se la asignan a los primeros 0,01 s. La función LPC fue implementada en *Octave* de la siguiente manera:

```

1 function [a,G] = funcionlpc(senial_w, cant_coefs)
2     M = cant_coefs;
3     r = xcorr(senial_w);
4     r_utiles = r(round(end/2):round(end/2+M));
5     r0 = r_utiles(1);
6     R = toeplitz(r_utiles(1:M-1));
7     r = r_utiles(2:M);
8
9
10    a = inv(R)*r;           % Calculo los coeficientes
11    G = sqrt(r0 - a' * r);
12 end

```

Tras el cálculo de coeficientes, utilizando la función `filter()` se reconstruye la señal ventaneada. En la Figura 1 se exponen la señal original y reconstruida por LPC.

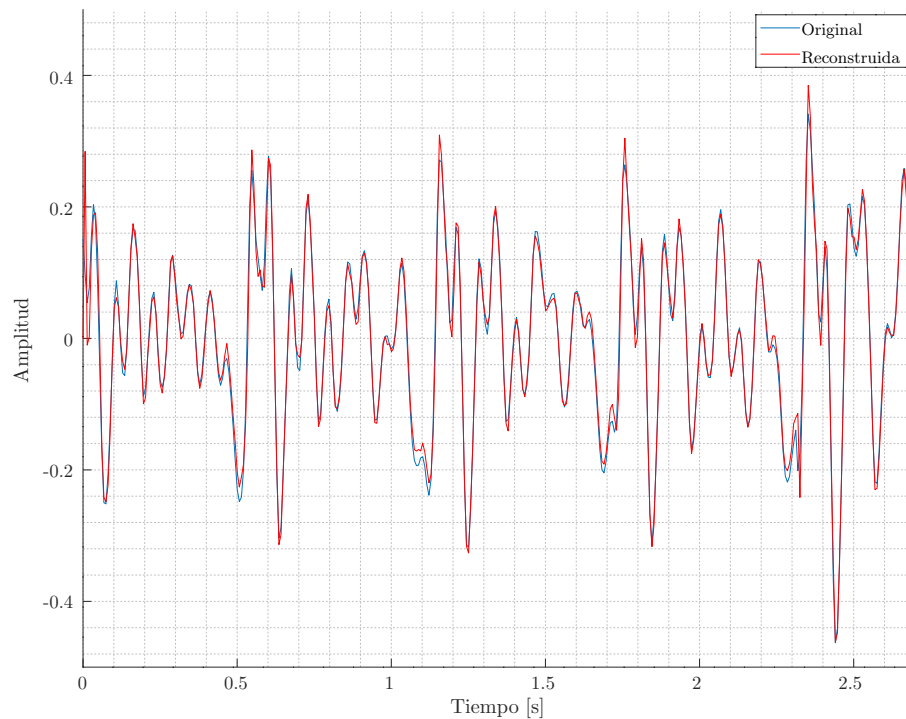


Figura 1: Señal original versus la reconstrucción por LPC.

Al igual que en la Figura 1, se puede ver en la Figura 2 que hay pequeñas diferencias cerca de los picos de la curva pero en general sigue a la original con precisión (difiere como mucho en 0,09). Otra característica importante que revela el análisis del error en la Figura 2 es la frecuencia glótica, dado que se supuso una entrada de ruido blanco y por tanto (al estar relacionado con el pulso glótico) se expone como error.

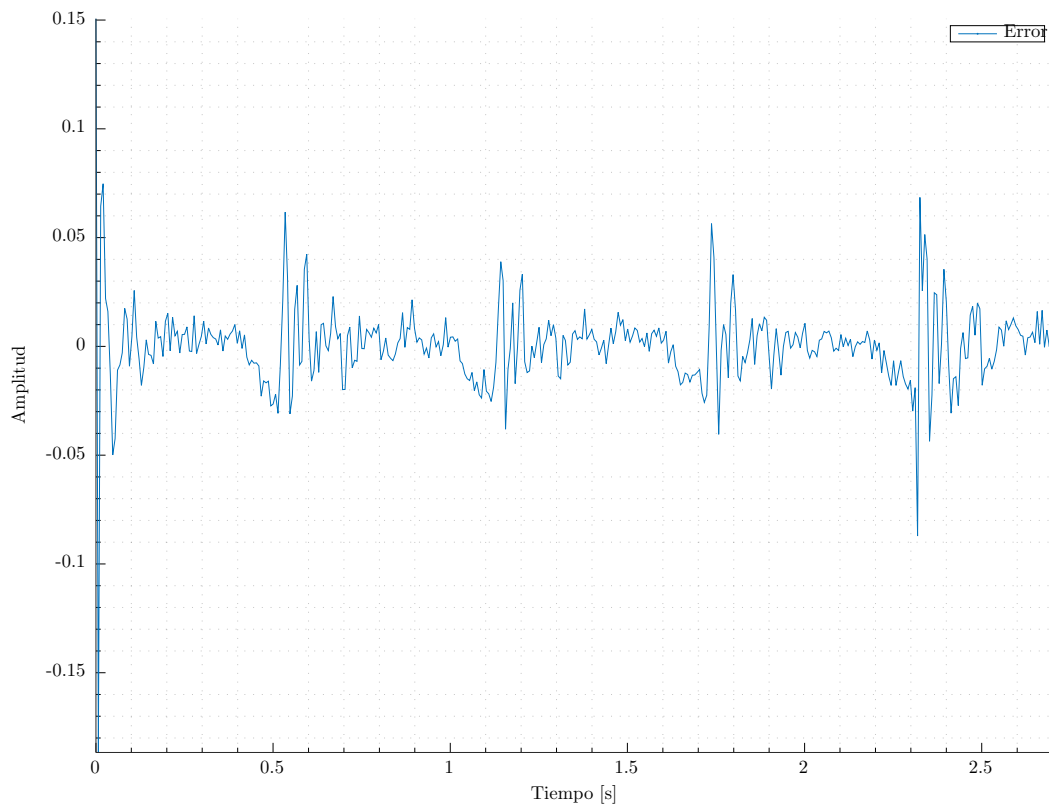


Figura 2: Error de estimación de la reconstrucción por LPC.

2. Análisis de envolventes

Al calcular la respuesta en frecuencia, se ve en la Figura 3 que la reconstrucción sigue con claridad a la envolvente de la original. Con dicho gráfico se puede afirmar que el primer formante se encuentra cerca de los 750 Hz y el segundo, 1800 Hz.

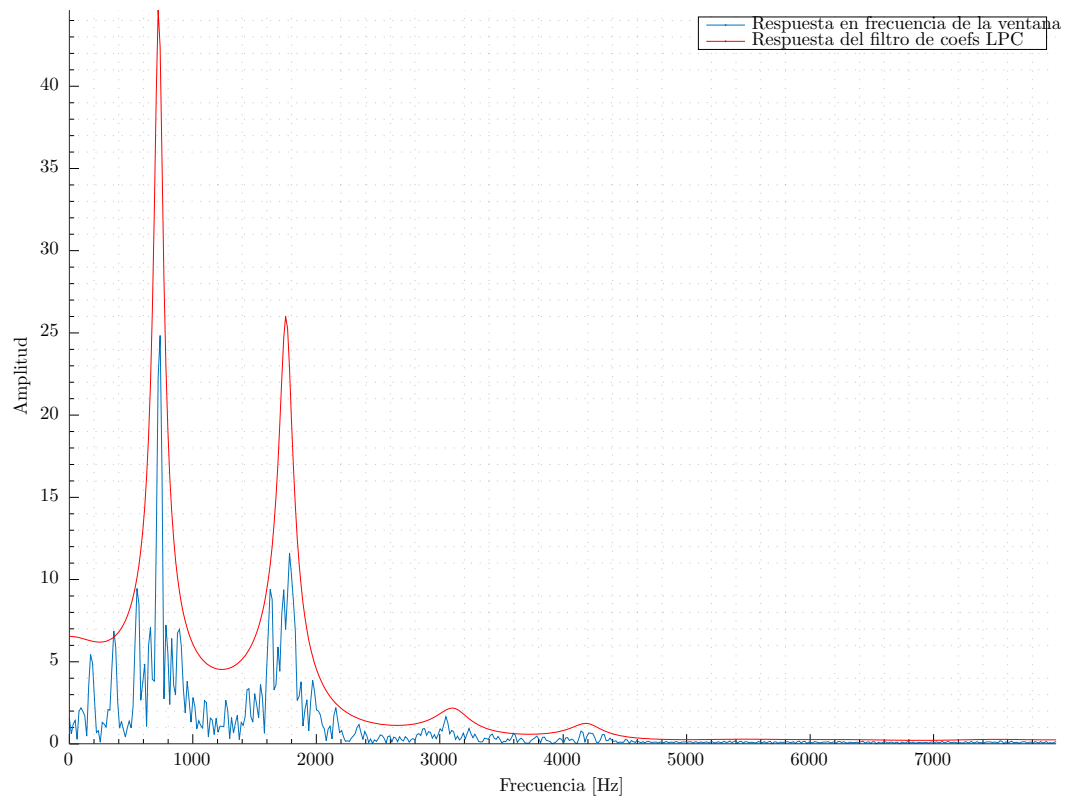


Figura 3: Comparación en frecuencia de las señal original y LPC.

En la Figura 4 se pueden ver las envolventes para las vocales de la señal de audio *fantasía*. Del gráfico se destaca que los formantes de la letra *a* se agrupan cerca de los 800 Hz y 1600 Hz mientras que la *i* tiene un formante de baja frecuencia (350 Hz) y otro más alto (3600 Hz), pudiéndose así distinguir claramente a partir del análisis de envolventes las vocales.

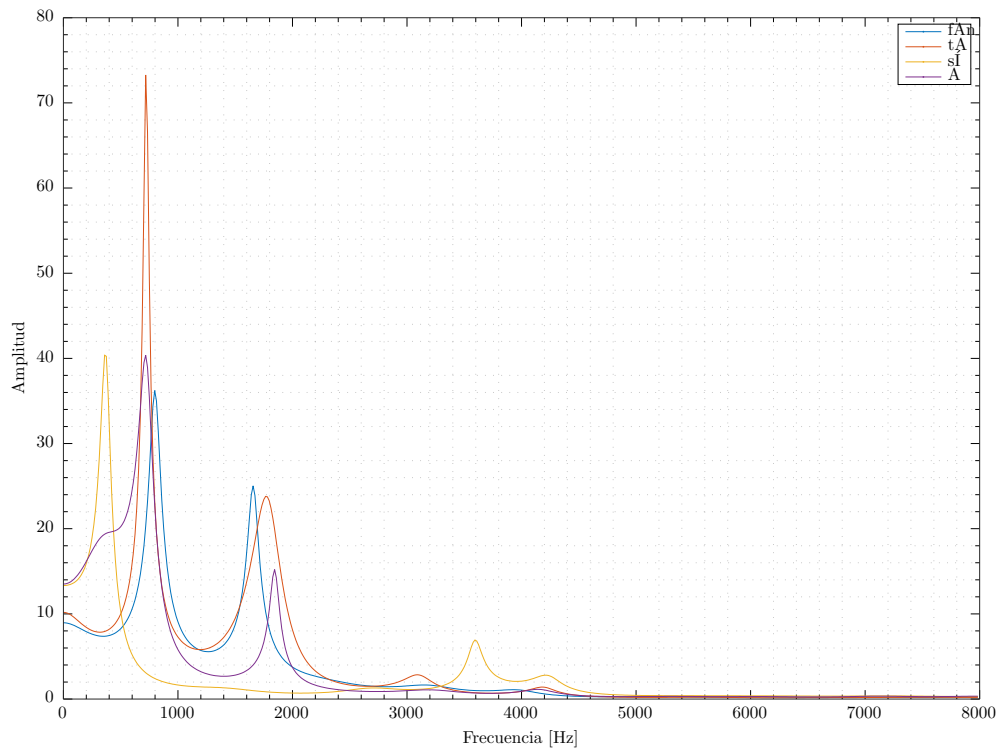


Figura 4: Comparación en frecuencia de las señal original y LPC.

3. Espectrograma

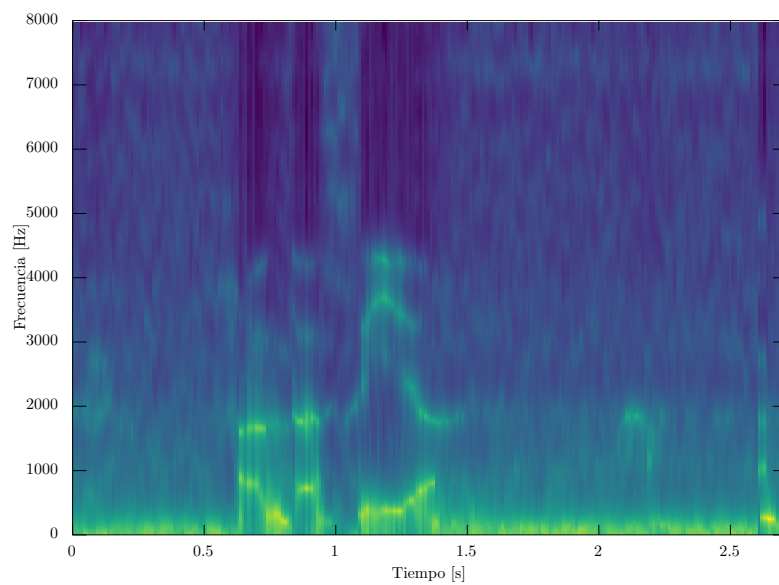


Figura 5: Envoltentes del filtro LPC a lo largo del tiempo.

La Figura 5 consta de las curvas envolventes a la largo del tiempo y la Figura 6 es el espectrograma de la señal original realizado con la función `specgram()`. Se puede ver que ambos gráficos son muy similares, salvando la diferencia que el de envolventes es más suave pudiéndose ver mejor los formantes de cada letra. Sin embargo, el gráfico del `specgram()` tiene como ventaja que se puede localizar mejor cuándo sucedió cada vocal. Ésto puede deberse al tamaño de ventana que usa la función por omisión.

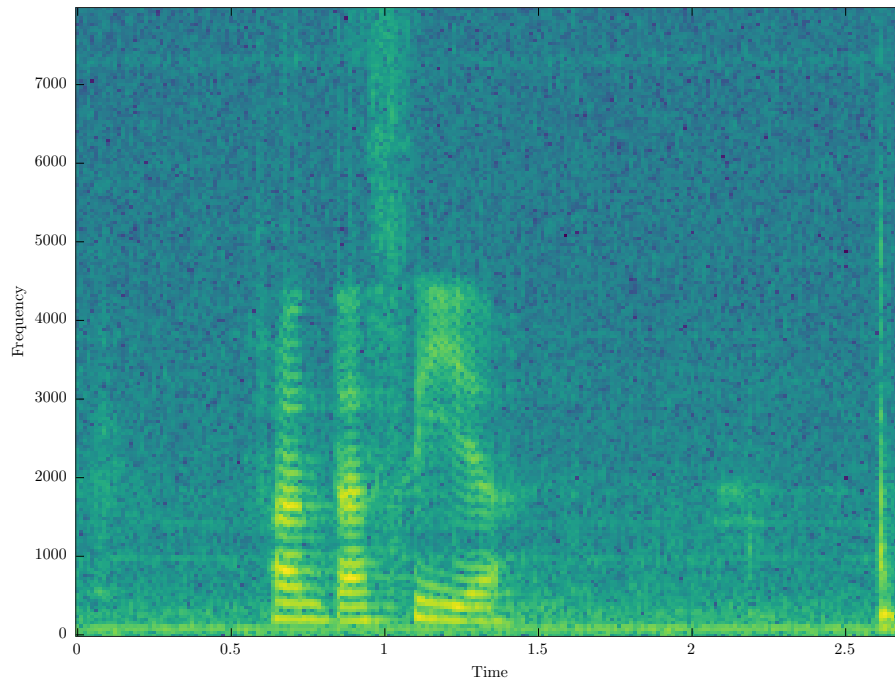


Figura 6: Espectrograma de la señal ventaneada.

Parte II

Codificación

El objetivo del algoritmo de LPC es la codificación de la señal de habla para facilitar su transmisión y recepción. Para ello se realiza el cálculo de coeficientes del filtro resultante de la modulación de las articulaciones. Con ésto y el error, se puede reconstruir la señal con cierta incertidumbre establecida por la precisión utilizada.

Cabe acalarar que el *script* `ej3.m` genera gráficos que no permiten ver con claridad el efecto de las operaciones realizadas. Es por ésto que las comparaciones y gráficos se generan con la ventana previamente analizada.

4. Reconstrucción con redondeo

La función de redondeo fue implementada de la siguiente manera:

```

1 % Función para redondear la señal con cierta cantidad de bits.
2
3 function redondeada = redondear(senial, bits)
4     offset = abs(min(senial));
5     senial = senial + offset;
```

```

6     niveles = 2^bits - 1;
7
8     redondeada = round(senial * niveles)/niveles;
9
10    redondeada = redondeada - offset;
11 end

```

A continuación se muestran los gráficos comparativos entre las distintas reconstrucciones.

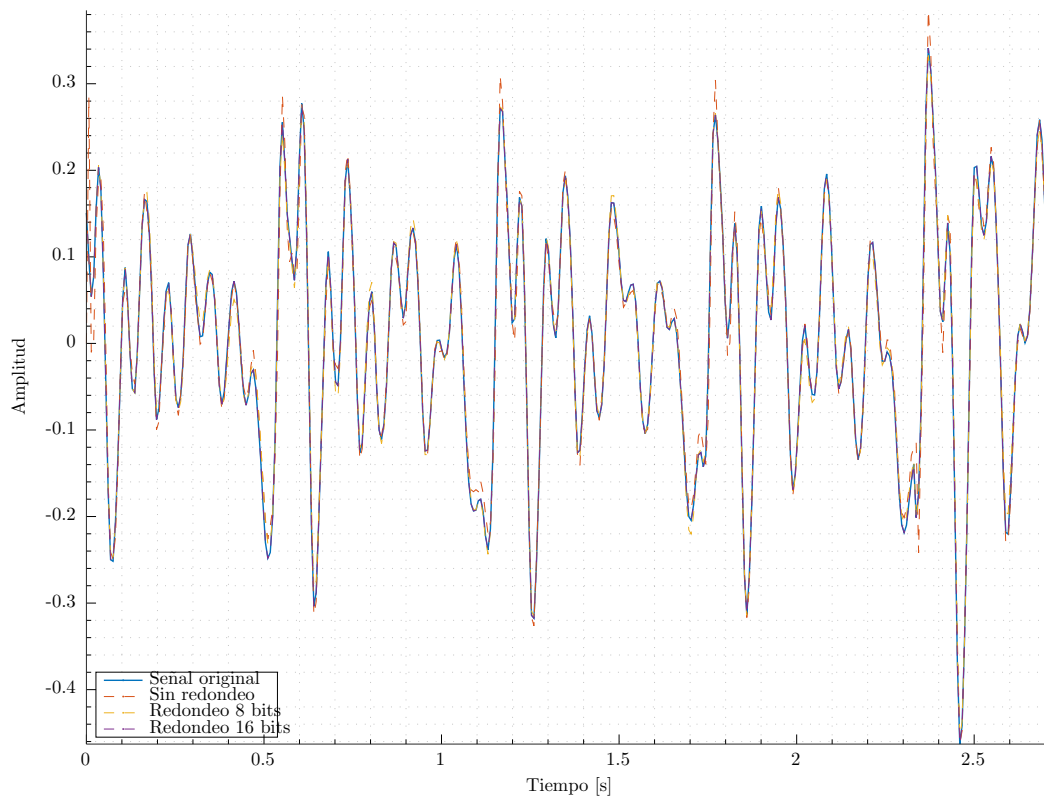


Figura 7: Comparación entre la señal original y las distintas reconstrucciones.

Se puede ver en la Figura 7 que la señal reconstruida con 8 bits tiene un error apreciable en los picos de la señal. Al analizar la Figura 8 se comprueba que el error de la codificación en 8 bits es apreciable pero muy pequeña, estando por debajo de 0,03.

Lo curioso es que la señal reconstruida sin redondeo presenta un error muy grande (llegando cerca de 0,8) cuando la misma debería ser la reconstrucción más precisa. Se puede ver que la curva del error se asemeja a la señal de error.

5. Compresión de la señal

El archivo de audio *fantasia.wav* requirió en memoria 348 800 bytes. Por lo tanto si se realizase una transmisión de dicho archivo completo, será muy costoso.

Si en cambio se utilizase una codificación de 8 bits y se enviasen los coeficientes LPC y la señal de error, el paquete pesaría:

$$\begin{aligned}
 Memoria_{LPC} &= size(LPC_{coef}) + size(err8) \\
 &= 41\,192\text{bytes} + 3208\text{bytes} = 44\,400\text{bytes}
 \end{aligned}$$

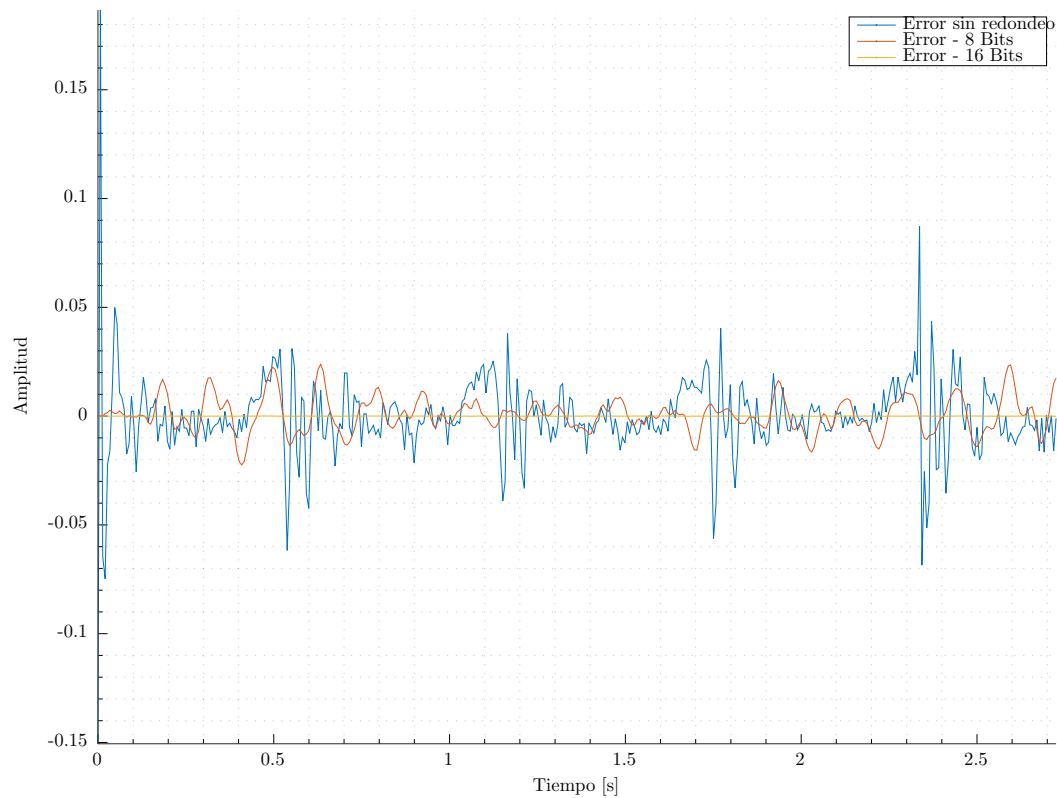


Figura 8: Comparación de las incertidumbres entre las distintas reconstrucciones.

De esta manera se logra comprimir un archivo de 348 800 bytes a uno de 44 400 (se comprimió a un 12 %).

Parte III

Conclusiones

Se pudo ver que la técnica LPC es muy útil tanto para el análisis de señales de habla como para su codificación. Con ella se pueden distinguir vocales a partir del análisis de formantes, como también detectar la variación de la frecuencia glótica a través de las envolventes en el tiempo. Como herramienta de compresión es muy poderosa porque puede lograr disminuir el tamaño en un orden de magnitud con una precisión muy alta.