



Universidad Nacional Autónoma de México

Posgrado en Ciencia e Ingeniería de la Computación

Asignatura: Reconocimiento de Patrones

Práctica 2: Procesamiento de señales de voz utilizando técnicas de LPC

Artemio Baños Gallardo

Introducción

El procesamiento de señales de voz es utilizado actualmente en la transmisión de estas en teléfonos celulares, reconocimiento de voz en sistemas inteligentes y speak-to-chat conversión.

Se trata del uso de múltiples filtros aplicados a la señal de entrada, en este caso voz, para evitar problemas de precisión finita.

En esta práctica aplicaremos los siguientes tratamientos para así poder tener una señal de voz que sea capaz de ser procesada, reconocida y predecida para prácticas futuras.

1. Filtro de Preénfasis
2. Bloques en marcos
3. Ventaneo usando Hamming
4. Aplicación de la función de correlación

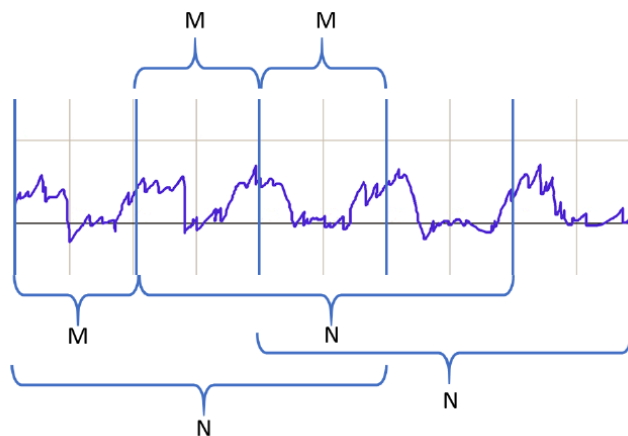
Filtro de Preénfasis

La señal de voz es pasada a través de un filtro paso altas para hacer que el espectro de la señal sea suavizado y así evitar problemas de precisión finita después del procesamiento.

$$\bar{s}(n) = s(n) - \bar{a}s(n-1)$$

Bloques en marcos y ventana de Hamming

Se aplican operaciones sobre ventanas de tamaño N para minimizar el efecto de las discontinuidades de la señal al principio y al final de cada marco. Cada ventana está separada por una fase de M muestras, vea la imagen de abajo



El segundo bloque de N muestras está separado M muestras del primero.

*Tomada de las diapositivas del curso de Reconocimiento de Patrones Primavera 2024 por el Dr. Jesús Savage

Y la siguiente es la función en el dominio del tiempo aplicada a cada una de las ventanas

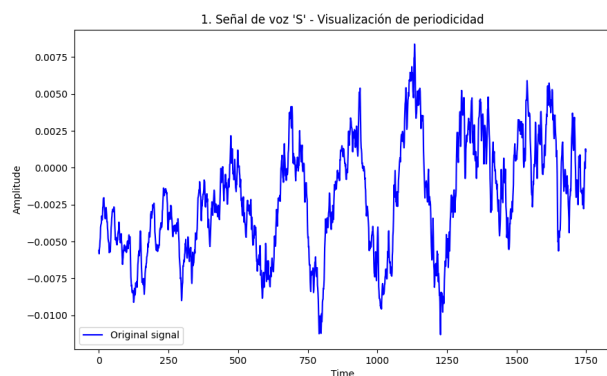
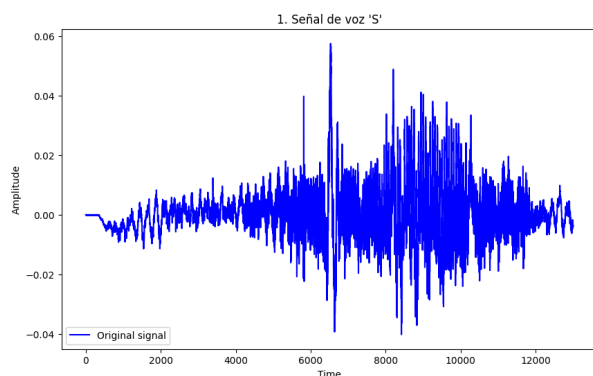
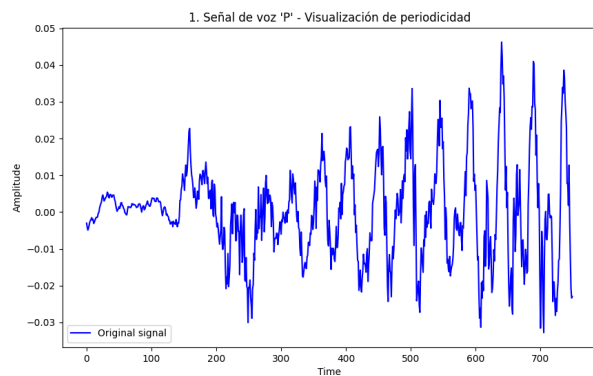
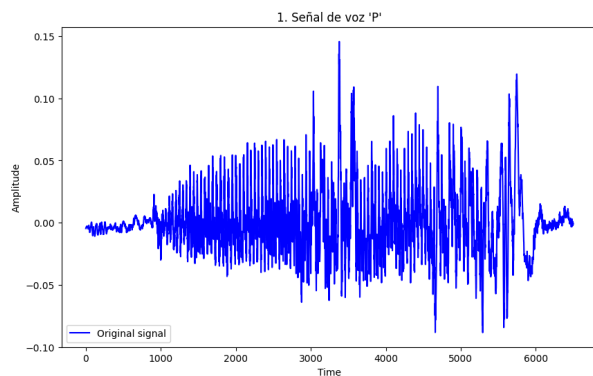
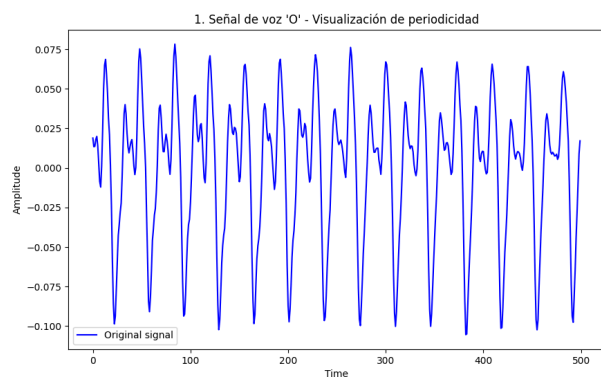
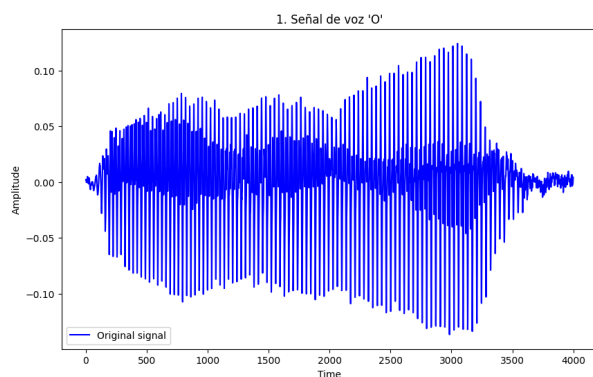
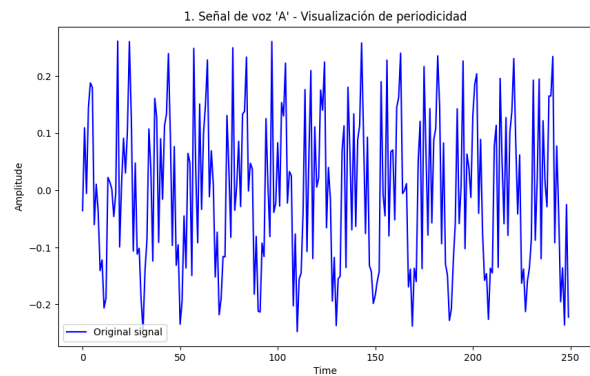
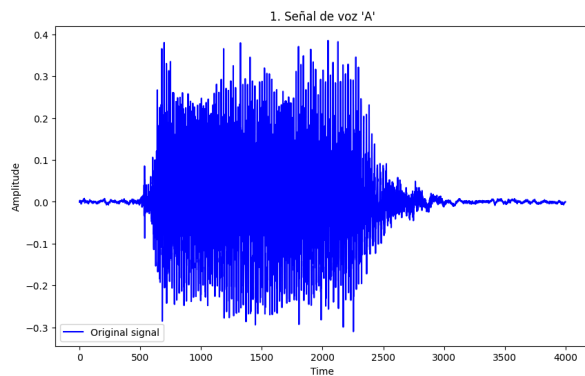
$$w(n) = 0.54 - 0.46 * \cos\left(\frac{2\pi * n}{N-1}\right), 0 \leq n \leq N-1$$

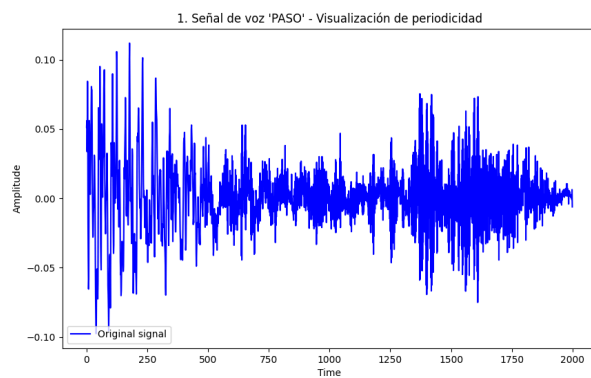
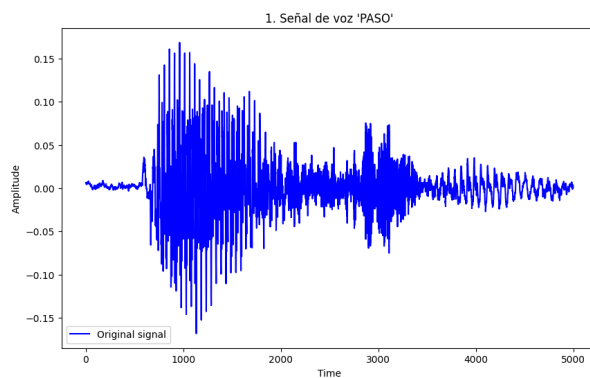
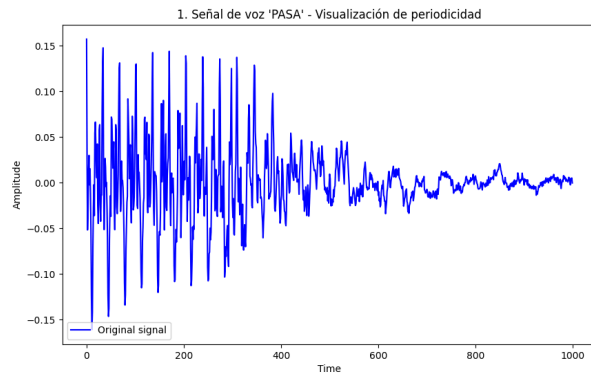
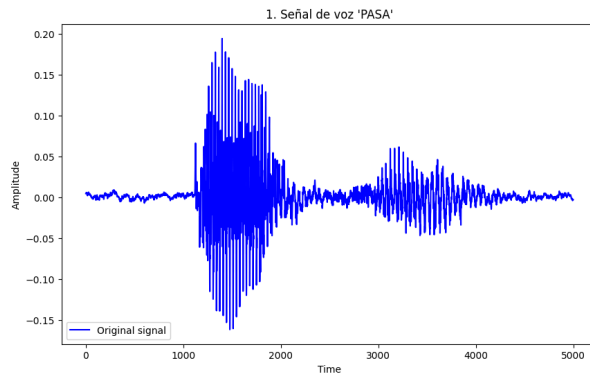
Aplicación de la función de correlación

Esto se usa para obtener coeficientes de LPC y así poder describir/predecir la señal con un número valor de coeficientes y no transmitir la señal completa pero si los coeficientes.

Desarrollo y Resultados

1. Muestrear a 16 KHZ las siguientes señales de voz: 'a', 'o', 'p', 's', 'pasa' y 'paso'



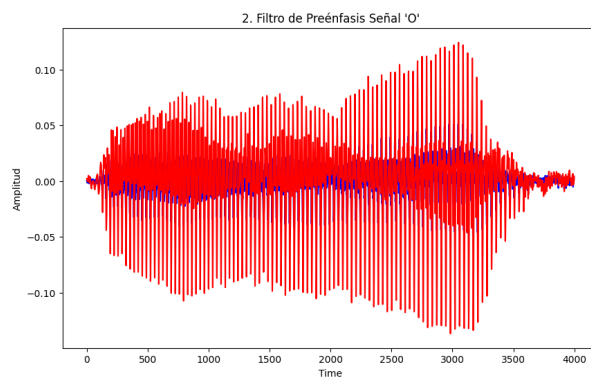
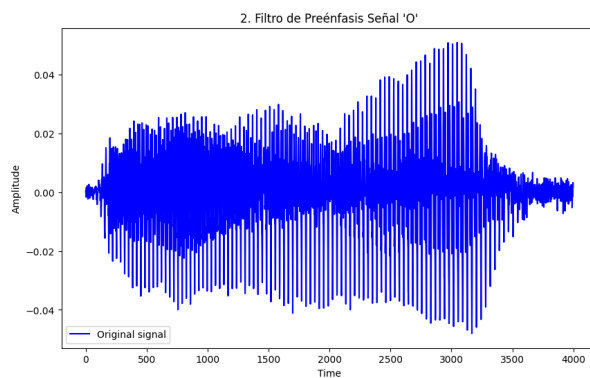


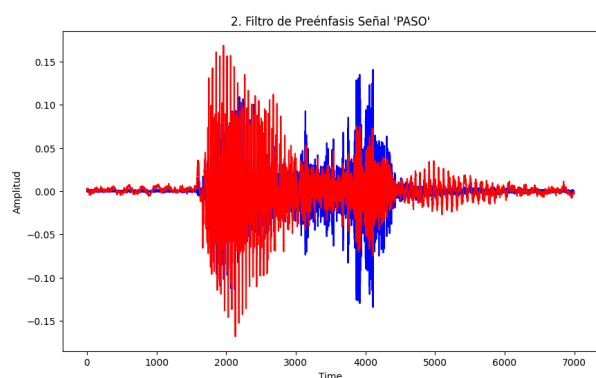
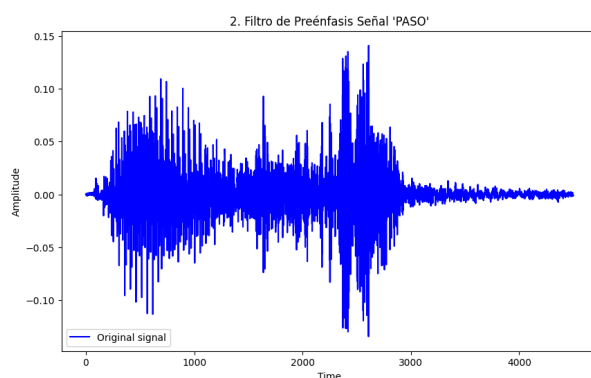
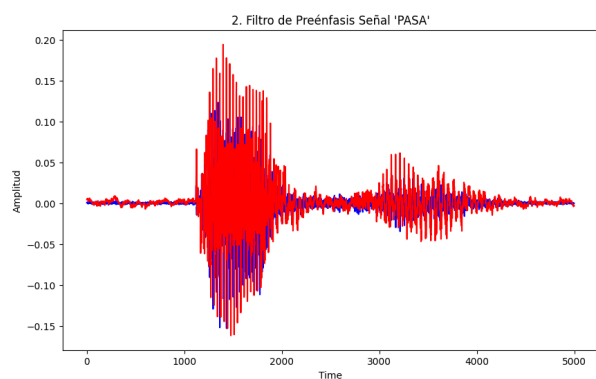
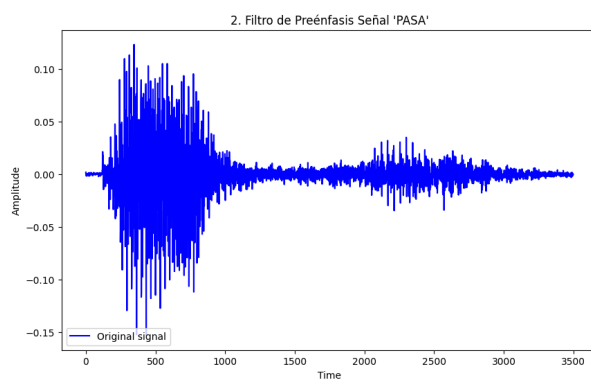
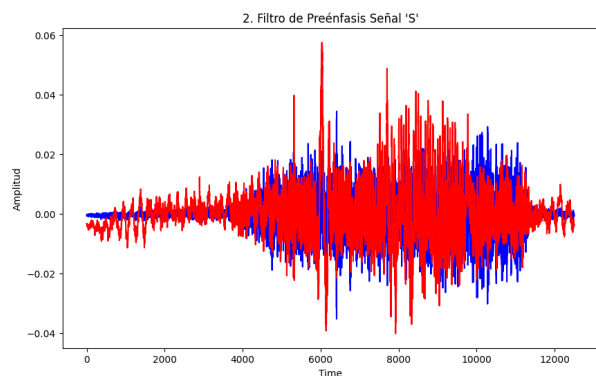
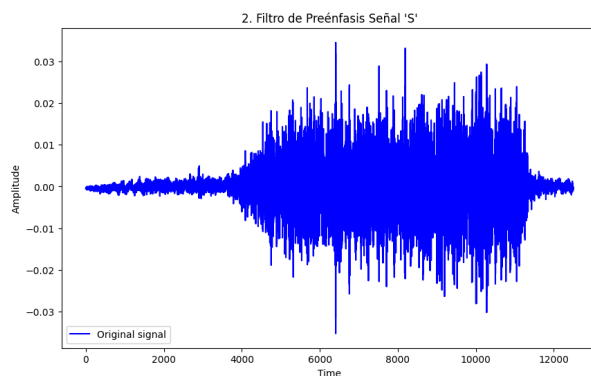
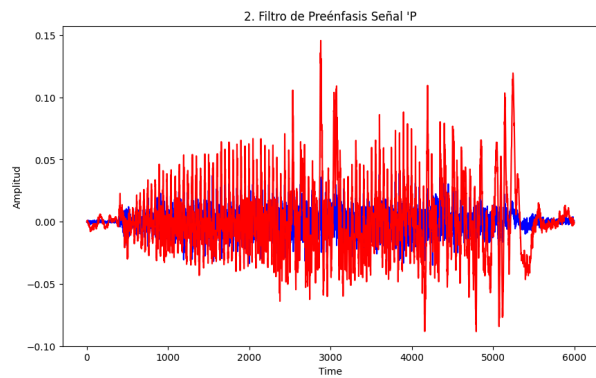
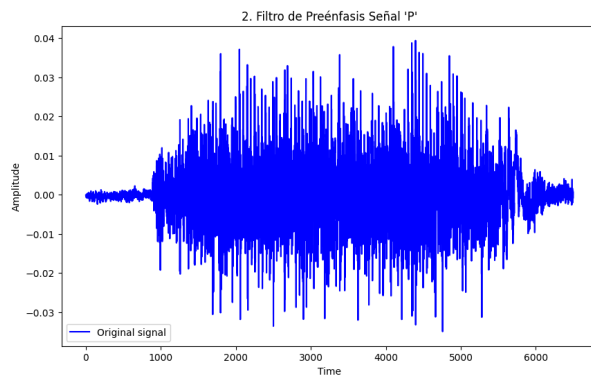
2. Filtro de preénfasis.

Puede notarse cómo el filtro de preénfasis elimina ciertas frecuencias, en el dominio del tiempo puede verse que las amplitudes han sido reducidas o suavizadas.

La señal roja es la señal original.

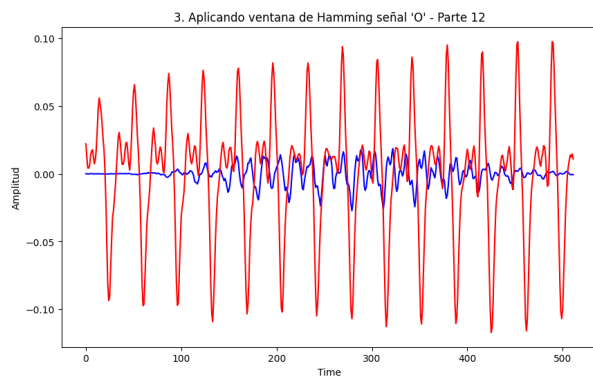
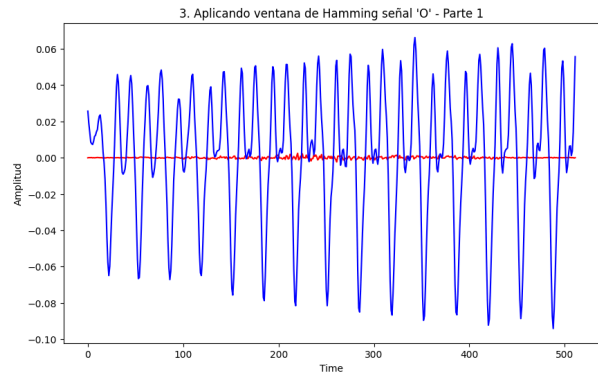
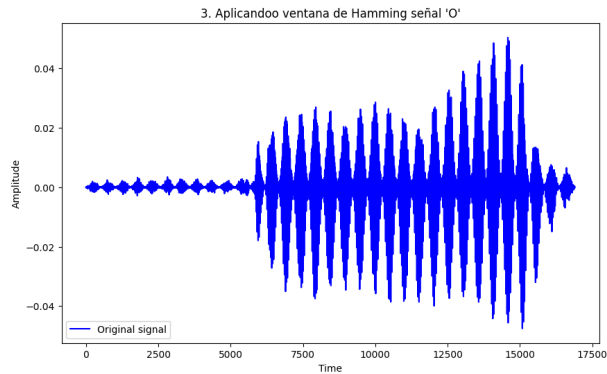
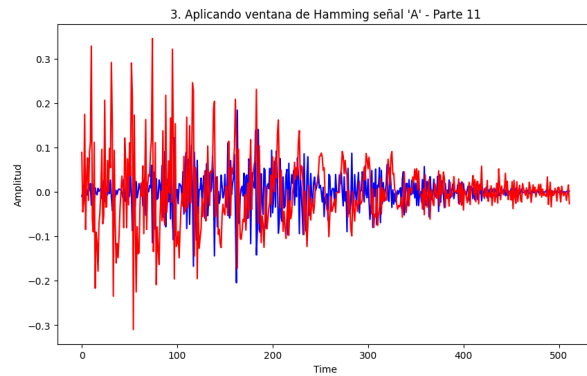
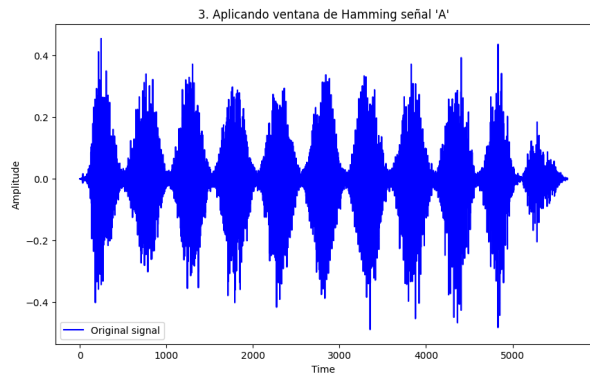
La señal azul es la señal a la que se aplicó el preénfasis.



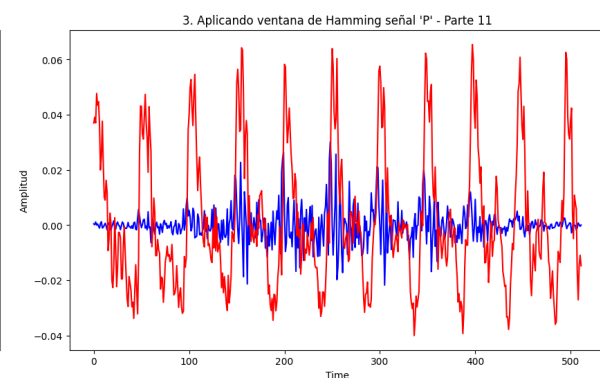
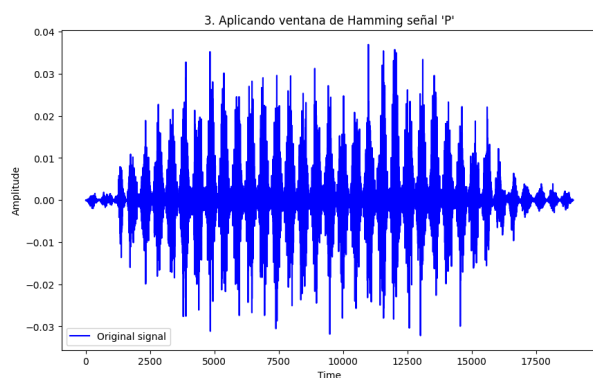


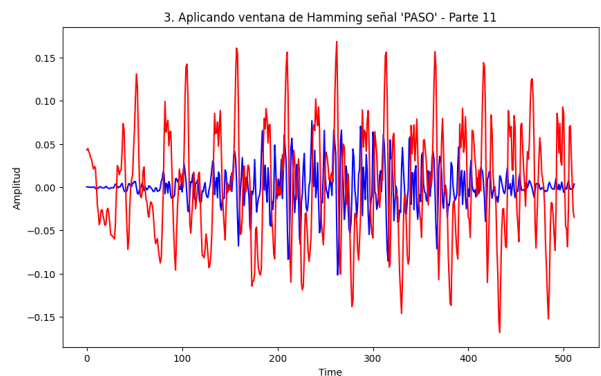
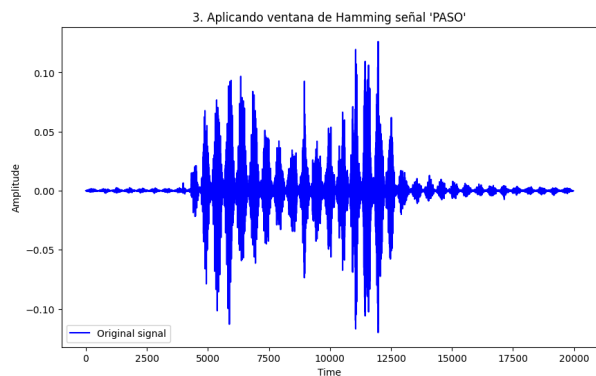
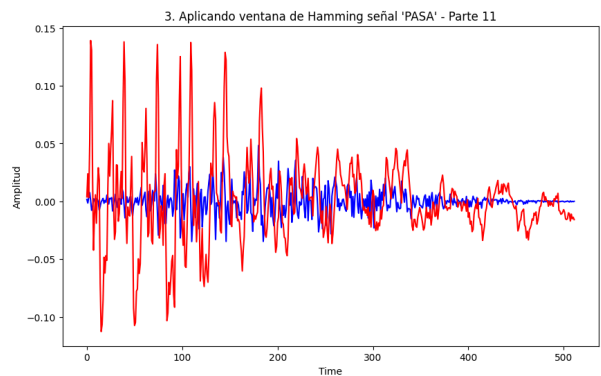
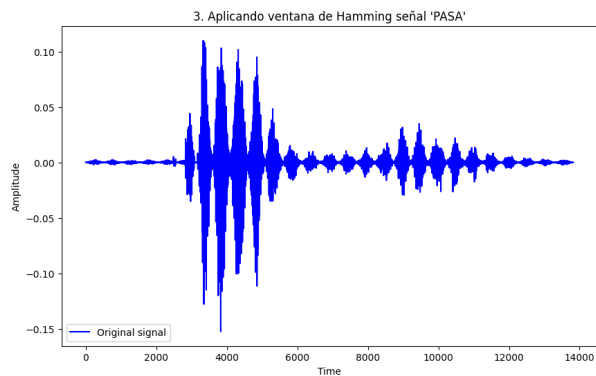
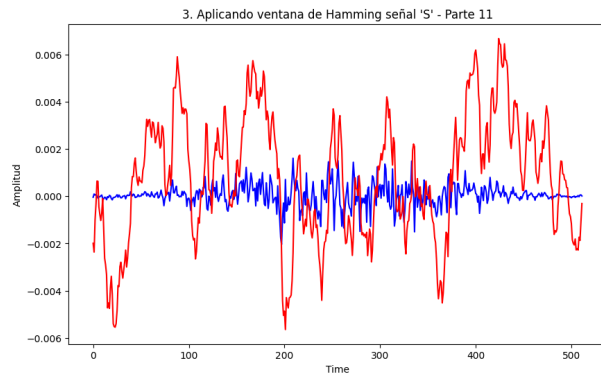
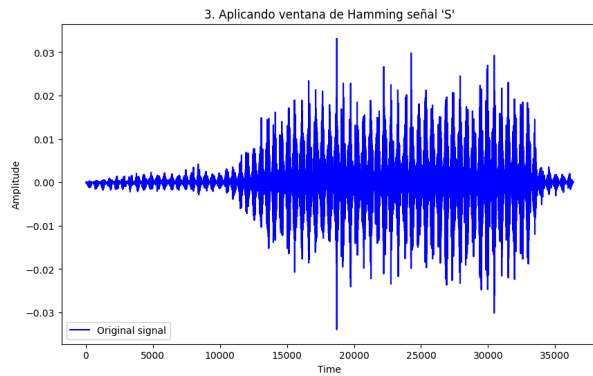
3. Ventana de Hamming

***La señal azul es la señal a la que se aplicó la ventana de Hamming. La señal roja es la señal original**

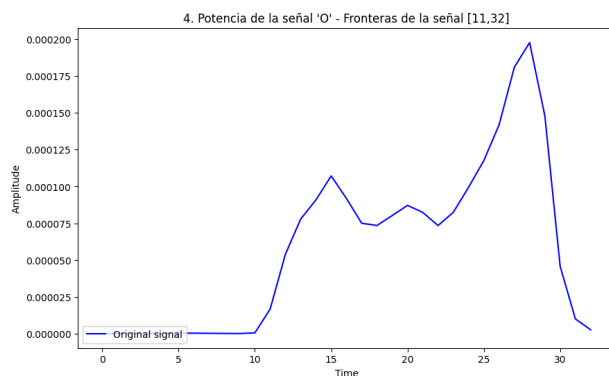


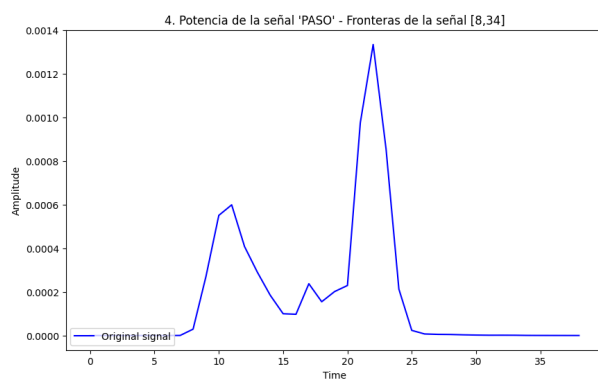
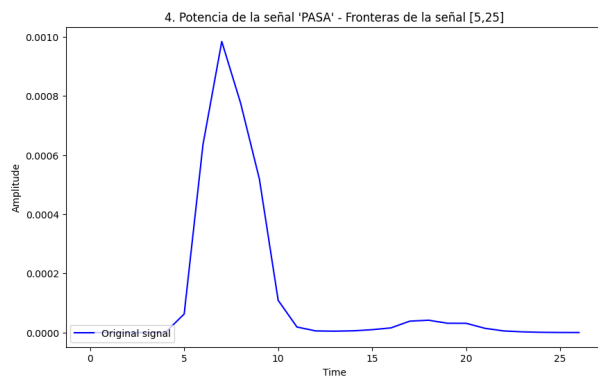
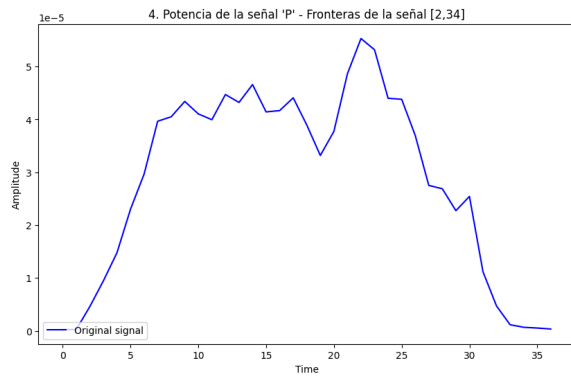
Aquí pueden verse todas las ventanas creadas





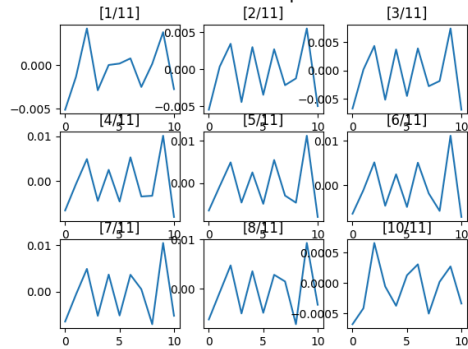
4. Calculando la potencia de la señal y determinando así el inicio y final de la señal.



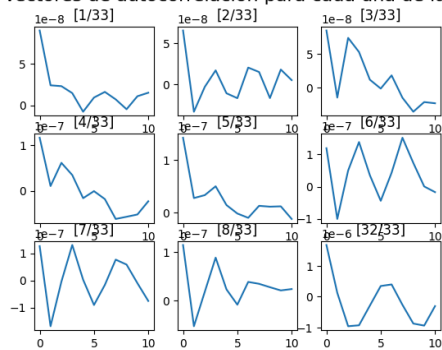


5. Vector de correlación de orden 12 para cada bloque

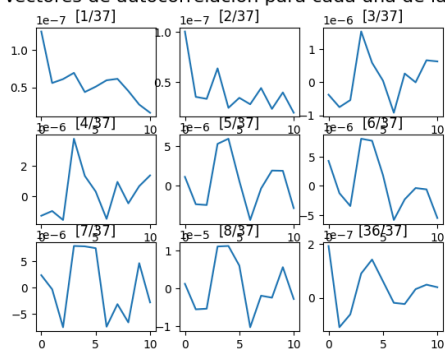
5. Señal 'A' - Vectores de autocorrelación para cada una de las 11 ventanas



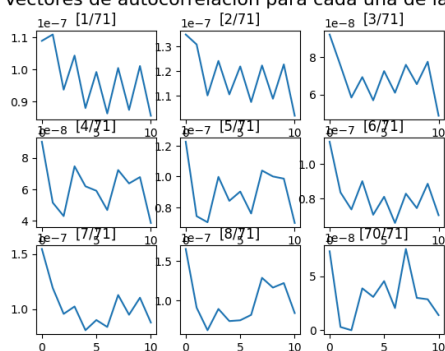
5. Señal 'O' - Vectores de autocorrelación para cada una de las 33 ventanas



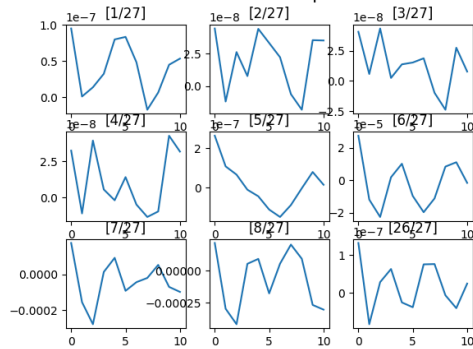
5. Señal 'P' - Vectores de autocorrelación para cada una de las 37 ventanas



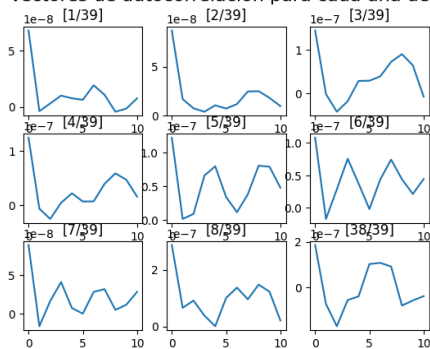
5. Señal 'S' - Vectores de autocorrelación para cada una de las 71 ventanas



5. Señal 'PASA' - Vectores de autocorrelación para cada una de las 27 ventanas

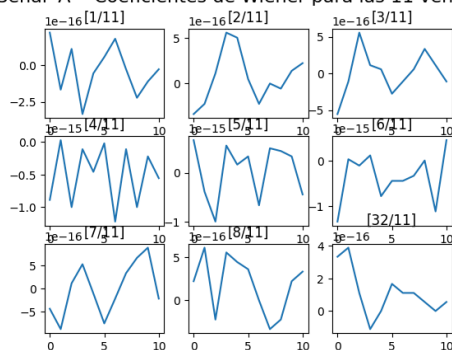


5. Señal 'PASO' - Vectores de autocorrelación para cada una de las 39 ventanas

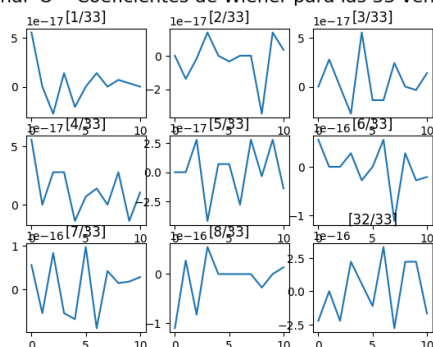


6. Coeficientes de predicción lineal usando filtro de Wiener de orden 12

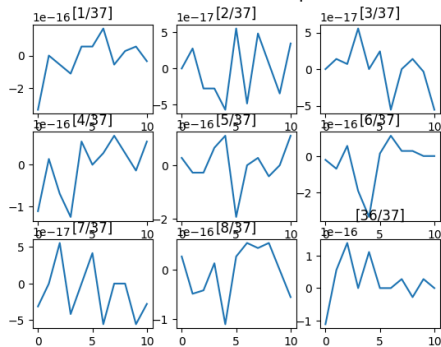
6. Señal 'A' - Coeficientes de Wiener para las 11 ventanas



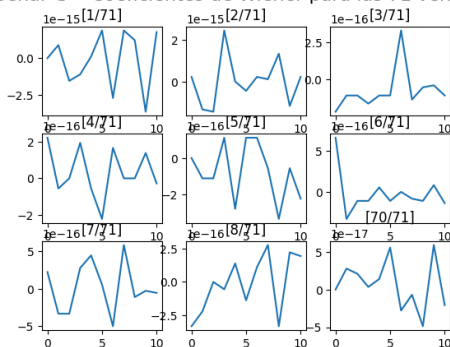
6. Señal 'O' - Coeficientes de Wiener para las 33 ventanas



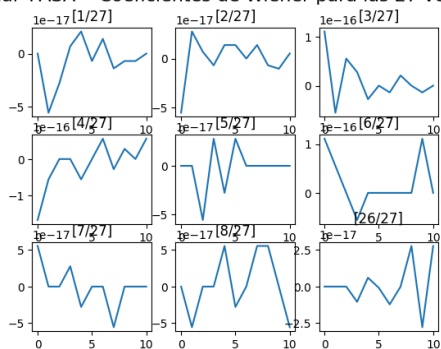
6. Señal 'P' - Coeficientes de Wiener para las 37 ventanas



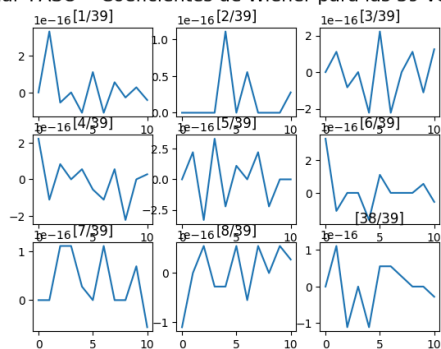
6. Señal 'S' - Coeficientes de Wiener para las 71 ventanas



6. Señal 'PASA' - Coeficientes de Wiener para las 27 ventanas



6. Señal 'PASO' - Coeficientes de Wiener para las 39 ventanas



Conclusiones

Las herramientas de PCL para procesamiento de voz realmente pueden darnos valiosa información de la misma, se puede calcular el inicio y final de una señal después de aplicar varios filtros. Esto se refiere a que el tratamiento de la señal es de sumo valor antes de poder obtener más información útil de la misma.

Finalmente, al aplicar los predictores de Wiener podemos así resumir una señal y reducir el impacto del consumo de energía al tener muchos menos parámetros para poder reconstruir una señal en el destino.

Referencias:

1. Jesús Savage [2024]. “Lección 3: Sistemas de Reconocimiento de Voz”. Clase Procesamiento de señales. UNAM, Primavera 2024.