

Procesamiento digital de señales de audio

Análisis por Predicción Lineal

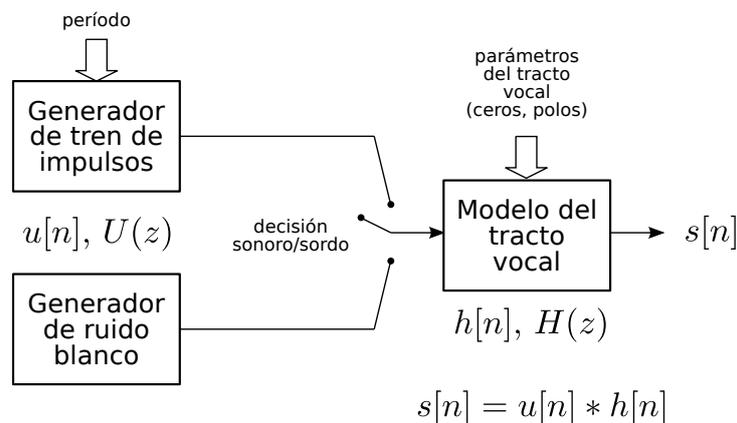
Instituto de Ingeniería Eléctrica
Facultad de Ingeniería



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

- ① Introducción
- ② Principios de LPC
 - Predicción lineal
 - Solución por mínimos cuadrados
- ③ Cálculo de los parámetros del modelo
 - Método de la Autocorrelación
 - Método de la Covarianza
 - Cálculo de la ganancia
- ④ Análisis de señales de voz
 - Consideraciones prácticas
 - LPC en sonidos sonoros
 - LPC en sonidos sordos
- ⑤ Análisis de los resultados
- ⑥ Aplicaciones
 - Análisis de señales de voz
 - Codificación de voz
 - Reconocimiento del habla

Modelo en tiempo discreto del mecanismo de producción de la voz



Modelo general

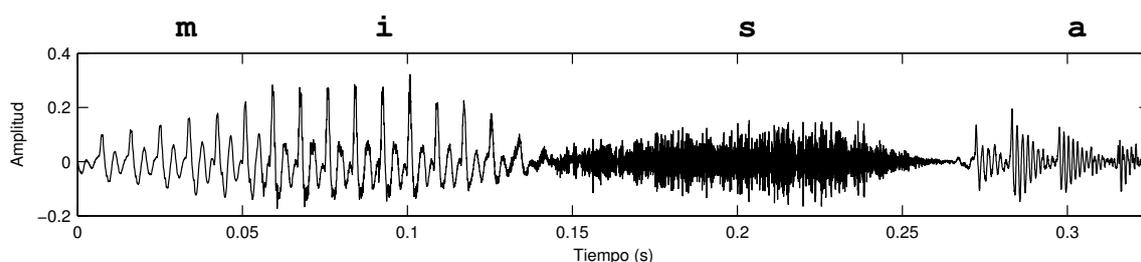
La señal de voz se representa como la salida de un sistema lineal variante en el tiempo.

Modelo en tiempo discreto del mecanismo de producción de la voz

Modelo básico

En períodos cortos de tiempo, cada fragmento de la señal de voz puede modelarse como la salida de un LTI alimentado con

- un tren de pulsos cuasi-periódico (sonidos sonoros)
- ruido blanco (sonidos sordos)



En la práctica los parámetros del modelo no se conocen, solo se conocen las muestras de la señal de voz.

Codificación por Predicción Lineal (LPC)

Orígenes

- LPC impone un modelo en $H(z)$ y estima los parámetros del modelo a partir de $s[n]$
 - $H(z)$ es todo polos
 - técnica paramétrica
- LPC es parte del problema mas general denominado **identificación de sistemas**, desarrollado en el área de la ingeniería de Teoría de Control para el estudio y la identificación de sistemas dinámicos.
- Encontrar un buen modelo del sistema subyacente a una serie temporal permite aplicaciones importantes como predicción, pronóstico, control, compresión.
 - Neurofísica, geofísica, astronomía, economía, sistemas de comunicación.

Codificación por Predicción Lineal

LPC [Makhoul, 1975] [Rabiner and Schafer, 2011]

- LPC es una técnica ampliamente empleada para la estimación de los parámetros del modelo del mecanismo de producción de la voz.
 - frecuencia y ancho de banda de formantes
 - envolvente espectral (estimación de $H(z)$)
- Su importancia radica en que
 - produce estimadores precisos de los parámetros
 - es eficiente computacionalmente

Idea

Una muestra de la señal de voz puede predecirse con bajo error a partir de una combinación lineal de muestras pasadas.

$$\tilde{s}[n] = \sum_{k=1}^p \alpha_k s[n-k] \quad \text{para ciertos valores de } \alpha_1, \dots, \alpha_p$$

Predicción lineal

Modelo general del mecanismo de producción de voz

- La señal de voz $s[n]$ es la salida del sistema con función de transferencia $H(z)$ y entrada $u[n]$ ($H(z)$ y $u[n]$ desconocidos).

$$H(z) = \frac{S(z)}{U(z)} = G \frac{1 + \sum_{l=1}^q b_l z^{-l}}{1 - \sum_{k=1}^p a_k z^{-k}}$$

- $H(z)$ tiene q ceros y p polos
- $S(z)$ y $U(z)$ transformadas \mathcal{Z} de la salida y la entrada respectivamente.

- En el dominio del tiempo, la relación se expresa como

$$s[n] = a_1 s[n-1] + \dots + a_p s[n-p] + G(u[n] + b_1 u[n-1] + \dots + b_q u[n-q])$$

- La salida actual $s[n]$ es una combinación lineal de muestras pasadas de la salida y la muestra actual y muestras pasadas de la entrada.
- $s[n]$ es **predecible** a partir de muestras pasadas de la entrada y la salida.

Predicción lineal

Clasificación del modelo según los parámetros

| Raíces | Parámetros | Modelo estadístico |
|-------------|--------------------------------|-----------------------------------|
| Polos-Ceros | $a_k \neq 0$ $b_l \neq 0$ | Media móvil autorregresivo (ARMA) |
| Todo-ceros | $a_k = 0$ $1 \leq k \leq p$ | Media móvil (MA) |
| Todo-polos | $b_l = 0$ $1 \leq l \leq q$ | Autorregresivo (AR) |

En procesamiento de voz se emplea el **modelo todo-polos**.

Predicción lineal

Modelo todo-polos

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (1)$$

- Los parámetros del modelo de producción de voz son

| Excitación | Función de transferencia |
|----------------------------|-------------------------------------|
| Clasificación sonoro/sordo | Ganancia G |
| Altura | Coeficientes $a_k, 1 \leq k \leq p$ |

- Mediante LPC se obtienen los parámetros de la función de transferencia.
- La excitación puede estimarse filtrando $s[n]$ con el filtro inverso $1/H(z)$.
- En el dominio del tiempo el modelo todo polos se expresa como

$$s[n] = a_1 s[n-1] + \dots + a_p s[n-p] + Gu[n]$$

La salida actual $s[n]$ es una combinación lineal de muestras pasadas de la salida y la muestra actual de la entrada $u[n]$.

Predicción lineal

Observaciones sobre el modelo todo-polos

- La simplificación a un modelo todo-polos implica asumir solo cavidades resonantes en el aparato fonador.
 - Esto es razonable para sonidos sonoros no nasales, donde el sistema todo-polos es un buen modelo del aparato fonador.
 - En el caso de sonidos nasales, fricativos o plosivos, la configuración del aparato fonador tiene cavidades antiresonantes que deben ser modeladas con ceros en la función de transferencia.
- En el caso en que el orden p del modelo es suficientemente grande, el modelo es razonable para todos los sonidos del habla.
- La virtud del modelo todo-polos es que el factor de ganancia G y los coeficientes a_k del filtro pueden estimarse de forma inmediata y eficiente mediante LPC.

Predicción lineal

Solución del problema por mínimos cuadrados

- Modelo todo-polos:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad s[n] = \sum_{k=1}^p a_k s[n-k] + Gu[n]$$

- Como la excitación $u[n]$ es desconocida, se define un estimador de $s[n]$ a partir únicamente de las muestras pasadas de la salida,

$$\tilde{s}[n] = \sum_{k=1}^p \alpha_k s[n-k] \quad \text{Predictor de orden } p$$

- Se define el error de predicción como

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

Predicción lineal

Solución del problema por mínimos cuadrados

Se busca el conjunto de coeficientes $\{\alpha_k\}$ que minimizan el error cuadrático medio en un fragmento de tiempo corto. Los coeficientes $\{\alpha_k\}$ se asumen son los coeficientes $\{a_k\}$ del modelo.

- Se considera el error cuadrático medio: $E_n = \sum_m e_n^2[m]$

El error cuadrático medio es función de $\{\alpha_k\}$ y las muestras de la señal en un fragmento de tiempo corto, por ejemplo en $0 \leq m \leq N-1$,

$$\begin{cases} e[0] &= s[0] - \alpha_1 s[-1] - \dots - \alpha_p s[-p] \\ e[1] &= s[1] - \alpha_1 s[0] - \dots - \alpha_p s[-p+1] \\ \vdots & \vdots \\ e[N-1] &= s[N-1] - \alpha_1 s[N-2] - \dots - \alpha_p s[N-p-1] \end{cases}$$

- Se calcula $\{\hat{\alpha}_k\} = \min_{\{\alpha_k\}} E_n$

Predicción lineal

Error de predicción

$$\begin{aligned}e[n] &= s[n] - \tilde{s}[n] \\ &= s[n] - \sum_{k=1}^p \alpha_k s[n-k] \\ &= \sum_{k=1}^p a_k s[n-k] + Gu[n] - \sum_{k=1}^p \alpha_k s[n-k] \\ &= Gu[n], \text{ si } \{\alpha_k\} = \{a_k\}.\end{aligned}\tag{2}$$

Se concluye que si $s[n]$ obedece exactamente el modelo y si $\{\alpha_k\} = \{a_k\}$,

$$e[n] = Gu[n].\tag{3}$$

Predicción lineal

Filtro de error de predicción

- De la ecuación 2, se puede ver que la función de transferencia entre la señal de voz y el error de predicción es el **filtro inverso**,

$$A(z) = \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^p \alpha_k z^{-k}\tag{4}$$

- Se denomina filtro inverso porque es el filtro inverso de $H(z)$,

$$H(z) = \frac{G}{A(z)}\tag{5}$$

- Se concluye que si se filtra la señal de voz $s[n]$ con el filtro inverso $A(z)$ se obtiene la excitación $e[n] = Gu[n]$.

Predicción lineal

Justificación de la minimización del error de predicción

El problema consiste en encontrar el conjunto de coeficientes $\{\alpha_k\}$ que minimicen el error de predicción $e[n]$.

- Si $\{\alpha_k\} = \{a_k\}$ se cumple que $e[n] = Gu[n]$. Esto significa que para sonidos sonoros, $e[n]$ será un tren de impulsos periódico, es decir, $e[n]$ será pequeño casi todo el tiempo.
- En el caso en que una señal $s[n]$ se genera con el modelo todo-polos excitado con un impulso o ruido blanco estacionario, se puede demostrar que la minimización del error de predicción conduce a coeficientes idénticos a los coeficientes del filtro.
- Una ventaja práctica del enfoque es que la minimización del error cuadrático medio conduce a un sistema lineal de ecuaciones que puede resolverse eficientemente.

Solución del problema

Error cuadrático medio de predicción

- Desarrollando el error cuadrático medio se ve que

$$\begin{aligned} E_n &= \sum_m e_n^2[m] \\ &= \sum_m (s_n[m] - \tilde{s}_n[m])^2 \\ &= \sum_m \left(s_n[m] - \sum_{k=1}^p \alpha_k s_n[m-k] \right)^2 \end{aligned} \quad (6)$$

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

$$s_n[m] = s[m+n].$$

- El rango de la sumatoria se deja momentáneamente sin especificar, pero como la técnica trabaja sobre fragmentos de tiempo corto será finita.

Solución del problema

Minimización del error cuadrático medio de predicción

- Los coeficientes que minimizan el error cuadrático E_n en la ecuación 6 se obtienen estableciendo la derivada respecto a cada uno en cero,

$$\frac{\partial E_n}{\partial \alpha_i} = 0, \quad i = 1, 2, \dots, p.$$

- Esto conduce al siguiente sistema de ecuaciones ([ejercicio](#)) con p ecuaciones y p incógnitas denominado [ecuaciones normales](#):

$$\sum_{k=1}^p \hat{\alpha}_k \sum_m s_n[m-i]s_n[m-k] = \sum_m s_n[m-i]s_n[m], \quad 1 \leq i \leq p \quad (7)$$

- Definiendo $\phi_n[i, k] = \sum_m s_n[m-i]s_n[m-k]$, la ec. 7 puede expresarse como

$$\sum_{k=1}^p \alpha_k \phi_n[i, k] = \phi_n[i, 0], \quad i = 1, 2, \dots, p. \quad (8)$$

Solución del problema

Minimización del error cuadrático medio de predicción

- Sustituyendo los coeficientes que minimizan el error, se obtiene la expresión del error mínimo,

$$E_n = \sum_m s_n^2[m] - \sum_{k=1}^p \alpha_k \sum_m s_n[m]s_n[m-k]$$

que expresado en función de $\phi_n[i, k]$ queda

$$E_n = \phi_n[0, 0] - \sum_{k=1}^p \alpha_k \phi_n[0, k]. \quad (9)$$

[Para obtener el error mínimo se multiplica la i -ésima ecuación del sistema de ecuaciones normales (ec. 7) por α_i y se suman. Luego se sustituye el resultado en la ecuación 6 del error cuadrático, [ejercicio](#)]

Solución del problema

Cálculo de los coeficientes del predictor

1. Calcular

$$\phi_n[i, k] = \sum_m s_n[m - i]s_n[m - k], \text{ para } 1 \leq i \leq p, 0 \leq k \leq p \quad (10)$$

2. Resolver el sistema lineal de ecuaciones

$$\sum_{k=1}^p \alpha_k \phi_n[i, k] = \phi_n[i, 0], \quad i = 1, 2, \dots, p.$$

Enfoques

- Hay dos enfoques para resolver el problema que difieren en la elección de los límites de la sumatoria en la ecuación 6 del error cuadrático medio.
- Los límites de la sumatoria en el error coinciden con los de la ecuación 10.
- Como el análisis es de tiempo corto, los límites de la sumatoria son finitos.

Método de la Autocorrelación

Cálculo de los coeficientes del predictor

- El error cuadrático medio se calcula entre $-\infty$ y ∞ ,

$$E_n = \sum_{m=-\infty}^{+\infty} e_n^2[m] \implies \phi_n[i, k] = \sum_{m=-\infty}^{+\infty} s_n[m - i]s_n[m - k], \quad \begin{matrix} 1 \leq i \leq p \\ 0 \leq k \leq p \end{matrix}$$

- Se asume que $s_n[m]$ es nula fuera del intervalo $0 \leq m \leq N - 1$,

$$s_n[m] = s[m + n]w[m],$$

con $w[m]$ ventana suavizante nula fuera del intervalo $0 \leq m \leq N - 1$.

Método de la Autocorrelación

Cálculo de los coeficientes del predictor

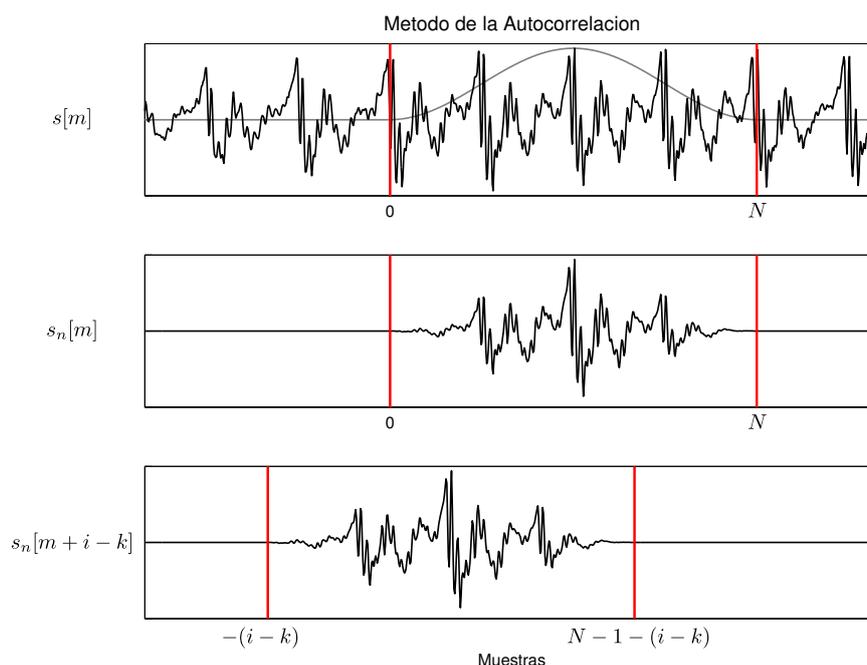
- Como los límites de la suma son entre $-\infty$ y ∞ , $\phi_n[i, k]$ es la autocorrelación de la señal enventanada:

$$\begin{aligned}\phi_n[i, k] &= \sum_{m=-\infty}^{+\infty} s_n[m - i]s_n[m - k] \\ &= \sum_{m=-\infty}^{+\infty} s_n[m]s_n[m + i - k] \\ &= R_n[|i - k|],\end{aligned}$$

- Como la señal enventanada es nula fuera del intervalo $0 \leq m \leq N - 1$, la suma tiene una cantidad de términos finita,

$$R_n[|i - k|] = \sum_{m=0}^{N-1-|i-k|} s_n[m]s_n[m + |i - k|] \quad (11)$$

Método de la Autocorrelación



Método de la Autocorrelación

Cálculo de los coeficientes del predictor

Sustituyendo $\phi_n[i, k] = R_n[|i - k|]$ en las ecuaciones 8 y 9 se obtiene las ecuaciones del método de la autocorrelación:

- Las ecuaciones normales quedan en este caso

$$\sum_{k=1}^p \alpha_k R_n[|i - k|] = R_n[i], \quad 1 \leq i \leq p. \quad (12)$$

- El error cuadrático medio mínimo es

$$E_n = R_n[0] - \sum_{k=1}^p \alpha_k R_n[k]. \quad (13)$$

Método de la Autocorrelación

Cálculo de los coeficientes del predictor

El sistema de ecuaciones expresado en forma matricial queda

$$\begin{pmatrix} R_n[0] & R_n[1] & R_n[2] & \cdots & R_n[p-1] \\ R_n[1] & R_n[0] & R_n[1] & \cdots & R_n[p-2] \\ R_n[2] & R_n[1] & R_n[0] & \cdots & R_n[p-3] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_n[p-1] & R_n[p-2] & R_n[p-3] & \cdots & R_n[0] \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} R_n[1] \\ R_n[2] \\ R_n[3] \\ \vdots \\ R_n[p] \end{pmatrix} \quad (14)$$

- Para calcular los coeficientes hay que invertir una matriz $p \times p$ Toeplitz.
- Existen algoritmos eficientes para hacerlo.

Método de la Covarianza

Cálculo de los coeficientes del predictor

- El error cuadrático medio se calcula entre 0 y $N - 1$ sin asumir nada sobre la señal $s[n]$:

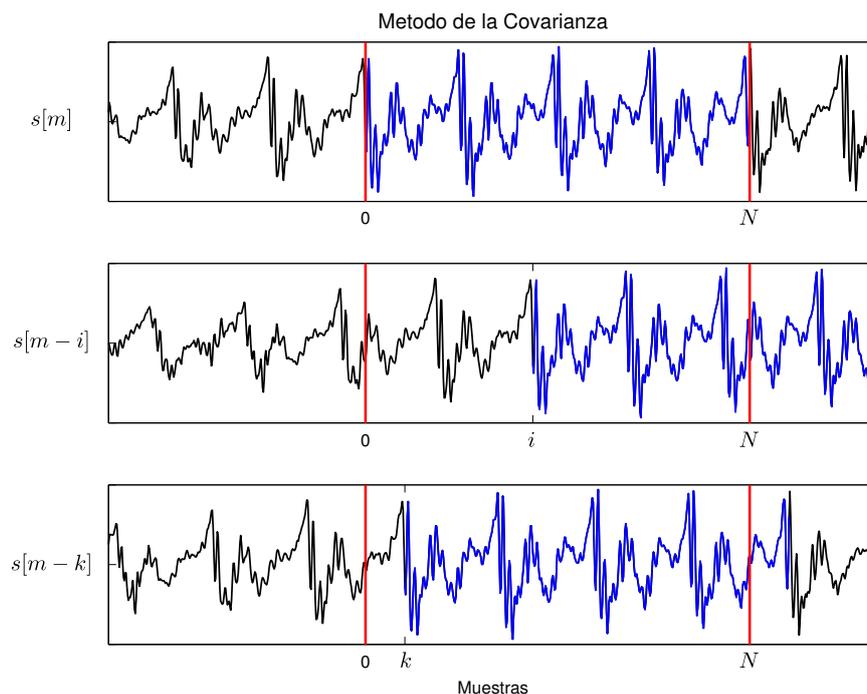
$$E_n = \sum_{m=0}^{N-1} e_n^2[m] \implies \phi_n[i, k] = \sum_{m=0}^{N-1} s_n[m - i]s_n[m - k], \quad \begin{matrix} 1 \leq i \leq p \\ 0 \leq k \leq p \end{matrix}$$

- Haciendo un cambio de variable se llega a que

$$\phi_n[i, k] = \sum_{m=-i}^{N-1-i} s_n[m]s_n[m + i - k], \quad \begin{matrix} 1 \leq i \leq p \\ 0 \leq k \leq p \end{matrix}$$

- Para evaluar $\phi_n[i, k]$ en los valores requeridos de i y k se necesitan valores de $s[m]$ en el intervalo $-p \leq m \leq N - 1$.

Método de la Covarianza



Método de la Covarianza

Cálculo de los coeficientes del predictor

- Las ecuaciones normales quedan en este caso

$$\sum_{k=1}^p \alpha_k \phi_n[i, k] = \phi_n[i, 0], \quad 1 \leq i \leq p.$$

- El error cuadrático medio mínimo es

$$E_n = \phi_n[0, 0] - \sum_{k=1}^p \alpha_k \phi_n[0, k].$$

Método de la Covarianza

Cálculo de los coeficientes del predictor

El sistema de ecuaciones expresado en forma matricial queda

$$\begin{pmatrix} \phi_n[1, 1] & \phi_n[1, 2] & \phi_n[1, 3] & \cdots & \phi_n[1, p] \\ \phi_n[2, 1] & \phi_n[2, 2] & \phi_n[2, 3] & \cdots & \phi_n[2, p] \\ \phi_n[3, 1] & \phi_n[3, 2] & \phi_n[3, 3] & \cdots & \phi_n[3, p] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_n[p, 1] & \phi_n[p, 2] & \phi_n[p, 3] & \cdots & \phi_n[p, p] \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} \phi_n[1, 0] \\ \phi_n[2, 0] \\ \phi_n[3, 0] \\ \vdots \\ \phi_n[p, 0] \end{pmatrix} \quad (15)$$

- La matriz $p \times p$ a invertir tiene las características de una matriz de covarianza: es simétrica y definida positiva.
- No es Toeplitz,

$$\phi_n[i + 1, k + 1] = \phi_n(i, k) + s[-i - 1]s[-k - 1] - s[N - 1 - i]s[N - 1 - k]$$

Planteo como problema de mínimos cuadrados

Predicción lineal

Se utiliza un predictor de orden p para predecir cada muestra de la señal $s[m]$ definida en el intervalo $0 \leq m \leq N - 1$.

$$\left\{ \begin{array}{l} \alpha_1 s[-1] + \alpha_2 s[-2] + \dots + \alpha_p s[-p] = s[0] \\ \alpha_1 s[0] + \alpha_2 s[-1] + \dots + \alpha_p s[-p+1] = s[1] \\ \vdots \\ \alpha_1 s[p-2] + \alpha_2 s[p-3] + \dots + \alpha_p s[-1] = s[p-1] \\ \alpha_1 s[p-1] + \alpha_2 s[p-2] + \dots + \alpha_p s[0] = s[p] \\ \vdots \\ \alpha_1 s[N-2] + \alpha_2 s[N-3] + \dots + \alpha_p s[N-p-1] = s[N-1] \end{array} \right.$$

- Para encontrar los coeficientes del predictor $\{\alpha_k\}$, el problema consiste en resolver un sistema de N ecuaciones y p incógnitas.
- Como en general $N \gg p$, el sistema es sobredeterminado y no tiene solución única.

Planteo como problema de mínimos cuadrados

Predicción lineal

Expresado en notación matricial, el sistema queda

$$\underbrace{\begin{pmatrix} s[-1] & s[-2] & s[-3] & \dots & s[-p] \\ s[0] & s[-1] & s[-2] & \dots & s[-p+1] \\ s[1] & s[0] & s[-1] & \dots & s[-p+2] \\ \vdots & \vdots & \vdots & & \vdots \\ s[p-2] & s[p-3] & s[p-4] & \dots & s[-1] \\ s[p-1] & s[p-2] & s[p-3] & \dots & s[0] \\ \vdots & \vdots & \vdots & & \vdots \\ s[N-2] & s[N-3] & s[N-4] & \dots & s[N-p-1] \end{pmatrix}}_A \underbrace{\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_p \end{pmatrix}}_{\alpha} = \underbrace{\begin{pmatrix} s[0] \\ s[1] \\ s[2] \\ \vdots \\ s[p-1] \\ s[p] \\ \vdots \\ s[N-1] \end{pmatrix}}_b$$

- A es $N \times p$, α es $p \times 1$ y b es $N \times 1$.
- El sistema $A\alpha = b$ es sobredeterminado.

Planteo como problema de mínimos cuadrados

Predicción lineal

- Se busca el vector α que “mejor” aproxima $A\alpha = b$.
- Se define el residuo como el vector $e = b - A\alpha$ y se busca

$$\hat{\alpha} = \min_{\alpha} \|e\|_2^2 = \min_{\alpha} e^T e$$

- La solución del problema son las **ecuaciones normales**,

$$A^T A \hat{\alpha} = A^T b.$$

- En el método de la **covarianza**,

$$(A^T A)_{ij} = \phi[i, j], \quad 1 \leq i, j \leq p$$

$$A^T b = \phi[i, 0], \quad 1 \leq i \leq p,$$

y se necesita conocer $s[m]$ en el intervalo $-p \leq m \leq N-1$ para construir la matriz A y el vector b .

Planteo como problema de mínimos cuadrados

Predicción lineal. Método de la Autocorrelación.

En el método de la **autocorrelación** el error $e[m]$ se minimiza para todo m .

$$\begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ s[0] & 0 & 0 & \dots & 0 \\ s[1] & s[0] & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s[p-2] & s[p-3] & s[p-4] & \dots & 0 \\ s[p-1] & s[p-2] & s[p-3] & \dots & s[0] \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ s[N-2] & s[N-3] & s[N-4] & \dots & s[N-p-1] \\ s[N-1] & s[N-2] & s[N-3] & \dots & s[N-p] \\ 0 & s[N-1] & s[N-2] & \dots & s[N-p+1] \\ 0 & 0 & s[N-1] & \dots & s[N-p+2] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & s[N-1] \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} s[0] \\ s[1] \\ s[2] \\ \vdots \\ s[p-1] \\ s[p] \\ \vdots \\ s[N-1] \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Consideraciones computacionales

Método de la autocorrelación

Se minimiza el error de predicción $e[n]$ en $-\infty < m < \infty$ pero se eventana la señal para realizar el análisis en tiempo corto.

- En el comienzo de la ventana se intenta predecir muestras no nulas a partir de muestras artificialmente nulas.
- Luego del final de la ventana se intenta predecir muestras artificialmente nulas a partir de muestras no nulas.

En el método de la autocorrelación la señal debe eventanarse con una ventana suavizante para hacer pequeñas las muestras de los bordes.

Método de la covarianza

Se minimiza el error de predicción $e[n]$ en un rango finito de valores, $0 \leq m \leq N - 1$.

- No es necesario eventanar la señal.
- Se necesitan muestras de la señal $s[n]$ en el rango $-p \leq m \leq N - 1$.

Cálculo de la ganancia

Cálculo del parámetro G

- Modelo:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \implies s[n] = Gu[n] + \sum_{k=1}^p a_k s[n-k]$$

- Del error de predicción (ec. 2): $s[n] = e[n] + \sum_{k=1}^p \alpha_k s[n-k]$
- La única entrada $u[n]$ que resulta en la salida $s[n]$ es $Gu[n] = e[n]$.
- Se busca que para cualquier entrada $u[n]$, la energía de la salida sea igual a la energía de la señal original $s[n]$.
- Para calcular G hay que imponer hipótesis sobre la entrada $u[n]$:
 - Sonidos sonoros: $u[n] = \delta[n]$
 - Sonidos sordos: $u[n]$ ruido blanco de media nula y varianza unidad.

Cálculo de la ganancia

Calculo de G para sonidos sonoros

- Si la entrada es $u[n] = \delta[n]$, la salida es la respuesta al impulso $h[n]$,

$$h[n] = \sum_{k=1}^p \alpha_k h[n-k] + G\delta[n]. \quad (16)$$

- Puede demostrarse (ejercicio) que la autocorrelación $\hat{R}[m]$ de $h[n]$ cumple que

$$\hat{R}[m] = \sum_{k=1}^p \alpha_k \hat{R}[|m-k|], \quad 1 \leq m \leq \infty. \quad (17)$$

$$\hat{R}[0] = \sum_{k=1}^p \alpha_k \hat{R}[k] + G^2. \quad (18)$$

[Multiplicar la ecuación 16 por $h[n-m]$ y sumar en n .]

Cálculo de la ganancia

Calculo de G para sonidos sonoros

- Imponiendo la condición de que la energía de la salida $h[n]$ sea igual a la energía de $s[n]$, se tiene que $\hat{R}[0] = R[0]$.
- Comparando las ecuaciones normales usando el método de la autocorrelación (ec. 12) con la ecuación 17, y empleando además lo indicado arriba, se tiene que

$$\hat{R}[m] = R[m], \quad 0 \leq m \leq p, \quad (19)$$

es decir, las primeras $p+1$ muestras de la autocorrelación de la respuesta al impulso y la autocorrelación de la señal son iguales.

- Sustituyendo este resultado en la ecuación 18 se llega a que,

$$G^2 = R[0] - \sum_{k=1}^p \alpha_k R[k] = E_n, \quad (20)$$

en donde en la segunda igualdad se empleó la ecuación 13.

Cálculo de la ganancia

Calculo de G para sonidos sordos

- La entrada es ahora ruido blanco de media nula y varianza unidad,

$$E\{u[n]\} = 0, \quad E\{u[n]u[n-m]\} = \delta[m]$$

- La salida del sistema $g[n] = \sum_{k=1}^p \alpha_k g[n-k] + Gg[n]$.

- Puede demostrarse (ejercicio) que la autocorrelación $\hat{R}[m]$ de $g[n]$ cumple que

$$\hat{R}[m] = \sum_{k=1}^p \alpha_k \hat{R}[|m-k|], \quad 1 \leq m \leq \infty \quad (21)$$

$$\hat{R}[0] = \sum_{k=1}^p \alpha_k \hat{R}[k] + G^2. \quad (22)$$

- Análogamente al caso anterior, se concluye que $G^2 = E_n$.

Cálculo de la ganancia

Conclusiones

- La ecuación 19 indica que las primeras $p + 1$ muestras de la autocorrelación de la respuesta al impulso $h[n]$ son iguales a las de la autocorrelación de la señal $s[n]$.
- El problema de predicción lineal consiste en encontrar un filtro que:
 - sea todo polos, como en la ecuación 1.
 - la primeras $p + 1$ muestras de la autocorrelación de la respuesta al impulso coincidan con las primeras $p + 1$ muestras de la autocorrelación de la señal.
 - la autocorrelación de la respuesta al impulso cumpla la ecuación 17.
- Las ecuaciones 21 y 22 de la autocorrelación de la salida cuando la entrada es ruido blanco se llaman ecuaciones de **Yule-Walker**.
- Las ecuaciones 17-18 y 21-22 indican que la autocorrelación de la salida es idéntica si la entrada es un impulso o ruido blanco. Esta dualidad se debe a que la autocorrelación de un impulso y de ruido blanco son iguales.

Comparación entre los métodos

Solución del sistema lineal de ecuaciones

- Hay que resolver un sistema lineal de ecuaciones con p ecuaciones y p incógnitas.
- Las particularidades de la matriz del sistema lineal hace que sea posible su solución de forma mas eficiente que el caso general.

Método de la autocorrelación

- La matriz del sistema es Toeplitz y simétrica (matriz de autocorrelación)
- El sistema lineal se resuelve eficientemente mediante la recursión de Durbin.
- Se calculan todas las soluciones de orden 1 hasta p .
- Los coeficientes obtenidos corresponden a un sistema estable.

Método de la covarianza

- La matriz del sistema es simétrica y definida positiva (matriz de covarianza)
- El sistema lineal se resuelve eficientemente mediante la descomposición de Cholesky.
- Se obtiene el error RMS para todos los valores hasta p .
- No hay garantía de que el sistema sea estable.

Consideraciones prácticas

Análisis LPC de una ventana con el método de la autocorrelación

1. Cálculo de la autocorrelación $R_n[k]$ de $s_n[m]$ para $0 \leq k \leq p$ (ec. 11).
2. Armado de la matriz de autocorrelación teniendo en cuenta que es Toeplitz y simétrica (ver ec. 14).
3. Solución del sistema de la ecuación 14 para obtener los coeficientes α .
4. Cálculo del error RMS E_n a partir de α y $R[k]$ usando la ecuación 13.
5. Cálculo de la ganancia con la ecuación 20: $G = \sqrt{E_n}$.

Se obtiene la función de transferencia del sistema

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^p \alpha_k z^{-k}}$$

y el filtro inverso

$$A(z) = \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^p \alpha_k z^{-k}.$$

Consideraciones prácticas

Análisis LPC de una ventana con el método de la autocorrelación

```
from scipy.linalg import toeplitz
from scipy.linalg import solve
import numpy as np

def lpc_analysis(s, p):

    # Calculo de la autocorrelacion
    N = s.shape[0]
    r = np.zeros((p+1, 1))
    for k in range(p+1):
        r[k] = np.dot(s[:N-k].T, s[k:])

    # Calculo de los coeficientes LPC
    ak = solve(toeplitz(r[:p]), r[1:p+1]).squeeze()
    e_rms = r[0] - np.dot(ak.T, r[1:p+1])
    e_rms_norm = e_rms/r[0]

    return ak, e_rms, e_rms_norm
```

Consideraciones prácticas

Análisis LPC de tiempo corto con la autocorrelación

1. Elección del orden p del sistema y el largo N de ventana.
2. Obtención de la señal de tiempo corto $s_n[m]$ mediante enventanado (Hanning, Hamming, etc).
3. Cálculo de los parámetros del filtro $H(z)$ con LPC.

Uso de los parámetros del filtro $H(z)$

- Cálculo de la respuesta al impulso $h[n]$ filtrando $u[n] = \delta[n]$.
- Cálculo de la excitación $e[n] = Gu[n]$ filtrando $s[n]$ con el filtro inverso $A(z)$.
- Cálculo de la frecuencia y ancho de banda de las formantes a partir de los polos de $H(z)$.
- Decisión sonoro/sordo a partir del error RMS normalizado.
- Detección de la frecuencia fundamental usando la excitación $e[n]$.

Consideraciones prácticas

Elección de p

Transferencia del modelo desde la fuente glotal hasta la salida a través de los labios

- Sonido sonoro

$$H(z) = A_v G(z) V(z) R(z)$$

- A_v : ganancia de la fuente glotal
- $G(z)$: modelo del pulso glotal
- $R(z)$: modelo de radiación
- $V(z)$: transferencia del tracto vocal

- Sonido sordo

$$H(z) = A_n V(z) R(z)$$

- A_n : ganancia de la fuente
- $R(z)$: modelo de radiación
- $V(z)$: transferencia del tracto vocal

Consideraciones prácticas

Elección de p

- Depende de la frecuencia de muestreo y es independiente del método usado.
- El tracto vocal $V(z)$ tiene aproximadamente una formante por kHz. Como se necesita un par de polos conjugados para crear una formante,

$$p \approx f_s / 1000.$$

- Se necesitan 3 o 4 polos mas para modelar $G(z)R(z)$.

Ejemplo: si $f_s = 16000$ Hz, $p \approx 20$.

- Para sonidos sordos en general se necesitan menos polos
 - No hay pulso glotal ($G(z) = 1$).
 - El tracto vocal $V(z)$ tiene menos formantes.

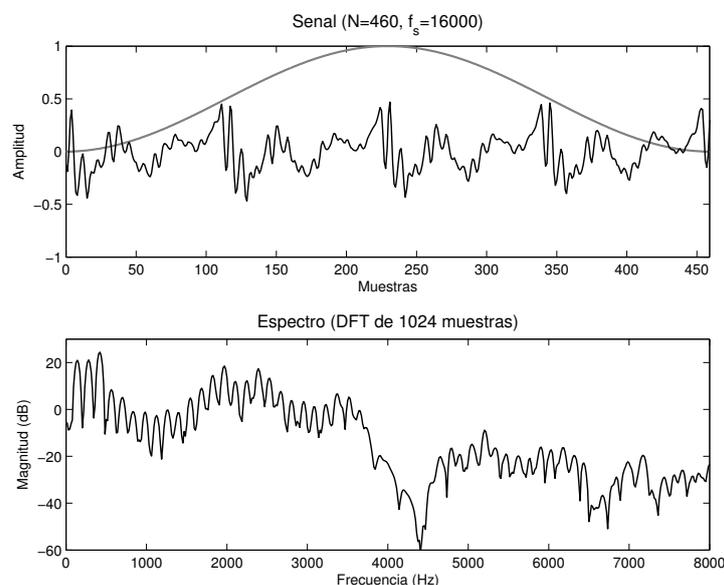
Consideraciones prácticas

Elección de N

- Es ventajoso mantener N lo mas pequeño posible
 - Análisis de tiempo corto efectivo
 - Mantener el costo computacional bajo
- En el método de la autocorrelación es necesario que N contenga varios períodos de la señal debido al enventanado.
- Típicamente se emplean ventanas de análisis de entre 10 a 40 milisegundos en implementaciones de LPC con el método de la autocorrelación.
- En el método de la covarianza no hay restricción en el tamaño de N , pero en la práctica se emplean ventanas del mismo orden para ambos métodos.

Ejemplo. Método de la autocorrelación.

Análisis de sonido sonoro. Enventanado.



Ejemplo. Método de la autocorrelación.

Estimación de $H(z)$

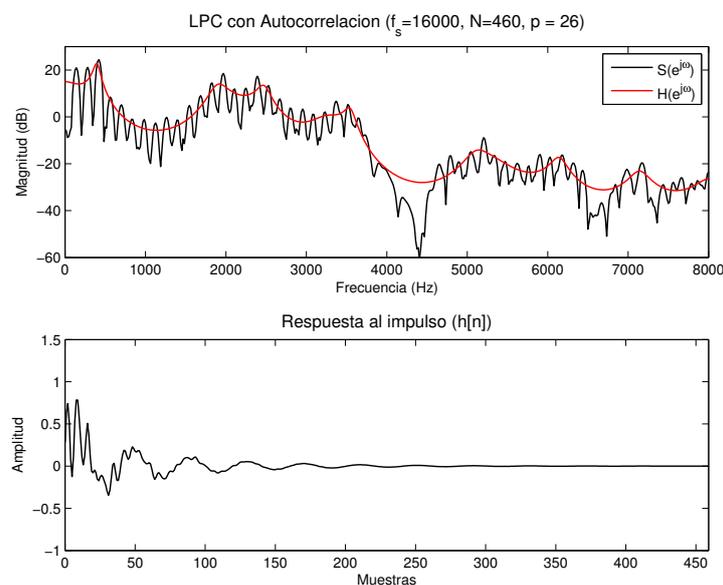
- Si se evalúa el modelo calculado en $z = e^{j\omega}$, se obtiene la respuesta en frecuencia del tracto vocal,

$$H(e^{j\omega}) = \frac{G}{1 - \sum_{k=1}^p \alpha_k e^{-jk\omega}}$$

- Los picos de la envolvente espectral, causados por cavidades de resonancia del tracto vocal, son modelados con los polos de $H(z)$.
- Los valles de la envolvente espectral, causadas por antiresonancias del tracto vocal, deberían ser modelados con ceros en la función de transferencia.
- Como el modelo es todo-polos, la respuesta en frecuencia no modela los valles de la envolvente espectral.
- Debido a que $H(e^{j\omega})$ tiene solo polos dentro del círculo unidad, es una representación de fase mínima del sistema verdadero.

Ejemplo. Método de la autocorrelación.

Estimación de $H(z)$.



Ejemplo. Método de la autocorrelación.

Estimación de la excitación $e[n] = Gu[n]$.

- Un subproducto del análisis LPC es la estimación del pulso glotal a partir del error de predicción,

$$e[n] = Gu[n].$$

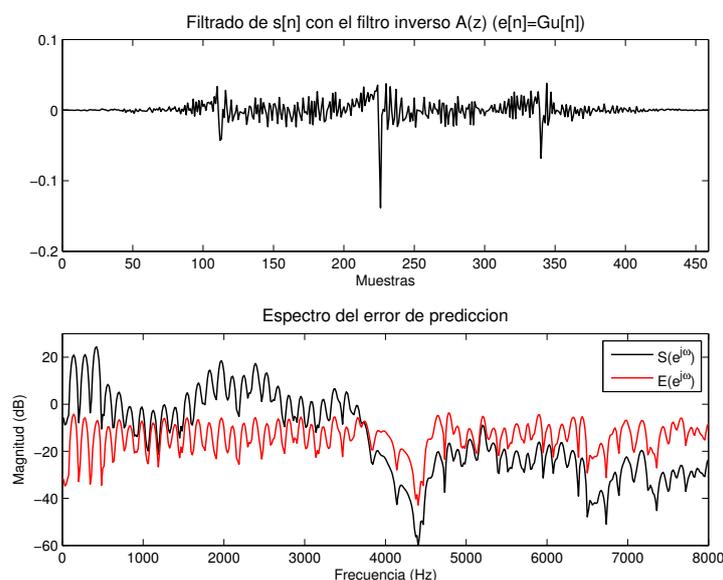
- El error de predicción se obtiene filtrando la señal $s[n]$ con el filtro inverso $A(z)$,

$$A(z) = 1 - \sum_{k=1}^p \alpha_k z^{-k}.$$

- La señal $s[n]$ correspondiente a un sonido sonoro es predecible excepto en los momentos en donde en la entrada ocurre un pulso glotal.
- Es esperable que el error de predicción sea en general pequeño salvo en los momentos en donde ocurre un pulso glotal.
- Como en la excitación no aparece la información de las formantes, es una señal mas apropiada para realizar una estimación de la altura.

Ejemplo. Método de la autocorrelación.

Estimación de la excitación $e[n] = Gu[n]$.



Ejemplo. Método de la autocorrelación.

Análisis de las formantes de la señal.

- LPC es una técnica paramétrica con la cual se obtienen explícitamente los coeficientes del filtro $H(z)$.
- Mediante el cálculo de las raíces del polinomio del denominador de $H(z)$ se obtienen los p polos del sistema.
- Cada polo corresponde a una formante y contiene la información sobre la frecuencia y el ancho de banda de la formante.

Ejemplo. Método de la autocorrelación.

Análisis de las formantes de la señal.

- Sea z_k el polo k -ésimo,

$$z_k = A_k e^{j\omega_k}.$$

La formante correspondiente tiene frecuencia f_k y ancho de banda BW_k dados por

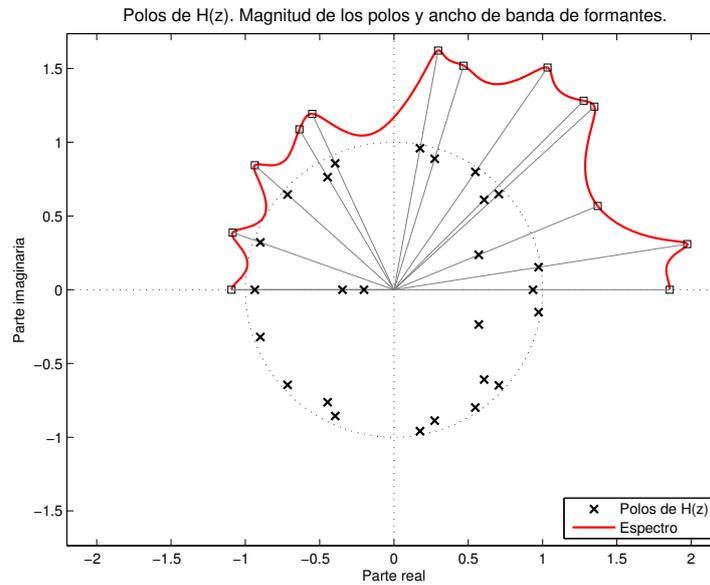
$$f_k = \frac{\omega_k f_s}{2\pi} \text{ Hz.}$$
$$BW_k \triangleq \frac{f_s}{\pi} \log \frac{1}{A_k} \text{ Hz.}$$

- No todos los polos corresponden a formantes genuinas. Los polos correspondientes a formantes relevantes pueden clasificarse imponiendo un umbral al ancho de banda. Por ejemplo, se establece que el polo k -ésimo es relevante si cumple que

$$BW_k \leq 200 \text{ Hz.}$$

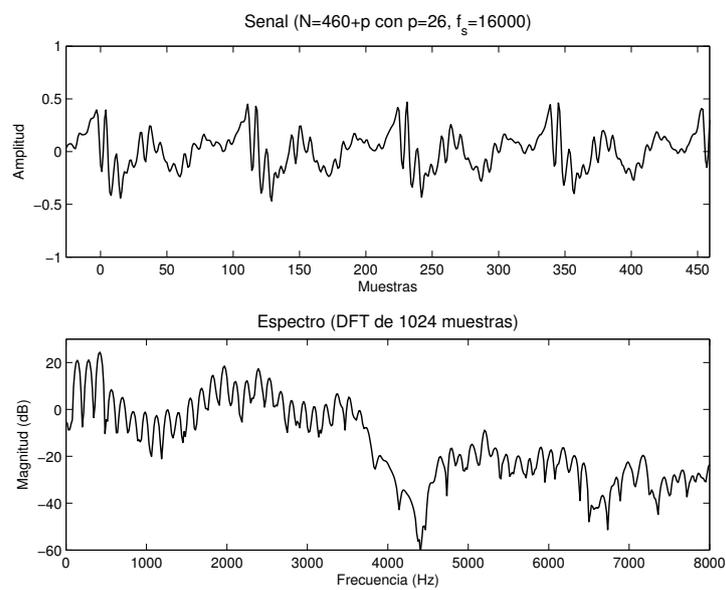
Ejemplo. Método de la autocorrelación.

Análisis de las formantes de la señal.



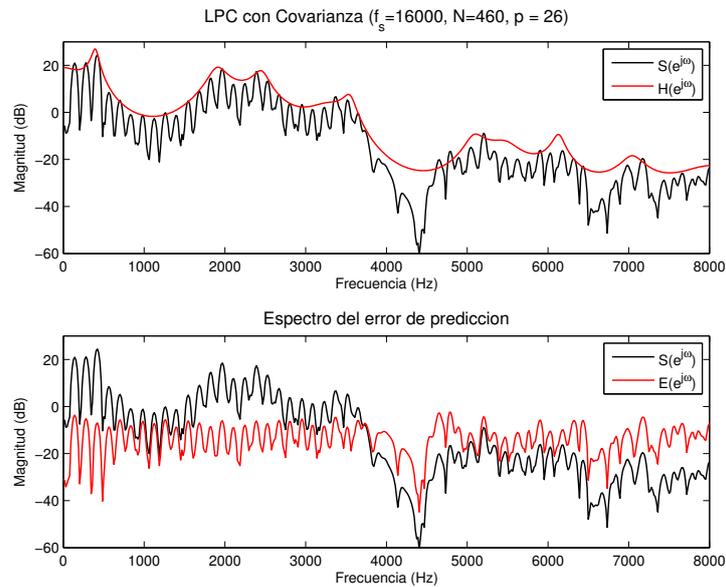
Ejemplo. Método de la covarianza.

Análisis de sonido sonoro.



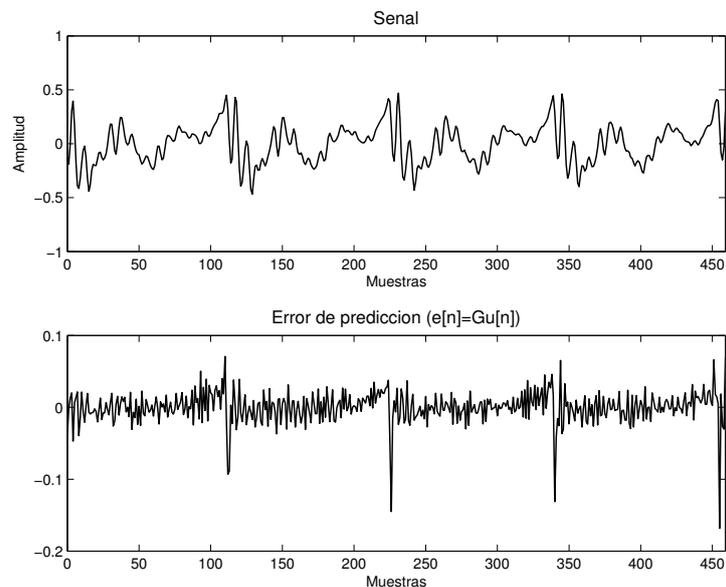
Ejemplo. Método de la covarianza.

Estimación de $H(z)$.



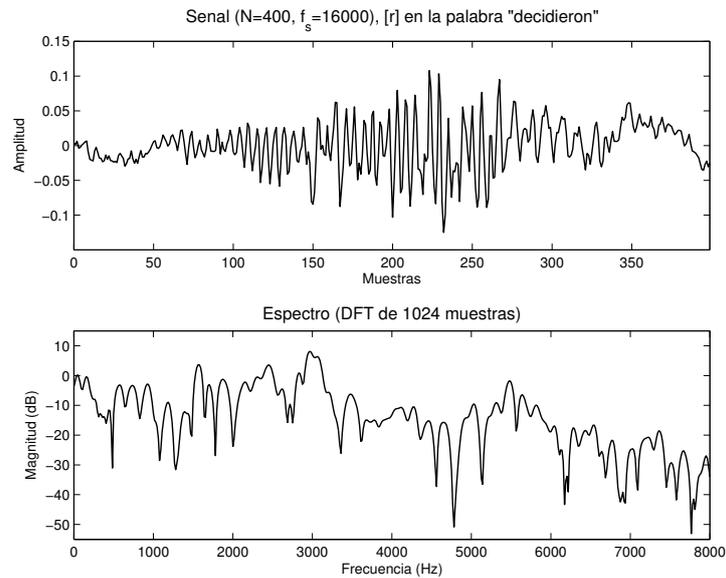
Ejemplo. Método de la covarianza

Estimación de la excitación $e[n] = Gu[n]$.



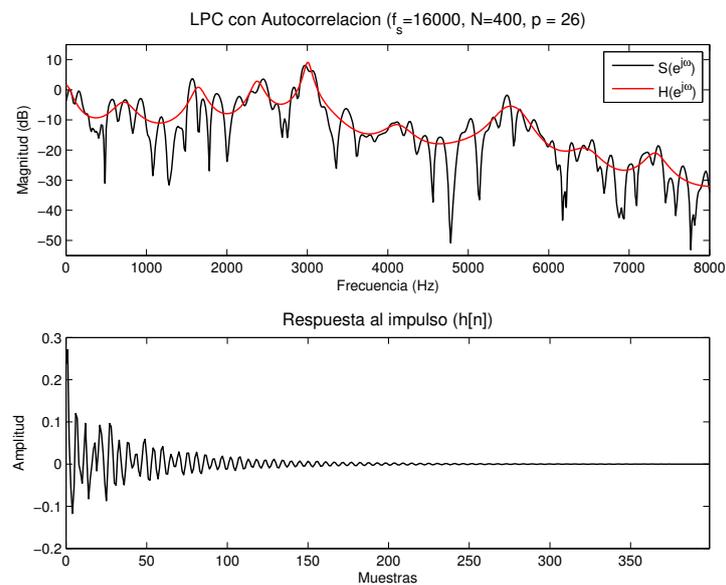
Ejemplo. Método de la autocorrelación.

Análisis de sonido sordo.



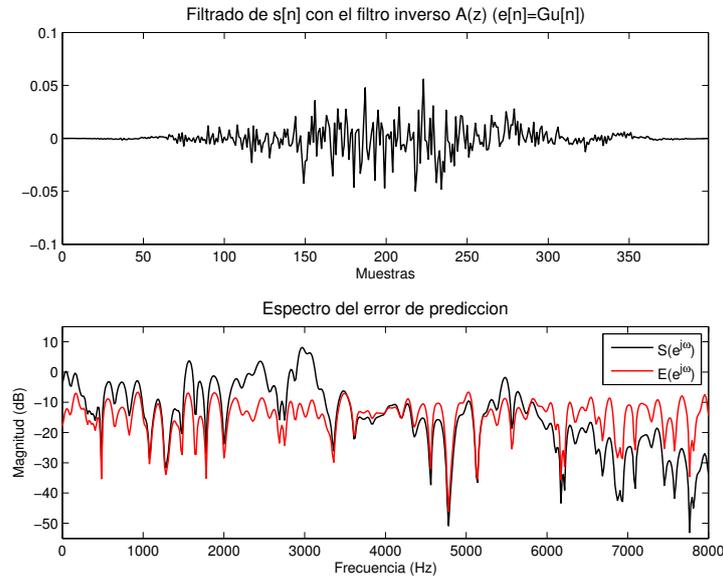
Ejemplo. Método de la autocorrelación.

Estimación de $H(z)$.



Ejemplo. Método de la autocorrelación

Estimación de la excitación $e[n] = Gu[n]$.



Interpretación en el dominio de la frecuencia.

Error RMS en el dominio de la frecuencia

De las ecuaciones 4 y 5 puede verse que el espectro del error de predicción es:

$$E(e^{j\omega}) = S_n(e^{j\omega})A(e^{j\omega}) = G \frac{S_n(e^{j\omega})}{H(e^{j\omega})}.$$

En el método de la autocorrelación, el error RMS es:

$$E_n = \sum_{m=-\infty}^{\infty} e_n^2[m]$$

Aplicando la identidad de Parseval, el error RMS expresado en el dominio de la frecuencia es:

$$E_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} |E(e^{j\omega})|^2 d\omega = \frac{G^2}{2\pi} \int_{-\pi}^{\pi} \frac{|S_n(e^{j\omega})|^2}{|H(e^{j\omega})|^2} d\omega \quad (23)$$

La ecuación 23 indica que la minimización del error RMS es equivalente a que el cociente $|S_n(e^{j\omega})|^2/|H(e^{j\omega})|^2$ tienda a uno. Recordar que $G^2 = E_n$ (ver ec. 20).

Interpretación en el dominio de la frecuencia.

Observaciones

- Como se mostró antes (ec. 19), los primeros $p + 1$ valores de la autocorrelación de $s_n[m]$ y la autocorrelación de $h[m]$ coinciden. Por lo tanto

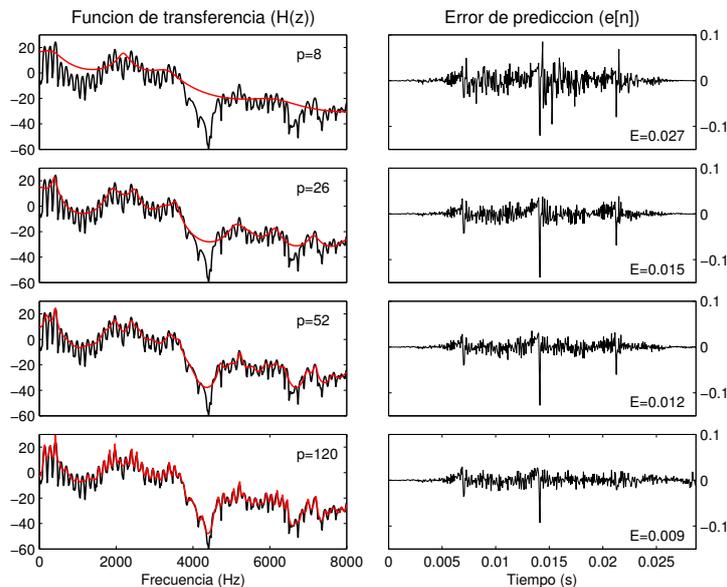
$$\lim_{p \rightarrow \infty} |H(e^{j\omega})|^2 = |S_n(e^{j\omega})|^2.$$

Esto implica que el error de predicción se puede hacer arbitrariamente pequeño al incrementar p . Mediante la elección del orden p del predictor se controla el grado de suavizado espectral.

- Cuando $p \rightarrow \infty$, si bien se cumple que $|H(e^{j\omega})|^2 = |S_n(e^{j\omega})|^2$, no necesariamente se cumple que $H(e^{j\omega}) = S_n(e^{j\omega})$. $H(e^{j\omega})$ es una representación de fase mínima de $S_n(e^{j\omega})$.
- A partir de la ec. 23, se observa que las regiones donde $|S_n(e^{j\omega})| > |H(e^{j\omega})|$ contribuyen mas al error total que las regiones donde $|S_n(e^{j\omega})| < |H(e^{j\omega})|$. Esto explica porque $|H(e^{j\omega})|$ interpola los picos espectrales.

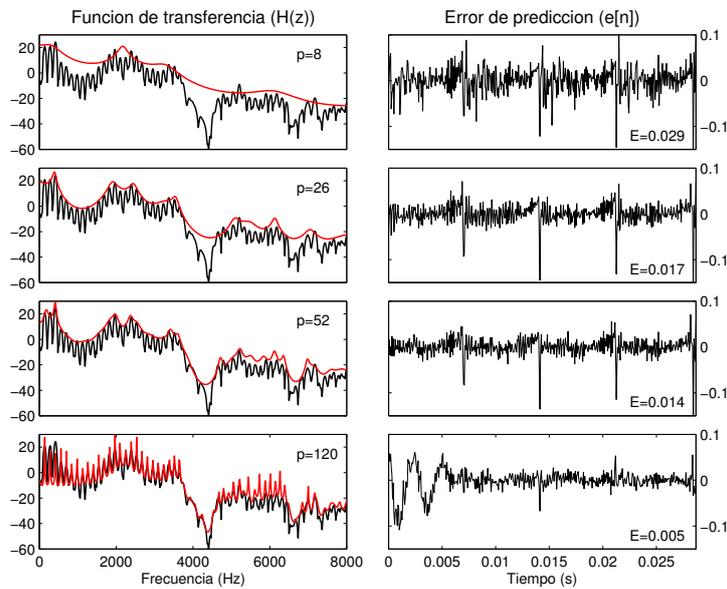
Suavizado espectral al variar el orden p .

Sonido sonoro. Método de la autocorrelación.



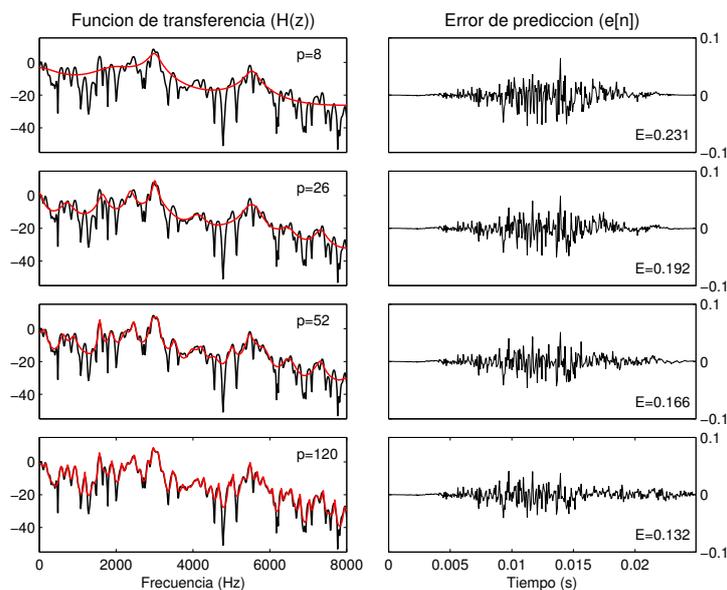
Suavizado espectral al variar el orden p .

Sonido sonoro. Método de la covarianza.

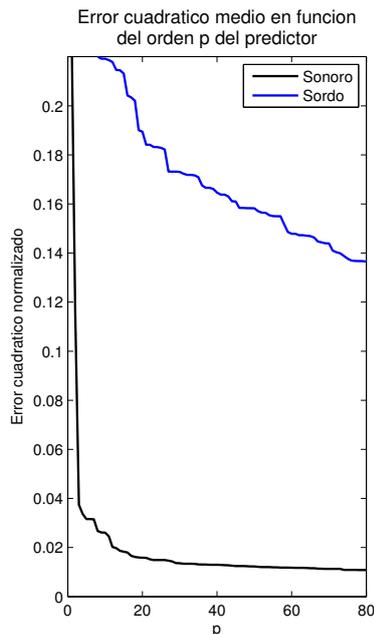


Suavizado espectral al variar el orden p .

Sonido sordo. Método de la autocorrelación.



Número óptimo de polos



Error normalizado

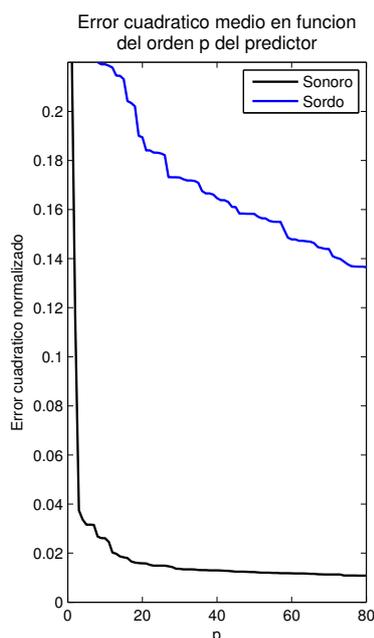
$$V_p = \frac{E_p}{R[0]} \text{ con } E_p = E_n \text{ con } p \text{ polos.}$$

- V_p decrece con p (al igual que E_p).
- Una forma de establecer el **número óptimo de polos** es analizar cuando la curva del error se hace plana,

$$1 - \frac{V_{p+1}}{V_p} < \delta.$$

- Esta prueba no aumenta el costo computacional porque al usar la recursión de Durbin o la descomposición de Cholesky se tienen todos los valores de E_p de 1 a p .

Decisión sonoro/sordo a partir del error normalizado

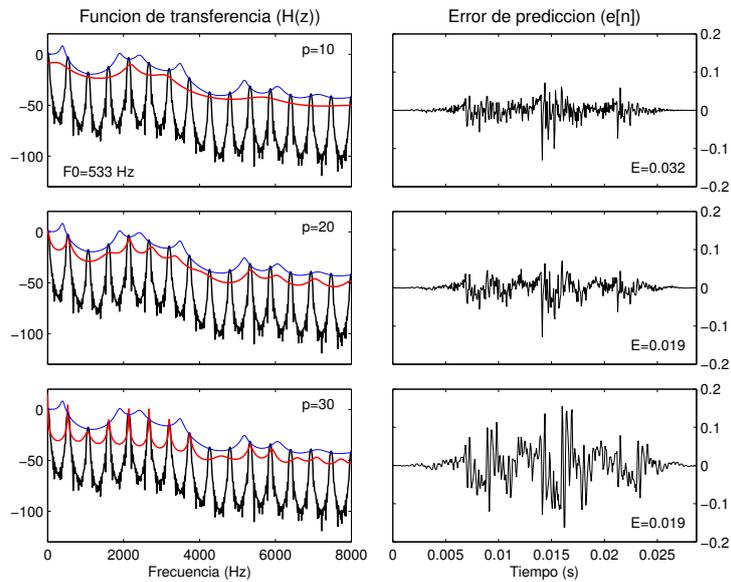


Error normalizado

- El modelo de señal predecible tiene mayor validez para sonidos sonoros que para sonidos sordos.
- El error normalizado es significativamente mayor para sonidos sordos que para sonidos sonoros.
- El error normalizado puede emplearse para discriminación sonoro/sordo.

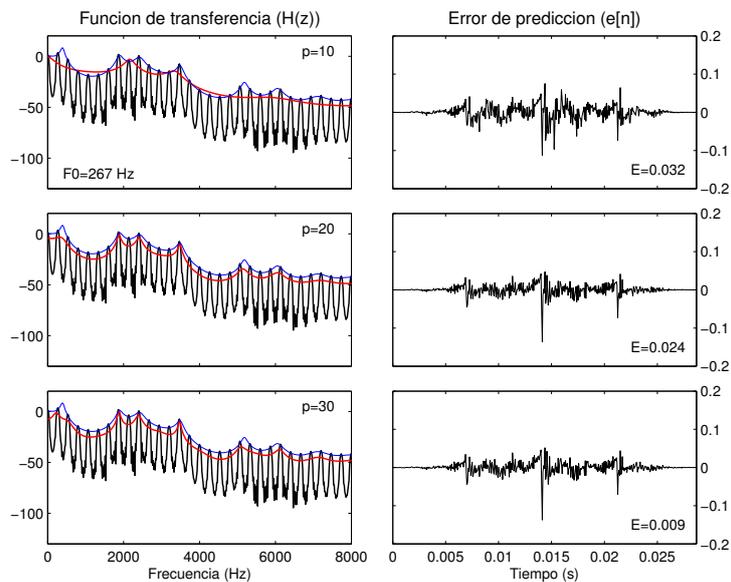
Dificultades con señales de F_0 alta.

Comparación entre el filtro y el modelo estimado con señal sintética.



Dificultades con señales de F_0 alta.

Comparación entre el filtro y el modelo estimado con señal sintética.



Dificultades con señales de F_0 alta.

Interpretación intuitiva

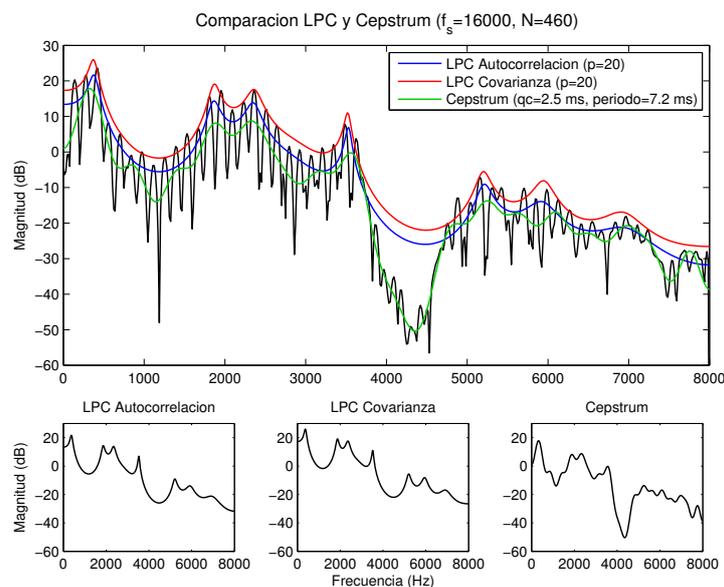
- En el espectro de la señal, la respuesta de la función de transferencia está “muestreada” por los picos espectrales.
- Si el espaciamiento en frecuencia de los picos espectrales es grande (F_0 alta), los picos espectrales representan pobremente a la función de transferencia.

Efectos en la función de transferencia estimada

- Unión o separación de polos de la función de transferencia verdadera.
- Alteración de la frecuencia y el ancho de banda de los polos. El desplazamiento de la frecuencia del polo es en general en la dirección del armónico mas cercano.
- A medida que la frecuencia fundamental decrece, también decrecen las discrepancias.

Comparación entre LPC y Cepstrum.

Estimación de la función de transferencia.



Comparación entre LPC y Cepstrum.

Interpretación intuitiva

- La estimación con LPC tiene menos picos extraños que el espectro homomórfico. Esto se debe a que en LPC se impuso $p = 20$ polos y por lo tanto puede haber a los sumo 10 frecuencias de resonancia.
- El ancho de banda de las resonancias es mucho menor en la estimación con LPC. El filtrado pasabajos del espectro en el análisis homomórfico incrementa el ancho de banda de las formantes.
- Es sencillo extraer la frecuencia y el ancho de banda de las formantes a partir de los polos del modelo LPC. En el espectro homomórfico puede ser difícil localizar las formantes de forma confiable.

El modelo de predicción lineal es muy bueno para sonidos vocálicos si se emplea el orden p correcto.

Análisis de señales de voz.

Aplicaciones de bajo nivel de LPC en análisis de señales

- Blanqueado del espectro.
 - Filtrado inverso de la señal de voz para obtener el error de predicción
 - En el error de predicción se suprimió la información de las formantes.
- Detección de formantes.
 - Los polos indican explícitamente la frecuencia y el ancho de banda de las formantes.
- Detección de sonoridad.
 - El error de predicción es significativamente mayor en sonidos sordos que en los sonoros.

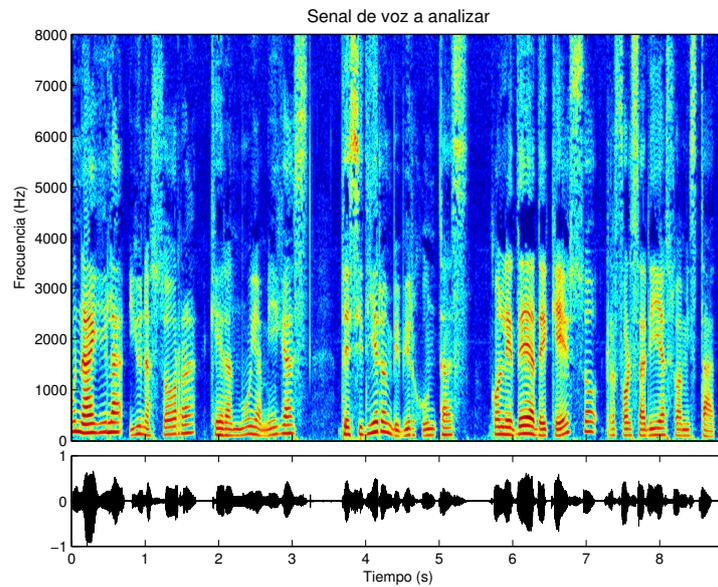
Ejemplo

Procesamiento de señal de voz masculina

- $f_s = 16000$ Hz.
- $N = 600$ muestras (≈ 38 ms)
- Salto = $N/4$ muestras (≈ 10 ms)
- Orden LPC: $p = 20$.

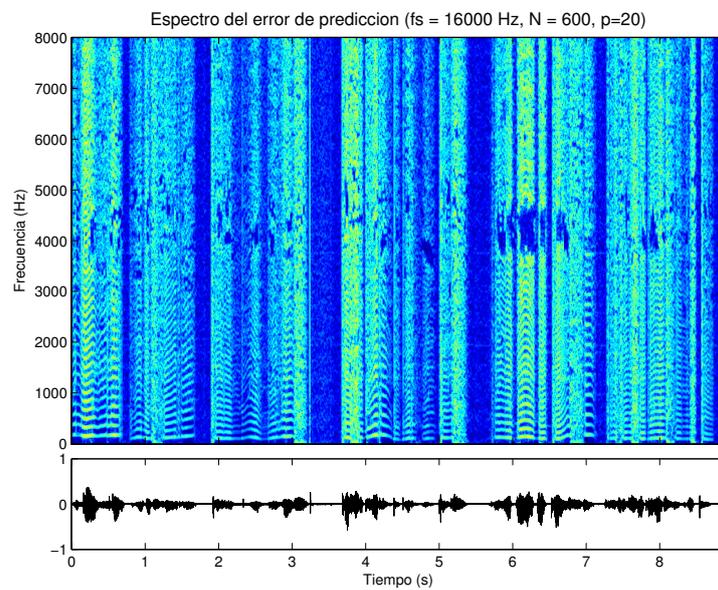
Análisis de señales de voz.

Blanqueado espectral.



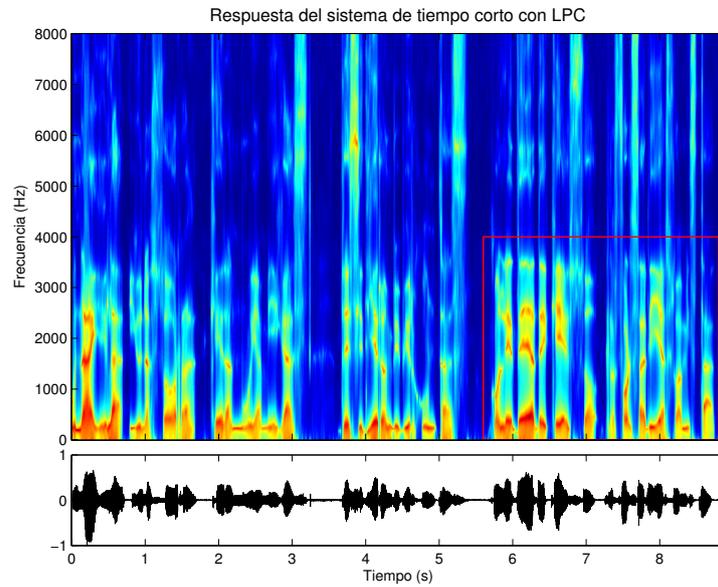
Análisis de señales de voz.

Blanqueado espectral.



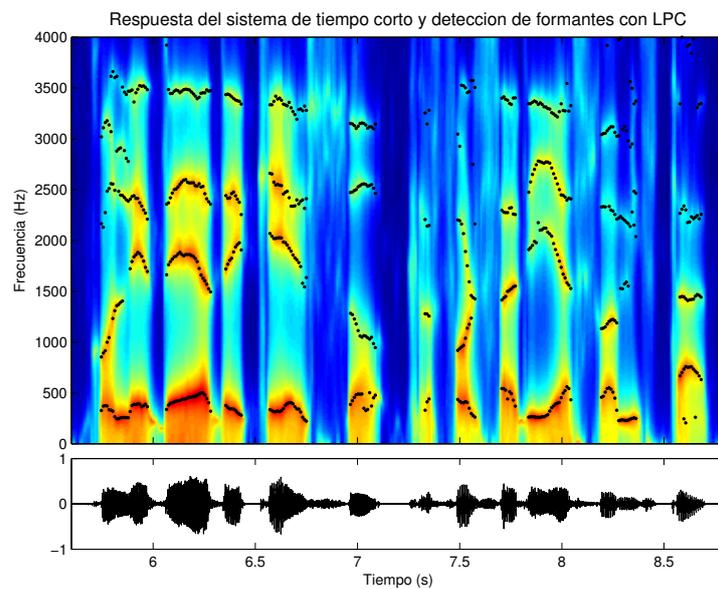
Análisis de señales de voz.

Detección de formantes.



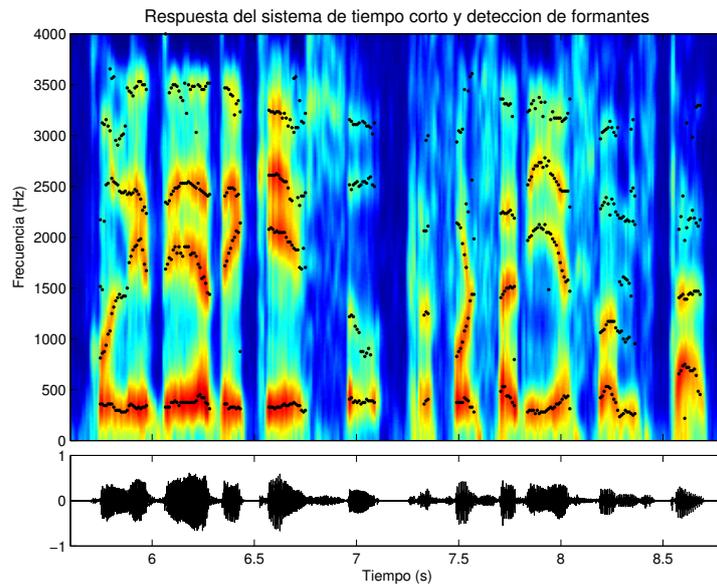
Análisis de señales de voz.

Detección de formantes con LPC.



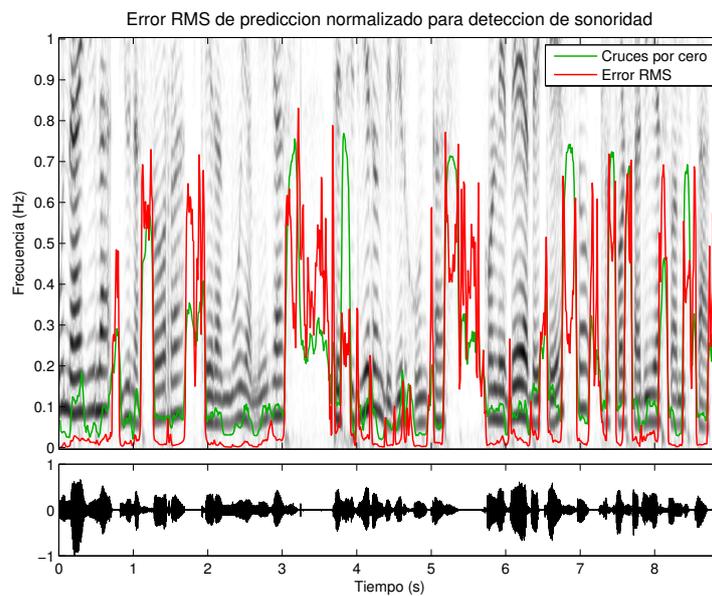
Análisis de señales de voz.

Detección de formantes con Cepstrum.



Análisis de señales de voz.

Detección de sonoridad a partir del error RMS mínimo.



Codificación de voz

Estructuras de análisis y síntesis

- A partir del modelo $H(e^{j\omega})$ obtenido mediante LPC se calcula la respuesta al impulso del sistema.
- Mediante el filtrado inverso de la señal original $s[n]$ se obtiene la excitación $e[n]$.
- La convolución entre la respuesta al impulso y la excitación reconstruye exactamente a la señal original de tiempo corto.
- Con una estructura de solapamiento y suma se reconstruye exactamente la señal original completa.

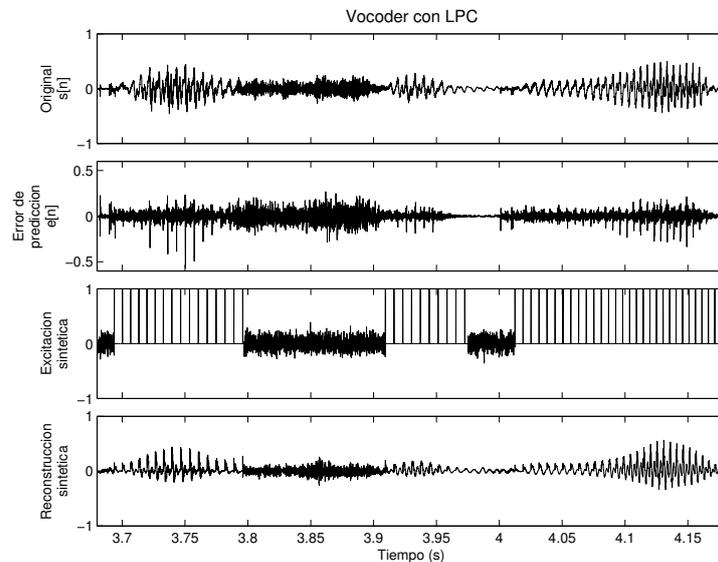
Codificación de voz

Vocoder con LPC

- Mediante la representación compacta obtenida con LPC es posible sintetizar de forma eficiente una estimación de la señal.
- **Etapas de análisis**
 - Cálculo de los coeficientes LPC de la señal de tiempo corto.
 - Estimación de la sonoridad y la frecuencia fundamental. En la práctica se emplea un algoritmo de detección de altura externo.
 - Representación de la señal: sonoridad, frecuencia fundamental coeficientes LPC $\{a_k\}$ y ganancia (G).
- **Etapas de síntesis**
 - Cálculo de la respuesta al impulso $h[n]$ a partir de los coeficientes y la ganancia.
 - Síntesis de la excitación a partir de la sonoridad y la frecuencia fundamental.
 - Síntesis de la señal de tiempo corto con la convolución de la excitación con la respuesta al impulso $h[n]$.
 - Solapamiento y suma de las reconstrucciones de tiempo corto.

Codificación de voz

Síntesis con el vocoder LPC.



Codificación de voz

Ejemplo: codificación de voz a baja tasa de bits

Procesamiento:

- $N = 512$ muestras
- Salto = 256 muestras
- Cálculo del modelo LPC de orden $p = 26$.

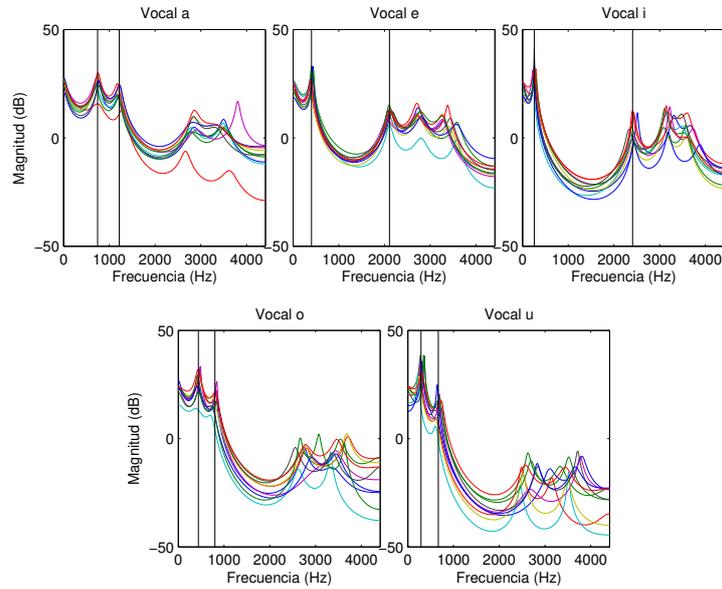
Codificación:

- 26 coeficientes LPC
- Ganancia G
- Estimación de f_0

Cada 256 muestras se transmiten 28: tasa de compresión ≈ 9

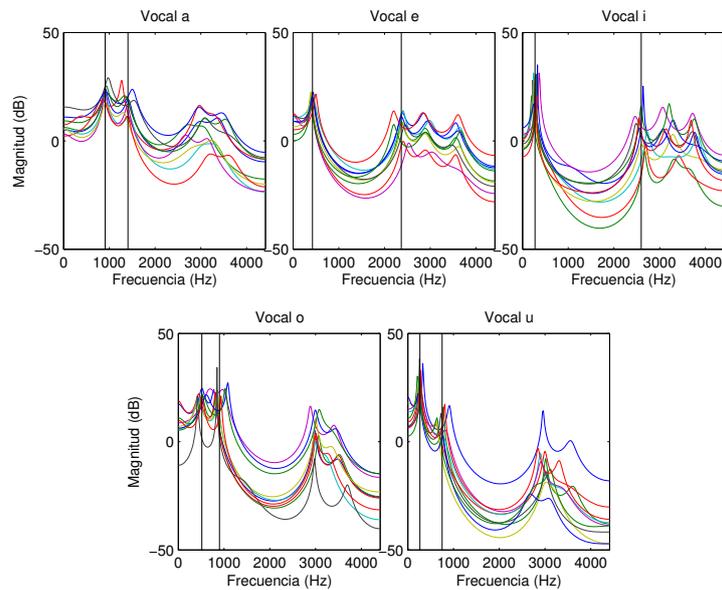
Reconocimiento del habla

Espectro LPC de vocales. Hablante masculino.



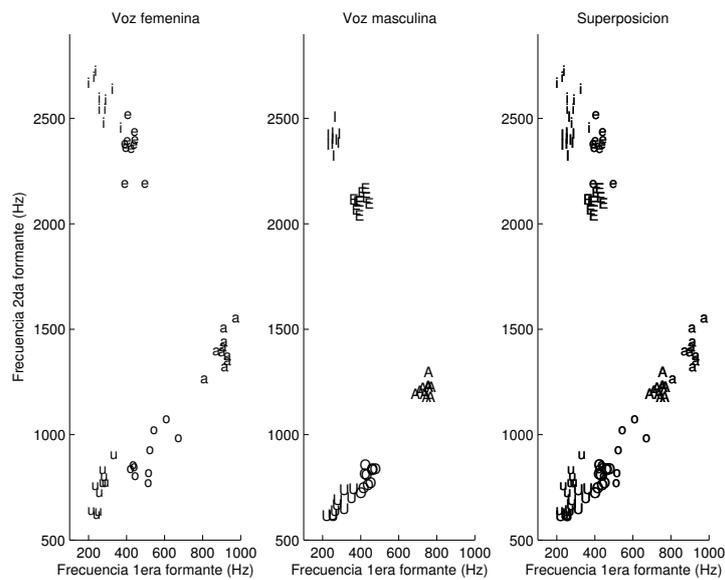
Reconocimiento del habla

Espectro LPC de vocales. Hablante femenino.



Reconocimiento del habla

Mapa de formantes.



Referencias I

-  Makhoul, J. (1975).
Linear prediction: A tutorial review.
Proceedings of the IEEE, 63(4):561 – 580.
-  Rabiner, L. R. and Schafer, R. W. (2011).
Theory and Applications of Digital Speech Processing.
Prentice Hall, 1st edition.
Chapter 7 - Frequency-domain representations.