ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

Ψηφιακή Επεξεργασία Φωνής

Ενότητα 4η: Γραμμική Πρόβλεψη: Ανάλυση και Σύνθεση

Στυλιανού Ιωάννης
Τμήμα Επιστήμης Υπολογιστών

CS578- Speech Signal Processing

Lecture 4: Linear Prediction of Speech; Analysis and Synthesis

Yannis Stylianou

University of Crete, Computer Science Dept., Multimedia Informatics Lab
yannis@csd.uoc.gr

Univ. of Crete, 2008 Winter Period

Outline

Transfer function from the glottis to the lips

- We shown that for voiced speech:

$$
\begin{aligned}
H(z) &= AG(z)V(z)R(z) \\
&= A\frac{(1-az^{-1})}{(1-bz)^2(1-\sum_{k=1}^{N} a_k z^{-k})}
\end{aligned}
$$

- However:

$$1-az^{-1} = \frac{1}{\sum_{k=0}^{\infty} a^k z^{-k}}, \quad \text{for } |z| > |a|$$

- Then:

$$H(z) = \frac{A}{1-\sum_{k=1}^{p} a_k z^{-k}}$$

Transfer function from the glottis to the lips

- We shown that for voiced speech:

$$\begin{aligned} H(z) &= AG(z)V(z)R(z) \\ &= A\frac{(1-az^{-1})}{(1-bz)^2(1-\sum_{k=1}^{N} a_k z^{-k})} \end{aligned}$$

- However:

$$1 - az^{-1} = \frac{1}{\sum_{k=0}^{\infty} a^k z^{-k}}, \quad \text{for } |z| > |a|$$

- Then:

$$H(z) = \frac{A}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

Transfer function from the glottis to the lips

- We shown that for voiced speech:

$$\begin{aligned} H(z) &= AG(z)V(z)R(z) \\ &= A\frac{(1-az^{-1})}{(1-bz)^2(1-\sum_{k=1}^{N} a_k z^{-k})} \end{aligned}$$

- However:

$$1-az^{-1} = \frac{1}{\sum_{k=0}^{\infty} a^k z^{-k}}, \quad \text{for } |z| > |a|$$

- Then:

$$H(z) = \frac{A}{1-\sum_{k=1}^{p} a_k z^{-k}}$$

Producing speech [1]

Assuming as input to $H(z)$ a train of unit samples, $u_g[n]$, with z-transform $U_g(z)$, then speech, $S(z)$ is given by:

$$H(z) = \frac{S(z)}{U_g(z)} = \frac{A}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

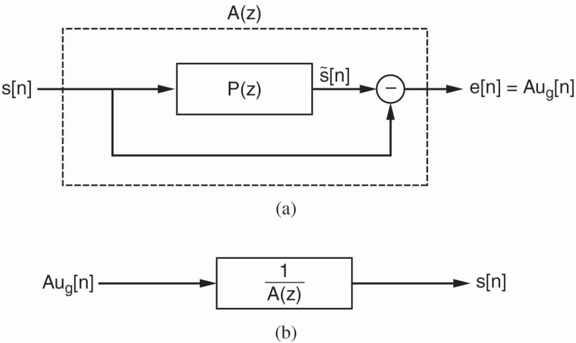
or

$$S(z) = \sum_{k=1}^{p} a_k z^{-k} S(z) + A U_g(z)$$

and in time domain:

$$s[n] = \sum_{k=1}^{p} a_k s[n-k] + A u_g[n]$$

Useful terms: *Linear prediction coefficients, Autoregressive (AR) model/process, Linear prediction analysis*



where

$$\begin{aligned} P(z) &= \sum_{k=1}^{p} a_k z^{-k} \quad \textit{prediction filter} \\ A(z) &= 1 - P(z) \quad \textit{prediction error filter} \end{aligned}$$

Outline

Justification of LP for speech

- If speech is (almost) an AR process, then:

$$s[n] = \sum_{k=1}^{p} a_k s[n-k] + Au_g[n]$$

- A p*th* linear predictor, means:

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

- *Prediction error:*

$$e[n] = s[n] - \tilde{s}[n]$$

- or:

$$e[n] \approx Au_g[n] \quad \text{if } a_k \approx l_k, \ \forall k$$

Justification of LP for speech

- If speech is (almost) an AR process, then:

$$s[n] = \sum_{k=1}^{p} a_k s[n-k] + A u_g[n]$$

- A p*th* linear predictor, means:

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

- *Prediction error:*

$$e[n] = s[n] - \tilde{s}[n]$$

- or:

$$e[n] \approx A u_g[n] \quad \text{if } a_k \approx l_k,\ \forall k$$

Justification of LP for speech

- If speech is (almost) an AR process, then:

$$s[n] = \sum_{k=1}^{p} a_k s[n-k] + A u_g[n]$$

- A p*th* linear predictor, means:

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

- *Prediction error:*

$$e[n] = s[n] - \tilde{s}[n]$$

- or:

$$e[n] \approx A u_g[n] \quad \text{if } a_k \approx l_k, \ \forall k$$

Justification of LP for speech

- If speech is (almost) an AR process, then:

$$s[n] = \sum_{k=1}^{p} a_k s[n-k] + A u_g[n]$$

- A p*th* linear predictor, means:

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

- *Prediction error:*

$$e[n] = s[n] - \tilde{s}[n]$$

- or:

$$e[n] \approx A u_g[n] \quad \text{if } a_k \approx l_k, \ \forall k$$

Outline

Error Minimization

- Over all time we wish to minimize the mean-squared prediction error:

$$E = \sum_{m=-\infty}^{\infty} (s[m] - \tilde{s}[m])^2$$

- Prediction error in the vicinity of n:

$$E_n = \sum_{m=n-M}^{n+M} (s[m] - \tilde{s}[m])^2$$

- *Prediction interval:* $[n - M,\ n + M]$
-

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

where

$$e_n[m] = s_n[m] - \sum_{k=1}^{p} l_k s_n[m-k], \quad n - M \leq m \leq n + M$$

Error Minimization

- Over all time we wish to minimize the mean-squared prediction error:

$$E = \sum_{m=-\infty}^{\infty} (s[m] - \tilde{s}[m])^2$$

- Prediction error in the vicinity of n:

$$E_n = \sum_{m=n-M}^{n+M} (s[m] - \tilde{s}[m])^2$$

- *Prediction interval:* $[n - M,\ n + M]$
-

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

where

$$e_n[m] = s_n[m] - \sum_{k=1}^{p} l_k s_n[m-k], \quad n - M \leq m \leq n + M$$

Error Minimization

- Over all time we wish to minimize the mean-squared prediction error:

$$E = \sum_{m=-\infty}^{\infty} (s[m] - \tilde{s}[m])^2$$

- Prediction error in the vicinity of n:

$$E_n = \sum_{m=n-M}^{n+M} (s[m] - \tilde{s}[m])^2$$

- *Prediction interval:* $[n - M,\ n + M]$
-

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

where

$$e_n[m] = s_n[m] - \sum_{k=1}^{p} l_k s_n[m-k], \quad n - M \leq m \leq n + M$$

Error Minimization

- Over all time we wish to minimize the mean-squared prediction error:

$$E = \sum_{m=-\infty}^{\infty} (s[m] - \tilde{s}[m])^2$$

- Prediction error in the vicinity of n:

$$E_n = \sum_{m=n-M}^{n+M} (s[m] - \tilde{s}[m])^2$$

- *Prediction interval:* $[n - M,\ n + M]$
-

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

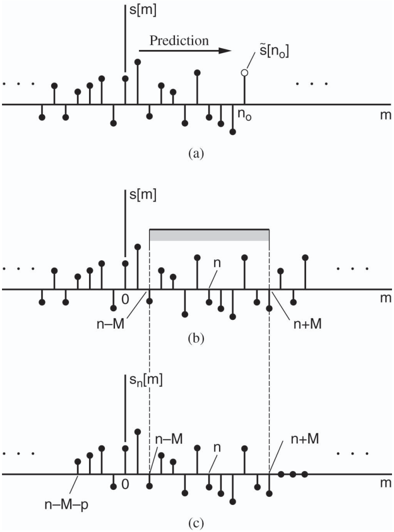
where

$$e_n[m] = s_n[m] - \sum_{k=1}^{p} l_k s_n[m-k], \quad n - M \leq m \leq n + M$$

Covariance Method

- Samples outside the prediction error interval are NOT zero
- Minimization of the mean-squared error in the prediction error interval

Short-time sequences: Covariance



Covariance Method: Formulation

- In matrix notation

$$\mathbf{e}_n^{(2M+1\times 1)} = \mathbf{s}_n^{(2M+1\times 1)} - \mathbf{S}_n^{(2M+1\times p)}\mathbf{l}^{(p\times 1)}$$

- Mean-squared error

$$\mathbf{e}_n^T\mathbf{e}_n = \mathbf{s}_n^T\mathbf{s}_n - 2\mathbf{s}_n^T\mathbf{S}_n\mathbf{l} + \mathbf{l}^T\mathbf{S}_n^T\mathbf{S}_n\mathbf{l}$$

- Solution:

$$\mathbf{l} = \left(\mathbf{S}_n^T\mathbf{S}_n\right)^{-1}\mathbf{S}_n^T\mathbf{s}_n$$

- Same solution by considering the *Projection Theorem*:

$$\mathbf{S}_n^T\mathbf{e}_n = \mathbf{0}$$

Covariance Method: Formulation

- In matrix notation

$$\mathbf{e}_n^{(2M+1\times 1)} = \mathbf{s}_n^{(2M+1\times 1)} - \mathbf{S}_n^{(2M+1\times p)}\mathbf{l}^{(p\times 1)}$$

- Mean-squared error

$$\mathbf{e}_n^T\mathbf{e}_n = \mathbf{s}_n^T\mathbf{s}_n - 2\mathbf{s}_n^T\mathbf{S}_n\mathbf{l} + \mathbf{l}^T\mathbf{S}_n^T\mathbf{S}_n\mathbf{l}$$

- Solution:

$$\mathbf{l} = \left(\mathbf{S}_n^T\mathbf{S}_n\right)^{-1}\mathbf{S}_n^T\mathbf{s}_n$$

- Same solution by considering the *Projection Theorem*:

$$\mathbf{S}_n^T\mathbf{e}_n = \mathbf{0}$$

Covariance Method: Formulation

- In matrix notation

$$\mathbf{e}_n^{(2M+1\times 1)} = \mathbf{s}_n^{(2M+1\times 1)} - \mathbf{S}_n^{(2M+1\times p)}\mathbf{l}^{(p\times 1)}$$

- Mean-squared error

$$\mathbf{e}_n^T\mathbf{e}_n = \mathbf{s}_n^T\mathbf{s}_n - 2\mathbf{s}_n^T\mathbf{S}_n\mathbf{l} + \mathbf{l}^T\mathbf{S}_n^T\mathbf{S}_n\mathbf{l}$$

- Solution:

$$\mathbf{l} = \left(\mathbf{S}_n^T\mathbf{S}_n\right)^{-1}\mathbf{S}_n^T\mathbf{s}_n$$

- Same solution by considering the *Projection Theorem*:

$$\mathbf{S}_n^T\mathbf{e}_n = \mathbf{0}$$

Covariance Method: Formulation

- In matrix notation

$$\mathbf{e}_n^{(2M+1\times 1)} = \mathbf{s}_n^{(2M+1\times 1)} - \mathbf{S}_n^{(2M+1\times p)}\mathbf{l}^{(p\times 1)}$$

- Mean-squared error

$$\mathbf{e}_n^T\mathbf{e}_n = \mathbf{s}_n^T\mathbf{s}_n - 2\mathbf{s}_n^T\mathbf{S}_n\mathbf{l} + \mathbf{l}^T\mathbf{S}_n^T\mathbf{S}_n\mathbf{l}$$

- Solution:

$$\mathbf{l} = \left(\mathbf{S}_n^T\mathbf{S}_n\right)^{-1}\mathbf{S}_n^T\mathbf{s}_n$$

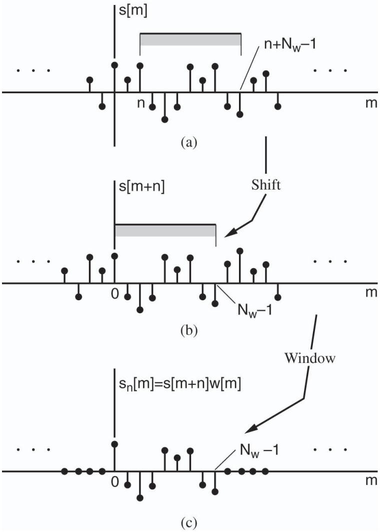
- Same solution by considering the *Projection Theorem*:

$$\mathbf{S}_n^T\mathbf{e}_n = \mathbf{0}$$

Autocorrelation Method

- Samples outside the prediction error interval are all zero
- Minimization of the mean-squared error in $\pm\infty$

Short-time sequences: Autocorrelation



Autocorrelation method: Formulation

- Error is nonzero in the interval $[0,\ N_w + p - 1]$:

$$E_n = \sum_{m=0}^{N_w+p-1} e_n^2[m]$$

- Normal equations:

$$\sum_{k=1}^{p} l_k \Phi_n[i,k] = \Phi_n[i,0], \quad i = 1,2,3,\ldots,p$$

where

$$\Phi_n[i,k] = \sum_{m=0}^{N_w+p-1} s_n[m-i]\, s_n[m-k], \;\; 1 \le i \le p, \;\; 0 \le k \le p$$

Autocorrelation method: Formulation

- Error is nonzero in the interval $[0,\ N_w + p - 1]$:

$$E_n = \sum_{m=0}^{N_w+p-1} e_n^2[m]$$

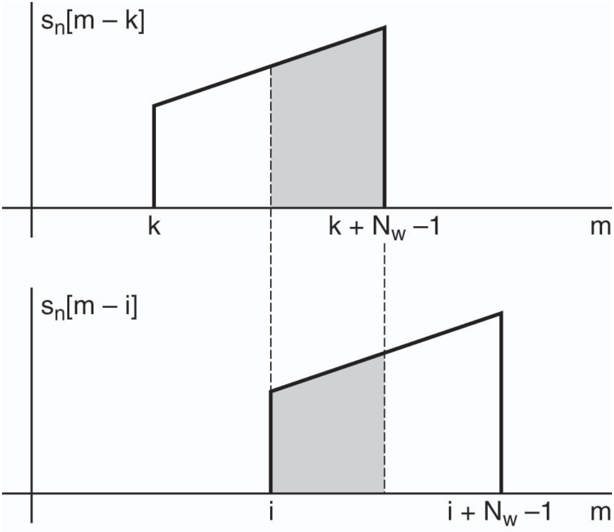
- Normal equations:

$$\sum_{k=1}^{p} l_k \Phi_n[i,k] = \Phi_n[i,0], \quad i = 1,2,3,\ldots,p$$

where

$$\Phi_n[i,k] = \sum_{m=0}^{N_w+p-1} s_n[m-i]\, s_n[m-k], \ \ 1 \le i \le p, \ \ 0 \le k \le p$$

Constructing the autocorrelation function



Using the autocorrelation function

- by denoting:

$$r_n[i-k] = \Phi_n[i,k]$$

- Then:

$$\sum_{k=1} (p l_k r_n[i-k] = r_n[i], \quad 1 \le i \le p$$

- In matrix notation:

$$\mathbf{R}_n^{(p\times p)} \mathbf{l}^{(p\times 1)} = \mathbf{r}_n^{(p\times 1)}$$

- Or (Toeplitz matrix):

$$\begin{bmatrix} r_n[0] & r_n[1] & \cdots & r_n[p-1] \\ r_n[1] & r_n[0] & \cdots & r_n[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_n[p-1] & r_n[p-2] & \cdots & r_n[0] \end{bmatrix} \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_p \end{bmatrix} = \begin{bmatrix} r_n[1] \\ r_n[2] \\ \vdots \\ r_n[p] \end{bmatrix}$$

Using the autocorrelation function

- by denoting:

$$r_n[i-k] = \Phi_n[i,k]$$

- Then:

$$\sum_{k=1} (p l_k r_n[i-k] = r_n[i], \quad 1 \leq i \leq p$$

- In matrix notation:

$$\mathbf{R}_n^{(p\times p)} \mathbf{l}^{(p\times 1)} = \mathbf{r}_n^{(p\times 1)}$$

- Or (Toeplitz matrix):

$$\begin{bmatrix} r_n[0] & r_n[1] & \cdots & r_n[p-1] \\ r_n[1] & r_n[0] & \cdots & r_n[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_n[p-1] & r_n[p-2] & \cdots & r_n[0] \end{bmatrix} \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_p \end{bmatrix} = \begin{bmatrix} r_n[1] \\ r_n[2] \\ \vdots \\ r_n[p] \end{bmatrix}$$

Using the autocorrelation function

- by denoting:

$$r_n[i-k] = \Phi_n[i,k]$$

- Then:

$$\sum_{k=1}(pl_k r_n[i-k] = r_n[i], \quad 1 \leq i \leq p$$

- In matrix notation:

$$\mathbf{R}_n^{(p\times p)}\mathbf{l}^{(p\times 1)} = \mathbf{r}_n^{(p\times 1)}$$

- Or (Toeplitz matrix):

$$\begin{bmatrix} r_n[0] & r_n[1] & \cdots & r_n[p-1] \\ r_n[1] & r_n[0] & \cdots & r_n[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_n[p-1] & r_n[p-2] & \cdots & r_n[0] \end{bmatrix} \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_p \end{bmatrix} = \begin{bmatrix} r_n[1] \\ r_n[2] \\ \vdots \\ r_n[p] \end{bmatrix}$$

Using the autocorrelation function

- by denoting:

$$r_n[i-k] = \Phi_n[i,k]$$

- Then:

$$\sum_{k=1} (pl_k r_n[i-k] = r_n[i], \quad 1 \leq i \leq p$$

- In matrix notation:

$$\mathbf{R}_n^{(p\times p)} \mathbf{l}^{(p\times 1)} = \mathbf{r}_n^{(p\times 1)}$$

- Or (Toeplitz matrix):

$$\begin{bmatrix} r_n[0] & r_n[1] & \cdots & r_n[p-1] \\ r_n[1] & r_n[0] & \cdots & r_n[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_n[p-1] & r_n[p-2] & \cdots & r_n[0] \end{bmatrix} \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_p \end{bmatrix} = \begin{bmatrix} r_n[1] \\ r_n[2] \\ \vdots \\ r_n[p] \end{bmatrix}$$

Levinson Recursion

▷ Build an order $i+1$ solution from an order i solution until the desired order p is reached:

- **Initial step:**

$$l_0^0 = 0, \quad E^0 = r[0]$$

- **Step 1:** Compute the *partial correlation coefficients*

$$k_i = \frac{r[i] - \sum_{j=1}^{i-1} l_j^{i-1} r[i-j]}{E^{i-1}}$$

- **Step 2:** Update prediction coefficients, l

$$\begin{aligned} l_i^i &= k_i \\ l_j^i &= l_j^{i-1} - k_i l_{i-j}^{i-1}, \quad 1 \leq j \leq i-1 \end{aligned}$$

- **Step 3:** Update the minimum squared prediction error

$$E^i = (1 - k_i^2) E^{i-1}$$

- **Step 4:** Repeat steps 1 to 3 for $i = 1, 2, \cdots, p$
- **Final Step:** at pth step, compute the optimal predictor coefficients, l_j^*,

$$l_j^* = l_j^p, \quad 1 \leq j \leq p$$

Lossless Tube Model and Linear Prediction

There is a strong resemblance to the recursions in the lossless tube model and in the Autocorrelation Method for Linear Prediction:

- Transfer functions:

$$\begin{aligned} V(z) &= \frac{A}{D(z)} \quad D(z) = 1 - \sum_{k=1}^{N} l_k z^{-k} \\ H(z) &= \frac{A}{A(z)} \quad A(z) = 1 - \sum_{k=1}^{p} l_k z^{-k} \end{aligned}$$

- Recursions:

$$\begin{array}{rcl|rcl} D_0(z) &=& 1 & A^0(z) &=& 1 \\ \text{For } k &=& 1, 2, \cdots, N & \text{For } i &=& 1, 2, \cdots, p \\ D_k(z) &=& D_{k-1}(z) + r_k z^{-k} D_{k-1}(z^{-1}) & A^i(z) &=& A^{i-1}(z) - k_i z^{-i} A^{i-1}(z^{-1}) \\ D(z) &=& D_N(z) & A(z) &=& A_p(z) \end{array}$$

- Identical recursions if: $k_i = -r_i = -\frac{Ai+1-Ai}{Ai+1+Ai}$

Lossless Tube Model and Linear Prediction

There is a strong resemblance to the recursions in the lossless tube model and in the Autocorrelation Method for Linear Prediction:

- Transfer functions:

$$V(z) = \frac{A}{D(z)} \quad D(z) = 1 - \sum_{k=1}^{N} l_k z^{-k}$$

$$H(z) = \frac{A}{A(z)} \quad A(z) = 1 - \sum_{k=1}^{p} l_k z^{-k}$$

- Recursions:

$D_0(z)$	$=$	1	|	$A^0(z)$	$=$	1
For k	$=$	$1, 2, \cdots, N$	|	For i	$=$	$1, 2, \cdots, p$
$D_k(z)$	$=$	$D_{k-1}(z) + r_k z^{-k} D_{k-1}(z^{-1})$	|	$A^i(z)$	$=$	$A^{i-1}(z) - k_i z^{-i} A^{i-1}(z^{-1})$
$D(z)$	$=$	$D_N(z)$	|	$A(z)$	$=$	$A_p(z)$

- Identical recursions if: $k_i = -r_i = -\frac{Ai+1-Ai}{Ai+1+Ai}$

There is a strong resemblance to the recursions in the lossless tube model and in the Autocorrelation Method for Linear Prediction:

- Transfer functions:

$$V(z) = \frac{A}{D(z)} \quad D(z) = 1 - \sum_{k=1}^{N} l_k z^{-k}$$

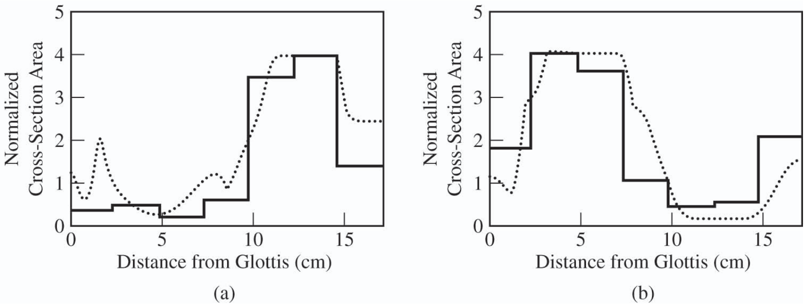
$$H(z) = \frac{A}{A(z)} \quad A(z) = 1 - \sum_{k=1}^{p} l_k z^{-k}$$

- Recursions:

$D_0(z)$	$=$	1	|	$A^0(z)$	$=$	1
For k	$=$	$1, 2, \cdots, N$	|	For i	$=$	$1, 2, \cdots, p$
$D_k(z)$	$=$	$D_{k-1}(z) + r_k z^{-k} D_{k-1}(z^{-1})$	|	$A^i(z)$	$=$	$A^{i-1}(z) - k_i z^{-i} A^{i-1}(z^{-1})$
$D(z)$	$=$	$D_N(z)$	|	$A(z)$	$=$	$A_p(z)$

- Identical recursions if: $k_i = -r_i = -\frac{Ai+1-Ai}{Ai+1+Ai}$

Estimating the vocal tract area functions via the autocorrelation method



Properties of the Autocorrelation method

- $|k_i| < 1, \quad \forall i$
- $H(z)$ is a minimum phase system (*stability*)
- Flip all maximum-phase poles inside the unit circle to their conjugate reciprocal locations
- One-to-One correspondence: $k_i \rightleftarrows l_i$, $l_i \rightleftarrows r_n[i]$.

$$\begin{aligned} k_i &= l_i^i \\ l_j^{i-1} &= \frac{l_j^i + k_i l_{i-j}^i}{1 - k_i^2} \end{aligned}$$

- *Autocorrelation matching:* If, $H(z)$ is an pth all-pole minimum phase system, and if $r_h[0] = r_n[0]$, then:

$$r_h[\tau] = r_n[\tau], \quad \text{for } |\tau| \leq p$$

Properties of the Autocorrelation method

- $|k_i| < 1, \quad \forall i$
- $H(z)$ is a minimum phase system (*stability*)
- Flip all maximum-phase poles inside the unit circle to their conjugate reciprocal locations
- One-to-One correspondence: $k_i \rightleftarrows l_i$, $l_i \rightleftarrows r_n[i]$.

$$\begin{aligned} k_i &= l_i^i \\ l_j^{i-1} &= \frac{l_j^i + k_i l_{i-j}^i}{1 - k_i^2} \end{aligned}$$

- *Autocorrelation matching:* If, $H(z)$ is an pth all-pole minimum phase system, and if $r_h[0] = r_n[0]$, then:

$$r_h[\tau] = r_n[\tau], \quad \text{for } |\tau| \leq p$$

Properties of the Autocorrelation method

- $|k_i| < 1, \quad \forall i$
- $H(z)$ is a minimum phase system (*stability*)
- Flip all maximum-phase poles inside the unit circle to their conjugate reciprocal locations
- One-to-One correspondence: $k_i \rightleftarrows l_i$, $l_i \rightleftarrows r_n[i]$.

$$k_i = l_i^i$$
$$l_j^{i-1} = \frac{l_j^i + k_i l_{i-j}^i}{1 - k_i^2}$$

- *Autocorrelation matching:* If, $H(z)$ is an pth all-pole minimum phase system, and if $r_h[0] = r_n[0]$, then:

$$r_h[\tau] = r_n[\tau], \quad \text{for } |\tau| \le p$$

Properties of the Autocorrelation method

- $|k_i| < 1, \quad \forall i$
- $H(z)$ is a minimum phase system (*stability*)
- Flip all maximum-phase poles inside the unit circle to their conjugate reciprocal locations
- One-to-One correspondence: $k_i \rightleftarrows l_i$, $l_i \rightleftarrows r_n[i]$.

$$\begin{aligned} k_i &= l_i^i \\ l_j^{i-1} &= \frac{l_j^i + k_i l_{i-j}^i}{1 - k_i^2} \end{aligned}$$

- *Autocorrelation matching:* If, $H(z)$ is an pth all-pole minimum phase system, and if $r_h[0] = r_n[0]$, then:

$$r_h[\tau] = r_n[\tau], \quad \text{for } |\tau| \leq p$$

Properties of the Autocorrelation method

- $|k_i| < 1, \quad \forall i$
- $H(z)$ is a minimum phase system (*stability*)
- Flip all maximum-phase poles inside the unit circle to their conjugate reciprocal locations
- One-to-One correspondence: $k_i \rightleftarrows l_i$, $l_i \rightleftarrows r_n[i]$.

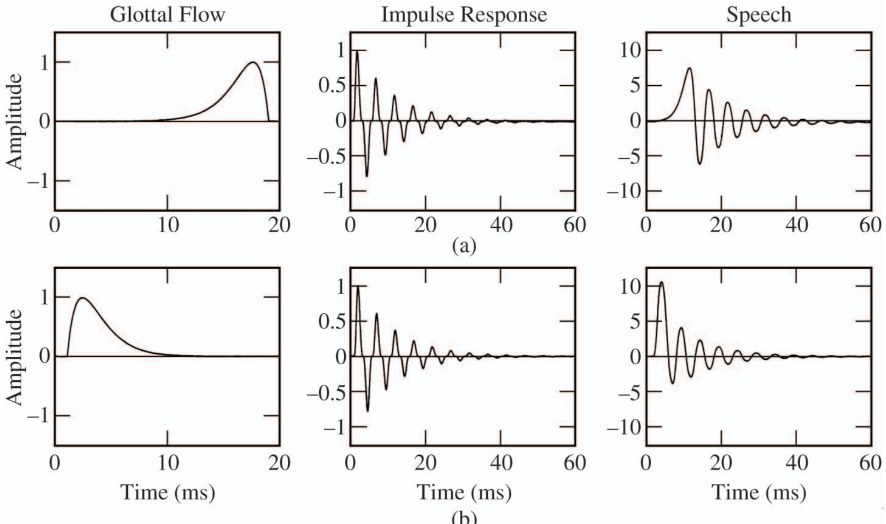
$$\begin{aligned} k_i &= l_i^i \\ l_j^{i-1} &= \frac{l_j^i + k_i l_{i-j}^i}{1 - k_i^2} \end{aligned}$$

- *Autocorrelation matching:* If, $H(z)$ is an pth all-pole minimum phase system, and if $r_h[0] = r_n[0]$, then:

$$r_h[\tau] = r_n[\tau], \quad \text{for } |\tau| \le p$$

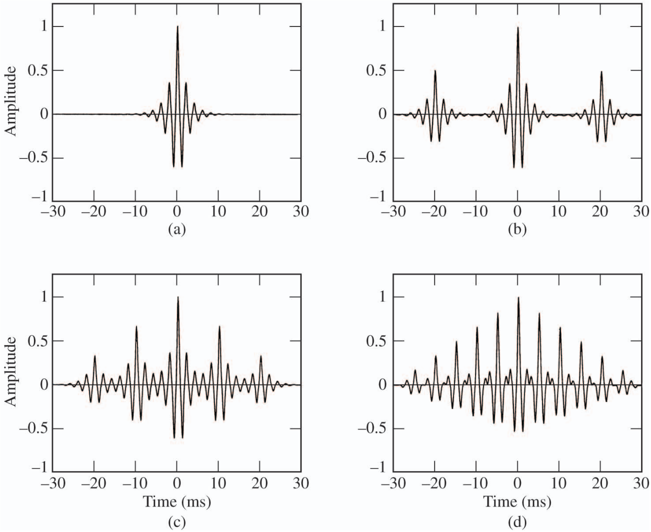
Consequence I

▷ Flip all maximum-phase poles inside the unit circle to their conjugate reciprocal locations



Consequence II

▷ Autocorrelation matching



Consequence III

▷ Autocorrelation matching:

$$A^2 = r_h[0] - \sum_{k=1}^{p} l_k r_h[k]$$

or

$$A^2 = r_n[0] - \sum_{k=1}^{p} l_k r_n[k] = E_n$$

Estimations in the frequency domain

- Let $|S(\omega)|$ be the magnitude spectrum of speech and $H(\omega) = A/A(\omega)$ be an all-pole model
- Define a frequency-domain error function

$$I = \frac{1}{2\pi}\int_{-\pi}^{\pi}[e^{Q(\omega)} - Q(\omega) - 1]d\omega$$

where

$$Q(\omega) = \log|S(\omega)|^2 - \log|H(\omega)|^2 = \log\left|\frac{E(\omega)}{A}\right|^2$$

- Minimizing I over the linear prediction coefficients, results in the minimization of:

$$\int_{-\pi}^{\pi}|E(\omega)|^2 d\omega$$

- Minimizing I over A it gives:

$$A^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}|E(\omega)|^2 d\omega$$

Estimations in the frequency domain

- Let $|S(\omega)|$ be the magnitude spectrum of speech and $H(\omega) = A/A(\omega)$ be an all-pole model
- Define a frequency-domain error function

$$I = \frac{1}{2\pi}\int_{-\pi}^{\pi}[e^{Q(\omega)} - Q(\omega) - 1]d\omega$$

where

$$Q(\omega) = \log|S(\omega)|^2 - \log|H(\omega)|^2 = \log\left|\frac{E(\omega)}{A}\right|^2$$

- Minimizing I over the linear prediction coefficients, results in the minimization of:

$$\int_{-\pi}^{\pi}|E(\omega)|^2 d\omega$$

- Minimizing I over A it gives:

$$A^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}|E(\omega)|^2 d\omega$$

Estimations in the frequency domain

- Let $|S(\omega)|$ be the magnitude spectrum of speech and $H(\omega) = A/A(\omega)$ be an all-pole model
- Define a frequency-domain error function

$$I = \frac{1}{2\pi}\int_{-\pi}^{\pi}[e^{Q(\omega)} - Q(\omega) - 1]d\omega$$

where

$$Q(\omega) = \log|S(\omega)|^2 - \log|H(\omega)|^2 = \log\left|\frac{E(\omega)}{A}\right|^2$$

- Minimizing I over the linear prediction coefficients, results in the minimization of:

$$\int_{-\pi}^{\pi}|E(\omega)|^2 d\omega$$

- Minimizing I over A it gives:

$$A^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}|E(\omega)|^2 d\omega$$

Estimations in the frequency domain

- Let $|S(\omega)|$ be the magnitude spectrum of speech and $H(\omega) = A/A(\omega)$ be an all-pole model
- Define a frequency-domain error function

$$I = \frac{1}{2\pi}\int_{-\pi}^{\pi}[e^{Q(\omega)} - Q(\omega) - 1]d\omega$$

where

$$Q(\omega) = \log|S(\omega)|^2 - \log|H(\omega)|^2 = \log\left|\frac{E(\omega)}{A}\right|^2$$

- Minimizing I over the linear prediction coefficients, results in the minimization of:

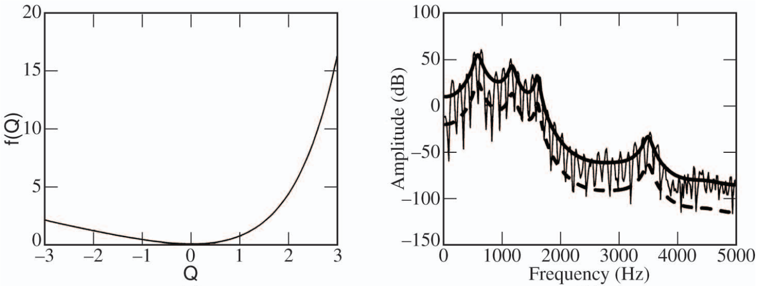
$$\int_{-\pi}^{\pi}|E(\omega)|^2 d\omega$$

- Minimizing I over A it gives:

$$A^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}|E(\omega)|^2 d\omega$$

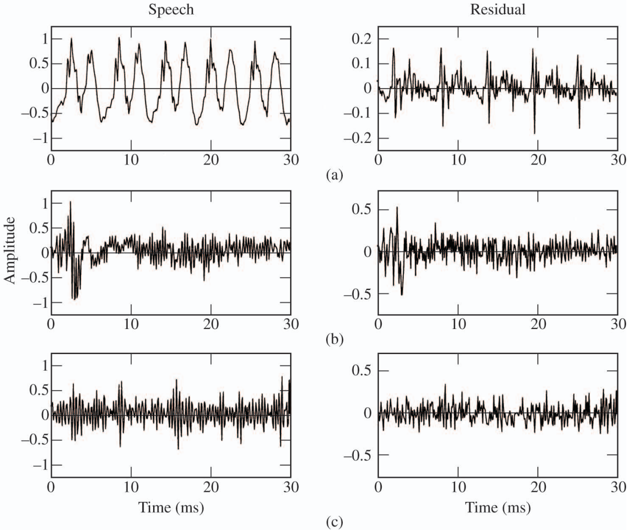
Favoring spectral peaks

A note on $f(Q) = e^Q - Q - 1$

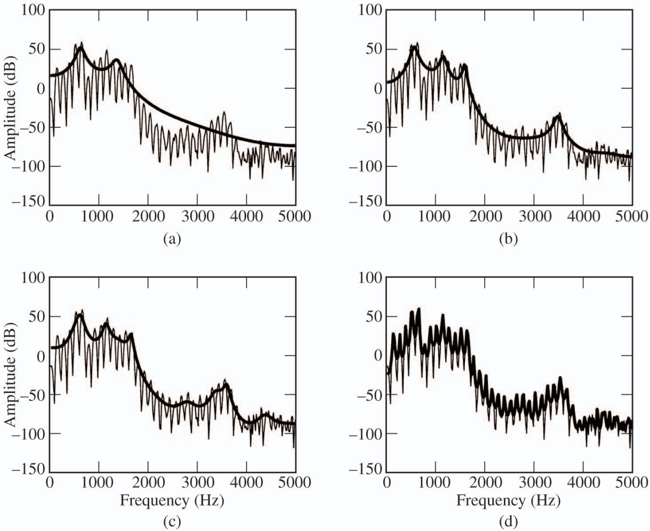


(a) (b)

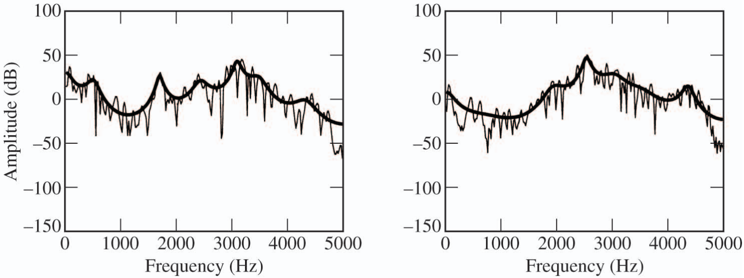
Time-domain



Frequency-domain: voiced



Frequency-domain: unvoiced



(a)

(b)

Comparing Covariance and Autocorrelation

- Simple test of estimation

$$s[n] = a^n u[n] \star \delta[n]$$

- Stability issues
- Sensitivity, pitch-synchronous analysis

Comparing Covariance and Autocorrelation

- Simple test of estimation

$$s[n] = a^n u[n] \star \delta[n]$$

- Stability issues
- Sensitivity, pitch-synchronous analysis

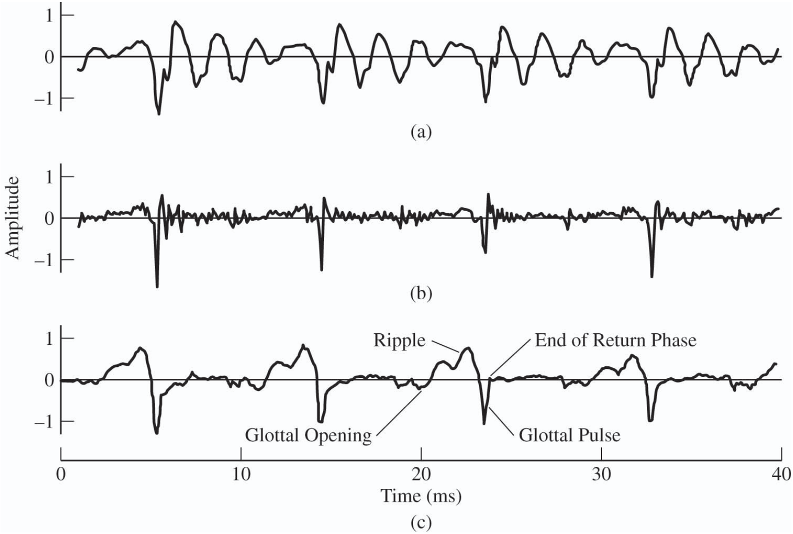
Comparing Covariance and Autocorrelation

- Simple test of estimation

$$s[n] = a^n u[n] \star \delta[n]$$

- Stability issues
- Sensitivity, pitch-synchronous analysis

Sensitivity, pitch-synchronous analysis



Outline

Synthesis

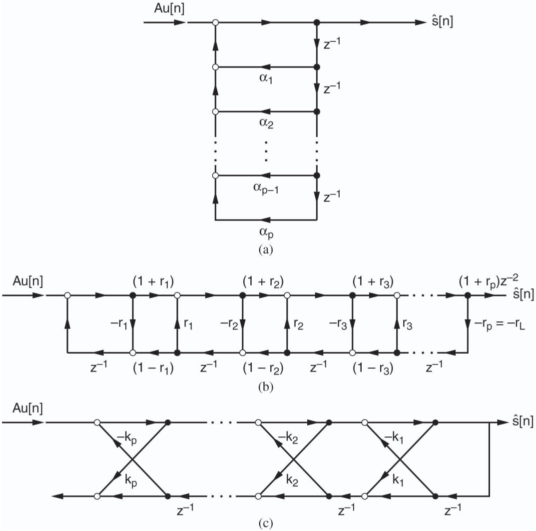
The synthesized speech is:

$$s[n] = \sum_{k=1}^{p} l_k s[n-k] + Au[n]$$

where $u[n]$ could be:

- A periodic impulse train
- An impulse
- White noise

Synthesis Structure



Consider ...

- Window duration
- Frame interval (frame rate)
- Model order
- Voiced/unvoiced state and pitch estimation
- Synthesis structure

Consider ...

- Window duration
- Frame interval (frame rate)
- Model order
- Voiced/unvoiced state and pitch estimation
- Synthesis structure

Consider ...

- Window duration
- Frame interval (frame rate)
- Model order
- Voiced/unvoiced state and pitch estimation
- Synthesis structure

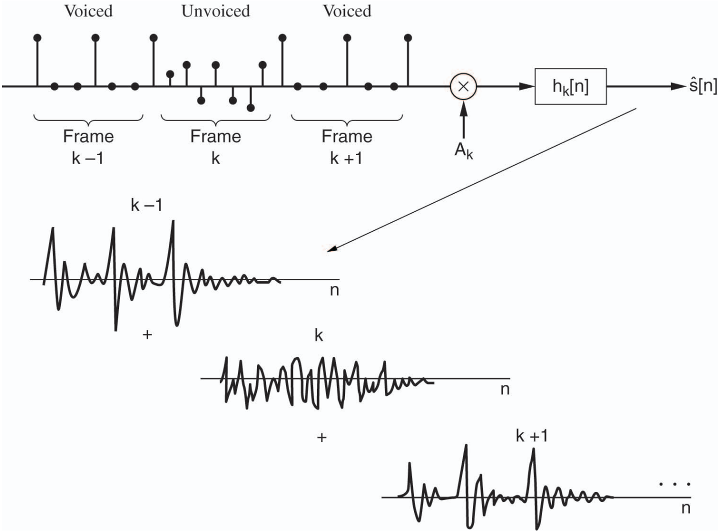
Consider ...

- Window duration
- Frame interval (frame rate)
- Model order
- Voiced/unvoiced state and pitch estimation
- Synthesis structure

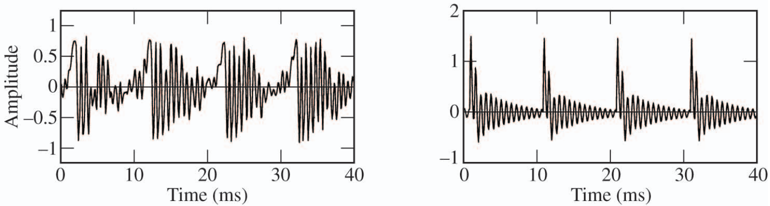
Consider ...

- Window duration
- Frame interval (frame rate)
- Model order
- Voiced/unvoiced state and pitch estimation
- Synthesis structure

Overlap and Add, OLA



Speech Reconstruction example



(a) (b)

How does it sound ...

- /a/
- /e/

Outline

Acknowledgments

Most, if not all, figures in this lecture are coming from the book:

T. F. Quatieri: Discrete-Time Speech Signal Processing, principles and practice
2002, Prentice Hall

and have been used after permission from Prentice Hall

Outline

J. Makhoul, “Linear Prediction: A Tutorial Review,” *Proceedings of the IEEE*, vol. 63, pp. 561–580, April 1975.

Τέλος Ενότητας

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημειώματα

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Κρήτης, Στυλιανού Ιωάννης. «Ψηφιακή Επεξεργασία Φωνής. Γραμμική Πρόβλεψη: Ανάλυση και Σύνθεση». Έκδοση: 1.0. Ηράκλειο/Ρέθυμνο 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: http://www.csd.uoc.gr/~hy578

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Εικόνες/Σχήματα/Διαγράμματα/Φωτογραφίες

Εικόνες/σχήματα/διαγράμματα/φωτογραφίες που περιέχονται σε αυτό το αρχείο προέρχονται από το βιβλίο:

Τίτλος: *Discrete-time Speech Signal Processing: Principles and Practice*

Prentice-Hall signal processing series, ISSN 1050-2769

Συγγραφέας: Thomas F. Quatieri

Εκδότης: Prentice Hall PTR, 2002

ISBN: 013242942X, 9780132429429

Μέγεθος: 781 σελίδες

και αναπαράγονται μετά από άδεια του εκδότη.