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

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:

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

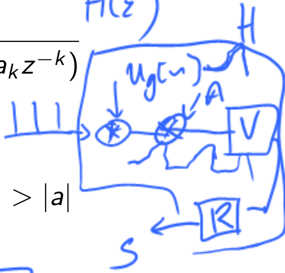
Transfer function from the glottis to the lips

$u_g[n] \to \boxed{h[n]} \to s[n]$

- We shown that for voiced speech:

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

$H(z)$



- However:

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

- Then:

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

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:

Έξοδος

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

Είσοδος

or

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

z^{-1}

and in time domain:

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

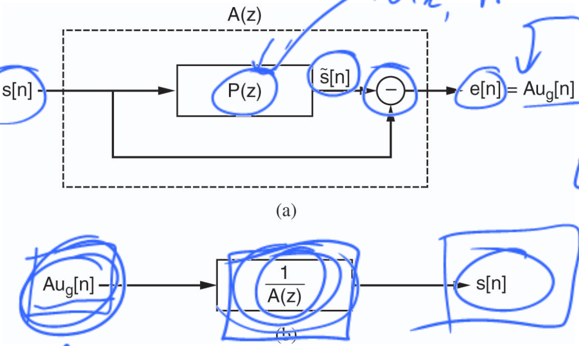
$\hookrightarrow s[n] - \sum_{k=1}^{p} a_k s[n-k]$

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

Filtering view of linear prediction

$\to a_k, A$

Analysis



Synthesis

LPC-10

Per.

Nonper.

where

$$P(z) = \sum_{k=1}^{p} a_k z^{-k} \quad \textit{prediction filter}$$

$$A(z) = 1 - P(z) \quad \textit{prediction error filter}$$

Outline

Justification of LP for speech

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

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

- 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:

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

- 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:

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

- 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:

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

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

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

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:

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

n

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

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

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

$m: n-M$
$n+M$

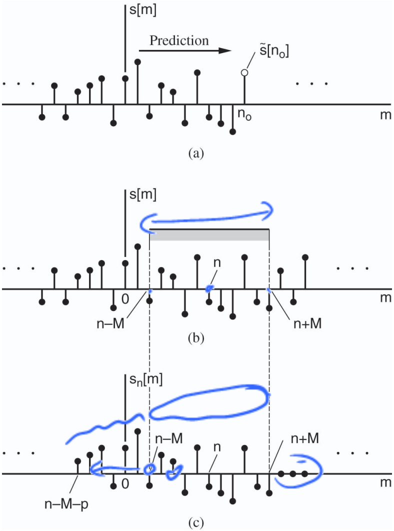
where

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

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

l $\quad s_{-p} \cdots s_{-4}\, s_{-3}\, s_{-2}\, s_{-1}\, s_0$ $\quad \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{2M+1} \end{bmatrix}$ $\quad [n-M : n+M]$ $\quad 2M+1$

- In matrix notation

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

$e = s - S l$ $\qquad \begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ \vdots \\ l_p \end{bmatrix}$

cy

- Mean-squared error

$[\quad]^T[\quad] \cdot e$

$cy + e = x \Rightarrow e = x - cy$ $\qquad e^T$

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

- Solution:

$\rightarrow e^T = s^T - l^T S^T$

$\langle e, y \rangle = 0$

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

- Same solution by considering the *Projection Theorem*:

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

$\Rightarrow S^T S \cdot l = S^T \cdot s$

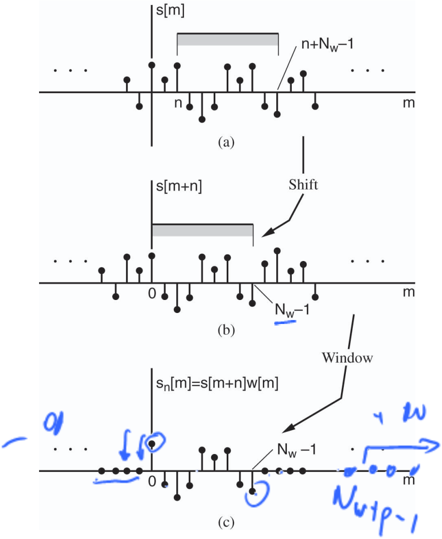
$\Rightarrow l = (S^T S)^{-1} \cdot S^T s$

$\dfrac{\partial e_n^T \cdot e_n}{\partial l} = 0 \Rightarrow -2 S^T \cdot s + 2 S^T S\, l = 0 \Rightarrow$

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 $[0,\ N_w + p - 1]$:

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

N_w

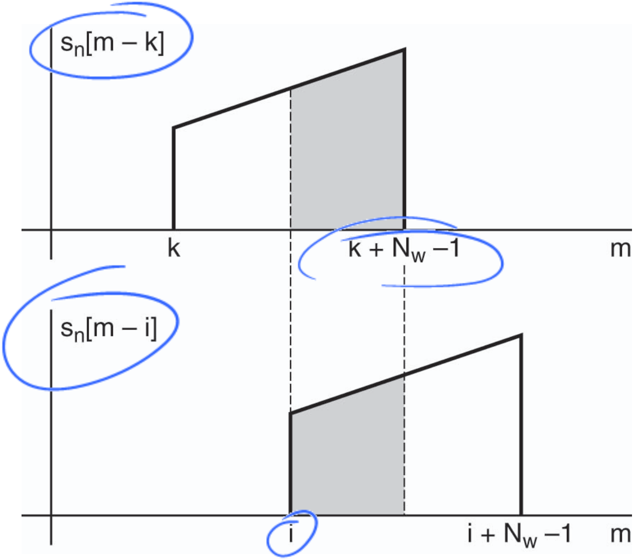
- Normal equations:

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

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:

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

- 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:

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

- In matrix notation:

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

$R \cdot l = r$

- Or (Toeplitz matrix):

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

$r_n[i]$

$r_n[i]$

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*

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

$$k_1 = \frac{r[1]}{r[0]}$$

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

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

$$l_1^1 = k_1$$

$$E^i < E^{i-1}$$

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

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

$$E^1 = (1-k_1^2)E^0$$

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

$$|k_i| < 1$$

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

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:

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

- 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:

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

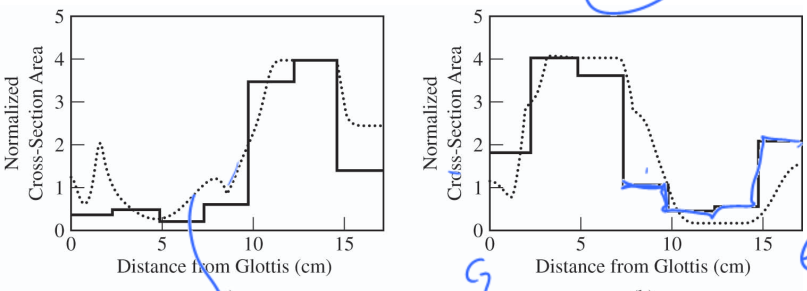
- Recursions:

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

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

Estimating the vocal tract area functions via the autocorrelation method

① ②



(a) (b)

$\boxed{S}$

$S[n] \rightarrow k_i \rightarrow r_i = \frac{A_{i+1} - A_i}{A_{i+1} + A_i}$

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: $k_i \rightleftarrows l_i$, $l_i \rightleftarrows r_n[i]$.

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

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

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

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: $k_i \rightleftarrows l_i$, $l_i \rightleftarrows r_n[i]$.

$\frac{1}{(1-bz)^2}$

$u_g[n]$

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

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

$z = re^{j\theta}$

$\frac{1}{z^*} = \frac{1}{r} e^{j\theta}$

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

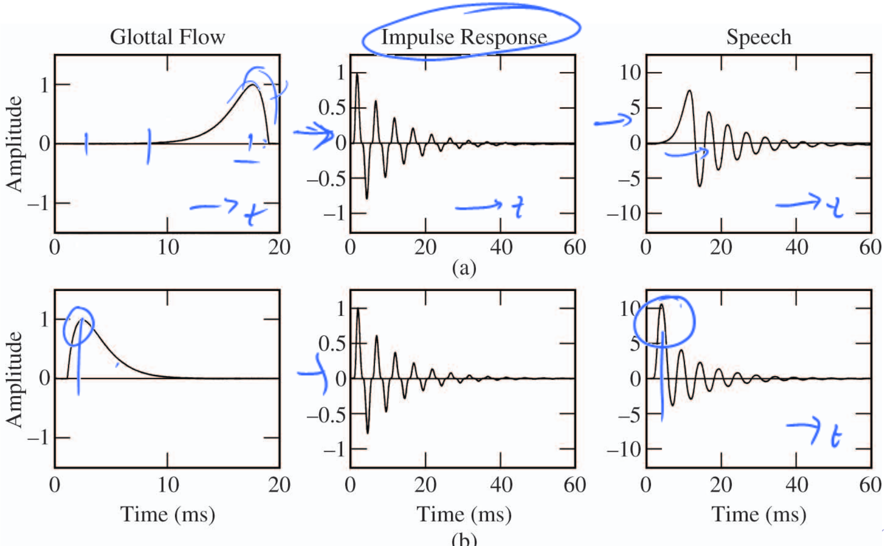
$$\boxed{r_h[\tau] = r_n[\tau],} \quad \text{for } |\tau| \le p$$

$u_g[n] \rightarrow \boxed{h[n]} \rightarrow s[n]$

$r_n[\tau] \rightarrow l_i \leftarrow$ συντ. γραμ. πρόβλεψης

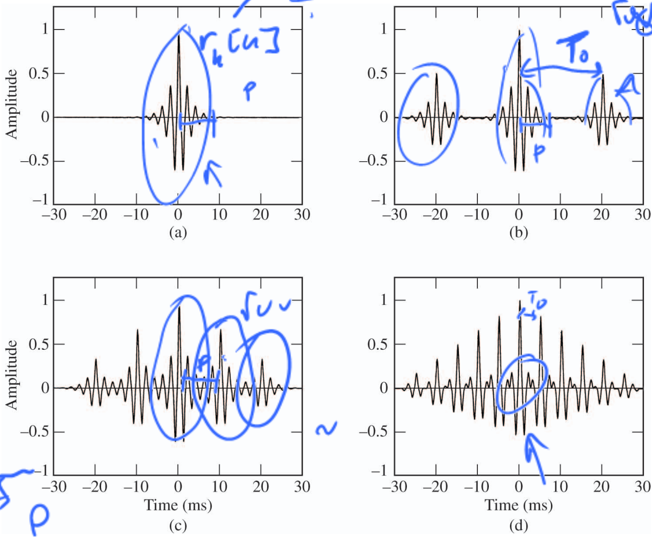
CONSEQUENCE I

▷ 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

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

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:

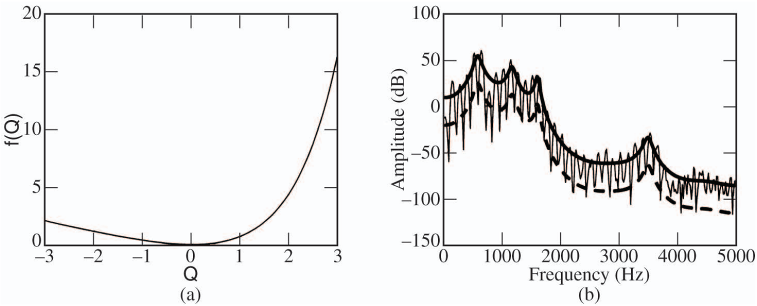
$$\int_{-\pi}^{\pi} |E(\omega)|^2 d\omega$$

- Minimizing I over A it gives:

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

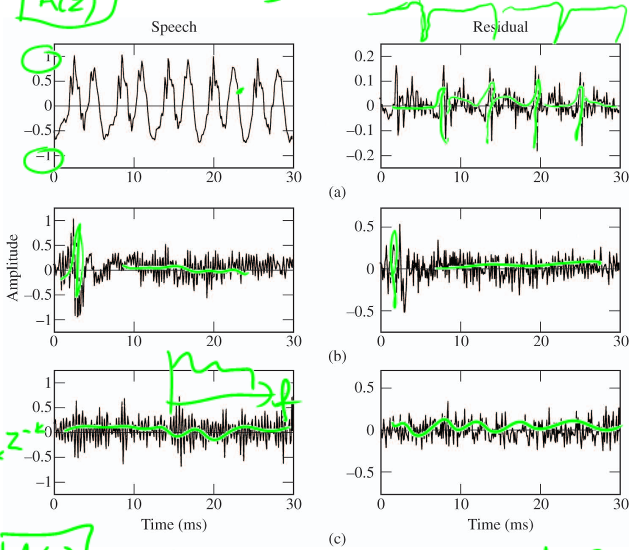
Favoring spectral peaks

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



Time-domain

$A u_g[n] \rightarrow \boxed{\frac{1}{A(z)}} \rightarrow s[n]$



$A(z) = 1 - \sum_{k=1}^{P} a_k z^{-k}$

$\rightarrow s[n] \rightarrow \boxed{A(z)} \rightarrow A u_g[n]$ excitation ≈ Residual

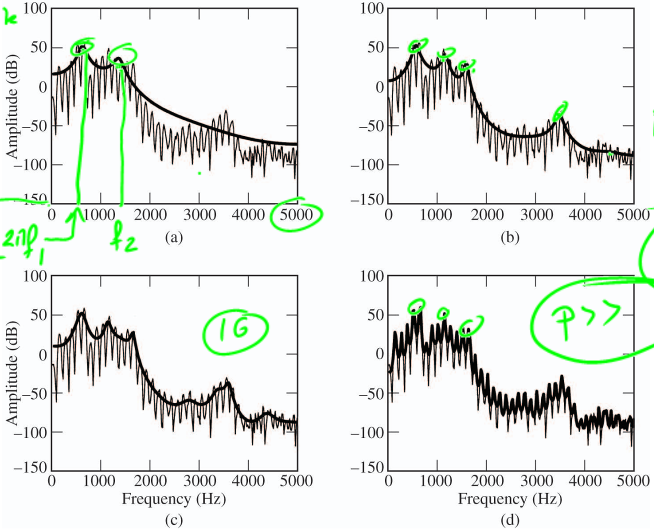
Frequency-domain: voiced

$p=4$ $\quad (1-r_1 z^{-1})(1-r_1^* z^{-1})\cdot(1-r_2 z^{-1})(1-r_2^* z^{-1})$

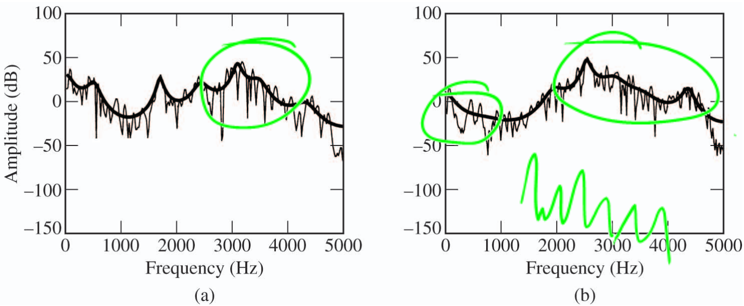
P

$\sum_{k=1}^{P} a_k z^{-k}$, $a_k \in \mathbb{R}$

$\omega_1 = 2\pi f_1$ $\quad f_2$



FREQUENCY-DOMAIN: UNVOICED



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

$P = 1$

(1) Cov. $\rightarrow a$ ✓

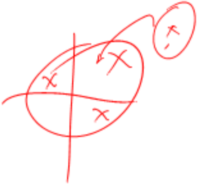
(2) Autocor. $\rightarrow a \dfrac{1-a^{2N-2}}{1-a^{2N}}$ $N \to \infty$ a

N πόσα δείγματα του σήματος χρειάζ. για τον υπολ. της αυτοσυσχέτισης

- ~~Simple test of estimation~~
- Stability issues
- Sensitivity, pitch-synchronous analysis

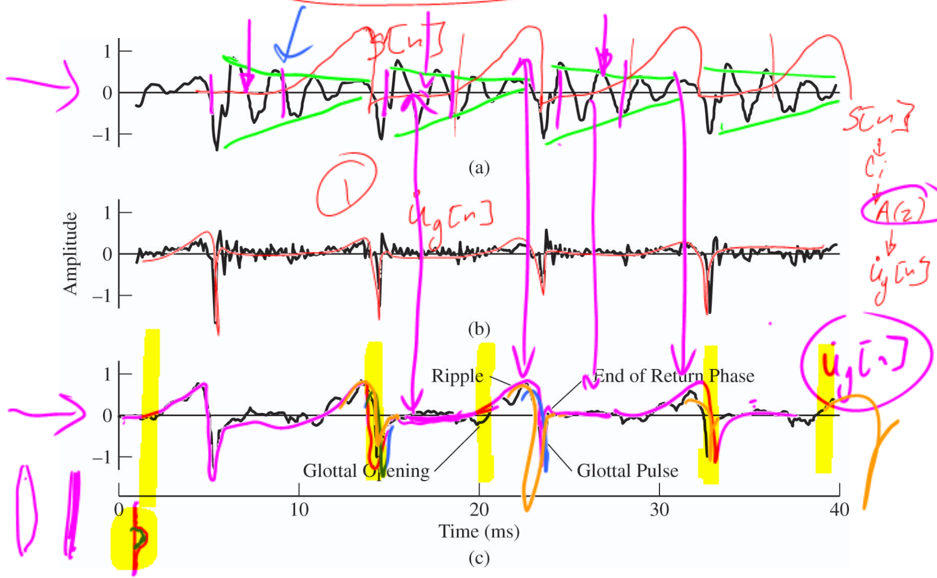
$$s[n] = a^n u[n] \star \delta[n] = a^n u[n] \rightarrow \frac{1}{1-az^{-1}}$$

$a < 1$



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

Sensitivity, pitch-synchronous analysis



Outline

Synthesis

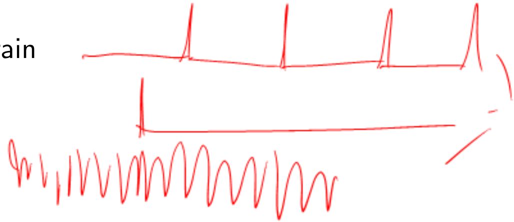
$$u_g[n] \rightarrow \boxed{l_k} \rightarrow s[n]$$

The synthesized speech is:

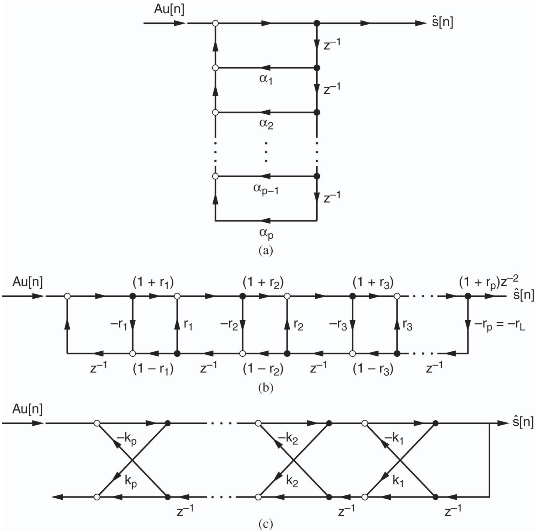
$$s[n] = \sum_{k=1}^{p} l_k s[n-k] + Au[n]$$

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

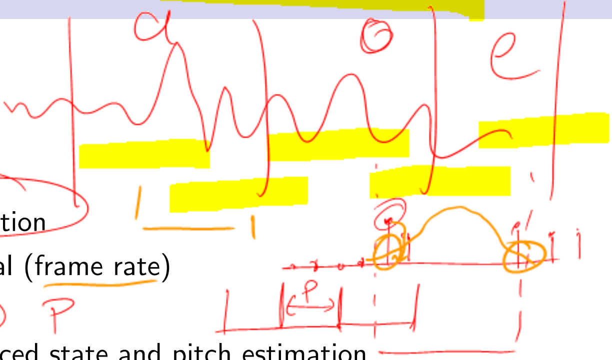
Consider ...

- 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 ...

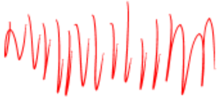


- Window duration
- Frame interval (frame rate)
- Model order

P

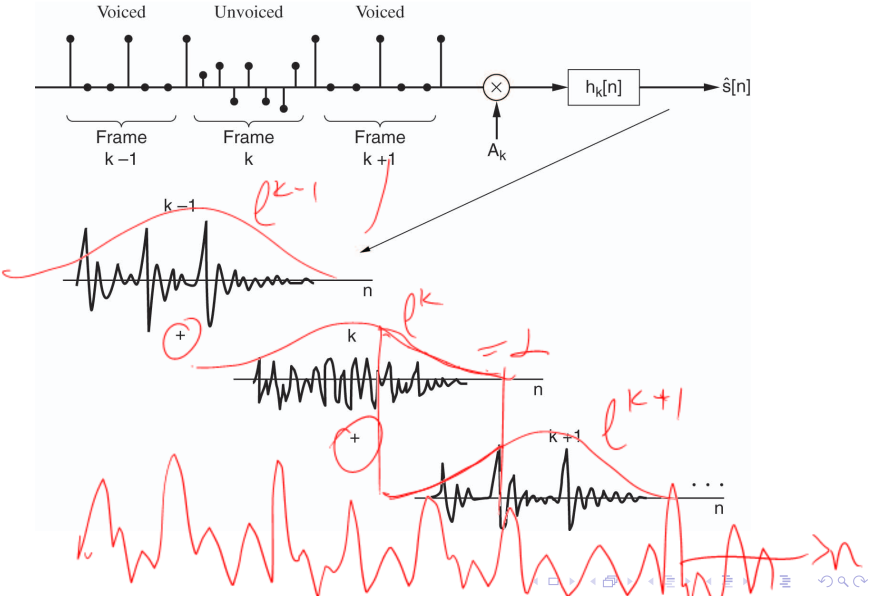
- Voiced/unvoiced state and pitch estimation
- Synthesis structure

8 kHz   P = 12

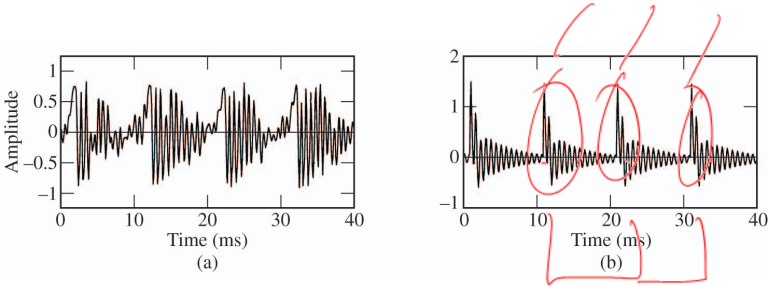


$f_s = 16\,\text{kHz}$ $p = f_s + 4$

Overlap and Add, OLA



Speech Reconstruction example



How does it sound ...

- /a/ pgflastimage
- /e/ pgflastimage

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.