Speech and Audio Coding Theory

Contents of lecture

- Solutions to LPC analysis
  - Covariance method (CM)
  - Lattice method (LM)
- Practical implementation of LPC analysis
- Interpretation of LPC analysis
- Pitch prediction
  - Pitch predictor (LTP) formulation

Covariance method (CM)

- Assumption: To consider only the fixed analysis frame, $0 \le m \le N\text{-}1$
  - That is, there is a constraint on analysis frame, but not on signal itself.
- Solution
  - $MSE = E\{e^2(n)\} = \sum_{m=0}^{N-1} e_n^2(m)$
  - $\phi_n(i, j) = \sum_{m=0}^{N-1} s_n(m-i)s_n(m-j), \quad 1 \le i \le p, \; 0 \le j \le p$
  - Now, let $m-i=m'$, then $m-j=m'+i-j$.
  - And, $m=0 \rightarrow m'=-i$ and $m=N-1 \rightarrow m'=N-1-i$.
  - Therefore, $\phi_n(i, j) = \sum_{m'=-i}^{N-i-1} s_n(m')s_n(m'+i-j), \quad 1 \le i \le p, \; 0 \le j \le p$

Covariance method (CM)

- Solution (cont.)
  - So, $\sum_{j=1}^{p} \alpha_j \phi_n(i,j) = \phi_n(i,0) \quad 1 \le i \le p$
  - Also, in matrix form, $\begin{bmatrix} \phi_n(1,1) & \phi_n(1,2) & . & \phi_n(1,p) \\ \phi_n(2,1) & \phi_n(2,2) & . & \phi_n(2,p) \\ \vdots & \vdots & \vdots & \vdots \\ \phi_n(p,1) & \phi_n(p,2) & . & \phi_n(p,p) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} \phi_n(1,0) \\ \phi_n(2,0) \\ \vdots \\ \phi_n(p,0) \end{bmatrix}$
  - The above matrix is also symmetric, but no longer Toeplitz.
  - So, we cannot use the Durbin's algorithm. → Cholesky decomposition method.
    - $\phi = VDV^T$ where V is a lower triangular matrix with 1's as diagonal elements and D is a diagonal matrix.
    - Refer to L.R. Rabiner and R.W. Schafer, *Digital processing of speech signals*, pp. 407-410.

Lattice method (LM)

- Another implementation of the autocorrelation-based solution
- Formulation
  - The i–th order inverse filter: $A^{(i)}(z) = 1 - \sum_{j=1}^{i} \alpha_j^{(i)} z^{-j}$ (from Durbin's algorithm)
  - Then, the prediction error of i–th order predictor:

$$e_n^{(i)}(m) = e^{(i)}(m) = s(m) - \sum_{j=1}^{i} \alpha_j^{(i)} s(m-j)$$

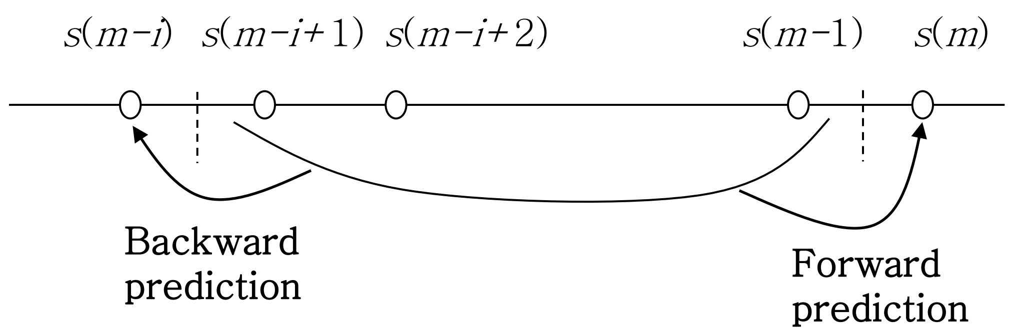
  - Its z–transform: $E^{(i)}(z) = A^{(i)}(z)S(z)$

Lattice method (LM)

- Formulation (cont.)
  - Substituting $\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \quad 1 \le j \le i-1$ to the inverse filter eq., we obtain
    - $$A^{(i)}(z) = 1 - \sum_{j=1}^{i-1} [\alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}] z^{-j} - \alpha_i^{(i)} z^{-i}$$
  - Rearranging the eq. and using $\alpha_i^{(i)} = k_i$
    - $$A^{(i)}(z) = 1 - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} z^{-j} + k_i \sum_{j=1}^{i-1} \alpha_{i-j}^{(i-1)} z^{-j} - k_i z^{-i}$$
    - $$A^{(i)}(z) = A^{(i-1)}(z) - k_i \left[z^{-i} - \sum_{j=1}^{i-1} \alpha_{i-j}^{(i-1)} z^{-j} \right]$$

Lattice method (LM)

- Formulation (cont.)
  - Therefore, $$E^{(i)}(z) = A^{(i-1)}(z)S(z) - k_i\left[z^{-i} - \sum_{j=1}^{i-1}\alpha_{i-j}^{(i-1)}z^{-j}\right]S(z)$$
    - The first term: the forward prediction error for $(i-1)$th order predictor.
    - The second term except k_i: the backward prediction error for $(i-1)$th order predictor.



Lattice method (LM)

- Formulation (cont.)
  - Modifying the eq.,

$$E^{(i)}(z) = A^{(i-1)}(z)S(z) - k_i z^{-1}\left[z^{-(i-1)} - \sum_{j=1}^{i-1} \alpha_{i-j}^{(i-1)} z^{-(j-1)} \right] S(z)$$

  - Here, since

$$\sum_{j=1}^{i-1} \alpha_{i-j}^{(i-1)} z^{-(j-1)} = \alpha_{i-1}^{(i-1)} z^{0} + \alpha_{i-2}^{(i-1)} z^{-1} + \cdots + \alpha_{1}^{(i-1)} z^{-i+2} = \sum_{j=1}^{i-1} \alpha_{j}^{(i-1)} z^{j-(i-1)}$$

  - Thus, the forward prediction error in terms of the lower-order inverse filter is

$$E^{(i)}(z) = A^{(i-1)}(z)S(z) - k_i z^{-1}\left[z^{-(i-1)} - \sum_{j=1}^{i-1} \alpha_{j}^{(i-1)} z^{j-(i-1)} \right] S(z)$$

Lattice method (LM)

- Formulation (cont.)
  - Let the backward prediction error of i–th order predictor be

$$B^{(i)}(z) = z^{-i} A^{(i)}(z^{-1}) S(z)$$

  - Then, since

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i \left[z^{-i} - \sum_{j=1}^{i-1} \alpha_{i-j}^{(i-1)} z^{-j} \right]$$

$$= A^{(i-1)}(z) - k_i \left[z^{-i} - \sum_{j=1}^{i-1} \alpha_{j}^{(i-1)} z^{j-i} \right]$$

$$= A^{(i-1)}(z) - k_i z^{-i} \left[1 - \sum_{j=1}^{i-1} \alpha_{j}^{(i-1)} z^{j} \right]$$

$$= A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1})$$

Lattice method (LM)

- Formulation (cont.)
  - Thus, we can get

$$B^{(i)}(z) = z^{-i}\left[A^{(i-1)}(z^{-1}) - k_i z^i A^{(i-1)}(z)\right]S(z)$$

$$= z^{-i} A^{(i-1)}(z^{-1})S(z) - k_i A^{(i-1)}(z)S(z)$$

$$= z^{-1} z^{-(i-1)} A^{(i-1)}(z^{-1})S(z) - k_i A^{(i-1)}(z)S(z)$$

  - Consequently, we can obtain the backward prediction error,

$$b^{(i)}(m) = b^{(i-1)}(m-1) - k_i e^{(i-1)}(m)$$

Lattice method (LM)

- Formulation (cont.)
  - Also since $\left[z^{-(i-1)} - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} z^{j-(i-1)}\right] S(z) = z^{-(i-1)}\left[1 - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} z^{j}\right] S(z)$

$$= z^{-(i-1)} A^{(i-1)}(z^{-1}) S(z)$$

$$= B^{(i-1)}(z)$$

  - From the previous eq., we can get

$$E^{(i)}(z) = A^{(i-1)}(z)S(z) - k_i z^{-1}\left[z^{-(i-1)} - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} z^{j-(i-1)}\right] S(z)$$

$$= A^{(i-1)}(z)S(z) - k_i z^{-1} B^{(i-1)}(z)$$

  - Therefore, we obtain $e^{(i)}(m) = e^{(i-1)}(m) - k_i b^{(i-1)}(m-1)$
    - Here, $e^{(0)}(m) = b^{(0)}(m) = s(m)$.

Lattice method (LM)

- Final solution for k_i without using α_j
  - Geometric mean of two solutions for minimum mean squared forward prediction error (MMSFE) and minimum mean squared backward prediction error (MMSBE)

$$k_i = \frac{\sum_{m=0}^{N-1} e^{(i-1)}(m) b^{(i-1)}(m-1)}{\sqrt{\sum_{m=0}^{N-1} [e^{(i-1)}(m)]^2 \times \sum_{m=0}^{N-1} [b^{(i-1)}(m-1)]^2}}$$

    - Refer to J. Makhoul, "Stable and efficient lattice methods for linear prediction," IEEE Trans. on ASSP, pp. 423-428, Oct. 1977.
  - k_i : normalized cross correlation function between $e^{(i-1)}(m)$ and $b^{(i-1)}(m)$ → PARtial CORrelation (PARCOR) coefficients

Lattice method (LM)

- Another solution for k_i
  - Burg implementation is based on the minimization of the sum of the mean squared forward and backward prediction errors, i.e.,

$$\hat{E}^{(i)} = \sum_{m=0}^{N-1}[(e^{(i)}(m))^2 + (b^{(i)}(m))^2]$$

  - Then,

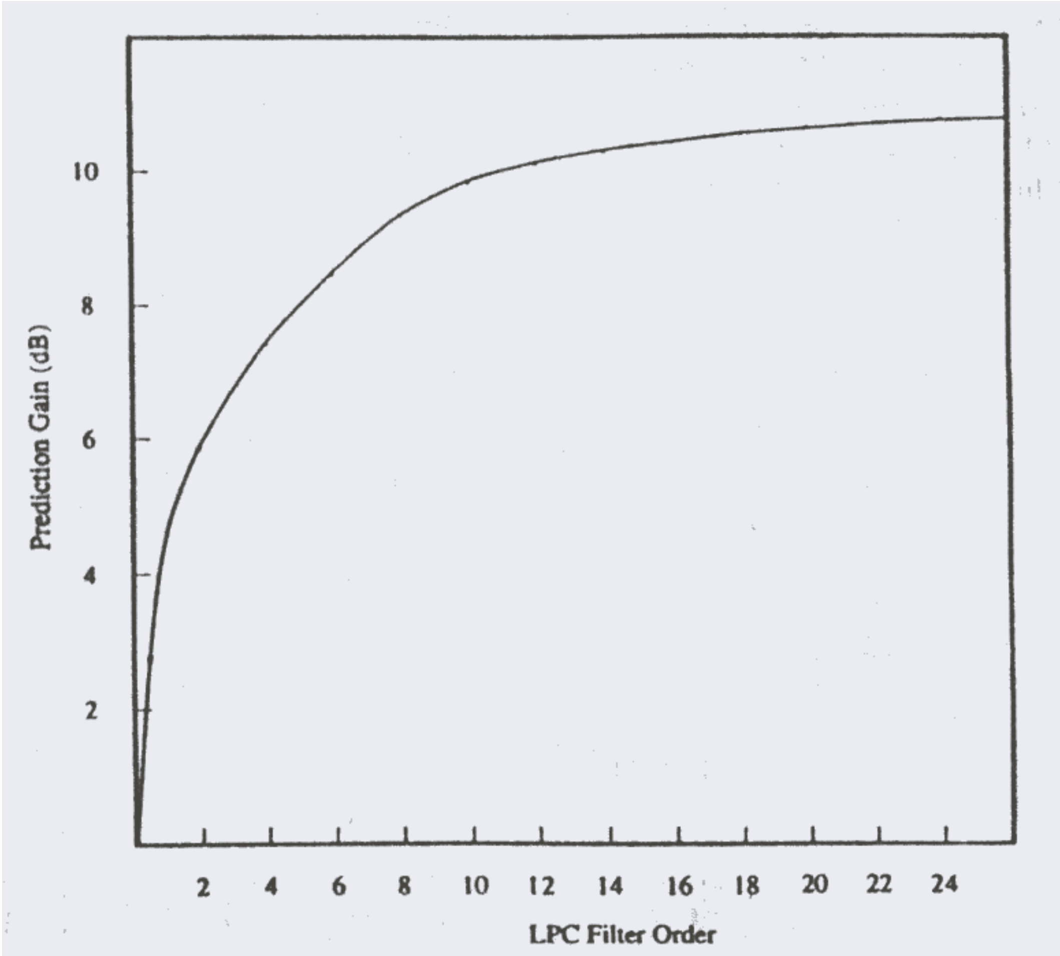
$$k_i = \frac{2\sum_{m=0}^{N-1} e^{(i-1)}(m) b^{(i-1)}(m-1)}{\sum_{m=0}^{N-1}[e^{(i-1)}(m)]^2 + \sum_{m=0}^{N-1}[b^{(i-1)}(m-1)]^2}$$

- All the solutions guarantee a stable filter since $|k_i| \leq 1$.

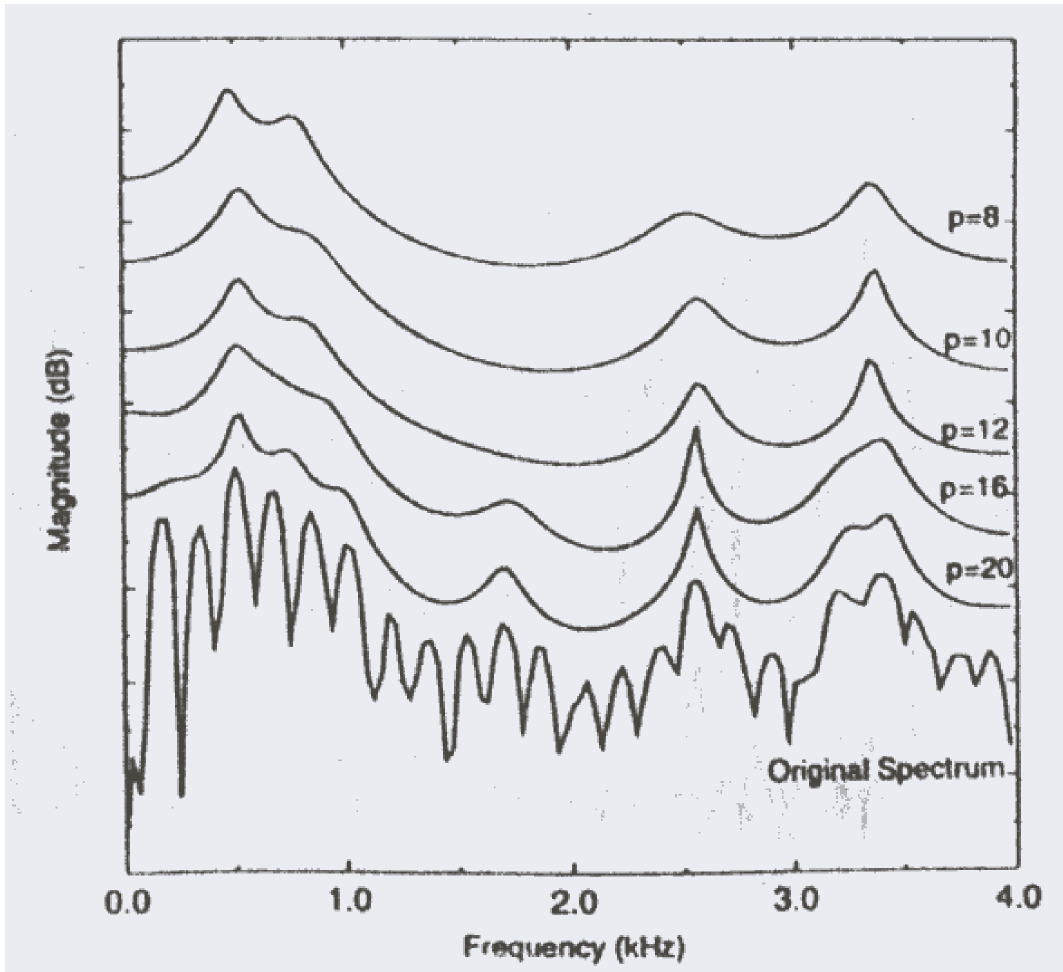
Practical implementation of LPC analysis

- Consideration factors
  - Performance, efficiency, stability
    - LM guarantees stability that is important in real implementation.
    - If a careful choice of windowing and fine precision arithmetic is performed, then both AM and CM are also stable.
  - Filter order (p), frame size (N)
    - In 8 kHz sampling, 4 kHz bandwidth → Usually 4 formants → At least, p = 8 → p = 10 for accuracy
      - Exception: In CCITT 16 kbps low-delay coder standard, p = 50
    - Frame size: 16-32msec to cover several pitch periods.
    - Results of the LPC analysis are different according to partitioning points of the analysis frame → No solution.

LPC prediction gain vs. LPC order



LPC envelops vs. LPC order



Practical implementation of LPC analysis

- Other consideration factors
  - Pre-emphasis: high-pass filtering for flattening the spectral envelop
  - Window overlapping: to overcome block-edge effects (10-20% of frame size)
  - Interpolation of LPC coefficients: in order to smooth out transitional effects

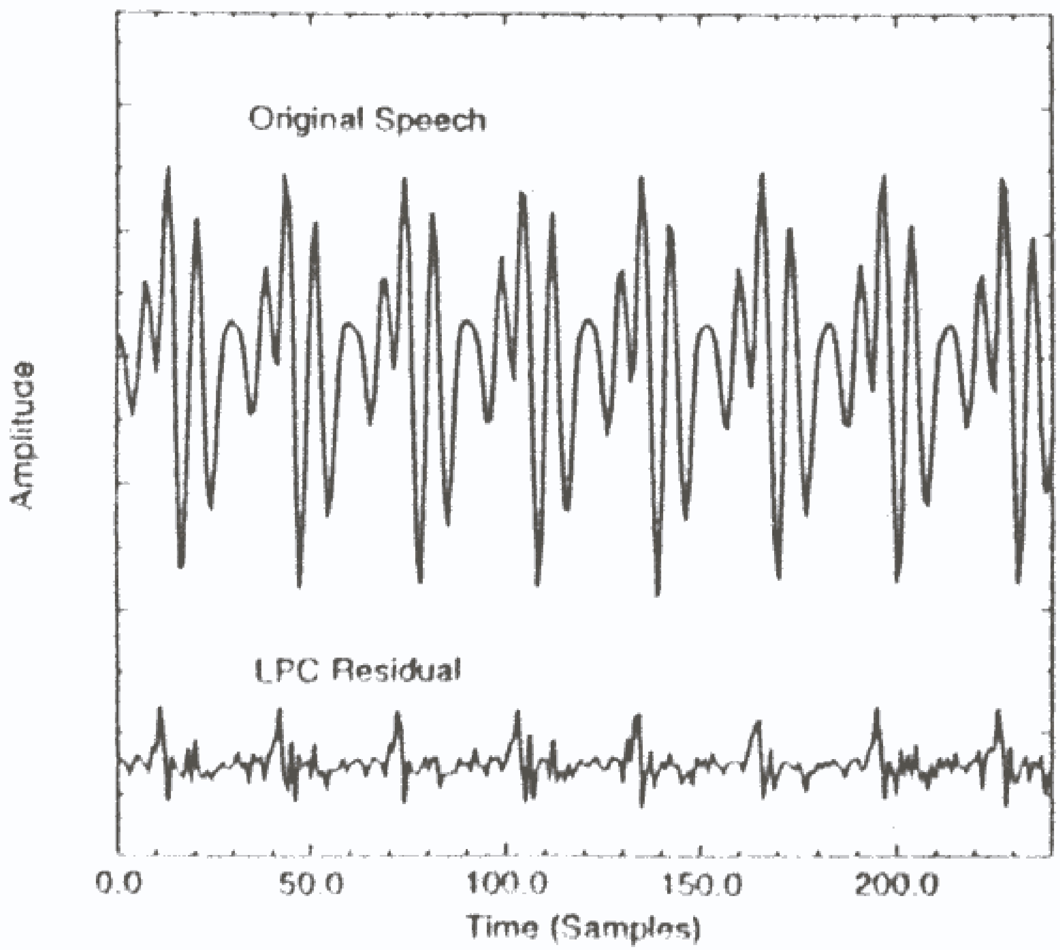
Interpretation of LPC analysis

- Residual signal
  - If $\alpha_j = a_j$, then the residual signal, $e(n) = Gx(n) = u(n)$ → excitation signal (which is assumed to be impulse train or the white noise)
  - $e(n)$ is obtained by the inverse filter $H^{-1}(z)$,

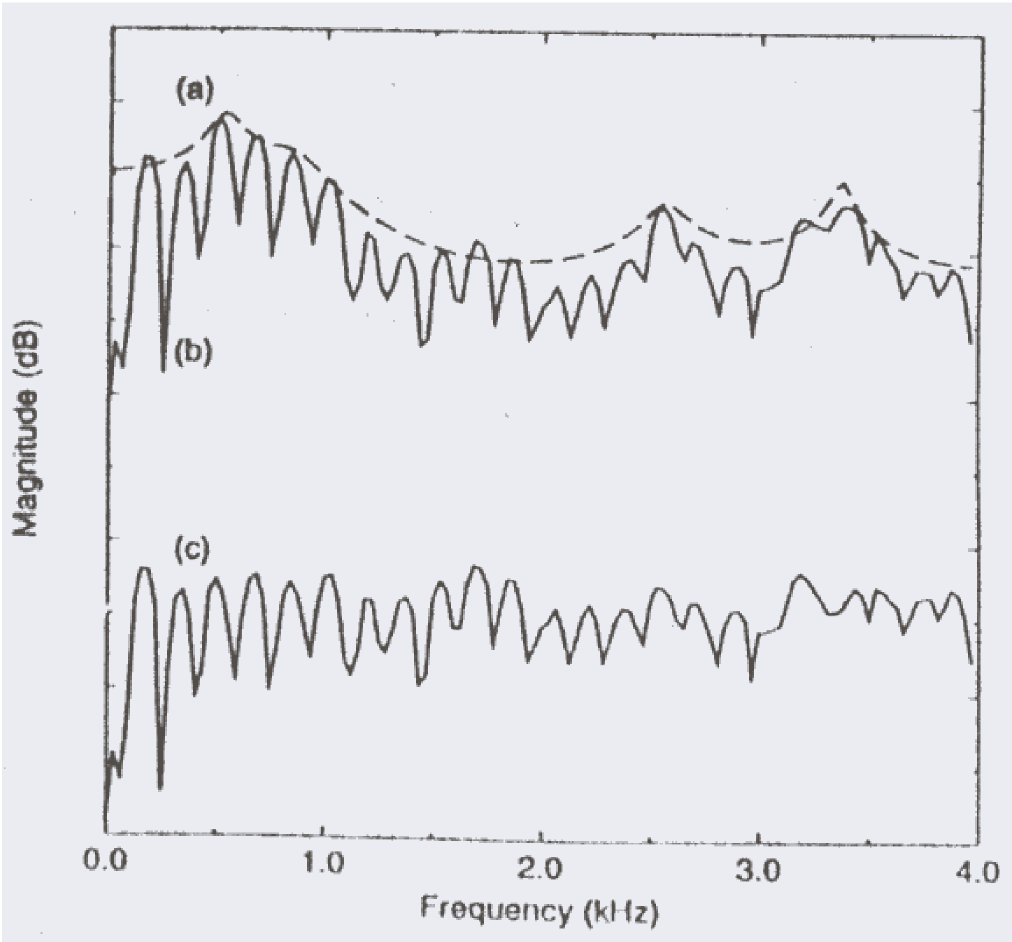
$$H^{-1}(z) = 1 - \sum_{j=1}^{p} \alpha_j z^{-j}$$

  - Therefore, $e(n)$ is expected to have pitch information.

Typical waveform of original and LPC residual signal



Typical spectra of original and LPC residual signal



(a) Original speech spectral envelope

(b) Original speech spectrum

(c) LPC residual spectrum

Some observations

- $H(z)$ has only poles, so cannot model the spectral valleys accurately.
  - One of drawbacks of LPC modeling
- Question
  - If we hear of the residual signal, can we understand the contents in original speech? → Maybe ...

Pitch prediction

- Periodicity in speech signals
  - After removing the spectral envelope by LPC filtering, a periodic component (long-term correlations) still exists, especially during voiced regions.
- Long-term prediction (LTP) filter, also called pitch prediction filter
  - To remove the periodic structure of the residual = to spectrally flatten the residual signal
  - Or, to change the signal to a white signal so as not to need to be transmitted. → Lower bit rate

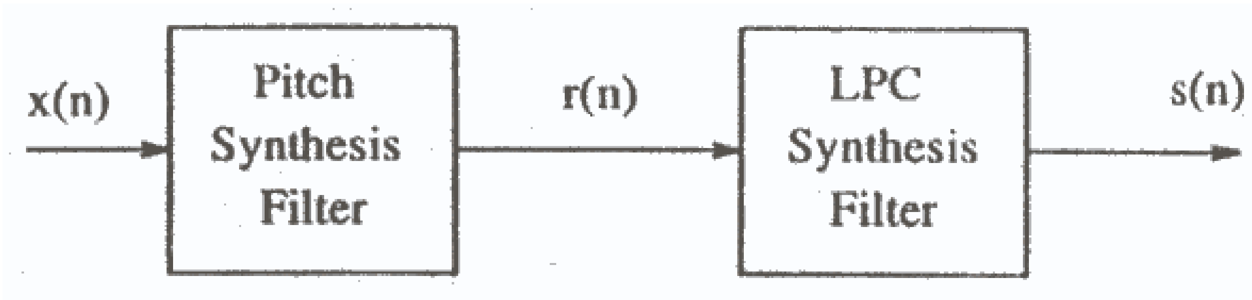
Pitch predictor formulation

- Pre-considerations
  - The operational order of LTP and STP is not too critical if the combination is carefully optimized. (Or, both STP → LTP and LTP → STP are possible.)

- LTP
  - General eq.: $P(z) = \dfrac{1}{1 - \sum_{j=-I}^{I} b_j z^{-(j+T)}}$
    - T : pitch period
    - b_j : pitch gain coefficients

Pitch predictor formulation

- Typical pitch-LPC formulation model



- Time domain difference equation of the combined model
  - $$s(n) = Gx(n) + \sum_{j=-I}^{I} b_j r(n - T - j) + \sum_{j=1}^{p} a_j s(n - j)$$
    - $r(n)$ is the past excitation (LPC residual) signal.
  - Goal: to determine the estimates (β_j, τ, α_j) of (b_j, T, a_j) so as to minimize the prediction error.
  - Here, the prediction error: $e(n) = s(n) - \sum_{j=-I}^{I} \beta_j r(n - \tau - j) - \sum_{j=1}^{p} \alpha_j s(n - j)$

Pitch predictor formulation

- MSE strategy
  - Not straightforward due to the delay factor τ
  - Two sub-optimal approaches
    - One-shot optimization
      - First STP, then LTP
      - If $\tau > N$ (analysis frame size in LTP), it results in near optimal.
    - Iterative sequential approach
      - Iteration of one-shot optimization, that is, STP → LTP → STP → LTP → …
    - Iterative sequential approach gives a better prediction gain and better perceptual performance, but the one-shot method is usually preferred due to termination criterion and complicate calculation of iterative sequential approach.

Pitch predictor formulation

- MSE solution for one-shot optimization
  - Prediction error signal of residual signal:

$$e(n) = r(n) - \sum_{j=-I}^{I} \beta_j r(n - \tau - j)$$

  - MSE of the error signal:

$$MSE = E\{e^2(n)\} = E\left\{\left[r(n) - \sum_{j=-I}^{I} \beta_j r(n - \tau - j)\right]^2\right\}$$

  - Expectation → finite time average:

$$MSE = \sum_m e_n^2(m) = \sum_m \left[r_n(m) - \sum_{j=-I}^{I} \beta_j r_n(m - \tau - j)\right]^2$$

Pitch predictor formulation

- MSE solution for one-shot optimization (cont.)
  - Minimization of MSE w.r.t. the long-term prediction coefficients
    - $$\frac{\partial MSE}{\partial \beta_i} = 2\sum_m \left[\left\{ r_n(m) - \sum_{j=-I}^{I} \beta_j r_n(m-\tau-j) \right\} r_n(m-\tau-i) \right] = 0$$
    - $$\sum_m r_n(m) r_n(m-\tau-i) = \sum_m \sum_{j=-I}^{I} \beta_j r_n(m-\tau-j) r_n(m-\tau-i)$$
    - $$\sum_{j=-I}^{I} \beta_j \sum_m r_n(m-\tau-i) r_n(m-\tau-j) = \sum_m r_n(m) r_n(m-\tau-i)$$
    - $$\sum_{j=-I}^{I} \beta_j V(i,j) = R(\tau+i,0), \quad -I \le i \le I$$

Pitch predictor formulation

- MSE solution for one-shot optimization (cont.)
  - Matrix equation to be solved

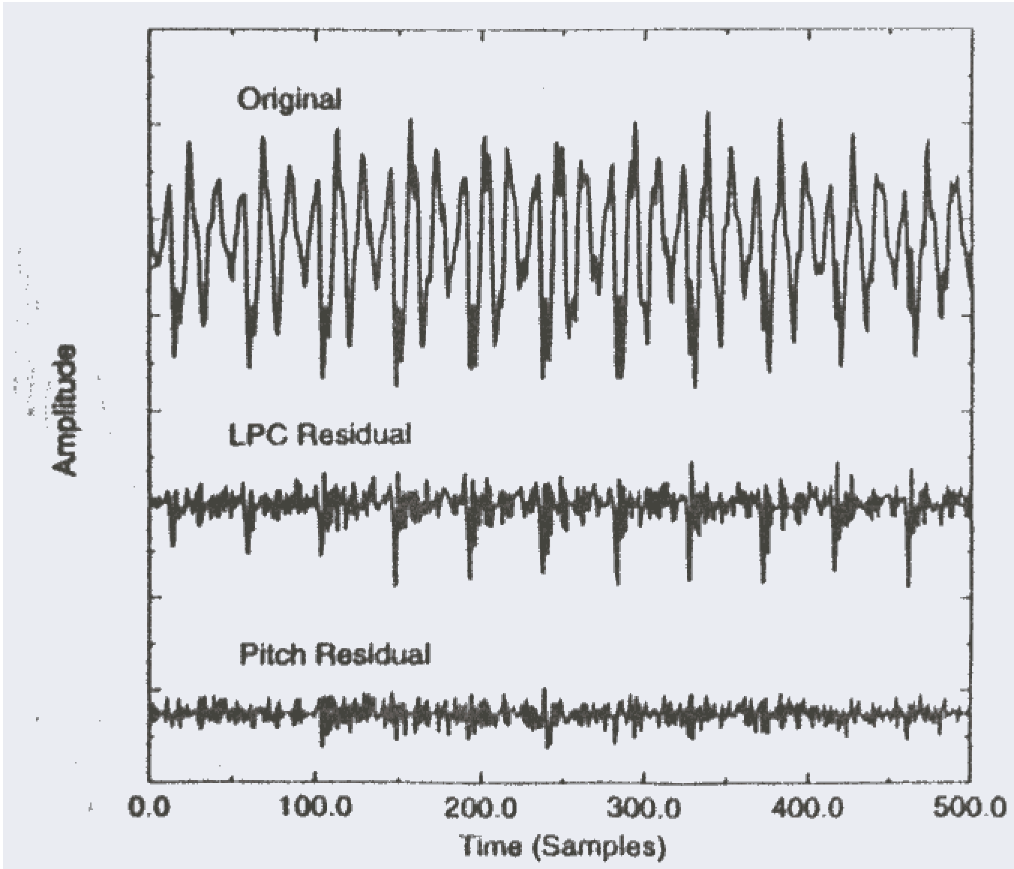
$$\begin{bmatrix} V(-I,-I) & \cdots & & V(-I,I) \\ \vdots & \ddots & & \\ & & \ddots & \vdots \\ V(I,-I) & & \cdots & V(I,I) \end{bmatrix} \begin{bmatrix} \beta_{-I} \\ \vdots \\ \\ \beta_I \end{bmatrix} = \begin{bmatrix} R(\tau - I,0) \\ \vdots \\ \\ R(\tau + I,0) \end{bmatrix}$$

    - β_j : can be solved by Cholesky's decomposition when the pitch lag τ is already given.
  - Determination of τ
    - Basic method: Exhaustive search for all the possible τ
    - More sophisticated methods: Pitch detection algorithms
    - Generally, window size ($N+\tau_{max} \geq 200$) > pitch value ($\tau_{min}(16) < \tau < \tau_{max}(160)$) > analysis frame size (N)

Example of pitch predictor

- In the case of 1-tap LTP,
  - $P_1(z) = \dfrac{1}{1 - \beta z^{-\tau}}$
  - $\beta = \dfrac{R(\tau,0)}{V(0,0)} = \dfrac{\sum_{m=0}^{N-1} r(m) r(m-\tau)}{\sum_{m=0}^{N-1} r^2(m-\tau)}, \quad \tau_{\min} \le \tau \le \tau_{\max}$
  - $MSE = \sum_{m=0}^{N-1} r^2(m) - \dfrac{\left[\sum_{m=0}^{N-1} r(m) r(m-\tau)\right]^2}{\sum_{m=0}^{N-1} r^2(m-\tau)}$
  - Testing τ between τ_{min} (16 samples) and τ_{max} (160 samples) to minimize E, then β minimizing E can be found.

Time domain plot of pitch residuals



- No longer the sharp pulse-like characteristics in pitch residual signal

Summary of lecture

- Solutions to LPC analysis
  - Covariance method (CM)
  - Lattice method (LM)
    - Two solutions guaranteeing the filter stability
- Practical implementation of LPC analysis
  - How to set filter order (p), frame size (N)
  - Pre-emphasis and window overlapping
- Interpretation of LPC analysis
  - Property of residual signal
  - Inaccurate capturing of spectral valley of speech signal
- Pitch prediction
  - Prediction of pitch predictor (LTP) gain
  - MSE solution for one-shot optimization