Speech and Audio Coding Theory

Contents of lecture

- Alternative representations of LPC parameters
  - PARCOR, LAR, Inverse Sine Function, LSF
- LPC-to-LSF transformation
  - Basic formulation
  - Various methods for getting the LSF parameters
- LSF-to-LPC transformation
- Summary of LSF properties

LPC parameter quantization for efficient transmission

- Objectives
  - To encode the LPC parameters with as few bits as possible without introducing additional spectral distortion, while maintaining the subjective transparency (quality)
- From results of conventional coding research (before 1990's)
  - Above 9.6 kbps: good quality but high coding capacity with 40-50 bits LPC scalar quantisation.
  - Below 4.8 kbps: low coding capacity but reasonable quality with 10-bits LPC codebook VQ.
- Another LPC parameter quantisation scheme was required for 4.8-9.6 kbps coding.
  - Line spectrum pairs (LSP) or line spectral frequency (LSF) quantisation having a number of advantages.

Alternative representation of LPC parameters

- LPC synthesis filter: $H(z) = \dfrac{1}{1 - \sum_{i=1}^{p} \alpha_i z^{-i}}$
  - Stable if all the roots (poles) of $H(z)$ are within the unit circle (or minimum phase)
  - If α_i is directly quantized, it is hard to check for stability.
    - If we want to check the stability, we must get the poles and evaluate whether the roots are inside unit circle.
    - But, it requires much computation, so it is desirable to avoid this procedure.
  - Alternative: PARCOR coefficient k_i

Alternative representation of LPC parameters

- PARCOR coefficient from Durbin's algorithm
  - LPC-to-PARCOR transformation for the use in encoder

$$a_j^p = \alpha_j \quad 1 \le j \le p$$

$$k_p = a_p^p = \alpha_p$$

For $i = p, p-1, \ldots, 1$

$$a_j^{i-1} = (a_j^i + a_i^i a_{i-j}^i)/(1-k_i^2), \quad 1 \le j \le i-1$$

$$k_{i-1} = a_{i-1}^{i-1}$$

  - PARCOR-to-LPC transformation for the use in decoder

For $i = 1, 2, \ldots, p$

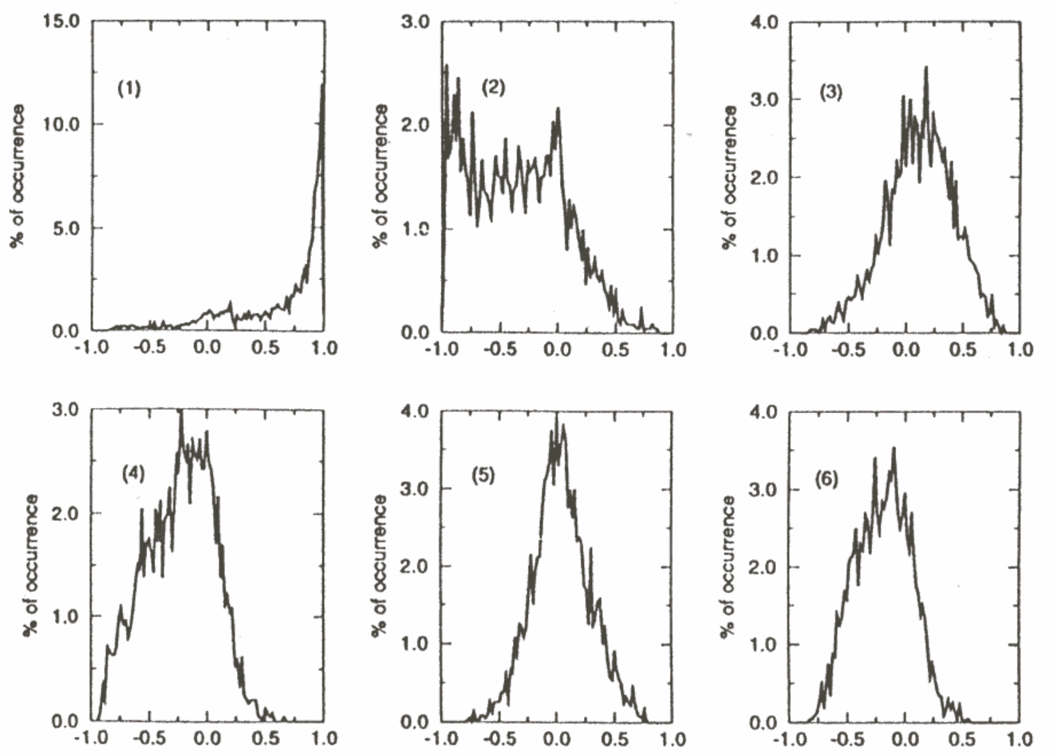
$$a_i^i = k_i$$

$$a_j^i = a_j^{i-1} - k_i a_{i-j}^{i-1}, \quad 1 \le j \le i-1$$

$$\alpha_j = a_j^p, \quad 1 \le j \le p$$

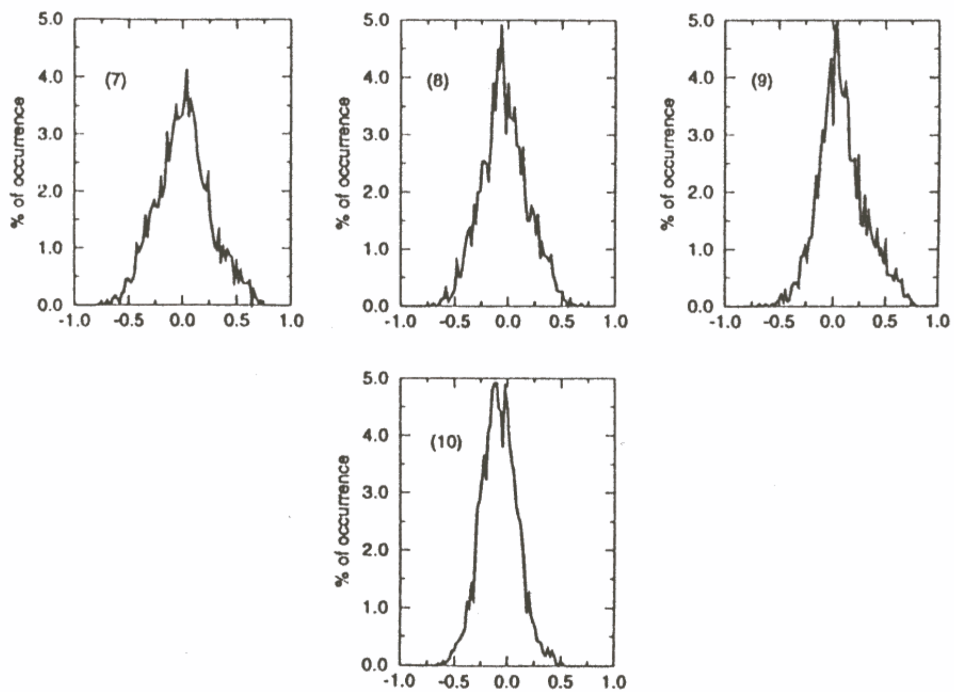
Alternative representation of LPC parameters

- Property of PARCOR coefficient
  - If $|k_i| \le 1.0$, the LPC filter is stable.
  - Distribution of PARCOR parameters (k_1 to k_6)



Alternative representation of LPC parameters

- Property of PARCOR coefficient (cont.)
  - Distribution of PARCOR parameters (k_7 to k_{10})



  - Non-flat spectral sensitivity: Values of k_1 and k_2 near 1 require more quantization accuracy.

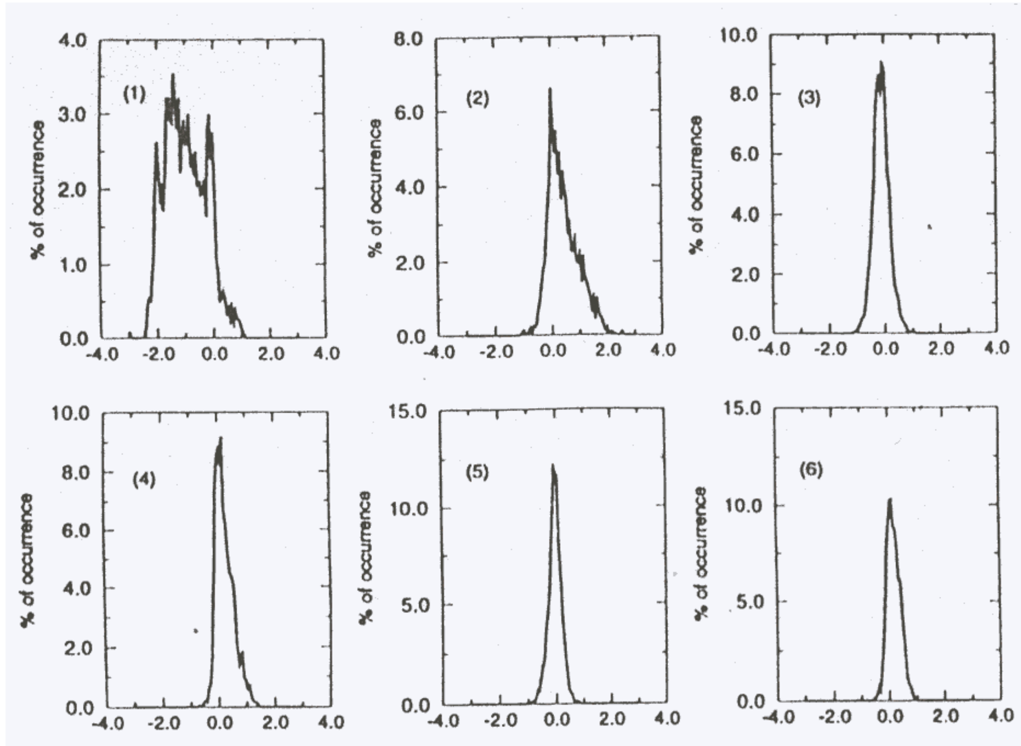
Alternative representation of LPC parameters

- Non-linear transformation of PARCOR coefficient
  - Log area ratio (LAR) (motivated from lossless tube model)
    - PARCOR-to-LAR

$$g_i = \log\left(\frac{1-k_i}{1+k_i}\right), \quad 1 \le i \le p$$

    - LAR-to-PARCOR

$$k_i = \left(\frac{1-10^{g_i}}{1+10^{g_i}}\right), \quad 1 \le i \le p$$



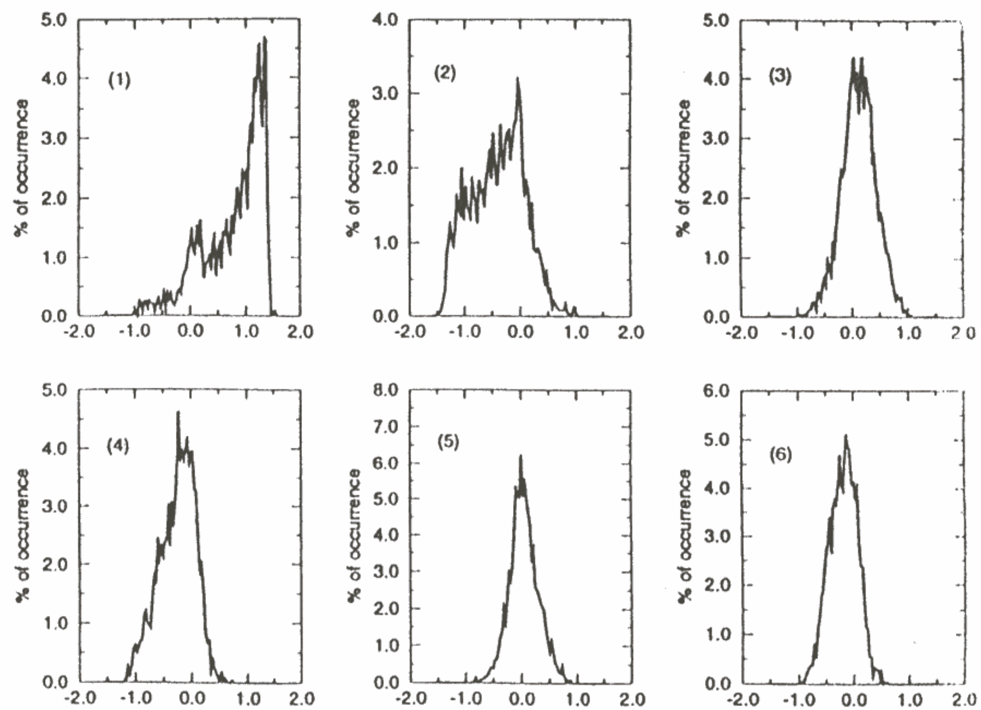
Alternative representation of LPC parameters

- Non-linear transformation of PARCOR coefficient (cont.)
  - Inverse sine (IS) function
    - PARCOR-to-IS

      $$s_i = \sin^{-1}(k_i), \quad 1 \le i \le p$$

    - IS-to-PARCOR

      $$k_i = \sin(s_i), \quad 1 \le i \le p$$



Alternative representation of LPC parameters

- Drawbacks of the above parameters (even though flatter spectral sensitivity)
  - In minimizing spectral distortion, too many bits (about 4 bits/coefficient) are required for coding the LPC parameters.
    - So, in coding below 8 kbps, the remained bits are not enough for coding the pitch period, etc.
  - The frame-to-frame correlation of LPC parameters is not highlighted, so we cannot further reduce bit-rate by rejecting the redundancies in frequency domain.

- Another good candidate: LSF proposed by Itakura
  - Encode speech spectral information in the frequency domain
  - Easy to incorporate well-known spectral features
  - Possible to interpolate the parameters in the level of frame-to-frame

Definition of LSF

- Derivation of LSF
  - All-pole synthesis filter: $H(z) = 1/A_p(z)$
    - Here, $A_p(z) = 1 + \sum_{k=1}^{p} \alpha_k z^{-k}$
  - From the PARCOR relations, we know that

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

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

  - Therefore, substituting p to i in the equations and transforming into z-domain, we can get the followings,

$$A_p(z) = Z\left[e^{(p)}(m)\right]/S(z) = A_{p-1}(z) - k_p V_{p-1}(z) z^{-1}$$

$$V_p(z) = Z\left[b^{(p)}(m)\right]/S(z) = V_{p-1}(z) z^{-1} - k_p A_{p-1}(z)$$

Definition of LSF

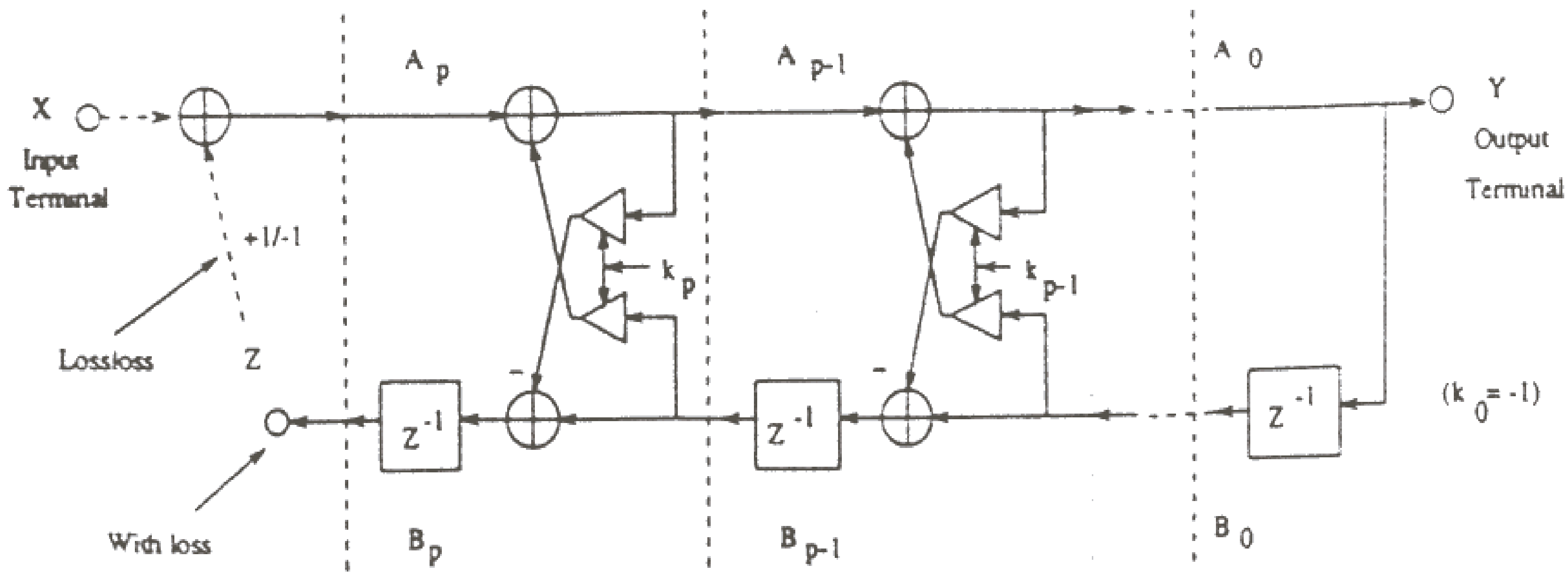
- Derivation of LSF (cont.)
  - Here, let $B_p(z) = z^{-1}V_p(z)$
    - Then, $V_p(z) = zB_p(z)$
  - Using this relation, we obtain that
    - $A_{p-1}(z) = A_p(z) + k_p B_{p-1}(z)$
    - $B_p(z) = z^{-1}B_{p-1}(z) - k_p z^{-1}A_{p-1}(z) = z^{-1}\left[B_{p-1}(z) - k_p A_{p-1}(z)\right]$
    - Here, $A_0(z) = 1$ and $B_0(z) = z^{-1}A_0(z^{-1}) = z^{-1}$ since

$$A_p(z) = 1 + \sum_{k=1}^{p} \alpha_k z^{-k}$$

$$B_p(z) = z^{-1}V_p(z) = z^{-1}z^{-p}A_p(z^{-1}) = z^{-(p+1)}A_p(z^{-1})$$

Definition of LSF

- PARCOR structure of LPC synthesis



Definition of LSF

- From the figure,
  - Transfer function from X to Y: $A_0(z)/A_p(z) = 1/A_p(z) = H_p(z)$
  - Transfer function from Y to Z: $B_p(z)/ A_0(z) = B_p(z)/1 = B_p(z)$
  - Transfer function from X to Z = $H_p(z) * B_p(z) = B_p(z) / A_p(z) = R_p(z)$: Ratio filter
  - In the viewpoint of an acoustic lossless tube model,
    - PARCOR coefficients = Reflection coefficients
- Consider a pair of artificial boundary conditions
  - Completely open at glottis: $k_{p+1} = 1$
  - Completely closed at glottis: $k_{p+1} = -1$
  - => perfectly lossless tube model
  - => Each resonance value becomes infinite. => The spectrum of distributed energy is **concentrated on several line spectra.**

Calculation of the LSFs

- If the PARCOR filter is stable and the order is even, then $A_p(z)$ may be decomposed into two transfer functions each with even symmetry and odd symmetry properties.
  - That is, the transfer functions for $H(z)$ become $1/P_{p+1}(z)$ or $1/Q_{p+1}(z)$ such that

$$\text{For } k_{p+1} = 1, \quad P_{p+1}(z) = A_p(z) - B_p(z) \text{ (Difference filter)}$$

$$\text{For } k_{p+1} = -1, \quad Q_{p+1}(z) = A_p(z) + B_p(z) \text{ (Sum filter)}$$

  - Therefore,

$$\Rightarrow A_p(z) = \frac{1}{2}[P_{p+1}(z) + Q_{p+1}(z)]$$

  - And, since

$$B_p(z) = z^{-(p+1)} A_p(z^{-1})$$

Calculation of the LSFs

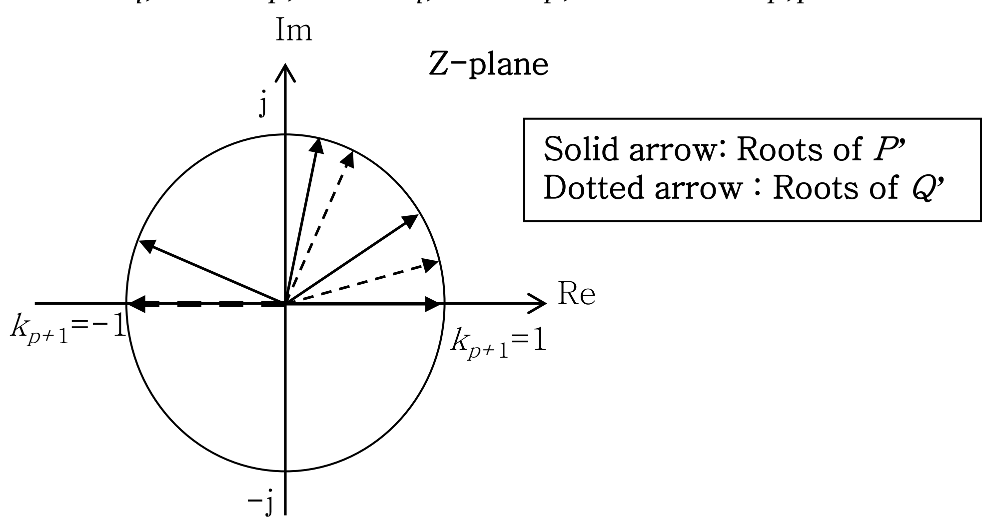
- (cont.)
  - $$\begin{aligned} P_{p+1}(z) &= A_p(z) - z^{-(p+1)} A_p(z^{-1}) \\ &= 1 + (\alpha_1 - \alpha_p) z^{-1} + \ldots + (\alpha_p - \alpha_1) z^{-p} - z^{-(p+1)} \\ &= z^{-(p+1)} \prod_{i=1}^{p+1} (z + a_i) \end{aligned}$$
  - $$\begin{aligned} Q_{p+1}(z) &= A_p(z) + z^{-(p+1)} A_p(z^{-1}) \\ &= 1 + (\alpha_1 + \alpha_p) z^{-1} + \ldots + (\alpha_p + \alpha_1) z^{-p} + z^{-(p+1)} \\ &= z^{-(p+1)} \prod_{i=1}^{p+1} (z + b_i) \end{aligned}$$

Calculation of the LSFs

- (cont.)
  - Since $k_{p+1} = \pm 1$, the order of $P_{p+1}(z)$ and $Q_{p+1}(z)$ can be reduced as follows.
    - $$P'(z) = \frac{P_{p+1}(z)}{(1-z)}\left(-z^{p+1}\right)$$
      $$= A_0 z^p + A_1 z^{(p-1)} + \ldots + A_p$$
    - $$Q'(z) = \frac{Q_{p+1}(z)}{(1+z)} \cdot z^{p+1}$$
      $$= B_0 z^p + B_1 z^{(p-1)} + \ldots + B_p$$
    - Here, $A_0 = 1 \qquad B_0 = 1$
      $$A_k = (\alpha_k - \alpha_{p+1-k}) + A_{k-1}$$
      $$B_k = (\alpha_k + \alpha_{p+1-k}) - B_{k-1} \text{ for } k = 1, \ldots, p$$
      $$A_p = 1 \qquad B_p = 1$$

Characteristics of LSFs

- Physical meaning and property of LSFs
  - LSFs: Angular positions of the roots of $P'(z)$ and $Q'(z)$ with $0 \le \omega_i \le \pi$.
  - The roots are on the unit circle.
  - The roots have complex conjugate pairs.
  - The roots alternate with each other on the unit circle, that is,
    $0 \le \omega_{q,0} < \omega_{p,0} < \omega_{q,1} < \omega_{p,1} < \ldots < \omega_{p,p/2-1} \le \pi$



LPC-to-LSF transformation

- Complex root method
  - The following equations can be solved directly by complex arithmetic.

$$P'(z) = A_0 z^{p} + A_1 z^{(p-1)} + \ldots + A_p = 0$$

$$Q'(z) = B_0 z^{p} + B_1 z^{(p-1)} + \ldots + B_p = 0$$

  - However, it requires non-deterministic time consumption due to the iteration procedure inherent in the method.
    - It is undesirable for real-time implementation.

LPC-to-LSF transformation

- Real root method
  - To use the property that the coefficients of $P'(z)$ and $Q'(z)$ are symmetrical.

$$P'(z) = A_0 z^{p} + A_1 z^{p-1} + \ldots + A_1 z^{1} + A_0$$
$$= z^{p/2}[A_0(z^{p/2} + z^{-p/2}) + A_1(z^{(p/2-1)} + z^{-(p/2-1)}) + \ldots + A_{p/2}]$$

$$Q'(z) = B_0 z^{p} + B_1 z^{p-1} + \ldots + B_1 z^{1} + B_0$$
$$= z^{p/2}[B_0(z^{p/2} + z^{-p/2}) + B_1(z^{(p/2-1)} + z^{-(p/2-1)}) + \ldots + B_{p/2}]$$

  - Since the roots are on the unit circle,

Let $z = e^{j\omega}$ then $z^{1} + z^{-1} = 2\cos(\omega)$

LPC-to-LSF transformation

- Real root method (cont.)

$$P'(z) = 2e^{jp\omega/2}\left[A_0\cos\left(\frac{p}{2}\omega\right) + A_1\cos\left(\frac{p-2}{2}\omega\right) + \ldots + \frac{1}{2}A_{p/2}\right]$$

$$Q'(z) = 2e^{jp\omega/2}\left[B_0\cos\left(\frac{p}{2}\omega\right) + B_1\cos\left(\frac{p-2}{2}\omega\right) + \ldots + \frac{1}{2}B_{p/2}\right]$$

  - When p = 10, substituting $x = \cos(\omega)$, then we obtain the followings.

$$P'_{10}(x) = 16A_0x^5 + 8A_1x^4 + (4A_2 - 20A_0)x^3 + (2A_3 - 8A_1)x^2 + (5A_0 - 3A_2 + A_4)x + (A_1 - A_3 + 0.5A_5)$$

$$Q'_{10}(x) = 16B_0x^5 + 8B_1x^4 + (4B_2 - 20B_0)x^3 + (2B_3 - 8B_1)x^2 + (5B_0 - 3B_2 + B_4)x + (B_1 - B_3 + 0.5B_5)$$

LPC-to-LSF transformation

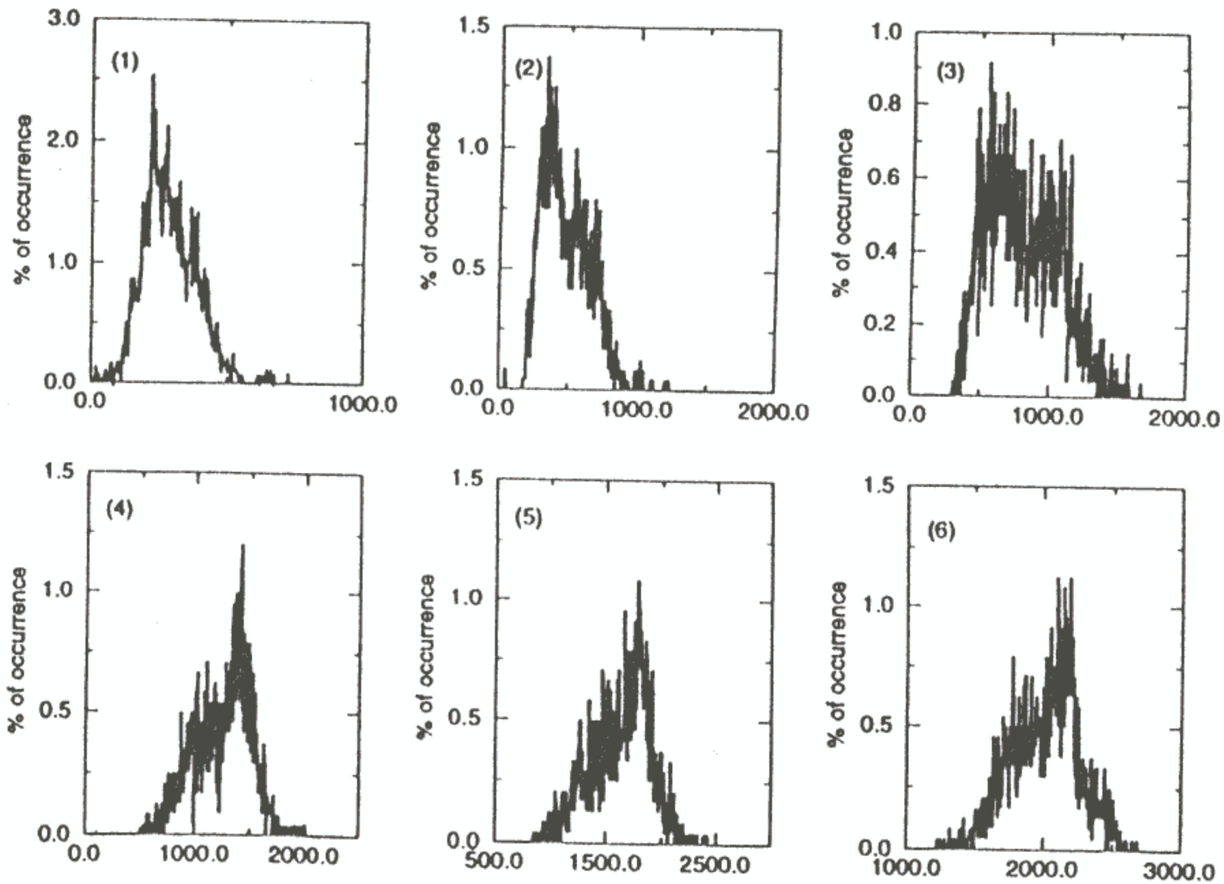
- Real root method (cont.)
  - Therefore, obtaining the roots x_i satisfying $P'_{10}(x) = 0$ and $Q'_{10}(x) = 0$, we can get

$$LSF(i) = \frac{\cos^{-1}(x_i)}{2\pi T}, \quad \text{for } 1 \le i \le p$$

  - This method still requires indeterministic computation time, but faster root search is possible by noting that the change from one LSF vector to the next is not too drastic in most cases.
    - So, the number of iterations required per root is considerably reduced, e.g. typically from 10 to 5, which is relatively smaller than the complex root method.

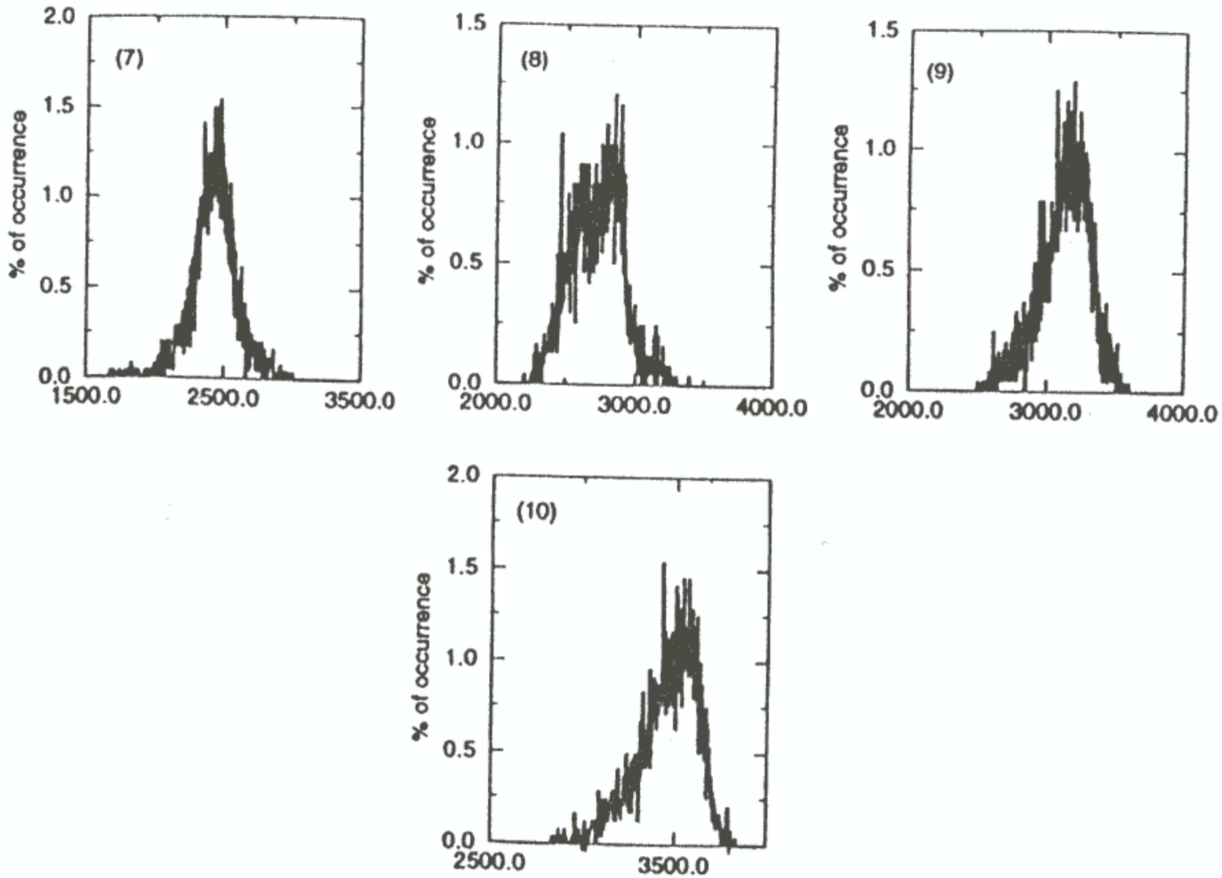
LPC-to-LSF transformation

- Distribution of LSF parameters (LSF(1) to LSF(6))

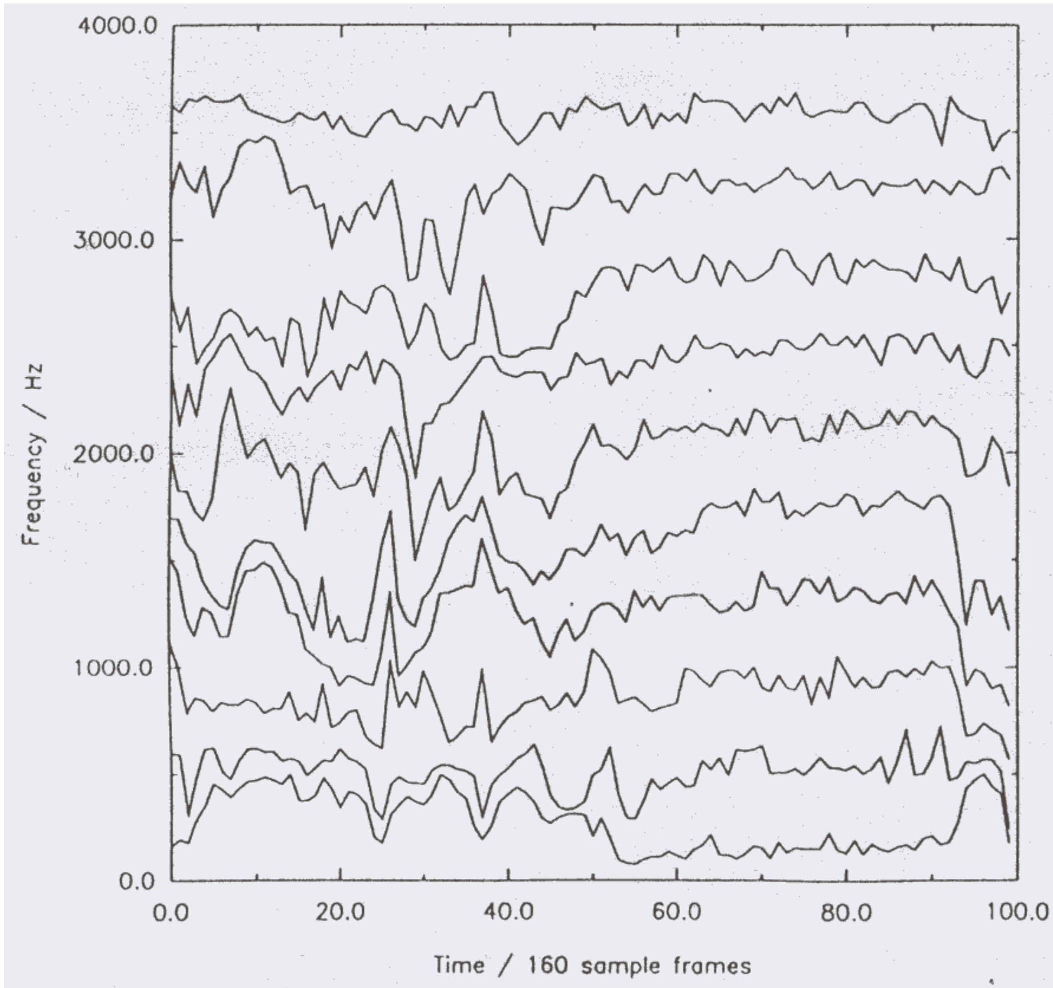


LPC-to-LSF transformation

- Distribution of LSF parameters (LSF(7) to LSF(10))



Typical LSF trajectories for voiced and unvoiced speech



LPC-to-LSF transformation

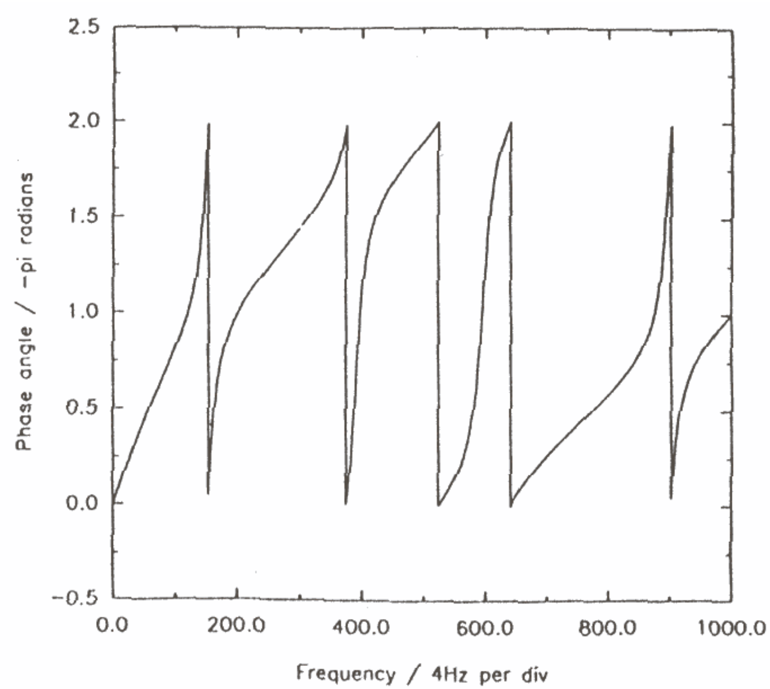
- Ratio filter method
  - Ratio filter is $R_p(z) = \dfrac{B_p(z)}{A_p(z)} = \dfrac{z^{-(p+1)}A_p(z^{-1})}{A_p(z)}$
    - Here, $A_p(z) = 1 - \sum_{i=1}^{p} \beta_i z^{-i}$
  - The phase response of the ratio filter is given by

$$\phi(kf_s) = -(p+1)(2\pi Tkf_s) - 2\tan^{-1}\left\{ \frac{\sum_{i=1}^{p} \beta_i \sin(2\pi iTkf_s)}{1 - \sum_{i=1}^{p} \beta_i \cos(2\pi iTkf_s)} \right\}$$

    - T is the sampling period.
    - f_s is the frequency step.
    - $k = 1, 2, \ldots, K_{\max}$

LPC-to-LSF transformation

- Ratio filter method (cont.)
  - Here, when $\phi(kf_s) = 0$ or π, $A_p(z)$ can be equal to $\pm B_p(z)$, that is, $P'(z)$ and $Q'(z)$ can be zero.
  - Therefore, the LSFs can be obtained by $\phi(kf_s) = 0$ or π for $k = 1,2,\ldots,K_{\max}$.
  - Typical phase response of ratio filter



LPC-to-LSF transformation

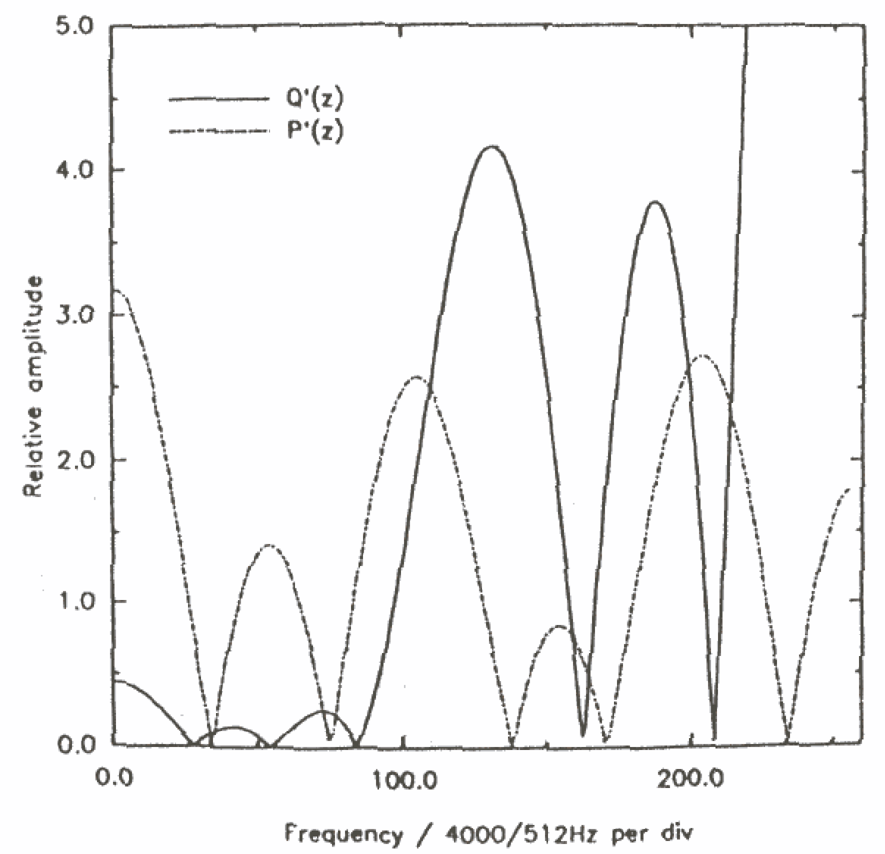
- DFT method
  - Since the LSFs are the poles of the equations,

$$\frac{1}{P'(z)} = \frac{1}{\sum_{k=0}^{p} A_k z^{p-k}} \qquad \frac{1}{Q'(z)} = \frac{1}{\sum_{k=0}^{p} B_k z^{p-k}}$$

  - Thus, we can regard the denominators as the z-transforms of the sequences A_k and B_k, with zero padding before $k = 0$ and after $k = p$.
  - Therefore if we perform a DFT on the sequences A_k and B_k, then the LSFs can be solved as the zero-valued frequencies of the power spectrum.

LPC-to-LSF transformation

- DFT method (cont.)
  - Zero-valued frequencies obtained from the DFT method



Using the ordering information of LSFs, the search space can be greatly reduced.

LSF-to-LPC transformation

- Direct expansion method
  - At first, obtain the LPC coefficients from LSFs, and then use the LPC synthesis filter for reconstructing the speech.
    - $$P_{p+1}(z) = -z^{-(p+1)}[P'(z)(1-z)]$$
      $$= -z^{-(p+1)}[(1-z)(z-r_0)(z-r_0^*)\ldots(z-r_{(p-2)/2})(z-r^*_{(p-2)/2})]$$
      $$= -z^{-(p+1)}[(1-z)(z^2-2u_0 z+t_0)\ldots(z^2-2u_{(p-2)/2}z+t_{(p-2)/2})]$$
      $$= S_0 + S_1 z^{-1} + \ldots + S_p z^{-p} + S_{p+1} z^{-(p+1)}$$
    - Similarly, $Q_{p+1}(z) = T_0 + T_1 z^{-1} + \ldots + T_p z^{-p} + T_{p+1} z^{-(p+1)}$
    - Here,

$$r_i = u_i + jv_i \quad \text{and} \quad r_i^* = u_i - jv_i$$

$$\Rightarrow \quad r_i + r_i^* = 2u_i \quad \text{and} \quad r_i \times r_i^* = u_i^2 + v_i^2 = t_i$$

LSF-to-LPC transformation

- Direct expansion method (cont.)
  - Therefore, using the terms of the above equations, we can obtain the followings.
    - $S_0 = 1 \qquad T_0 = 1$
    - $\alpha_i = \frac{1}{2}(T_i + S_i) \qquad \alpha_{p+1-i} = \frac{1}{2}(T_i - S_i) \quad \text{for } i = 1, \ldots, p/2$
    - $S_{p+1} = -1 \qquad T_{p+1} = 1$
  - Consequently, the computation process is
    - LSFs → r_i → u_i & t_i → S_i & T_i → α_i

LSF-to-LPC transformation

- LPC inverse filter method
  - To directly implement the inverse filter using the LSF parameters
  - $H(z) = 1/A_p(z) = 1/[1 + A_p(z) - 1]$
  - $A_p(z) - 1 = 1/2[(P_{p+1}(z) - 1) + (Q_{p+1}(z) - 1)]$

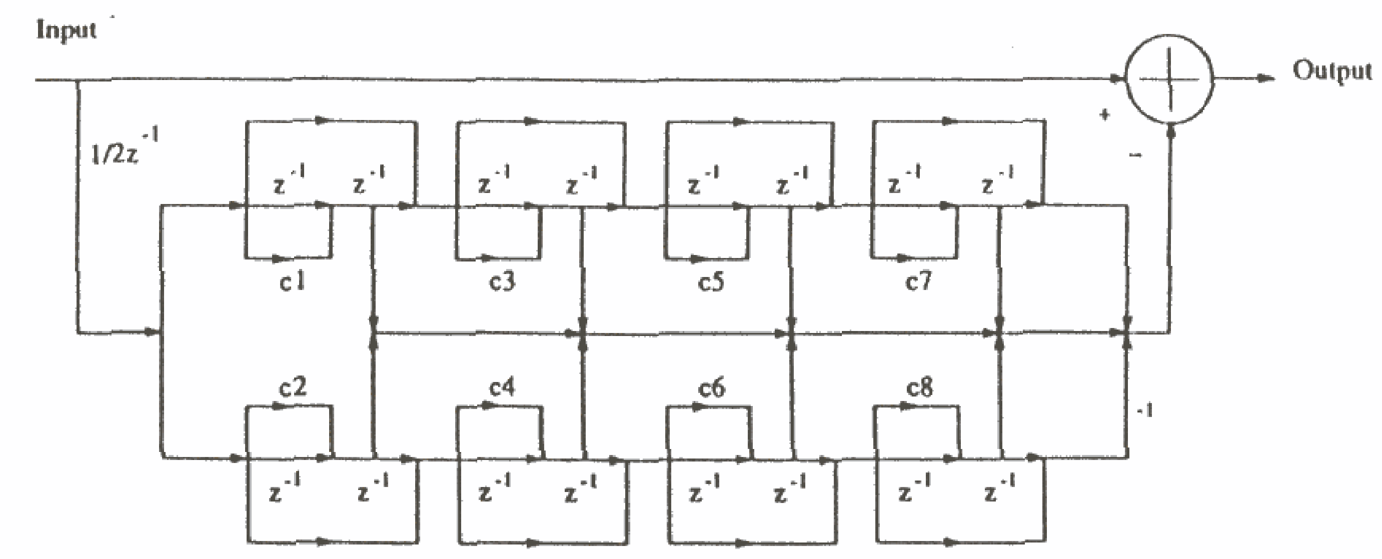
$$= 1/2\left\{(1-z)(-z^{-(p+1)})\prod_{i=1}^{p/2}(1 - 2\cos\omega_i z + z^2) - 1 + (1+z)z^{-(p+1)}\prod_{i=1}^{p/2}(1 - 2\cos\theta_i z + z^2) - 1\right\}$$

  - Here, let $u_i = -2\cos\omega_i, \quad v_i = -2\cos\theta_i$
  - Then, $A_p(z) - 1 = z^{-1}/2\left\{(u_1 + z^{-1}) - \prod_{j=1}^{p/2}(1 + u_j z^{-1} + z^{-2}) + \sum_{i=1}^{p/2-1}(u_{i+1} + z^{-1})\prod_{j=1}^{i}(1 + u_j z^{-1} + z^{-2})\right\}$

$$+ z^{-1}/2\left\{(v_1 + z^{-1}) + \prod_{j=1}^{p/2}(1 + v_j z^{-1} + z^{-2}) + \sum_{i=1}^{p/2-1}(v_{i+1} + z^{-1})\prod_{j=1}^{i}(1 + v_j z^{-1} + z^{-2})\right\}$$

LSF-to-LPC transformation

- LPC synthesis filter method (cont.)
  - So, using the final expression, we can implement the LSF parameter-based inverse filter as follows.



    - Here, $c_{2i} = u_i = -2\cos\omega_i$, and $c_{2i+1} = v_i = -2\cos\theta_i$.
  - Finally, the LPC coefficients are simply the impulse response of the inverse filter.

Summary of LSF properties

- Experimental conditions for estimating the following coefficients: 8 kHz, 10 ms update, 20 ms Hamming, p=10, 6000 frames
- Intra-frame correlation coefficients: $\Omega_{ij} = \omega_{n,i} * \omega_{n,j}$

	j									
i	1	2	3	4	5	6	7	8	9	10
1	1.00	0.65	0.30	0.35	0.41	0.49	0.39	0.40	0.36	0.20
2	0.65	1.00	0.28	0.11	0.07	0.13	0.07	0.05	0.06	0.07
3	0.30	0.28	1.00	0.72	0.50	0.53	0.46	0.54	0.39	0.28
4	0.35	0.11	0.72	1.00	0.72	0.62	0.46	0.42	0.45	0.21
5	0.41	0.07	0.50	0.72	1.00	0.79	0.52	0.47	0.34	0.26
6	0.49	0.13	0.53	0.62	0.79	1.00	0.71	0.61	0.49	0.28
7	0.39	0.07	0.46	0.46	0.52	0.71	1.00	0.73	0.58	0.41
8	0.40	0.05	0.54	0.42	0.47	0.61	0.73	1.00	0.58	0.46
9	0.36	0.06	0.39	0.45	0.34	0.49	0.58	0.58	1.00	0.41
10	0.20	0.07	0.28	0.21	0.26	0.28	0.41	0.46	0.41	1.00

Summary of LSF properties

- Inter-frame correlation coefficients: $\Psi_{ik} = \omega_{n,i} * \omega_{n-k,i}$

	k									
i	1	2	3	4	5	6	7	8	9	10
1	0.93	0.84	0.76	0.68	0.61	0.55	0.50	0.45	0.41	0.36
2	0.89	0.75	0.63	0.54	0.46	0.38	0.32	0.27	0.22	0.18
3	0.92	0.80	0.70	0.60	0.51	0.43	0.36	0.30	0.24	0.20
4	0.92	0.82	0.73	0.64	0.56	0.49	0.43	0.37	0.32	0.27
5	0.95	0.88	0.81	0.74	0.67	0.61	0.54	0.48	0.43	0.37
6	0.94	0.85	0.77	0.69	0.62	0.56	0.49	0.44	0.38	0.33
7	0.93	0.83	0.75	0.66	0.58	0.50	0.43	0.37	0.31	0.26
8	0.91	0.81	0.72	0.64	0.56	0.49	0.43	0.37	0.32	0.28
9	0.87	0.73	0.64	0.55	0.48	0.42	0.37	0.33	0.29	0.25
10	0.82	0.66	0.57	0.50	0.44	0.38	0.34	0.30	0.27	0.24

Summary of LSF properties

- Ordering information of LSF parameters → Speed-up of LPC-to-LSF transformation
- Comparison of PARCOR with LSF

Parameter	Advantages	Disadvantages	Ratio of information*
PARCOR	❑ Direct extraction by LM ❑ k_i independent of p ❑ Stable for $\lvert k_i \rvert < 1$	❑ Linear interpolation → large spectrum distortion ❑ Spectrum sensitivity is not uniform ❑ Indirect correspondence to spectrum envelope	100
LSF	❑ Small spectrum distortion by quantization & interpolation ❑ Direct correspondence to spectrum envelope ❑ Stable for $\omega_1 < \omega_2 < \ldots < \omega_p$	❑ Little more complicated ❑ ω_i dependent on p	60

* Ratio of bps necessary to get equal voice quality by synthesis

Summary of lecture

- Alternative representations of LPC parameters
  - PARCOR, LAR, Inverse Sine Function, LSF
  - Their properties
- LPC-to-LSF transformation
  - Definition, physical meaning, and property of LSF
  - Two methods for getting the LSF parameters
    - Complex root method and real root method
  - General distribution of LSF parameters
  - Other methods to get the LSF parameters
    - Ratio filter method and DFT method
- LSF-to-LPC transformation
  - Direct expansion method and LPC synthesis filter method
- Summary of LSF properties
  - Comparison of PARCOR with LSF