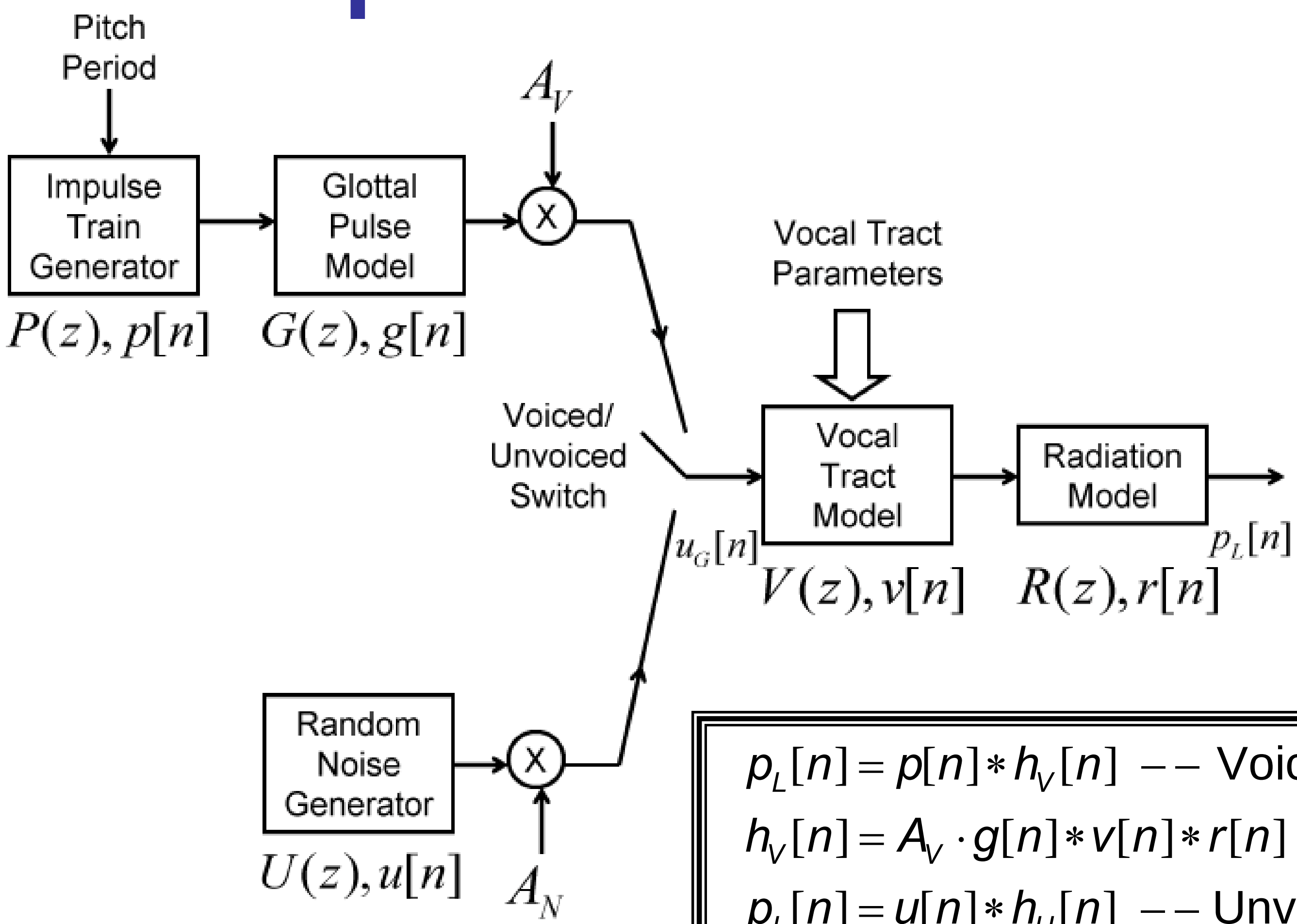
Digital Speech Processing—Lecture 12

Homomorphic Speech Processing

General Discrete-Time Model of Speech Production



$$p_L[n] = p[n] * h_V[n] \quad \text{-- Voiced Speech}$$

$$h_V[n] = A_V \cdot g[n] * v[n] * r[n]$$

$$p_L[n] = u[n] * h_U[n] \quad \text{-- Unvoiced Speech}$$

$$h_U[n] = A_N \cdot v[n] * r[n]$$

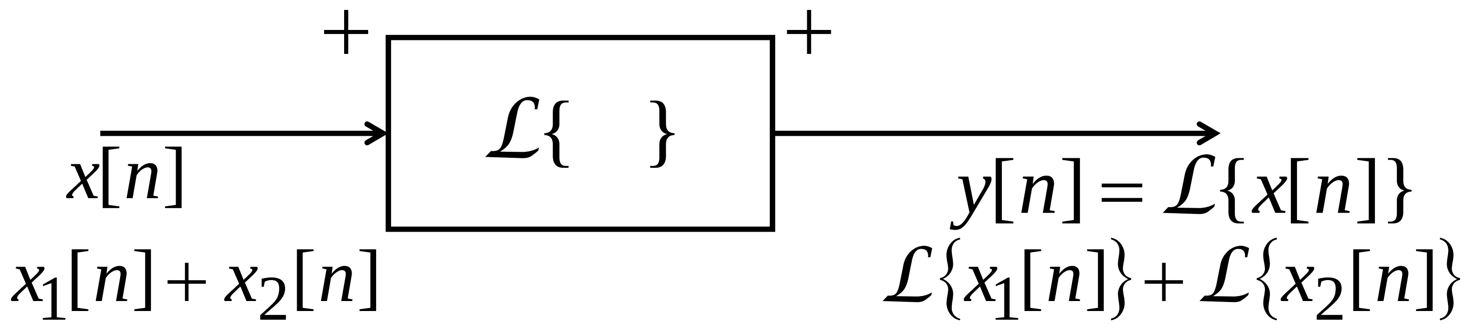
Basic Speech Model

- short segment of speech can be modeled as having been generated by exciting an LTI system either by a quasi-periodic impulse train, or a random noise signal
- speech analysis => estimate parameters of the speech model, measure their variations (and perhaps even their statistical variabilites-for quantization) with time

- speech = excitation $\ast$ system response

=> want to deconvolve speech into excitation and system response => do this using homomorphic filtering methods

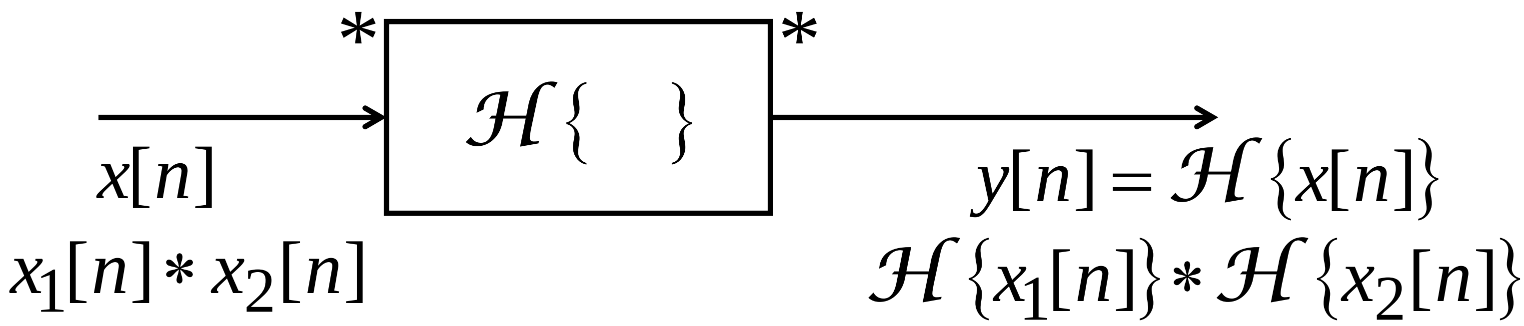
Superposition Principle



$$x[n] = ax_1[n] + bx_2[n]$$

$$y[n] = \mathcal{L}\{x[n]\} = a\mathcal{L}\{x_1[n]\} + b\mathcal{L}\{x_2[n]\}$$

Generalized Superposition for Convolution



- for LTI systems we have the result

$$y[n] = x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k]$$

- "generalized" superposition => addition replaced by convolution

$$x[n] = x_1[n] * x_2[n]$$

$$y[n] = \mathcal{H}\{x[n]\} = \mathcal{H}\{x_1[n]\} * \mathcal{H}\{x_2[n]\}$$

- homomorphic system for convolution

Homomorphic Filter

- homomorphic filter => homomorphic system $\left[\mathcal{H}\right]$ that passes the desired signal unaltered, while removing the undesired signal

$$x(n) = x_1[n] * x_2[n] \text{ - with } x_1[n] \text{ the "undesired" signal}$$

$$\mathcal{H}\{x[n]\} = \mathcal{H}\{x_1[n]\} * \mathcal{H}\{x_2[n]\}$$

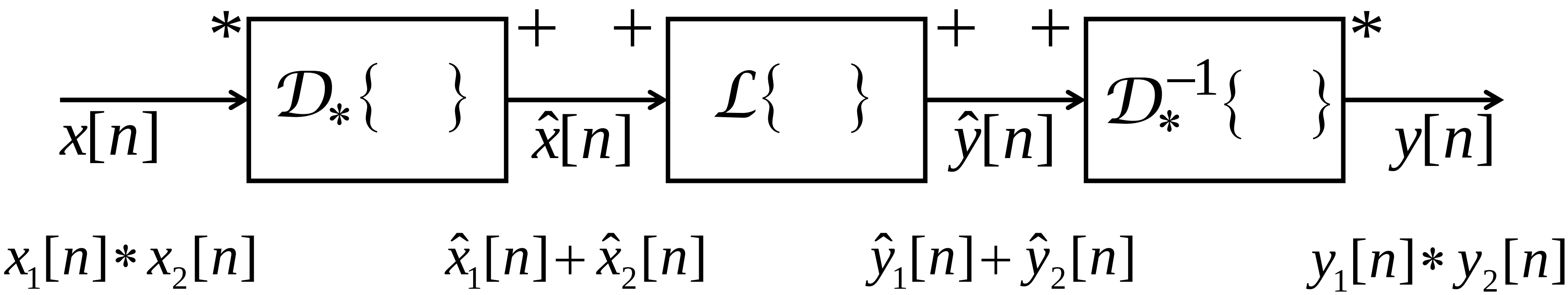
$$\mathcal{H}\{x_1[n]\} \rightarrow \delta(n) \text{ - removal of } x_1[n]$$

$$\mathcal{H}\{x_2[n]\} \rightarrow x_2[n]$$

$$\mathcal{H}\{x[n]\} = \delta[n] * x_2[n] = x_2[n]$$

- for linear systems this is analogous to additive noise removal

Canonic Form for Homomorphic Deconvolution



• any homomorphic system can be represented as a cascade of three systems, e.g., for convolution

1. system takes inputs combined by convolution and transforms them into additive outputs
2. system is a conventional linear system
3. inverse of first system--takes additive inputs and transforms them into convolutional outputs

Canonic Form for Homomorphic Convolution

$$x[n] \xrightarrow{\quad *\quad} \boxed{\mathcal{D}_*\{\quad\}} \xrightarrow[\hat{x}[n]]{+\qquad +} \boxed{\mathcal{L}\{\quad\}} \xrightarrow[\hat{y}[n]]{+\qquad +} \boxed{\mathcal{D}_*^{-1}\{\quad\}} \xrightarrow[y[n]]{\quad *\quad}$$

$$x_1[n]*x_2[n] \qquad \hat{x}_1[n]+\hat{x}_2[n] \qquad \hat{y}_1[n]+\hat{y}_2[n] \qquad y_1[n]*y_2[n]$$

$x[n] = x_1[n] * x_2[n]$ - convolutional relation

$\hat{x}[n] = \mathcal{D}_*\{x[n]\} = \hat{x}_1[n] + \hat{x}_2[n]$ - additive relation

$\hat{y}[n] = \mathcal{L}\{\hat{x}_1[n] + \hat{x}_2[n]\} = \hat{y}_1[n] + \hat{y}_2[n]$ - conventional linear system

$y[n] = \mathcal{D}_*^{-1}\{\hat{y}_1[n] + \hat{y}_2[n]\} = y_1[n] * y_2[n]$ - inverse of convolutional relation

=> design converted back to linear system, $\mathcal{L}$

$\mathcal{D}_*[\quad]$ - fixed (called the characteristic system for homomorphic deconvolution)

$\mathcal{D}_*^{-1}[\quad]$ - fixed (characteristic system for inverse homomorphic deconvolution)

Properties of Characteristic Systems

$$\begin{aligned}\hat{x}[n] &= \mathcal{D}_*\{x[n]\} = \mathcal{D}_*\{x_1[n] * x_2[n]\} \\ &= \mathcal{D}_*\{x_1[n]\} + \mathcal{D}_*\{x_2[n]\} \\ &= \hat{x}_1[n] + \hat{x}_2[n]\end{aligned}$$

$$\begin{aligned}\mathcal{D}_*^{-1}\{\hat{y}[n]\} &= \mathcal{D}_*^{-1}\{\hat{y}_1[n] + \hat{y}_2[n]\} \\ &= \mathcal{D}_*^{-1}\{\hat{y}_1[n]\} * \mathcal{D}_*^{-1}\{\hat{y}_2[n]\} \\ &= y_1[n] * y_2[n] = y[n]\end{aligned}$$

Discrete-Time Fourier Transform Representations

Canonic Form for Deconvolution Using DTFTs

- need to find a system that converts convolution to addition

$$x[n] = x_1[n] * x_2[n]$$

$$X(e^{j\omega}) = X_1(e^{j\omega}) \cdot X_2(e^{j\omega})$$

- since

$$\mathcal{D}_*\{x[n]\} = \hat{x}_1[n] + \hat{x}_2[n] = \hat{x}[n]$$

$$\mathcal{D}_*\left[X(e^{j\omega})\right] = \hat{X}_1(e^{j\omega}) + \hat{X}_2(e^{j\omega}) = \hat{X}(e^{j\omega})$$

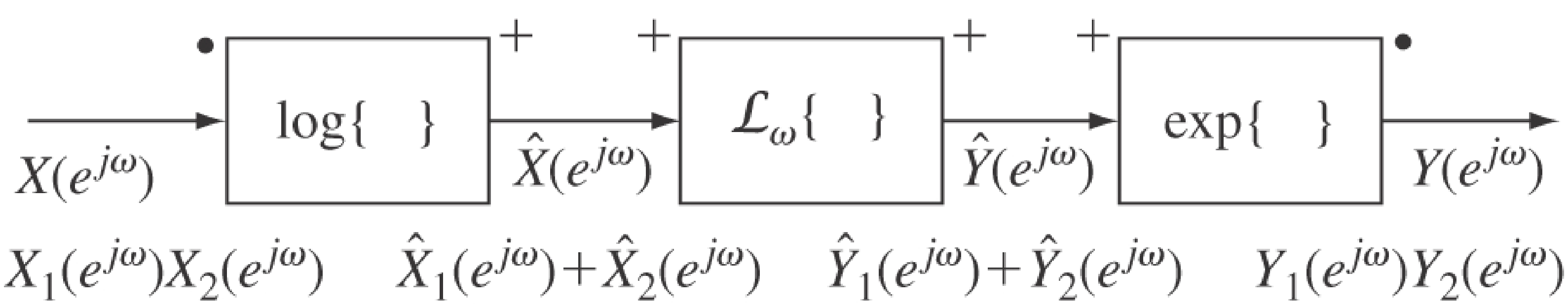
=> use log function which converts products to sums

$$\hat{X}(e^{j\omega}) = \log\left[X(e^{j\omega})\right] = \log\left[X_1(e^{j\omega}) \cdot X_2(e^{j\omega})\right]$$

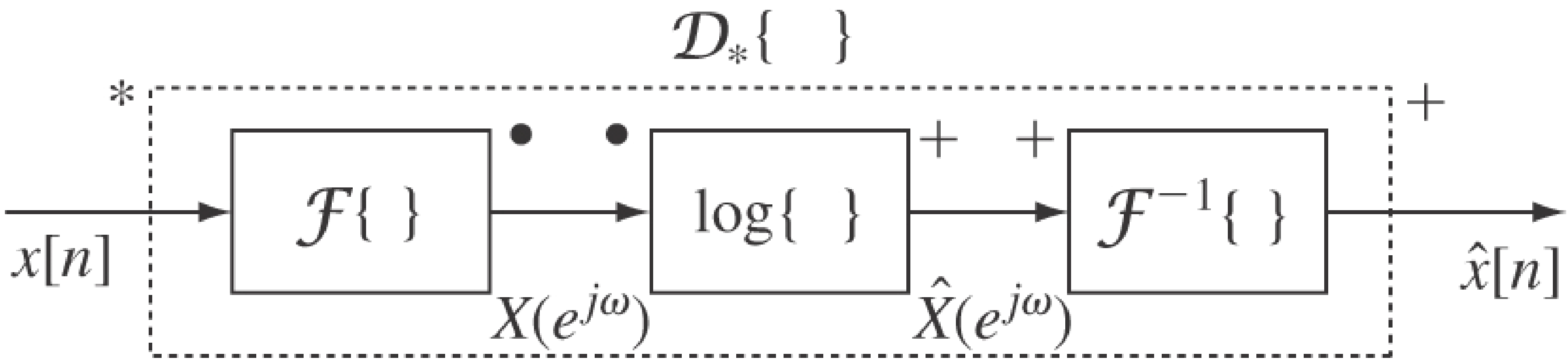
$$= \log\left[X_1(e^{j\omega})\right] + \log\left[X_2(e^{j\omega})\right] = \hat{X}_1(e^{j\omega}) + \hat{X}_2(e^{j\omega})$$

$$\hat{Y}(e^{j\omega}) = \mathcal{L}\left[\hat{X}_1(e^{j\omega}) + \hat{X}_2(e^{j\omega})\right] = \hat{Y}_1(e^{j\omega}) + \hat{Y}_2(e^{j\omega})$$

$$Y(e^{j\omega}) = \exp\left[\hat{Y}_1(e^{j\omega}) + \hat{Y}_2(e^{j\omega})\right] = Y_1(e^{j\omega}) \cdot Y_2(e^{j\omega})$$



Characteristic System for Deconvolution Using DTFTs

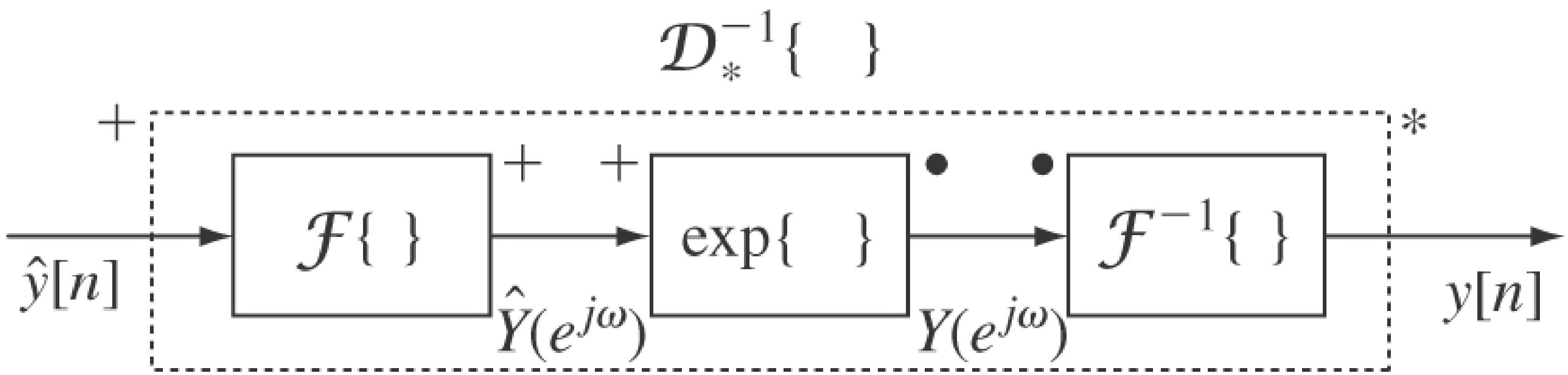


$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n}$$

$$\hat{X}(e^{j\omega}) = \log\left[X(e^{j\omega})\right] = \log\left|X(e^{j\omega})\right| + j\arg\left[X(e^{j\omega})\right]$$

$$\hat{x}[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{X}(e^{j\omega}) e^{j\omega n} d\omega$$

Inverse Characteristic System for Deconvolution Using DTFTs



$$\hat{Y}(e^{j\omega}) = \sum_{n=-\infty}^{\infty} \hat{y}[n] e^{-j\omega n}$$

$$Y(e^{j\omega}) = \exp\left[\hat{Y}(e^{j\omega})\right]$$

$$y[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} Y(e^{j\omega}) e^{j\omega n} d\omega$$

Issues with Logarithms

- it is essential that the logarithm obey the equation

$$\log\left[X_1(e^{j\omega})\cdot X_2(e^{j\omega})\right]=\log\left[X_1(e^{j\omega})\right]+\log\left[X_2(e^{j\omega})\right]$$

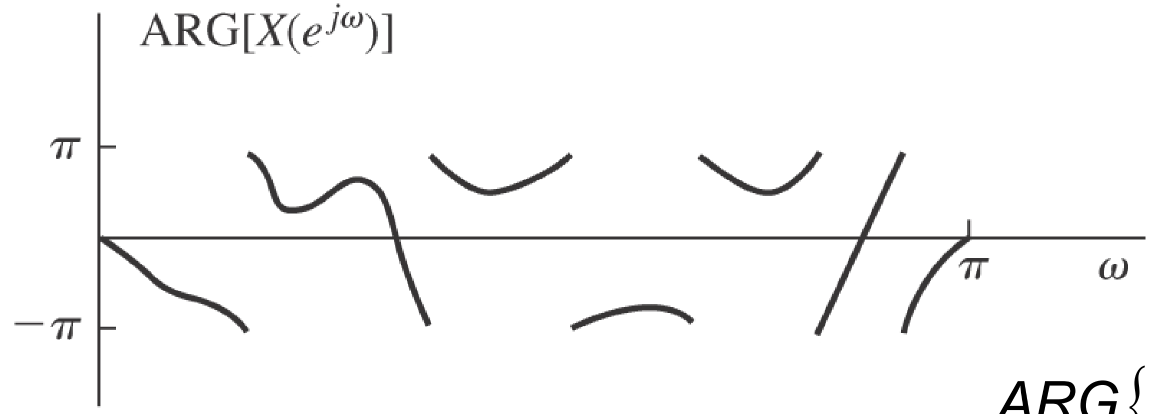
- this is trivial if $X_1(e^{j\omega})$ and $X_2(e^{j\omega})$ are <u>real</u> -- however usually $X_1(e^{j\omega})$ and $X_2(e^{j\omega})$ are <u>complex</u>
- on the unit circle the complex log can be written in the form:

$$X(e^{j\omega})=|X(e^{j\omega})|e^{j\arg\left[X(e^{j\omega})\right]}$$

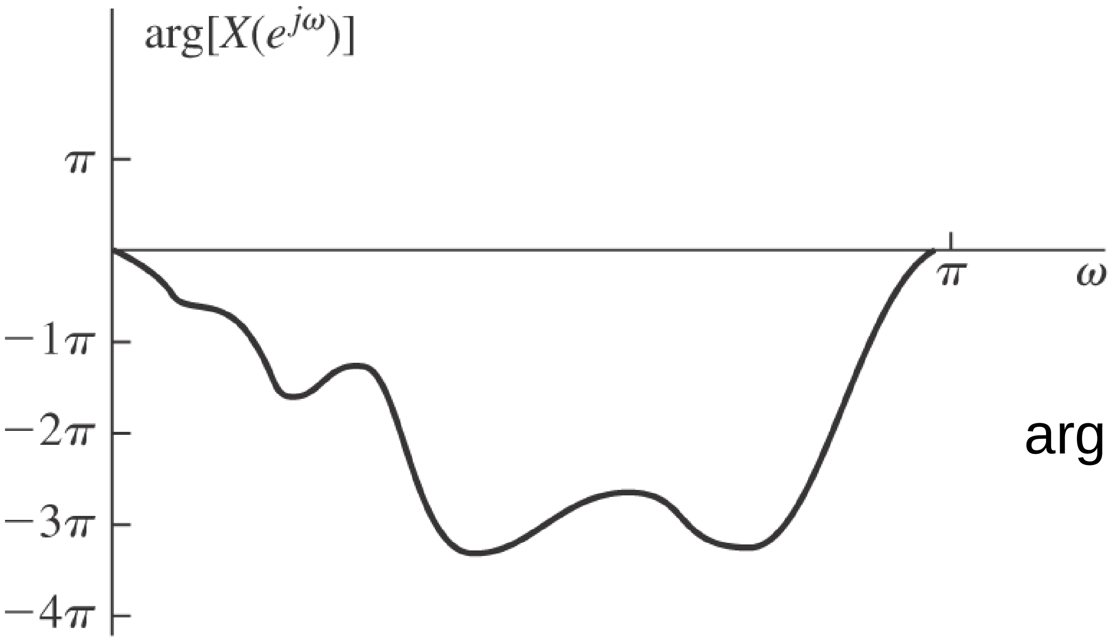
$$\log\left[X(e^{j\omega})\right]=\hat{X}(e^{j\omega})=\log\left[|X(e^{j\omega})|\right]+j\arg\left[X(e^{j\omega})\right]$$

- no problems with log magnitude term; uniqueness problems arise in defining the imaginary part of the log; can show that the imaginary part (the phase angle of the z-transform) needs to be a continuous odd function of ω

Problems with arg Function



$$ARG\{X(e^{j\omega})\} \neq ARG\{X_1(e^{j\omega})\} + ARG\{X_2(e^{j\omega})\}$$



$$\arg\{X(e^{j\omega})\} = \arg\{X_1(e^{j\omega})\} + \arg\{X_2(e^{j\omega})\}$$

Complex Cepstrum Properties

• Given a complex logarithm that satisfies the phase continuity condition, we have:

$$\hat{x}[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} (\log|X(e^{j\omega})| + j\arg\{X(e^{j\omega})\})\, e^{j\omega n}\, d\omega$$

• If $x[n]$ real, then $\log|X(e^{j\omega})|$ is an even function of ω and $\arg\{X(e^{j\omega})\}$ is an odd function of ω. This means that the real and imaginary parts of the complex log have the appropriate symmetry for $\hat{x}[n]$ to be a real sequence, and $\hat{x}[n]$ can be represented as:

$$\hat{x}[n] = c[n] + d[n]$$

where $c[n]$ is the inverse DTFT of $\log|X(e^{j\omega})|$ and the even part of $\hat{x}[n]$, and $d[n]$ is the inverse DTFT of $\arg\{X(e^{j\omega})\}$ and the odd part of $\hat{x}[n]$:

$$c[n] = \frac{\hat{x}[n] + \hat{x}[-n]}{2}; \qquad d[n] = \frac{\hat{x}[n] - \hat{x}[-n]}{2}$$

Complex and Real Cepstrum

- define the inverse Fourier transform of $\hat{X}(e^{j\omega})$ as

$$\hat{x}[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{X}(e^{j\omega}) e^{j\omega n}\, d\omega$$

- where $\hat{x}[n]$ called the "complex cepstrum" since a complex logarithm is involved in the computation
- can also define a "real cepstrum" using just the real part of the logarithm, giving

$$c[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{Re}\left[\hat{X}(e^{j\omega})\right] e^{j\omega n}\, d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|X(e^{j\omega})| e^{j\omega n}\, d\omega$$

- can show that $c[n]$ is the even part of $\hat{x}[n]$

Terminology

- **Spectrum** – Fourier transform of signal autocorrelation
- **Cepstrum** – inverse Fourier transform of log spectrum
- **Analysis** – determining the spectrum of a signal
- **Alanysis** – determining the cepstrum of a signal
- **Filtering** – linear operation on time signal
- **Liftering** – linear operation on cepstrum
- **Frequency** – independent variable of spectrum
- **Quefrency** – independent variable of cepstrum
- **Harmonic** – integer multiple of fundamental frequency
- **Rahmonic** – integer multiple of fundamental frequency

z-Transform Representation

- The $z-$transform of the signal:

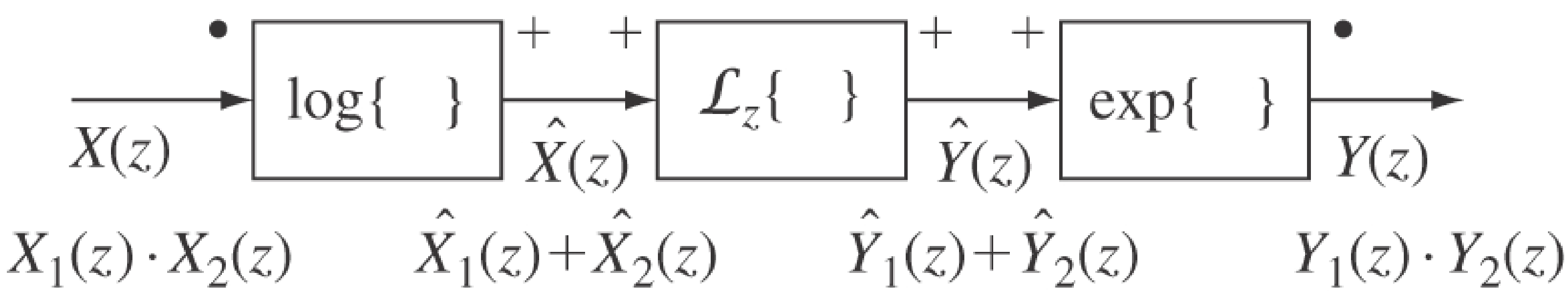
$$x[n] = x_1[n] * x_2[n]$$

is of the form:

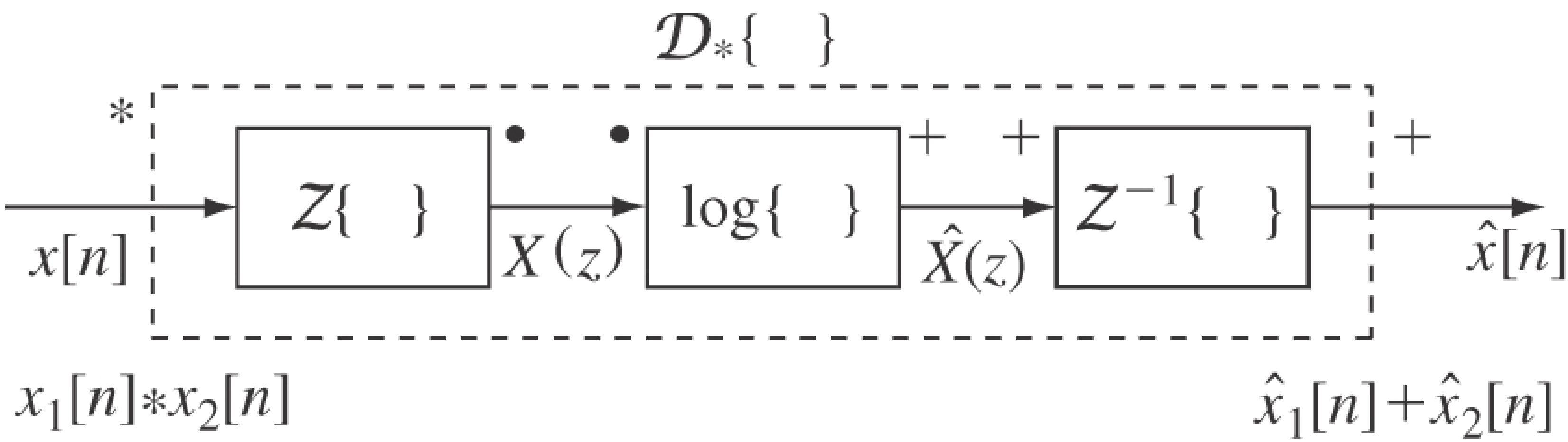
$$X(z) = X_1(z) \cdot X_2(z)$$

- With an appropriate definition of the complex log, we get:

$$\begin{aligned}\hat{X}(z) &= \log\{X(z)\} = \log\{X_1(z) \cdot X_2(z)\} \\ &= \log\{X_1(z)\} + \log\{X_2(z)\} \\ &= \hat{X}_1(z) + \hat{X}_2(z)\end{aligned}$$



Characteristic System for Deconvolution

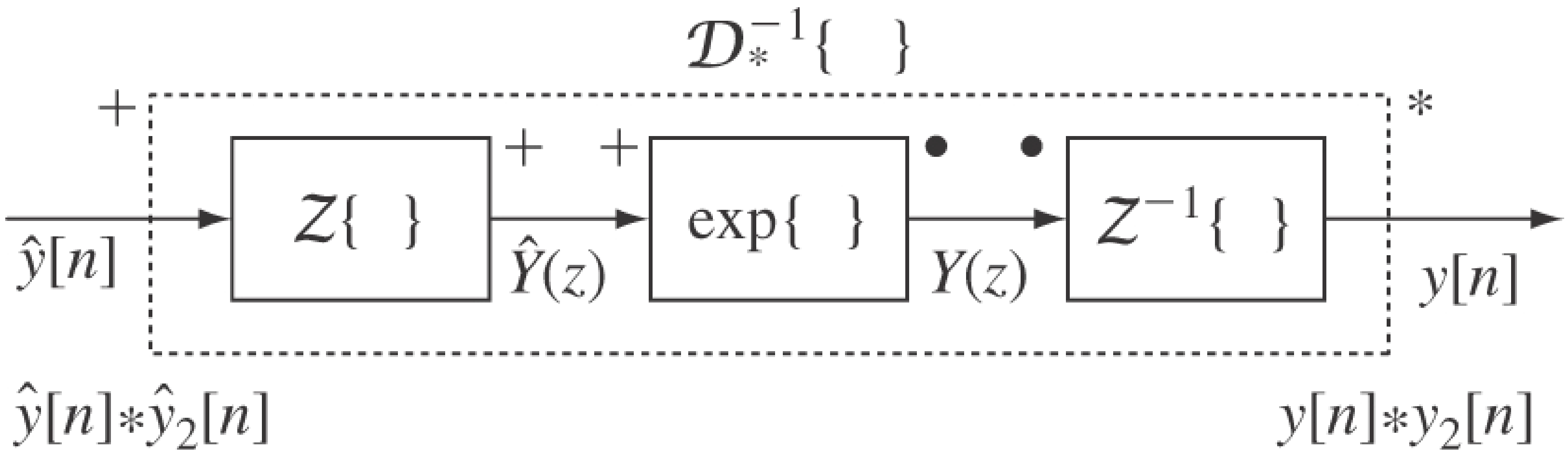


$$X(z)=\sum_{n=-\infty}^{\infty}x[n]z^{-n}=\left|X(z)\right|e^{j\arg\{X(z)\}}$$

$$\hat{X}(z)=\log\left[X(z)\right]=\log\left|X(z)\right|+j\arg\left[X(z)\right]$$

$$\hat{x}[n]=\frac{1}{2\pi j}\oint\hat{X}(z)z^{n}dz$$

Inverse Characteristic System for Deconvolution



$$\hat{Y}(z) = \sum_{n=-\infty}^{\infty} \hat{y}[n] z^{-n}$$

$$Y(z) = \exp\left[\hat{Y}(z)\right] = \log|Y(z)| + j\arg[Y(z)]$$

$$y[n] = \frac{1}{2\pi j}\oint Y(z) z^n dz$$

z-Transform Cepstrum *Alanysis*

- consider digital systems with rational z-transforms of the general type

$$X(z) = \frac{A \prod_{k=1}^{M_i} \left(1 - a_k z^{-1}\right) \prod_{k=1}^{M_0} \left(1 - b_k^{-1} z^{-1}\right)}{\prod_{k=1}^{N_i} \left(1 - c_k z^{-1}\right)}$$

- we can express the above equation as:

$$X(z) = \frac{z^{-M_0} A \prod_{k=1}^{M_0} \left(-b_k^{-1}\right) \prod_{k=1}^{M_i} \left(1 - a_k z^{-1}\right) \prod_{k=1}^{M_0} \left(1 - b_k z\right)}{\prod_{k=1}^{N_i} \left(1 - c_k z^{-1}\right)}$$

- with all coefficients $a_k, b_k, c_k < 1$ => all c_k poles and a_k zeros are inside the unit circle; all b_k zeros are outside the unit circle;

z-Transform Cepstrum *Alanysis*

- express $X(z)$ as product of minimum-phase and maximum-phase signals, i.e.,

$$X(z) = X_{\min}(z) \cdot z^{-M_0} X_{\max}(z)$$

- where

$$X_{\min}(z) = \frac{A\prod_{k=1}^{M_i}\left(1-a_k z^{-1}\right)}{\prod_{k=1}^{N_i}\left(1-c_k z^{-1}\right)}$$

- all poles and zeros inside unit circle

$$X_{\max}(z) = \prod_{k=1}^{M_i}\left(-b_k^{-1}\right)\prod_{k=1}^{M_i}\left(1-b_k z\right)$$

- all zeros outside unit circle

z-Transform Cepstrum *Alanysis*

- can express $x[n]$ as the convolution:

$$x[n] = x_{\min}[n] * x_{\max}[n - M_0]$$

- minimum-phase component is causal

$$x_{\min}[n] = 0, \quad n < 0$$

- maximum-phase component is anti-causal

$$x_{\max}[n] = 0, \quad n > 0$$

- factor z^{-M_0} is the shift in time origin by M_0 samples required so that the overall sequence, $x[n]$ be causal

z-Transform Cepstrum *Alanysis*

- the complex logarithm of $X(z)$ is

$$\hat{X}(z) = \log\left[X(z)\right] = \log|A| + \sum_{k=1}^{M_0} \log|b_k^{-1}| + \log[z^{-M_0}] +$$

$$\sum_{k=1}^{M_i} \log\left(1 - a_k z^{-1}\right) + \sum_{k=1}^{M_0} \log\left(1 - b_k z\right) - \sum_{k=1}^{N_i} \log\left(1 - c_k z^{-1}\right)$$

- evaluating $\hat{X}(z)$ on the unit circle we can ignore the term related to $\log\left[e^{j\omega M_0}\right]$ (as this contributes only to the imaginary part and is a linear phase shift)

z-Transform Cepstrum *Alanysis*

- we can then evaluate the remaining terms, use power series expansion for logarithmic terms (and take the inverse transform to give the complex cepstrum) giving:

$$\hat{x}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{X}(e^{j\omega})e^{j\omega n}\,d\omega$$

$$\boxed{\log(1-Z) = -\sum_{n=1}^{\infty}\frac{Z^n}{n}, \quad |Z|<1}$$

$$= \log|A| + \sum_{k=1}^{M_0}\log|b_k^{-1}| \qquad n = 0$$

$$= \sum_{k=1}^{N_i}\frac{c_k^n}{n} - \sum_{k=1}^{M_i}\frac{a_k^n}{n} \qquad n > 0$$

$$= \sum_{k=1}^{M_0}\frac{b_k^{-n}}{n} \qquad n < 0$$

Cepstrum Properties

1.complex cepstrum is non-zero and of infinite extent for both positive and negative n, even though $x[n]$ may be causal, or even of finite duration ($X(z)$ has only zeros).

2. complex cepstrum is a decaying sequence that is bounded by:

$$|\hat{x}[n]| < \beta \frac{\alpha^{|n|}}{|n|}, \quad \text{for } |n| \to \infty$$

3. zero-quefrency value of complex cepstrum (and the cepstrum) depends on the gain constant and the zeros outside the unit circle. Setting $\hat{x}[0] = 0$ (and therefore $c[0] = 0$) is equivalent to normalizing the log magnitude spectrum to a gain constant of:

$$A \prod_{k=1}^{M_0} (-b_k^{-1}) = 1$$

4. If $X(z)$ has no zeros outside the unit circle (all $b_k = 0$), then:

$$\hat{x}[n] = 0, \quad n < 0 \quad \text{(minimum-phase signals)}$$

5. If $X(z)$ has no poles or zeros inside the unit circle (all $a_k, c_k = 0$), then:

$$\hat{x}[n] = 0, \quad n > 0 \quad \text{(maximum-phase signals)}$$

z-Transform Cepstrum *Alanysis*

- The main z-transform formula for cepstrum alanysis is based on the power series expansion:

$$\boxed{\log(1+x)=\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}x^n \qquad |x|<1}$$

- ***Example 1***--Apply this formula to the exponential sequence

$$x_1(n)=a^n u(n) \Leftrightarrow X_1(z)=\frac{1}{1-az^{-1}}$$

$$\hat{X}_1(z)=\log[X_1(z)]=-\log(1-az^{-1})=-\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}(-a)^n z^{-n}$$

$$\hat{x}_1(n)=\frac{a^n}{n}u(n-1) \Leftrightarrow \hat{X}_1(z)=-\log(1-az^{-1})=\sum_{n=1}^{\infty}\left(\frac{a^n}{n}\right)z^{-n}$$

z-Transform Cepstrum *Alanysis*

- ***Example 2***--consider the case of a digital system with a single zero outside the unit circle ($|b| < 1$)

$$x_2(n) = \delta(n) + b\delta(n+1)$$

$$X_2(z) = 1 + bz \quad \text{(zero at } z = -1/b)$$

$$\hat{X}_2(z) = \log\left[X_2(z)\right] = \log\left(1 + bz\right)$$

$$= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} (b)^n z^n$$

$$\hat{x}_2(n) = \frac{(-1)^{n+1} b^n}{n} u(-n-1)$$

z-Transform Cepstrum Alanysis for 2 Pulses

- ***Example 3***--an input sequence of two pulses of the form

$$x_3(n) = \delta(n) + \alpha\delta(n - N_p) \qquad (0 < \alpha < 1)$$

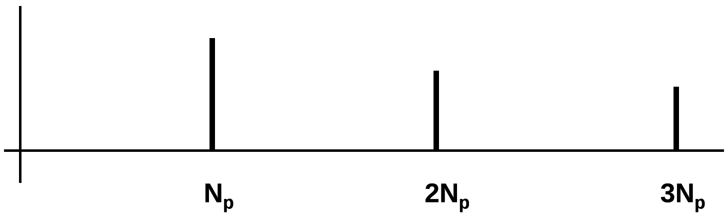
$$X_3(z) = 1 + \alpha z^{-N_p}$$

$$\hat{X}_3(z) = \log\left[X_3(z)\right] = \log\left(1 + \alpha z^{-N_p}\right)$$

$$= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \alpha^n z^{-nN_p}$$

$$\hat{x}_3(n) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\alpha^k}{k} \delta(n - kN_p)$$

- the cepstrum is an impulse train with impulses spaced at N_p samples



Cepstrum for Train of Impulses

- an important special case is a train of impulses of the form:

$$x(n) = \sum_{r=0}^{M} \alpha_r \delta(n - rN_p)$$

$$X(z) = \sum_{r=0}^{M} \alpha_r z^{-rN_p}$$

- clearly $X(z)$ is a polynomial in z^{-N_p} rather than z^{-1}; thus $X(z)$ can be expressed as a product of factors of the form $(1 - az^{-N_p})$ and $(1 - bz^{N_p})$, giving a complex cepstrum, $\hat{x}(n)$, that is non-zero only at integer multiples of N_p

z-Transform Cepstrum Alanysis for Convolution of 2 Sequences

- ***Example 4***--consider the convolution of sequences 1 and 3, i.e.,

$$x_4(n) = x_1(n) * x_3(n) = \left[a^n u(n) \right] * \left[\delta(n) + \alpha\delta(n - N_p) \right]$$
$$= a^n u(n) + \alpha\, a^{n-N_p} u(n - N_p)$$

- The complex cepstrum is therefore the sum of the complex cepstra of the two sequences (since convolution in the time domain is converted to addition in the cepstral domain)

$$\hat{x}_4(n) = \hat{x}_1(n) + \hat{x}_3(n)$$
$$= \frac{a^n}{n} u(n-1) + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\alpha^k}{k}\, \delta(n - kN_p)$$

z-Transform Cepstrum Alanysis for Convolution of 3 Sequences

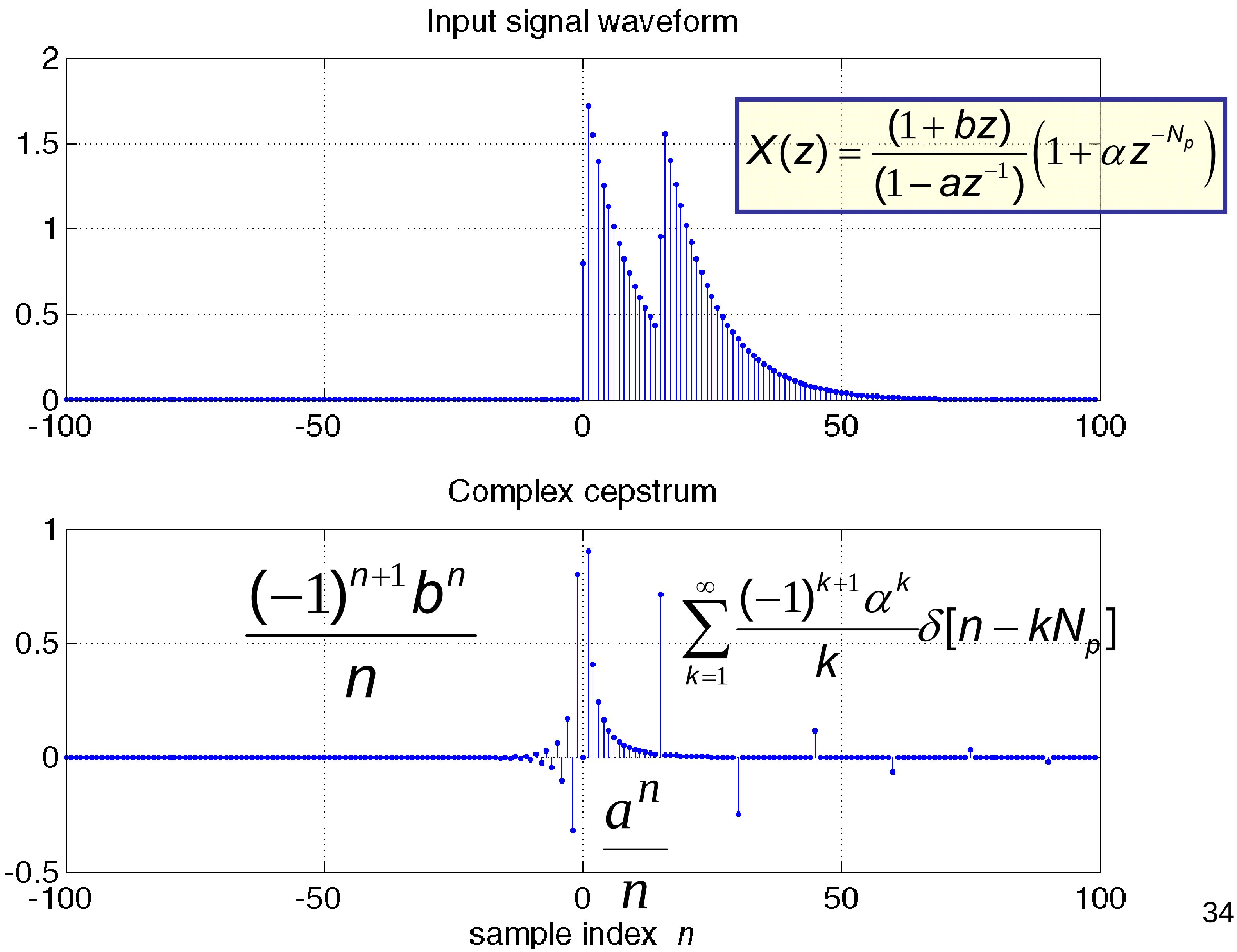
- ***Example 5***--consider the convolution of sequences 1, 2 and 3, i.e.,

$$
\begin{aligned}
x_5(n) &= x_1(n) * x_2(n) * x_3(n) \\
&= \left[a^n u(n)\right] * \left[\delta(n) + b\delta(n+1)\right] * \left[\delta(n) + \alpha\delta(n-N_p)\right] \\
&= a^n u(n) + \alpha\, a^{n-N_p} u(n-N_p) + b a^n u(n+1) + \alpha b a^{n-N_p+1} u(n-N_p+1)
\end{aligned}
$$

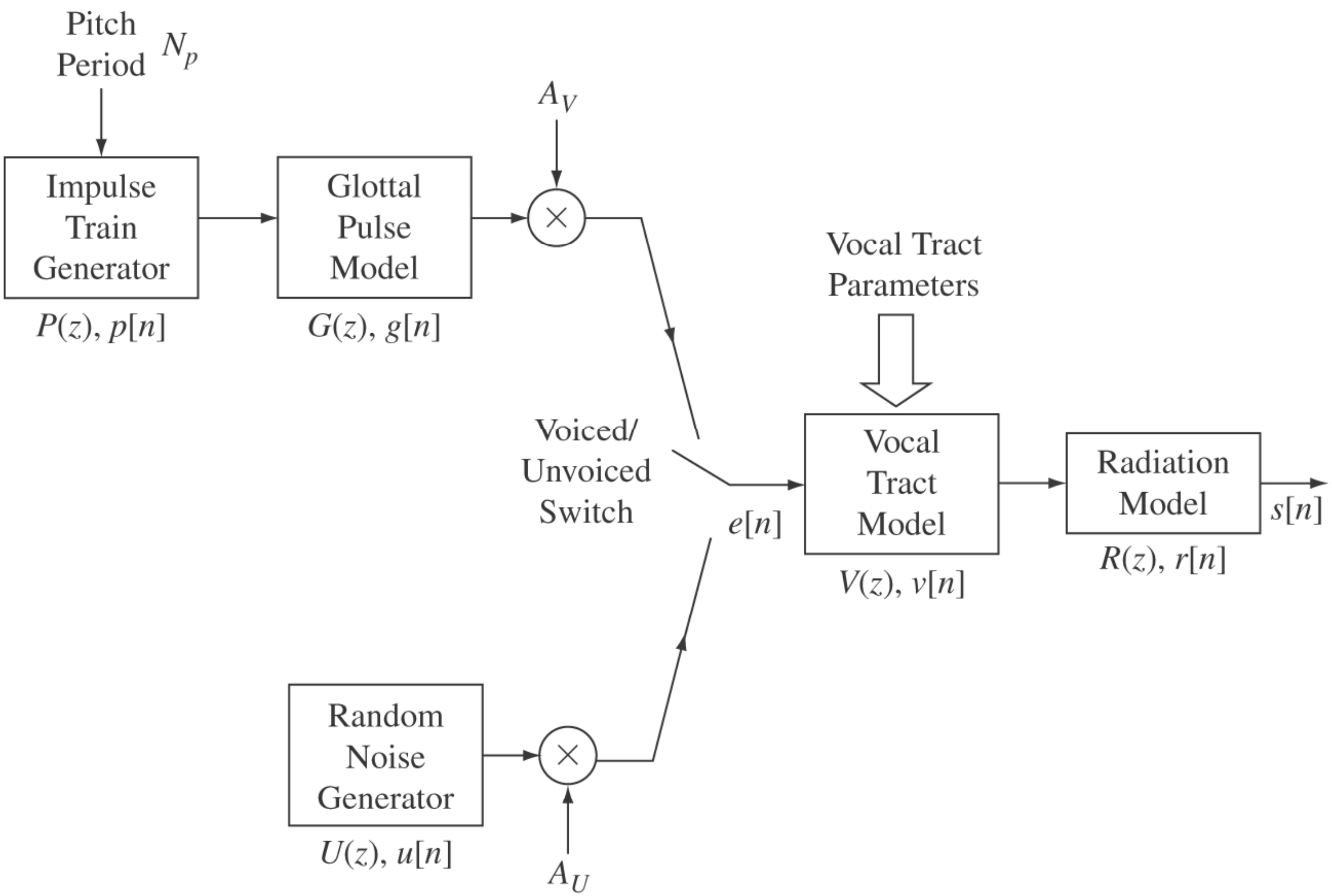
- The complex cepstrum is therefore the sum of the complex cepstra of the three sequences

$$
\begin{aligned}
\hat{x}_5(n) &= \hat{x}_1(n) + \hat{x}_2(n) + \hat{x}_3(n) \\
&= \frac{a^n}{n} u(n-1) + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\alpha^k}{k}\, \delta(n-kN_p) + \frac{(-1)^{n+1} b^n}{n} u(-n-1)
\end{aligned}
$$

Example: a=.9, b=.8, a=.7, Np=15



Homomorphic Analysis of Speech Model



Homomorphic Analysis of Speech Model

- the transfer function for voiced speech is of the form

$$H_V(z) = A_V \cdot G(z)V(z)R(z)$$

- with effective impulse response for voiced speech

$$h_V[n] = A_V \cdot g[n] * v[n] * r[n]$$

- similarly for unvoiced speech we have

$$H_U(z) = A_U \cdot V(z)R(z)$$

- with effective impulse response for unvoiced speech

$$h_U[n] = A_U \cdot v[n] * r[n]$$

Complex Cepstrum for Speech

- the models for the speech components are as follows:

1. vocal tract: $$V(z) = \frac{Az^{-M} \prod_{k=1}^{M_i} (1 - a_k z^{-1}) \prod_{k=1}^{M_0} (1 - b_k z)}{\prod_{k=1}^{N_i} (1 - c_k z^{-1})}$$

--for voiced speech, only poles => $a_k = b_k = 0$, all k

--unvoiced speech and nasals, need pole-zero model but all poles are inside the unit circle => $c_k < 1$

--all speech has complex poles and zeros that occur in complex conjugate pairs

2. radiation model: $R(z) \approx 1 - z^{-1}$ (high frequency emphasis)

3. glottal pulse model: finite duration pulse with transform

$$G(z) = B \prod_{k=1}^{L_i} (1 - \alpha_k z^{-1}) \prod_{k=1}^{L_0} (1 - \beta_k z)$$

with zeros both inside and outside the unit circle

Complex Cepstrum for Voiced Speech

- combination of vocal tract, glottal pulse and radiation will be non-minimum phase => complex cepstrum exists for all values of *n*
- the complex cepstrum will decay rapidly for large *n* (due to polynomial terms in expansion of complex cepstrum)
- effect of the voiced source is a periodic pulse train for multiples of the pitch period

Simplified Speech Model

- short-time speech model

$$x[n] = w[n] \cdot \left[p[n] * g[n] * v[n] * r[n] \right]$$

$$\approx p_w[n] * h_v[n]$$

- short-time complex cepstrum

$$\hat{x}[n] = \hat{p}_w[n] + \hat{g}[n] + \hat{v}[n] + \hat{r}[n]$$

Analysis of Model for Voiced Speech

- Assume sustained /AE/ vowel with fundamental frequency of 125 Hz
- Use glottal pulse model of the form:

$$g[n]=\begin{cases} 0.5\,[1-\cos(\pi(n+1)/N_1)] & 0\le n\le N_1-1 \\ \cos(0.5\pi(n+1-N_1)/N_2) & N_1\le n\le N_1+N_2-2 \\ 0 & \text{otherwise} \end{cases}$$

$N_1=25,\ \ N_2=10 \Rightarrow$ 34 sample impulse response, with transform

$$G(z)=z^{-33}\prod_{k=1}^{33}(-b_k^{-1})\prod_{k=1}^{33}(1-b_k z)\ \Rightarrow \text{all roots outside unit circle} \Rightarrow \text{maximum phase}$$

- Vocal tract system specified by 5 formants (frequencies and bandwidths)

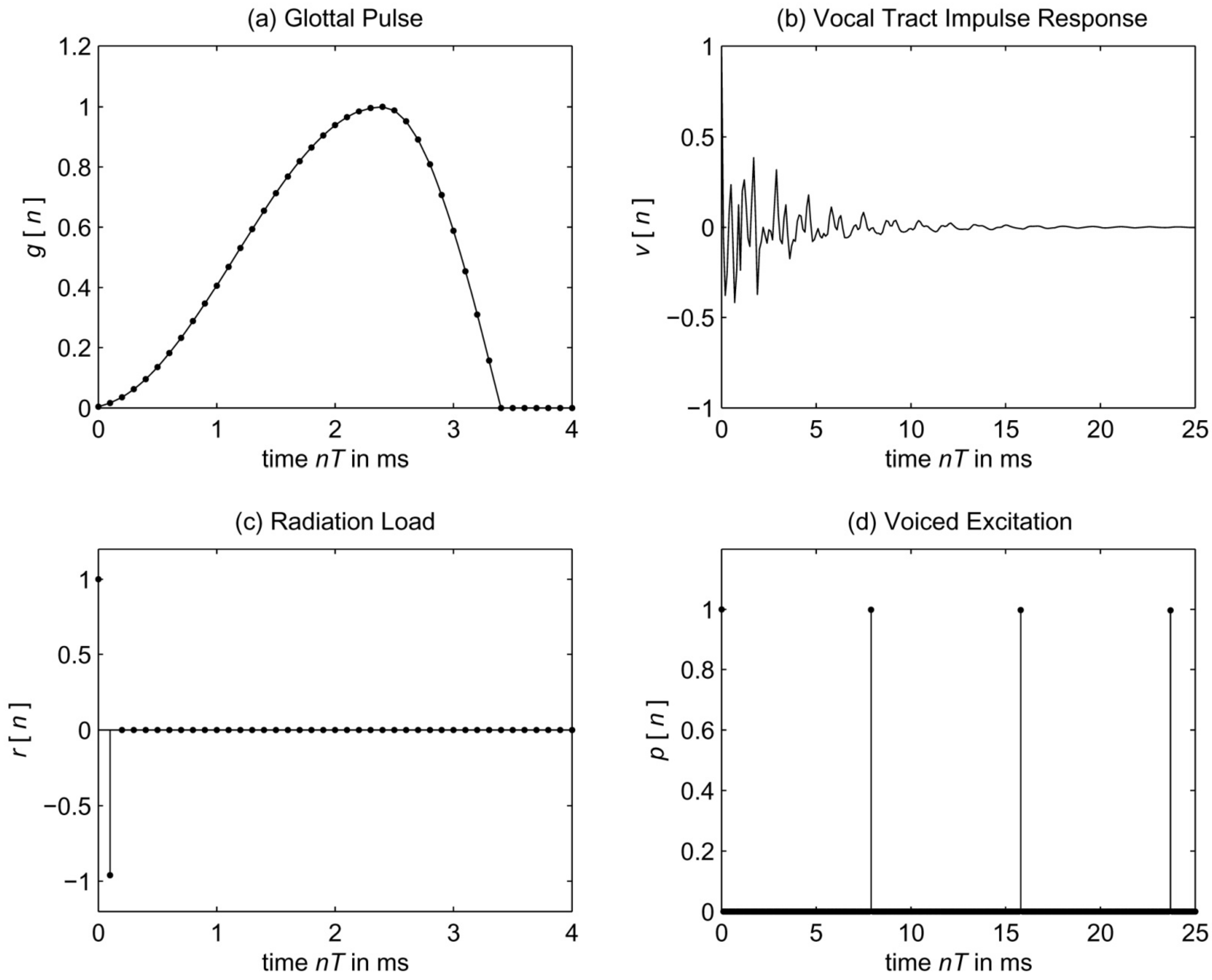
$$V(z)=\frac{1}{\prod_{k=1}^{5}(1-2e^{-2\pi\sigma_k T}\cos(2\pi F_k T)z^{-1}+e^{-4\pi\sigma_k T}z^{-2})}$$

$$\{F_k,\sigma_k\}=[(660,60),(1720,100),(2410,120),(3500,175),(4500,250)]$$

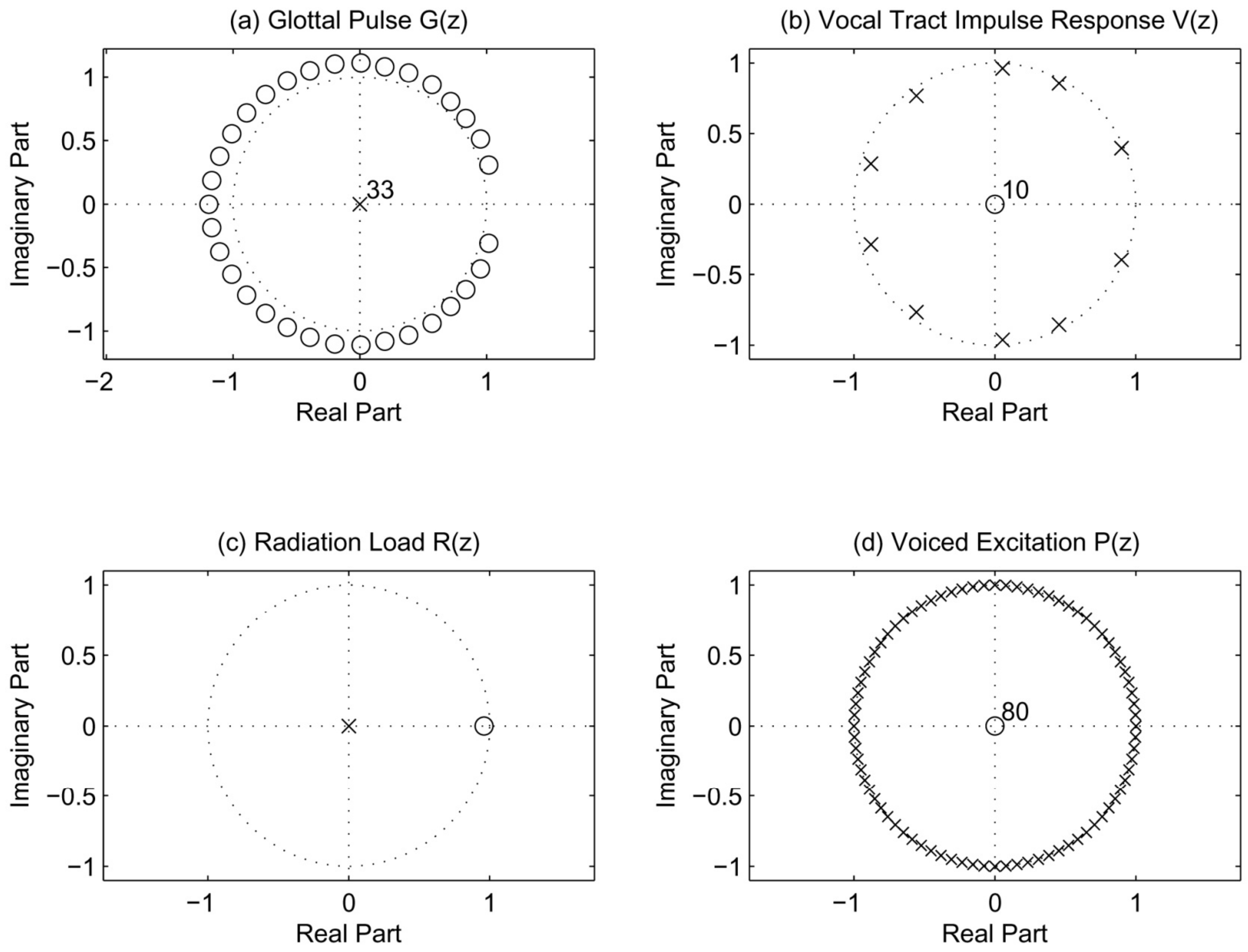
- Radiation load is simple first difference

$$R(z)=1-\gamma z^{-1},\ \ \gamma=0.96$$

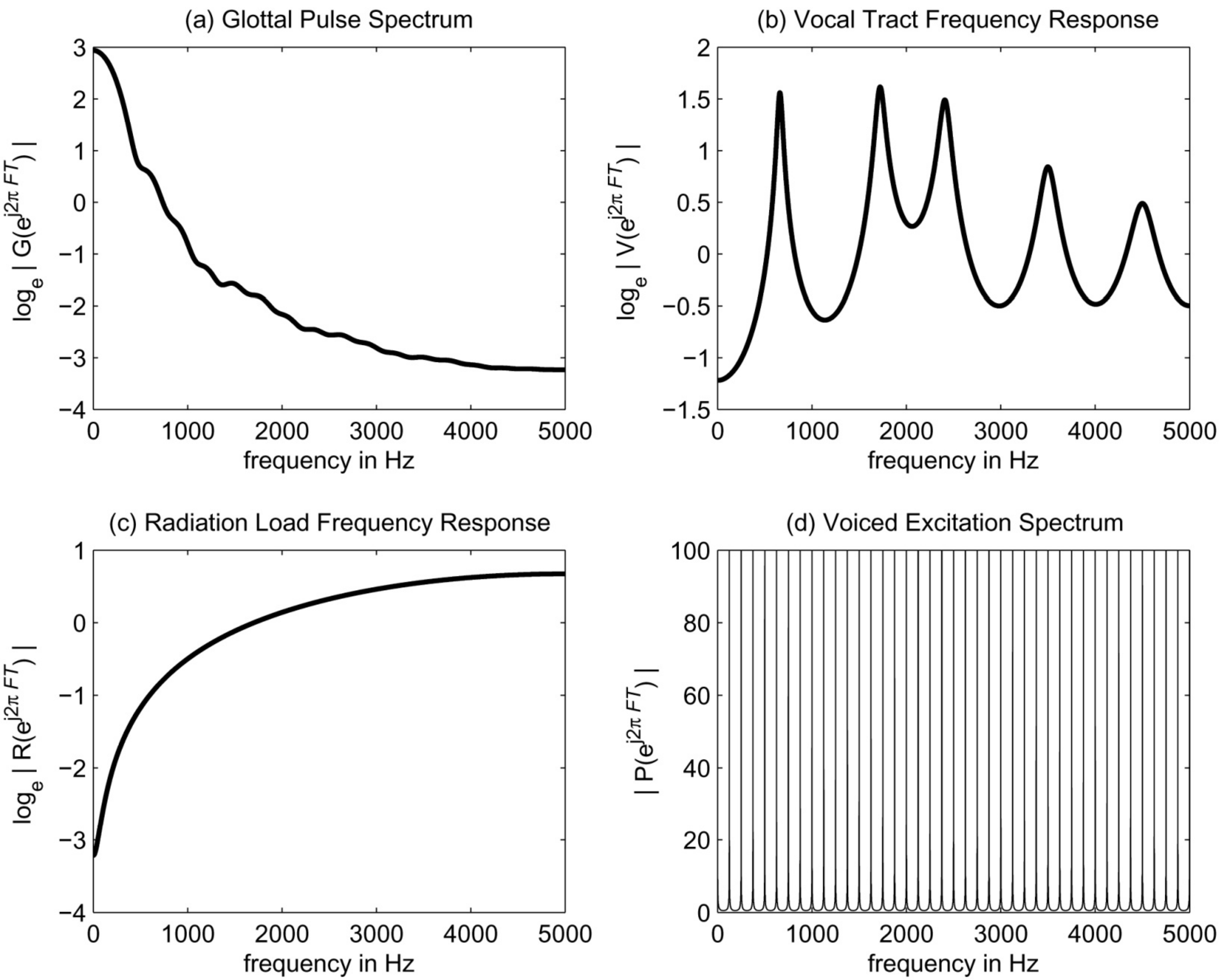
Time Domain Analysis



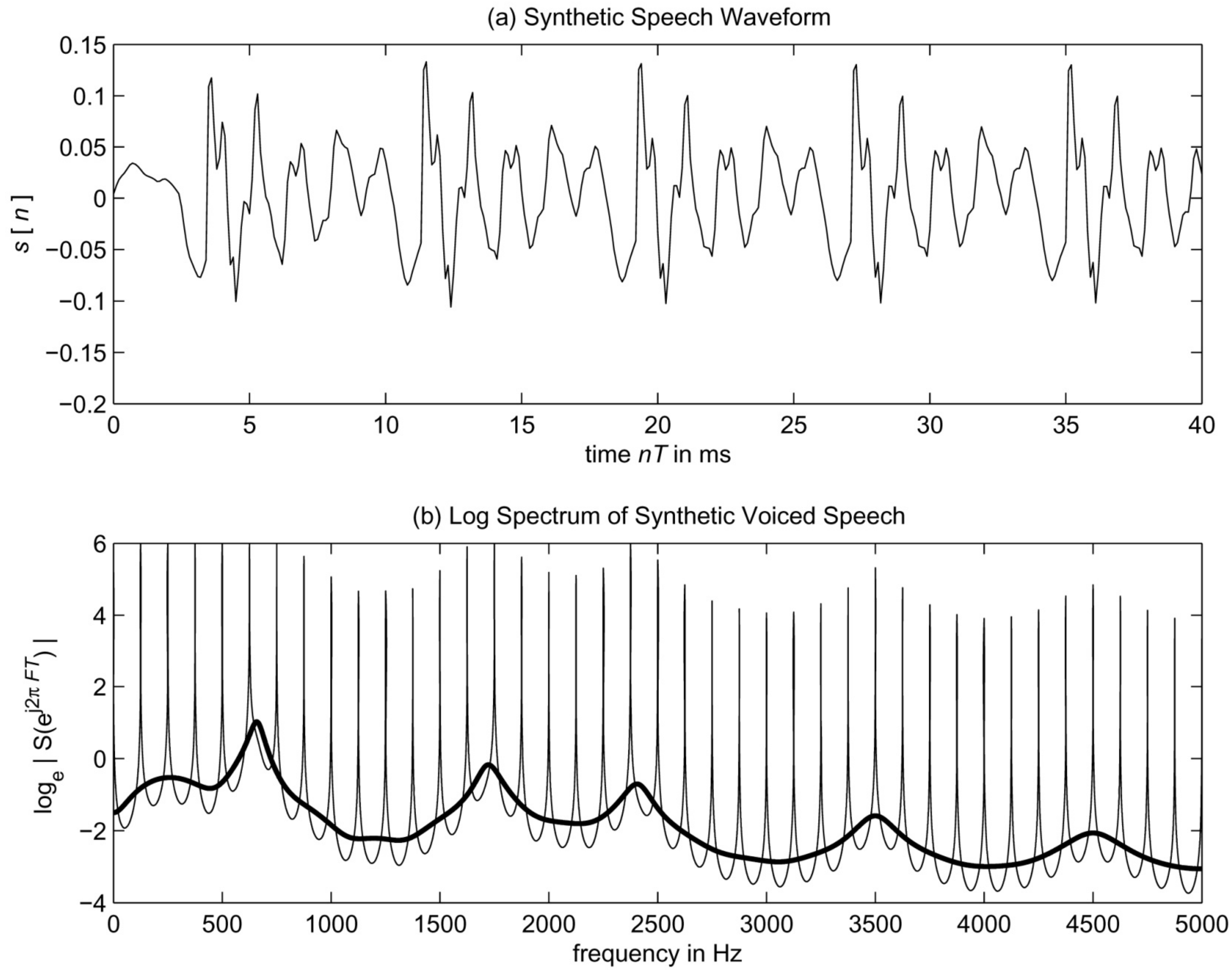
Pole-Zero Analysis of Model Components



Spectral Analysis of Model



Speech Model Output



Complex Cepstrum of Model

- The voiced speech signal is modeled as:

$$x[n] = A_V \cdot g[n] * v[n] * r[n] * p[n]$$

- with complex cepstrum:

$$\hat{s}[n] = \log|A_V|\delta[n] + \hat{g}[n] + \hat{v}[n] + \hat{r}[n] + \hat{p}[n]$$

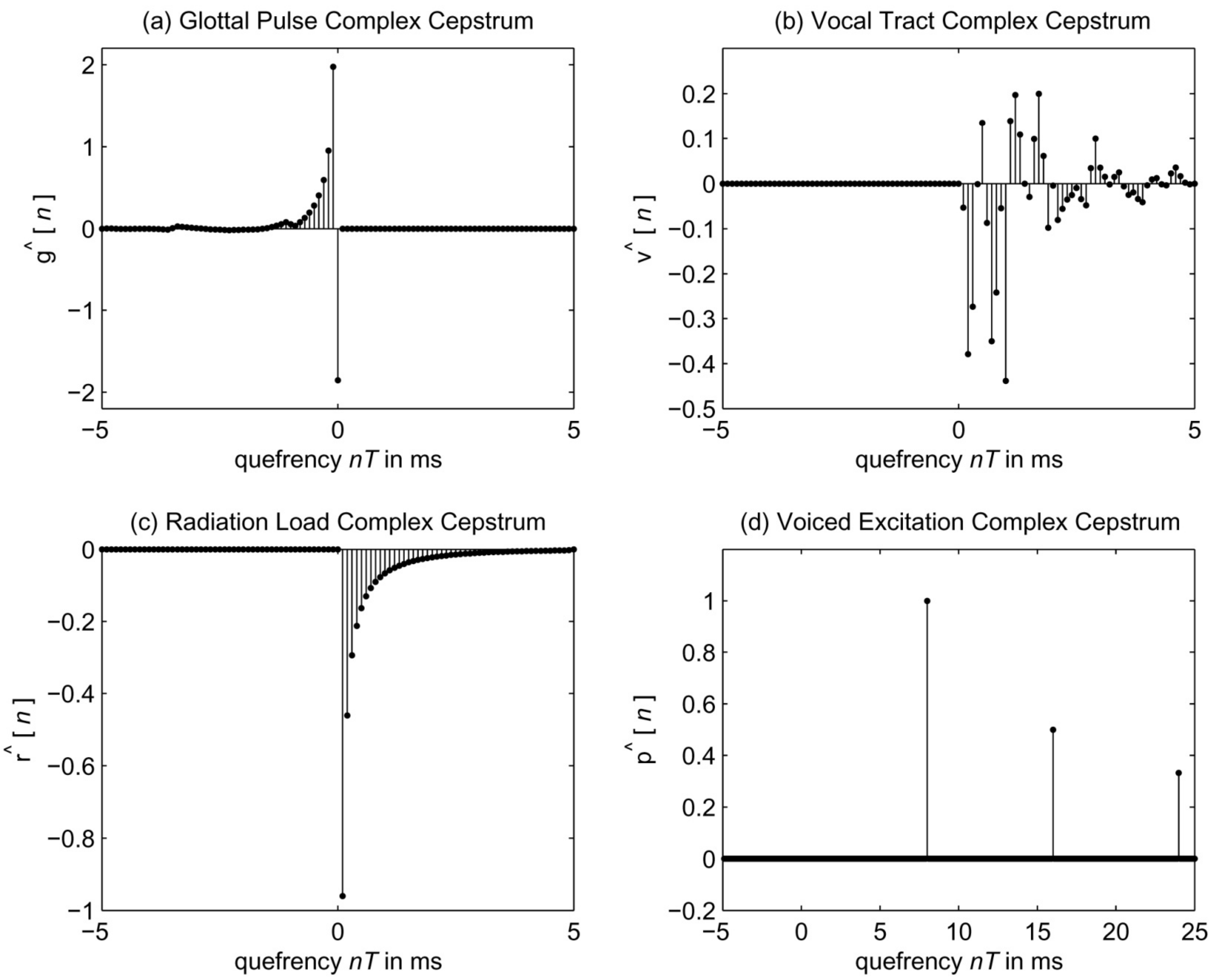
- glottal pulse is maximum phase $\Rightarrow \hat{g}[n] = 0,\ n > 0$
- vocal tract and radiation systems are minimum phase

$\Rightarrow \hat{v}[n] = 0,\ n < 0,\ \hat{r}[n] = 0,\ n < 0$

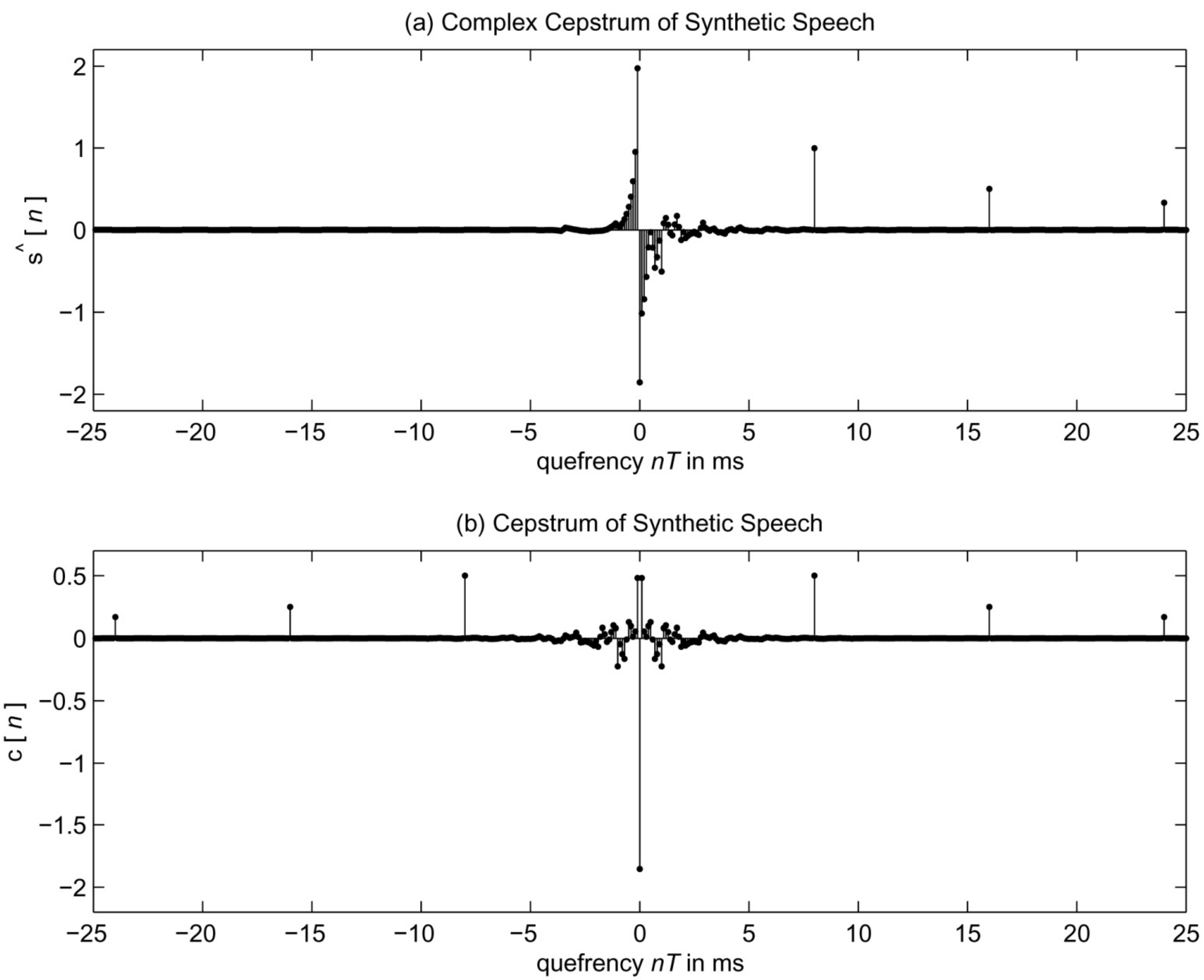
$$\hat{P}(z) = -\log(1 - \beta z^{-N_p}) = \sum_{k=1}^{\infty} \frac{\beta^k}{k} z^{-kN_p}$$

$$\hat{p}[n] = \sum_{k=1}^{\infty} \frac{\beta^k}{k} \delta[n - kN_p]$$

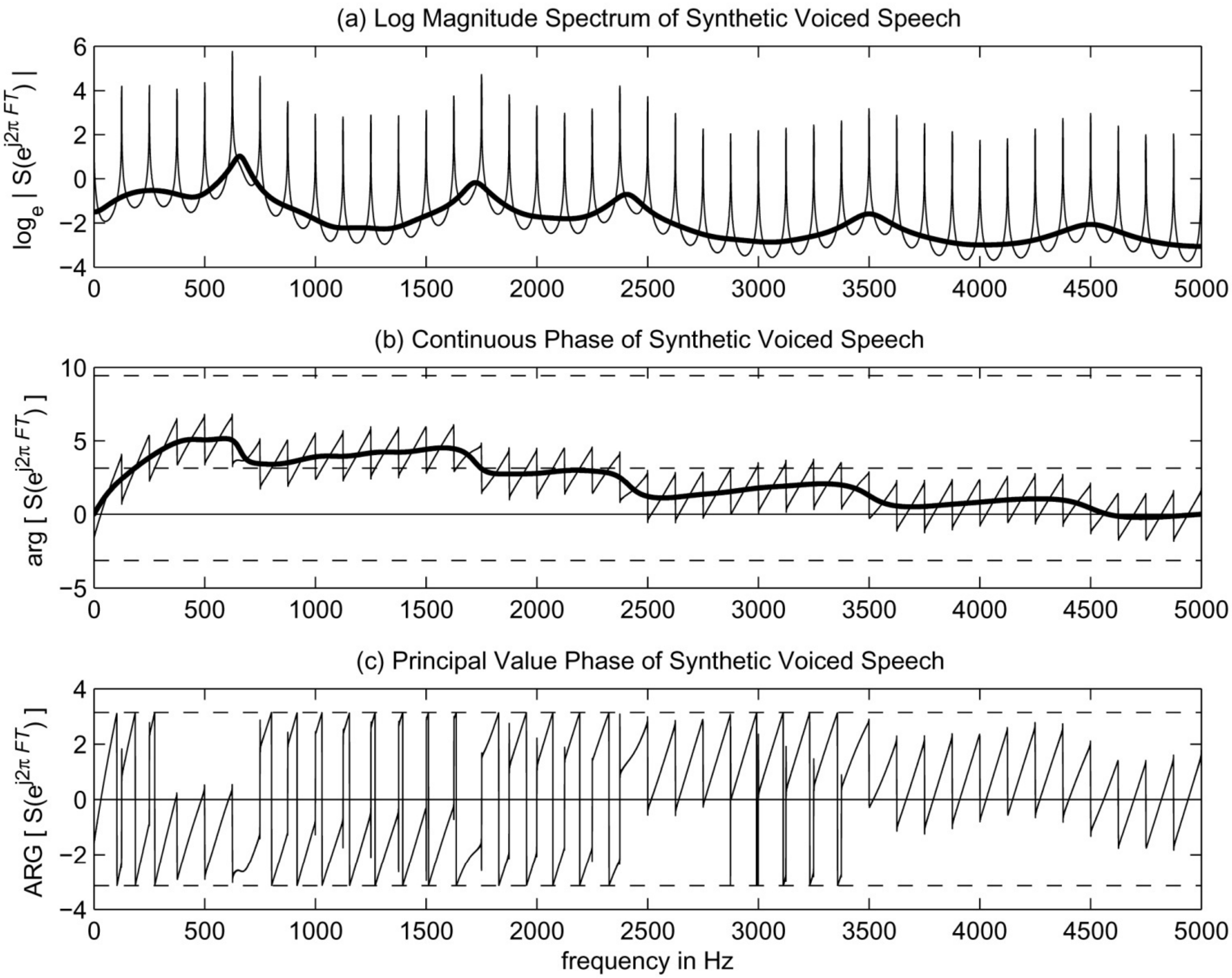
Cepstral Analysis of Model



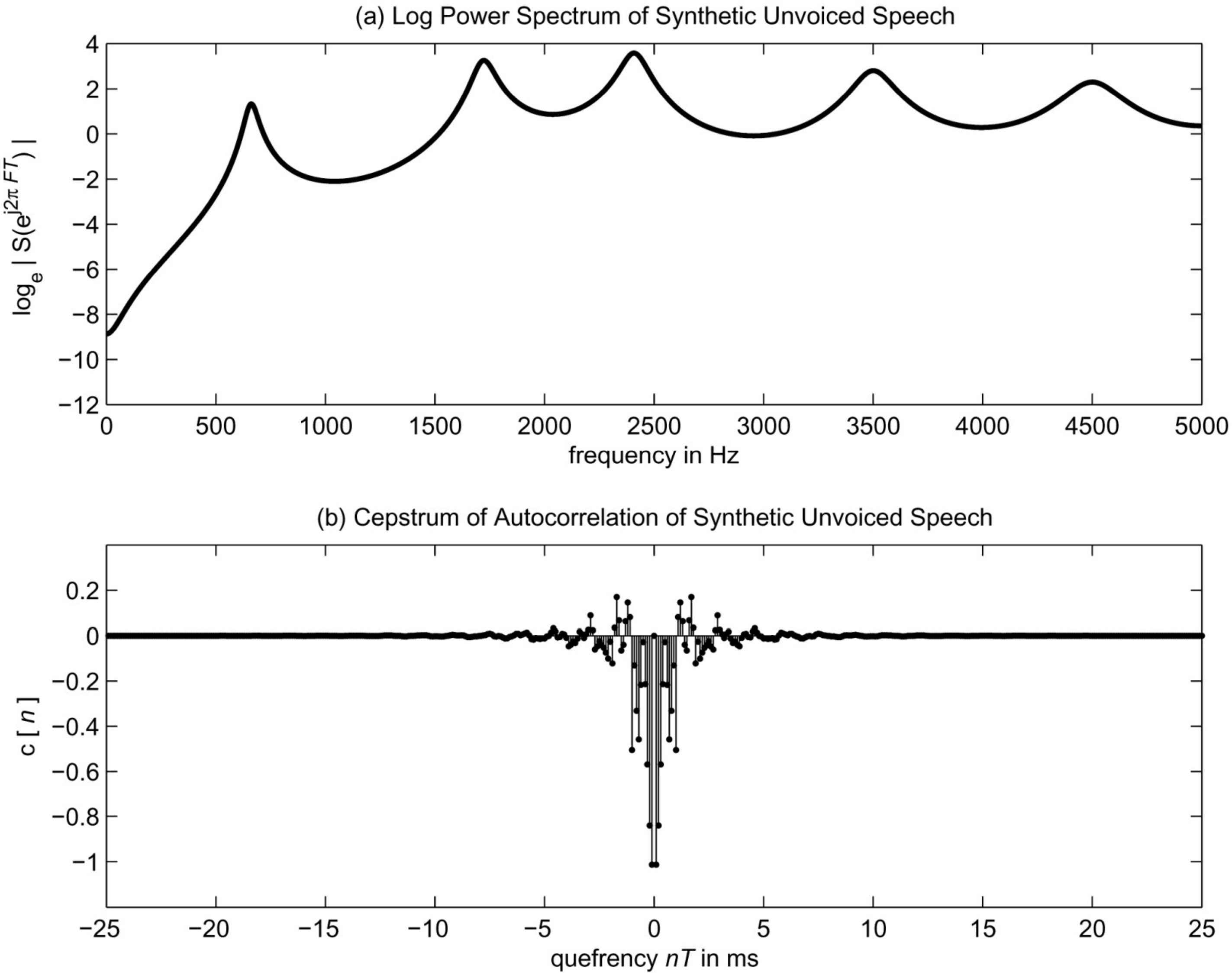
Resulting Complex and Real Cepstra



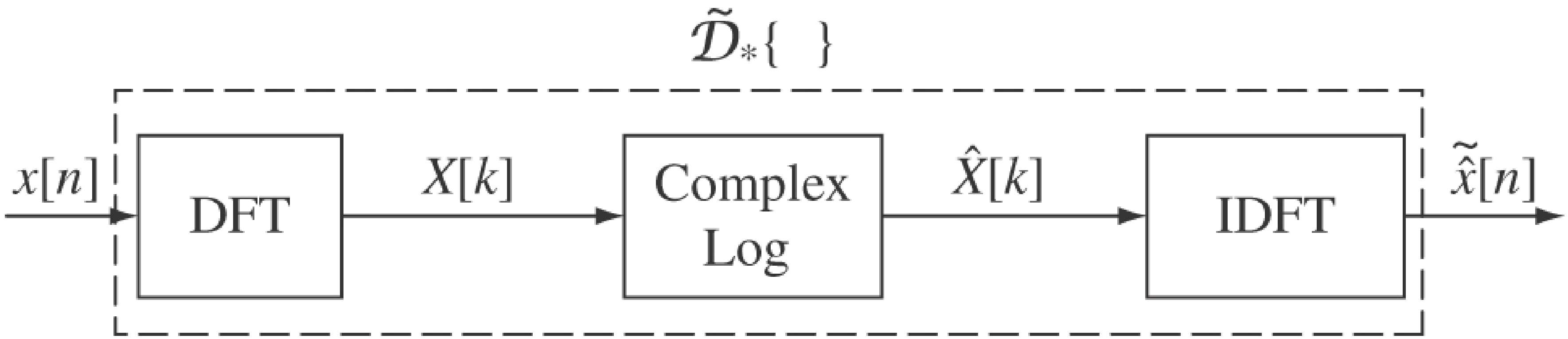
Frequency Domain Representations



Frequency-Domain Representation of Complex Cepstrum



The Complex Cepstrum-DFT Implementation



$$X[k] = X(e^{j\frac{2\pi}{N}k}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\frac{2\pi}{N}kn} \quad k = 0,1,\ldots,N-1,$$

- $X_p[k]$ is the N point DFT corresponding to $X(e^{j\omega})$

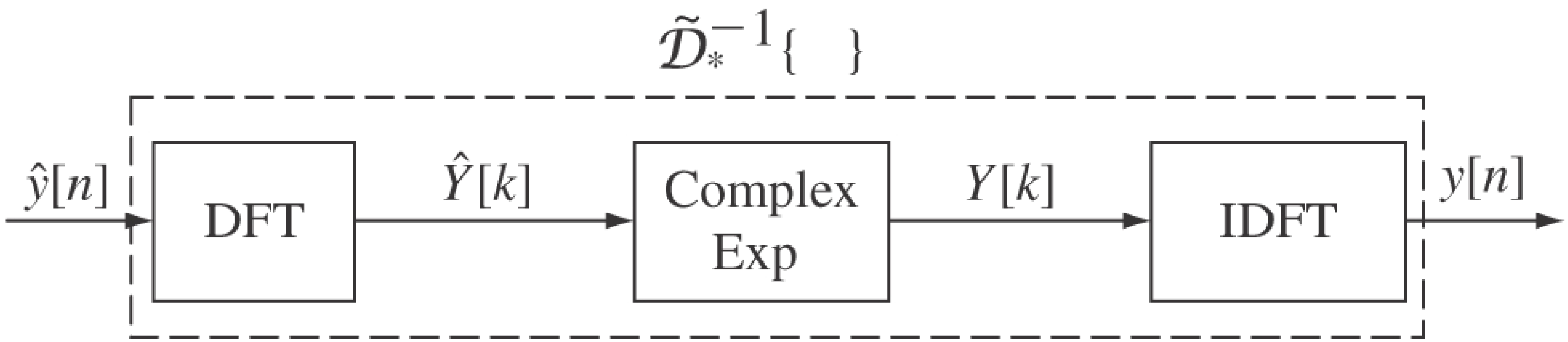
$$\hat{X}[k] = \hat{X}(e^{j2\pi k/N}) = \log\{X[k]\} = \log|X[k]| + j\arg\{X[k]\}$$

$$\tilde{\hat{x}}[n] = \frac{1}{N}\sum_{k=0}^{N-1}\hat{X}[k]e^{j\frac{2\pi}{N}kn} = \sum_{r=-\infty}^{\infty}\hat{x}[n+rN] \quad n = 0,1,\ldots,N-1$$

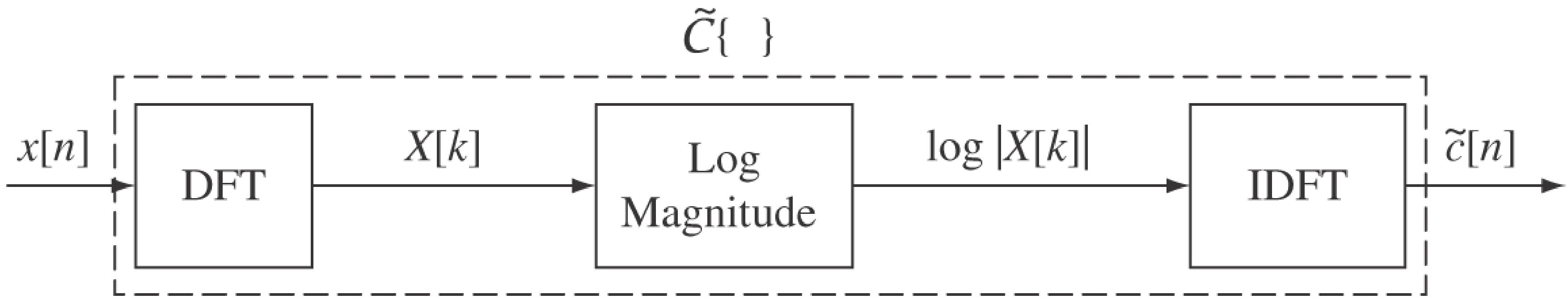
- $\tilde{\hat{x}}[n]$ is an aliased version of $\hat{x}[n]$

$\Rightarrow$ <u>use as large a value of N as possible to minimize aliasing</u>

Inverse System- DFT Implementation



The Cepstrum-DFT Implementation



$$c[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|X(e^{j\omega})| e^{j\omega n} d\omega \qquad -\infty < n < \infty$$

- Approximation to cepstrum using DFT:

$$X[k] = X(e^{j\frac{2\pi}{N}k}) = \sum_{n=-\infty}^{\infty} x[n] e^{-j\frac{2\pi}{N}kn} \quad k = 0,1,\ldots,N-1,$$

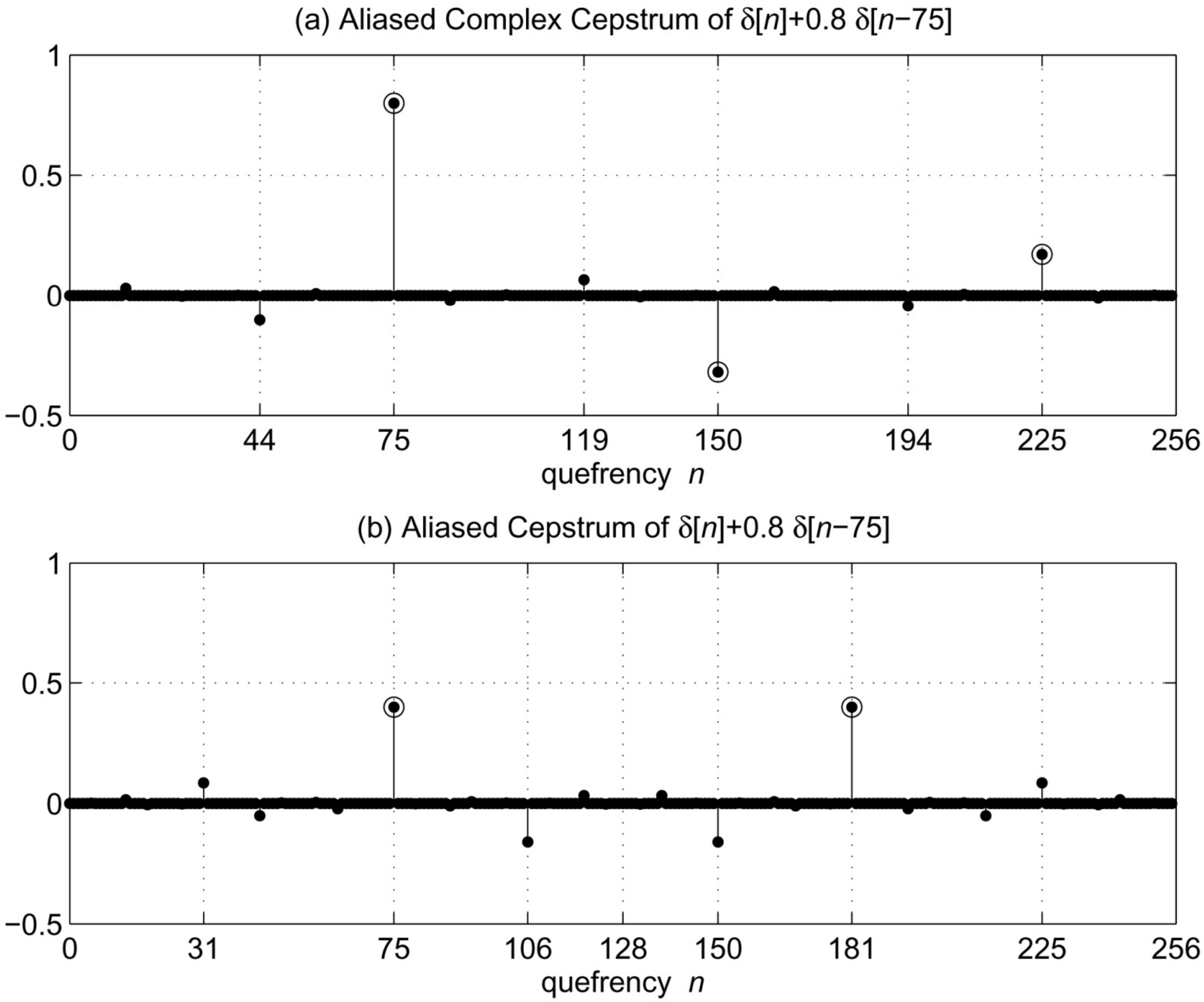
$$\tilde{c}[n] = \frac{1}{N}\sum_{k=0}^{N-1} \log|X[k]| e^{j2\pi kn/N}, \qquad 0 \le n \le N-1$$

$$\tilde{c}(n) = \sum_{r=-\infty}^{\infty} c[n+rN] \qquad n = 0,1,\ldots,N-1$$

- $\tilde{c}[n]$ is an aliased version of $c[n] \Rightarrow$ use as large a value of N as possible to minimize aliasing

$$\tilde{c}(n) = \frac{\tilde{\hat{x}}[n] + \tilde{\hat{x}}[-n]}{2}$$

Cepstral Computation Aliasing

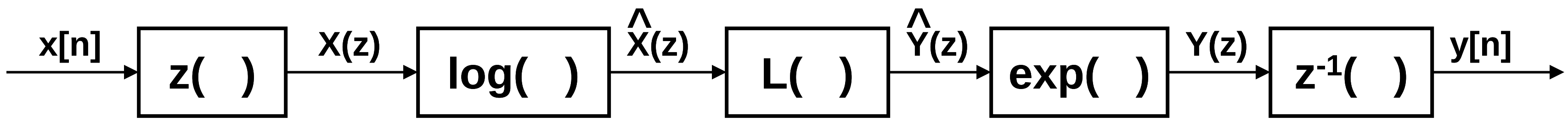


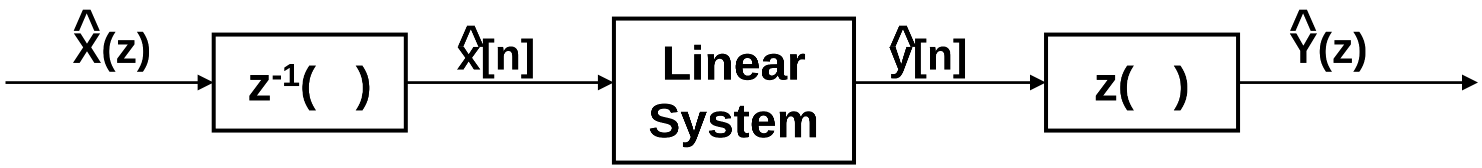
$N=256$, $N_p=75$, $\alpha=0.8$

Circle dots are cepstrum values in correct locations; all other dots are results of aliasing due to finite range computations

Summary

1. Homomorphic System for Convolution:





2. Practical Case:

$$z(\ ) \rightarrow DFT$$

$$z^{-1}(\ ) \rightarrow IDFT$$

$$X(e^{j\omega}) = \left| X(e^{j\omega}) \right| e^{j \arg\left\{X(e^{j\omega})\right\}}$$

$$\log\left[X(e^{j\omega})\right] = \log\left|X(e^{j\omega})\right| + j \arg\left\{X(e^{j\omega})\right\}$$

Summary

3. Complex Cepstrum:

$$\hat{x}[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{X}(e^{j\omega})e^{j\omega n}d\omega$$

4. Cepstrum:

$$c[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|X(e^{j\omega})\right|e^{j\omega n}d\omega$$

$$c[n] = \frac{\hat{x}[n] + \hat{x}[-n]}{2} = \text{even part of } \hat{x}[n]$$

Summary

$$x(n) \longrightarrow \boxed{\text{DFT}} \xrightarrow{X(k)} \boxed{\text{Complex Log}} \xrightarrow{\hat{X}(k)} \boxed{\text{IDFT}} \xrightarrow{\tilde{\hat{x}}(n)}$$

5. Practical Implementation of Complex Cepstrum:

$$X[k] = X(e^{j(2\pi/N)k}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j(2\pi/N)kn})$$

$$\hat{X}[k] = \log\left\{X_p(k)\right\} = \log\left|X_p(k)\right| + j\arg\left\{X_p(k)\right\}$$

$$\tilde{\hat{x}}[n] = \frac{1}{N}\sum_{k=0}^{N-1}\hat{X}[k]e^{j(2\pi/N)kn} = \sum_{r=-\infty}^{\infty}\hat{x}[n+rN] \Rightarrow \text{aliasing}$$

$$\tilde{\hat{x}}[n] = \text{aliased version of } \hat{x}[n]$$

6. Examples:

$$X(z) = \frac{1}{1-az^{-1}} \Leftrightarrow \hat{x}[n] = \sum_{r=1}^{\infty}\frac{a^r}{r}\delta[n-r]$$

$$X(z) = 1-bz \Leftrightarrow \hat{x}[n] = -\sum_{r=1}^{\infty}\frac{b^r}{r}\delta[n+r]$$

Complex Cepstrum Without Phase Unwrapping

- short-time analysis uses finite-length windowed segments, $x[n]$

$$X(z)=\sum_{n=0}^{M} x[n]\,z^{-n}, \quad M^{th}\text{-order polynomial}$$

- Find polynomial roots

$$X(z)=x[0]\prod_{m=1}^{M_i}(1-a_m z^{-1})\prod_{m=1}^{M_0}(1-b_m^{-1}z^{-1})$$

- a_m roots are inside unit circle (minimum-phase part)
- b_m roots are outside unit circle (maximum-phase part)
- Factor out terms of form $-b_m^{-1}z^{-1}$ giving:

$$X(z)=Az^{-M_0}\prod_{m=1}^{M_i}(1-a_m z^{-1})\prod_{m=1}^{M_0}(1-b_m z)$$

$$A=x[0](-1)^{M_0}\prod_{m=1}^{M_0} b_m^{-1}$$

- Use polynomial root finder to find the zeros that lie inside and outside the unit circle and solve directly for $\hat{x}[n]$.

Cepstrum for Minimum Phase Signals

- for minimum phase signals (no poles or zeros outside unit circle) the complex cepstrum can be completely represented by the real part of the Fourier transforms
- this means we can represent the complex cepstrum of minimum phase signals by the log of the magnitude of the FT alone
- since the real part of the FT is the FT of the even part of the sequence

$$\mathrm{Re}\left[\hat{X}(e^{j\omega})\right] = FT\left[\frac{\hat{x}(n)+\hat{x}(-n)}{2}\right]$$

$$FT\left[c(n)\right] = \log\left|X(e^{j\omega})\right|$$

$$c(n) = \frac{\hat{x}(n)+\hat{x}(-n)}{2}$$

- giving

$$\begin{aligned} \hat{x}(n) &= 0 & n<0 \\ &= c(n) & n=0 \\ &= 2c(n) & n>0 \end{aligned}$$

- thus the complex cepstrum (for minimum phase signals) can be computed by computing the cepstrum and using the equation above

Recursive Relation for Complex Cepstrum for Minimum Phase Signals

- the complex cepstrum for minimum phase signals can be computed recursively from the input signal, $x(n)$ using the relation

$$
\begin{aligned}
\hat{x}(n) &= 0 && n < 0 \\
&= \log\left[x(0)\right] && n = 0 \\
&= \frac{x(n)}{x(0)} - \sum_{k=0}^{n-1} \left(\frac{k}{n}\right) \hat{x}(k) \frac{x(n-k)}{x(0)} && n > 0
\end{aligned}
$$

Recursive Relation for Complex Cepstrum for Minimum Phase Signals

$x(n) \longleftrightarrow X(z)$ — 1. basic z-transform

$nx(n) \longleftrightarrow -z\dfrac{dX(z)}{dz} = -zX'(z)$ — 2. scale by n rule

$\hat{x}(n) \longleftrightarrow \hat{X}(z) = \log\left[X(z)\right]$ — 3. definition of complex cepstrum

$\dfrac{d\hat{X}(z)}{dz} = \dfrac{d}{dz}\left[\log[X(z)]\right] = \dfrac{X'(z)}{X(z)}$ — 4. differentiation of z-transform

$-z\dfrac{d\hat{X}(z)}{dz}X(z) = -zX'(z)$ — 5. multiply both sides of equation

Recursive Relation for Complex Cepstrum for Minimum Phase Signals

$$n\hat{x}(n) * x(n) \longleftrightarrow -z\frac{d\hat{X}(z)}{dz}X(z) = -zX'(z) \longleftrightarrow nx(n)$$

$$nx(n) = \sum_{k=-\infty}^{\infty} \hat{x}(k)x(n-k)(k)$$

- for minimum phase systems we have $\hat{x}(n) = 0$ for $n < 0$,

$x(n) = 0$ for $n < 0$, giving:

$$x(n) = \sum_{k=0}^{n} \hat{x}(k)x(n-k)\left(\frac{k}{n}\right)$$

- separating out the term for $k = n$ we get:

$$x(n) = \sum_{k=0}^{n-1} \hat{x}(k)x(n-k)\left(\frac{k}{n}\right) + x(0)\hat{x}(n)$$

$$\hat{x}(n) = \frac{x(n)}{x(0)} - \sum_{k=0}^{n-1} \hat{x}(k)\frac{x(n-k)}{x(0)}\left(\frac{k}{n}\right), \qquad n > 0$$

$$\hat{x}(0) = \log\left[x(0)\right], \quad \hat{x}(n) = 0, \; n < 0$$

Recursive Relation for Complex Cepstrum for Minimum Phase Signals

- why is $\hat{x}(0) = \log[x(0)]$?
- assume we have a finite sequence $x(n)$, $n = 0,1,\ldots,N-1$
- we can write $x(n)$ as:

$$x(n) = x(0)\delta(n) + x(1)\delta(n-1) + \ldots + x(N-1)\delta(N-1)$$

$$= x(0)\left[\delta(n) + \frac{x(1)}{x(0)}\delta(n-1) + \ldots + \frac{x(N-1)}{x(0)}\delta(N-1)\right]$$

- taking $z-$transforms, we get:

$$X(z) = \sum_{n=0}^{N-1} x(n) z^{-n} = G\prod_{k=1}^{NZ1}(1-a_k z^{-1})\prod_{k=1}^{NZ2}(1-b_k z)$$

- where the first term is the gain, $G = x(0)$, and the two product terms are the zeros inside and outside the unit circle.
- for minimum phase systems we have all zeros inside the unit circle so the second product term is gone, and we have the result that

$$\hat{x}(0) = \log[G] = \log[x(0)]; \quad \hat{x}(n) = 0,\ n < 0$$

$$\hat{x}(n) = -\sum_{k=1}^{NZ1}\left(\frac{a_k^n}{n}\right) \qquad n > 0$$

Cepstrum for Maximum Phase Signals

- for maximum phase signals (no poles or zeros inside unit circle)

$$c(n) = \frac{\hat{x}(n) + \hat{x}(-n)}{2}$$

- giving

$$\begin{aligned} \hat{x}(n) &= 0 && n > 0 \\ &= c(n) && n = 0 \\ &= 2c(n) && n < 0 \end{aligned}$$

- thus the complex cepstrum (for maximum phase signals) can be computed by computing the cepstrum and using the equation above

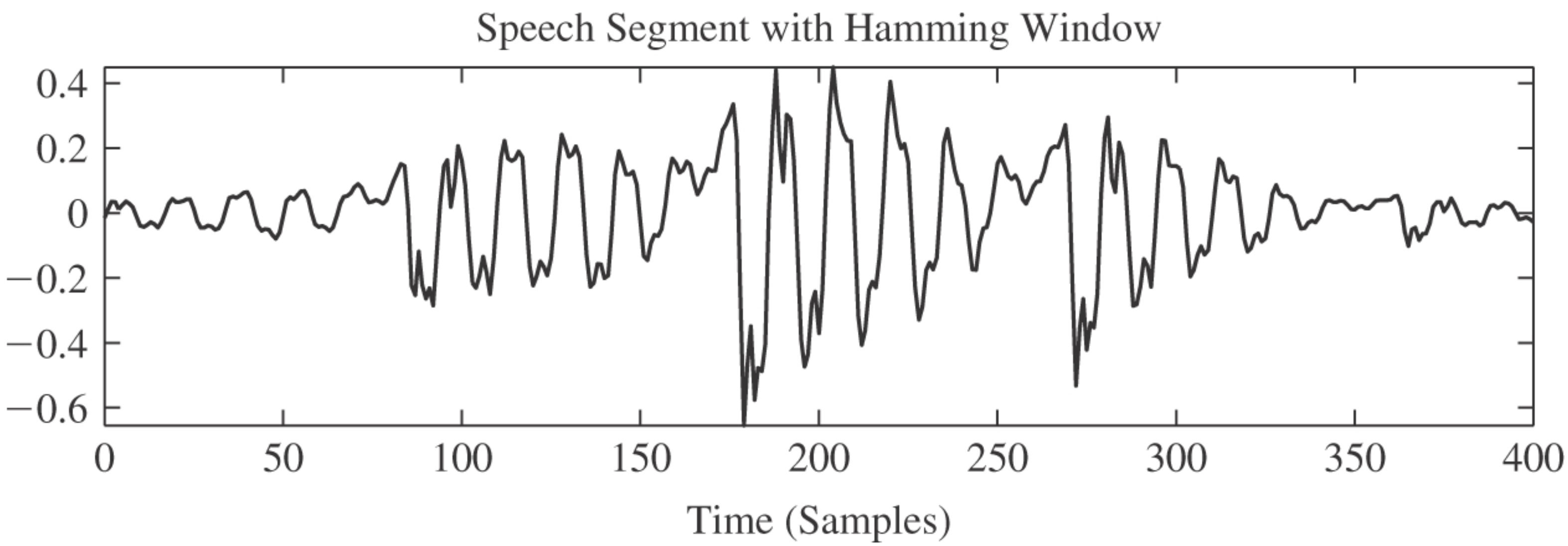
Recursive Relation for Complex Cepstrum for Maximum Phase Signals

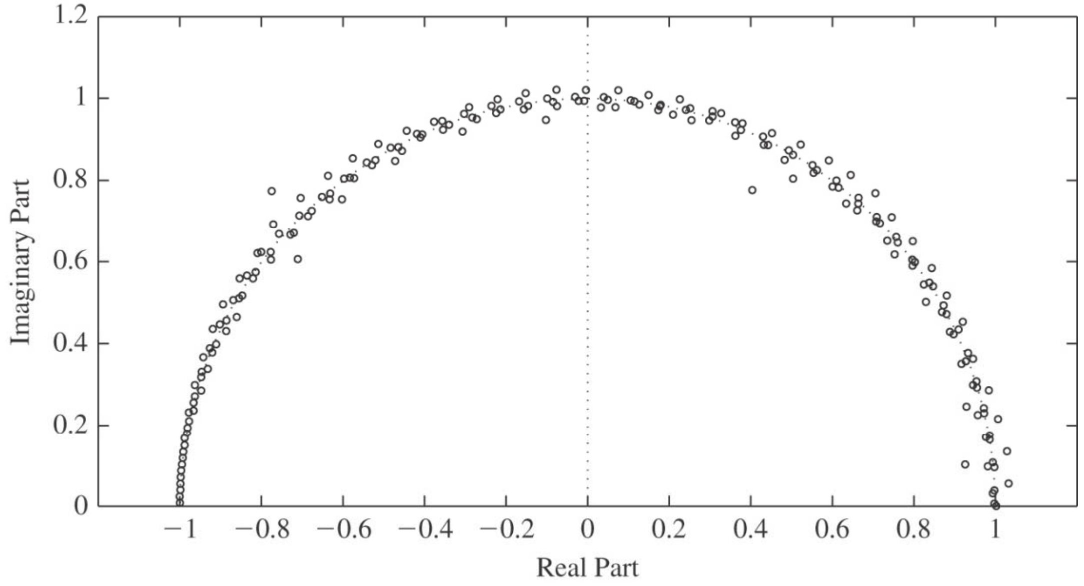
- the complex cepstrum for maximum phase signals can be computed recursively from the input signal, $x(n)$ using the relation

$$
\begin{aligned}
\hat{x}(n) &= 0 && n > 0 \\
&= \log\left[x(0)\right] && n = 0 \\
&= \frac{x(n)}{x(0)} - \sum_{k=n+1}^{0} \left(\frac{k}{n}\right) \hat{x}(k) \frac{x(n-k)}{x(0)} && n < 0
\end{aligned}
$$

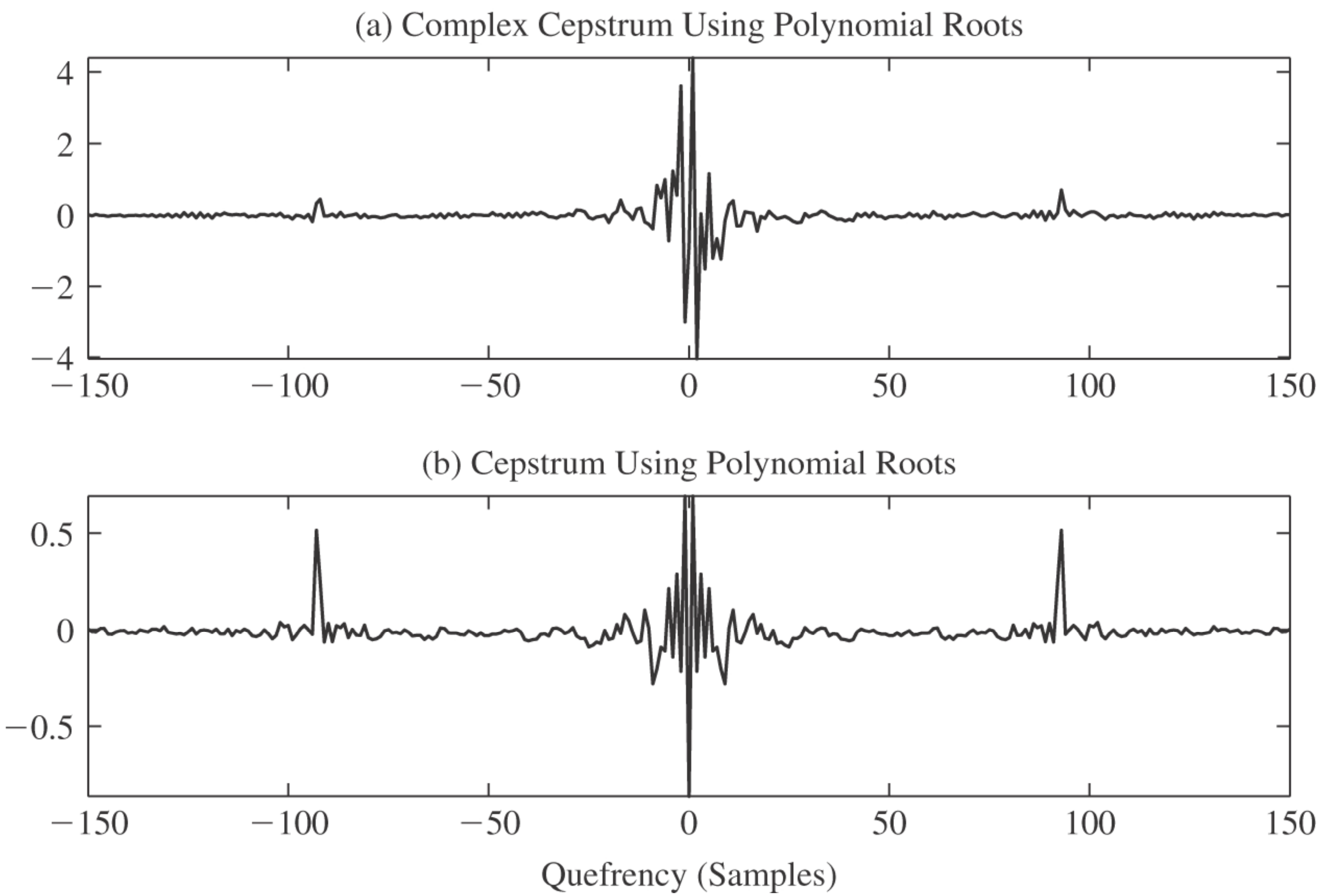
Computing Short-Time Cepstrums from Speech Using Polynomial Roots

Cepstrum From Polynomial Roots





Cepstrum From Polynomial Roots

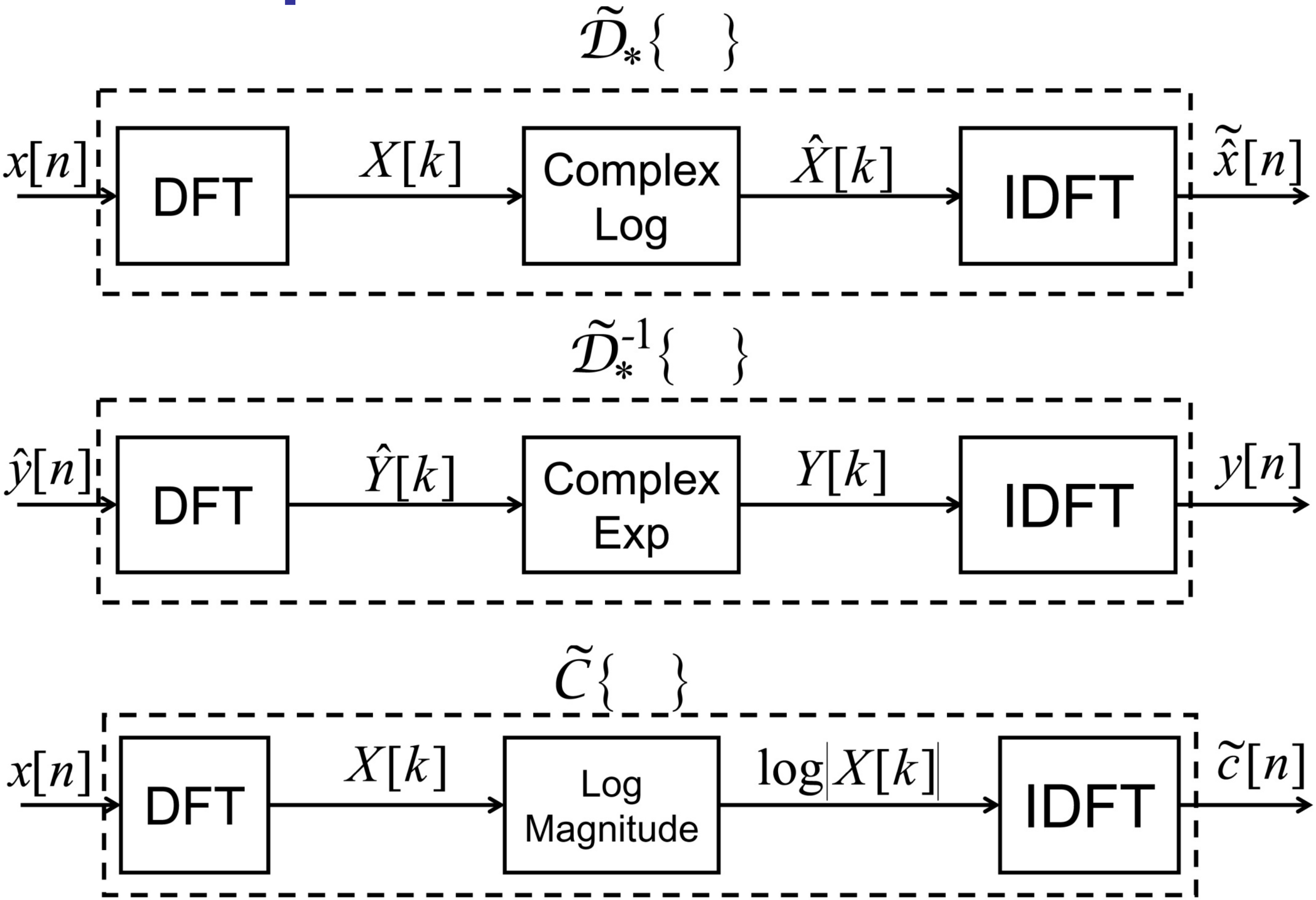


Computing Short-Time Cepstrums from Speech Using the DFT

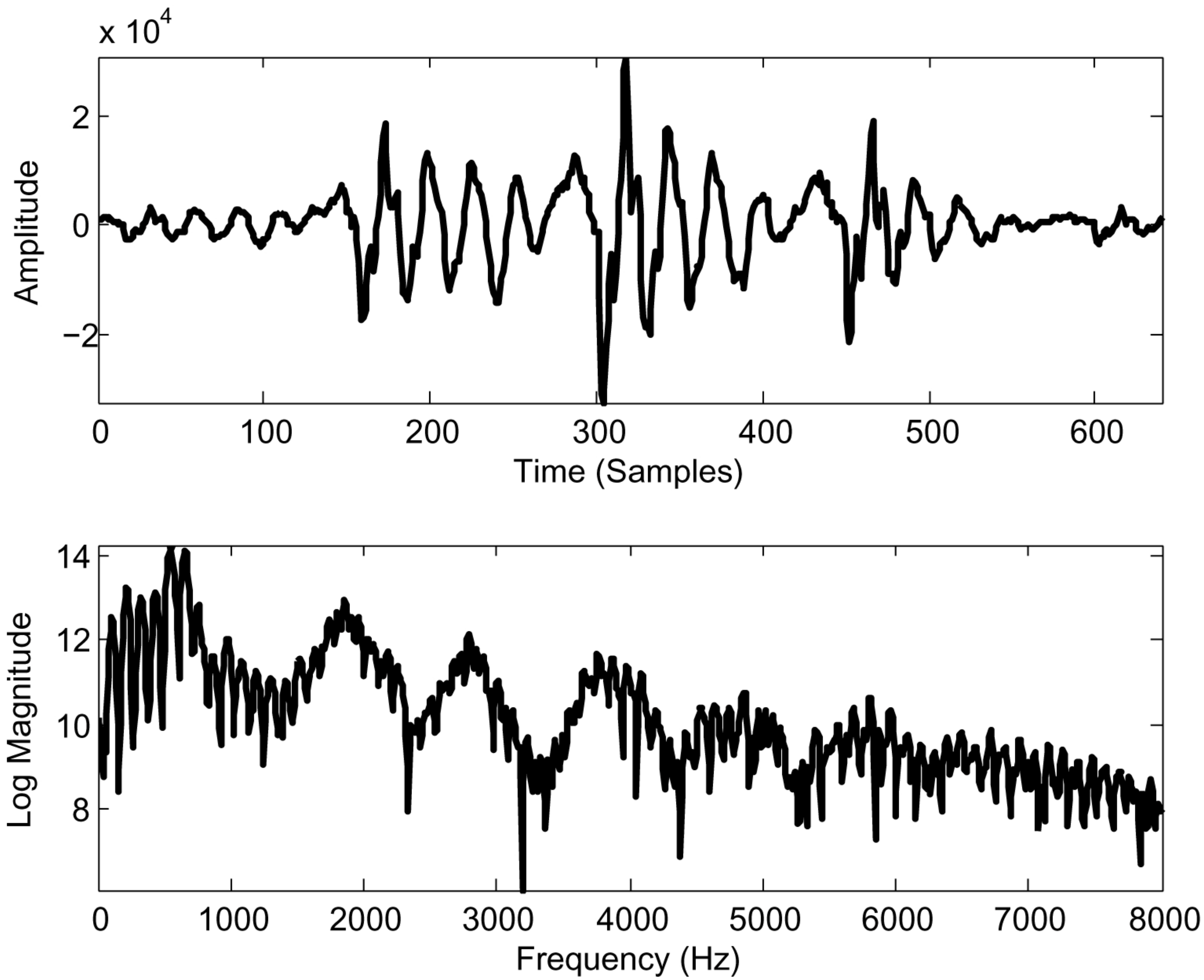
Practical Considerations

- window to define short-time analysis
- window duration (should be several pitch periods long)
- size of FFT (to minimize aliasing)
- elimination of linear phase components (positioning signals within frames)
- cutoff quefrency of lifter
- type of lifter (low/high quefrency)

Computational Considerations



Voiced Speech Example

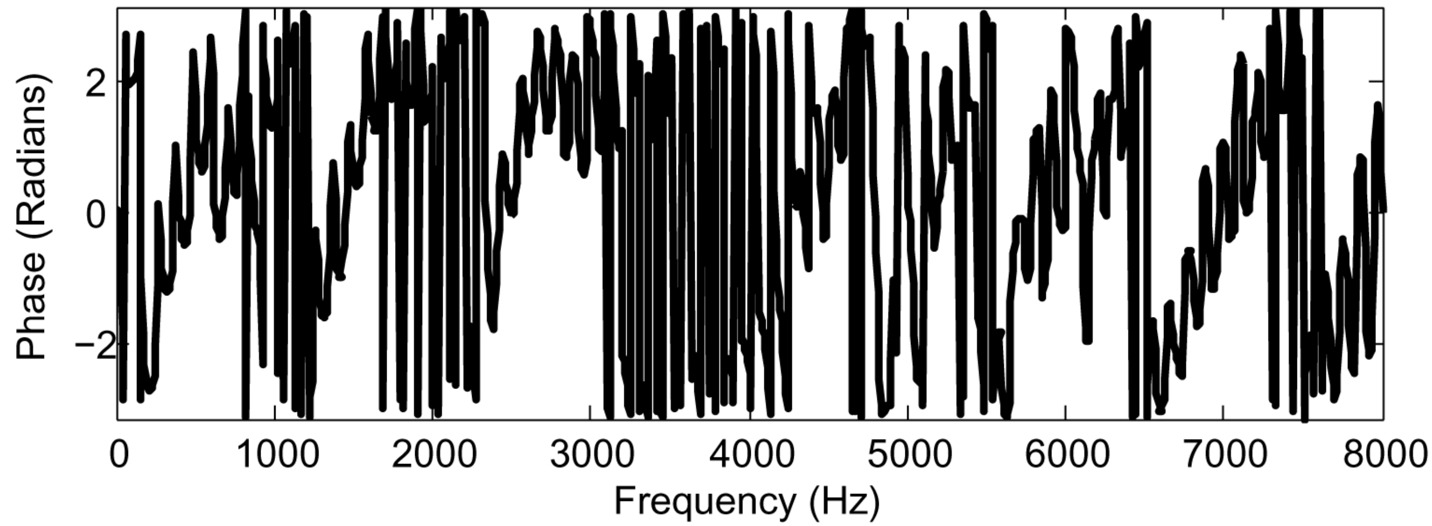


Hamming window

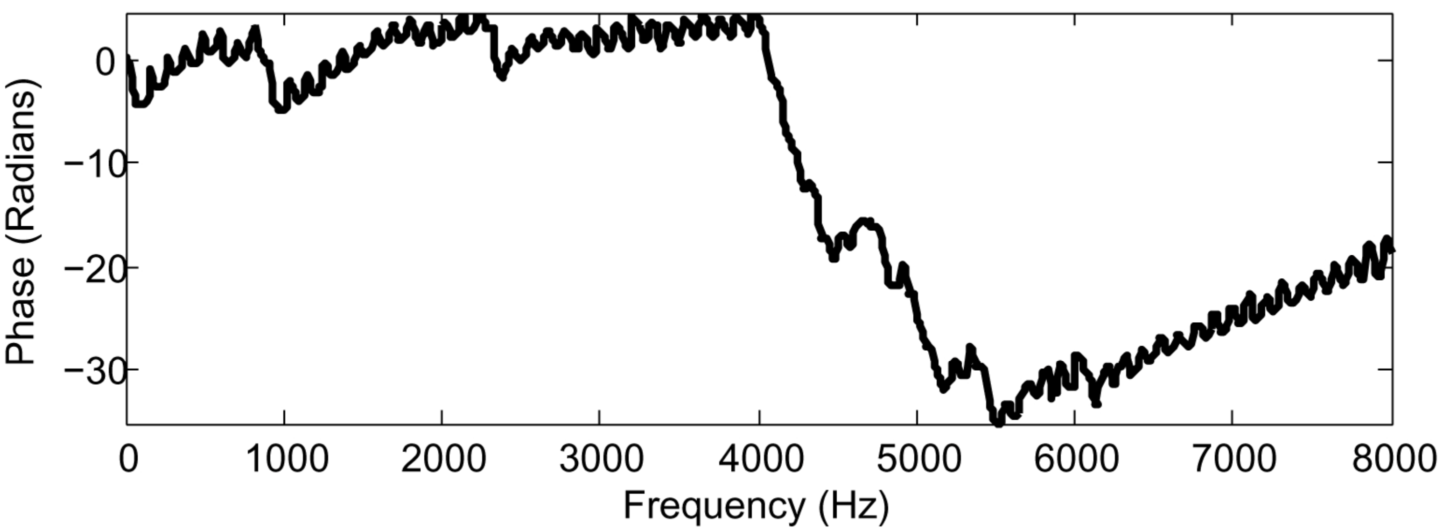
40 msec duration

(section beginning at sample 13000 in file test_16k.wav)

Voiced Speech Example

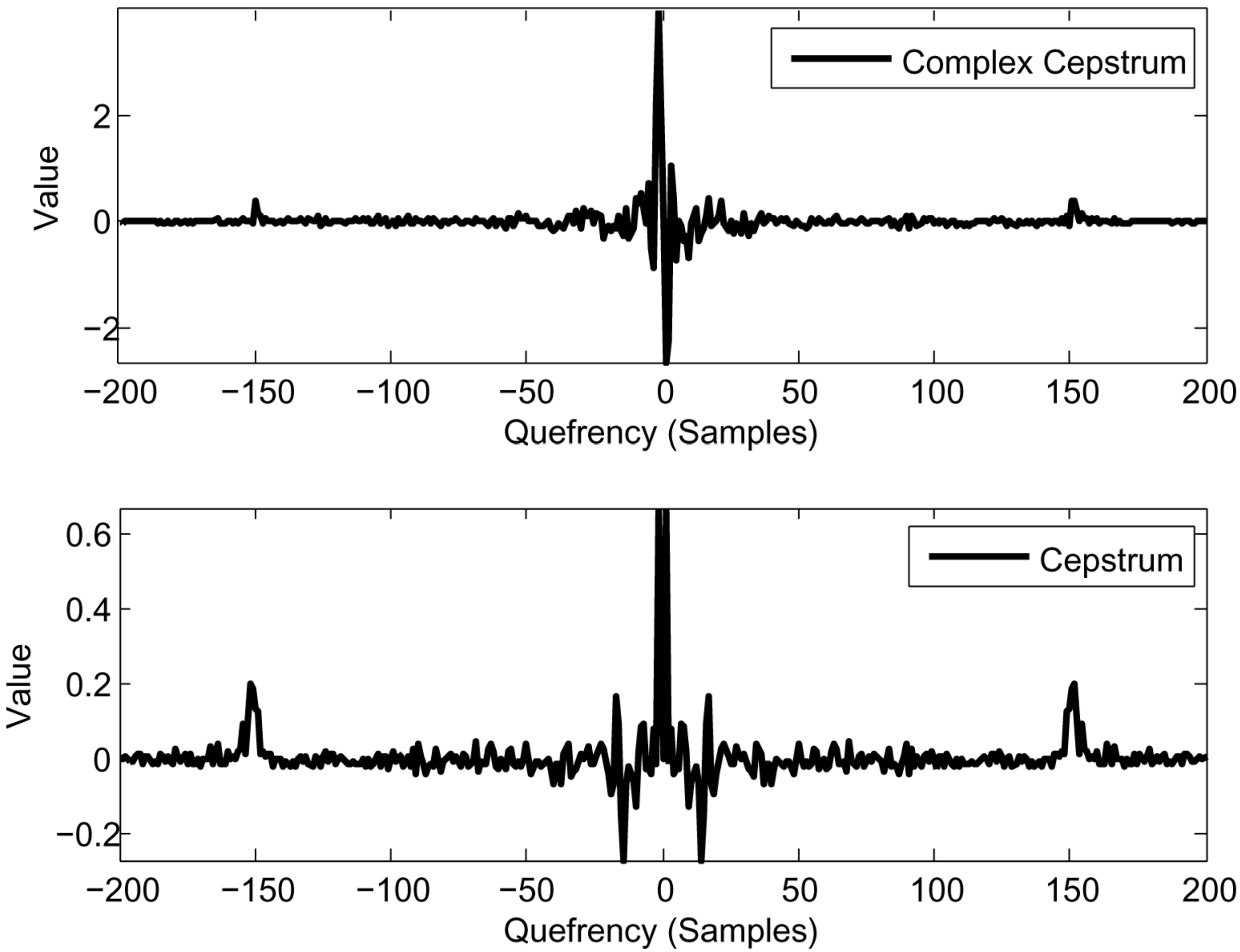


wrapped phase



unwrapped phase

Voiced Speech Example



Characteristic System for Homomorphic Convolution

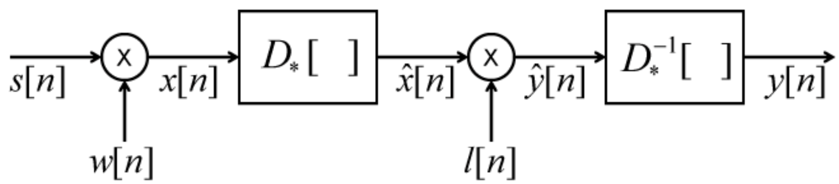
- still need to define (and design) the *L* operator part (the linear system component) of the system to completely define the characteristic system for homomorphic convolution for speech
  - to do this properly and correctly, need to look at the properties of the complex cepstrum for speech signals

Complex Cepstrum of Speech

- model of speech:
  - voiced speech produced by a quasi-periodic pulse train exciting slowly time-varying linear system => $p[n]$ convolved with $h_v[n]$
  - unvoiced speech produced by random noise exciting slowly time-varying linear system => $u[n]$ convolved with $h_v[n]$
- time to examine full model and see what the complex cepstrum of speech looks like

Homomorphic Filtering of Voiced Speech

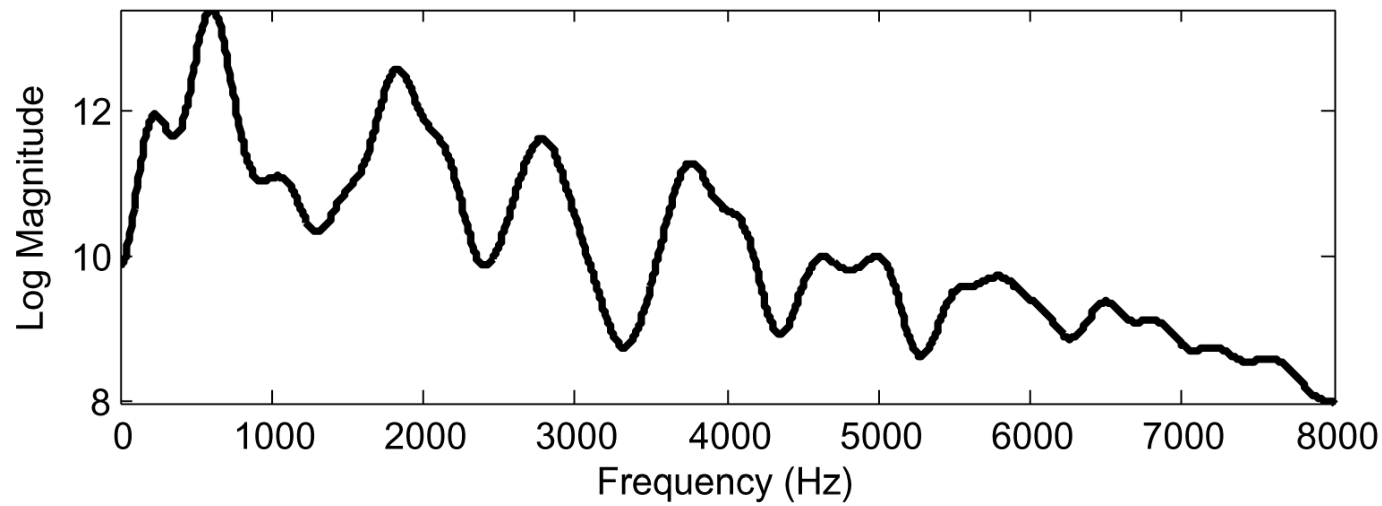
- goal is to separate out the excitation impulses from the remaining components of the complex cepstrum
- use cepstral window, *l(n),* to separate excitation pulses from combined vocal tract
  - $l(n)=1$ for $|n|<n_0<N_p$
  - $l(n)=0$ for $|n|\geq n_0$
  - this window removes excitation pulses
  - $l(n)=0$ for $|n|<n_0<N_p$
  - $l(n)=1$ for $|n|\geq n_0$
  - this window removes combined vocal tract
- the filtered signal is processed by the inverse characteristic system to recover the combined vocal tract component



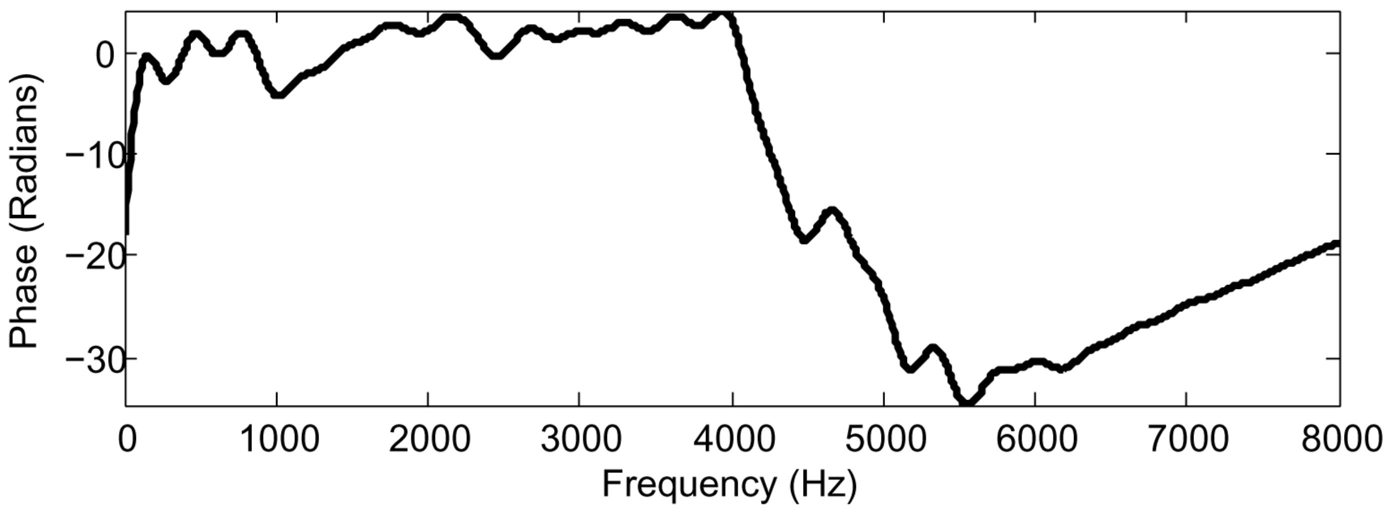
$$\hat{y}(n) = \ell(n) \cdot \hat{x}(n)$$

$$\hat{Y}(e^{j\omega}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{X}(e^{j\theta}) L(e^{j(\omega-\theta)}) d\theta$$

Voiced Speech Example

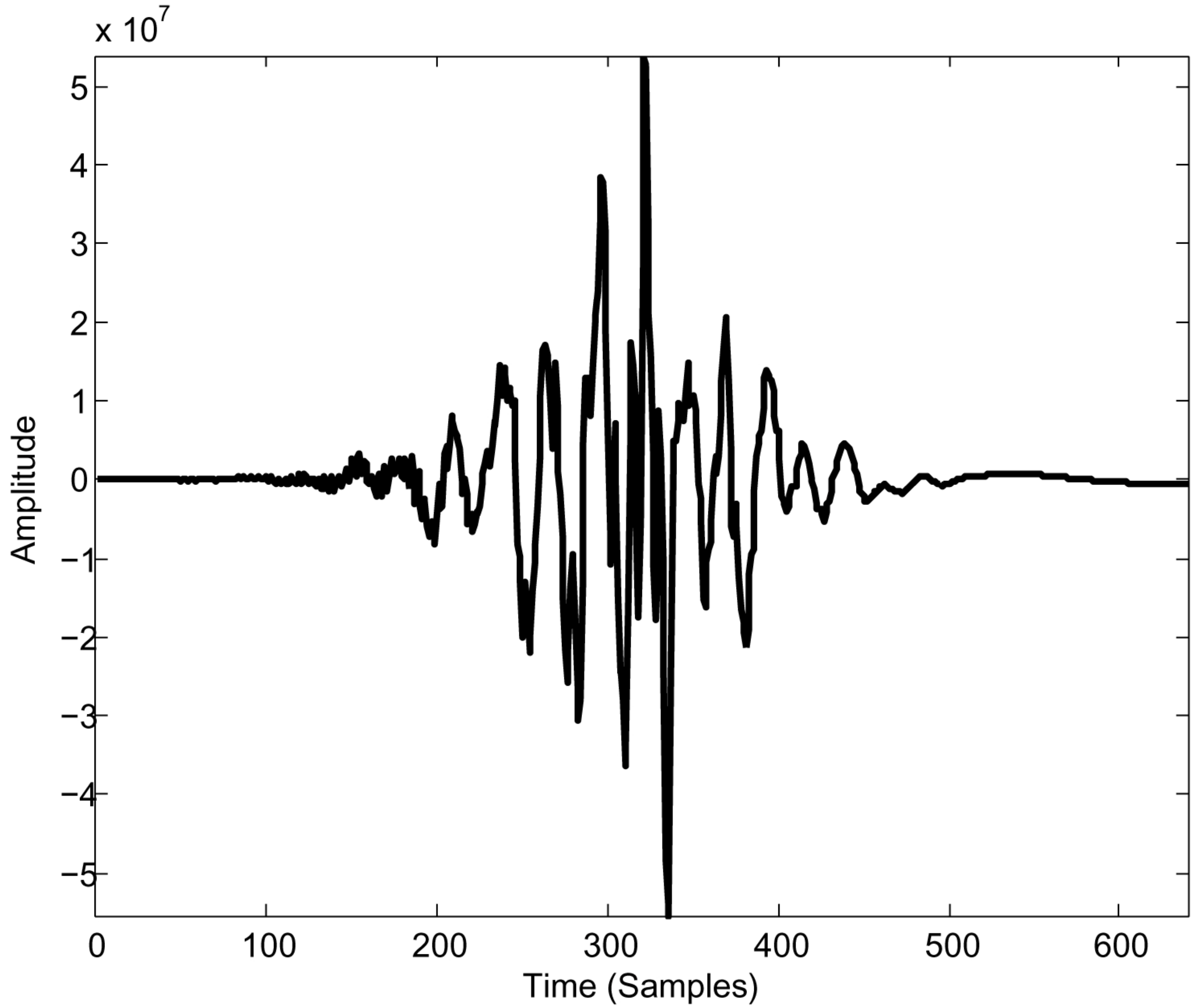


Cepstrally smoothed log magnitude, 50 quefrencies cutoff



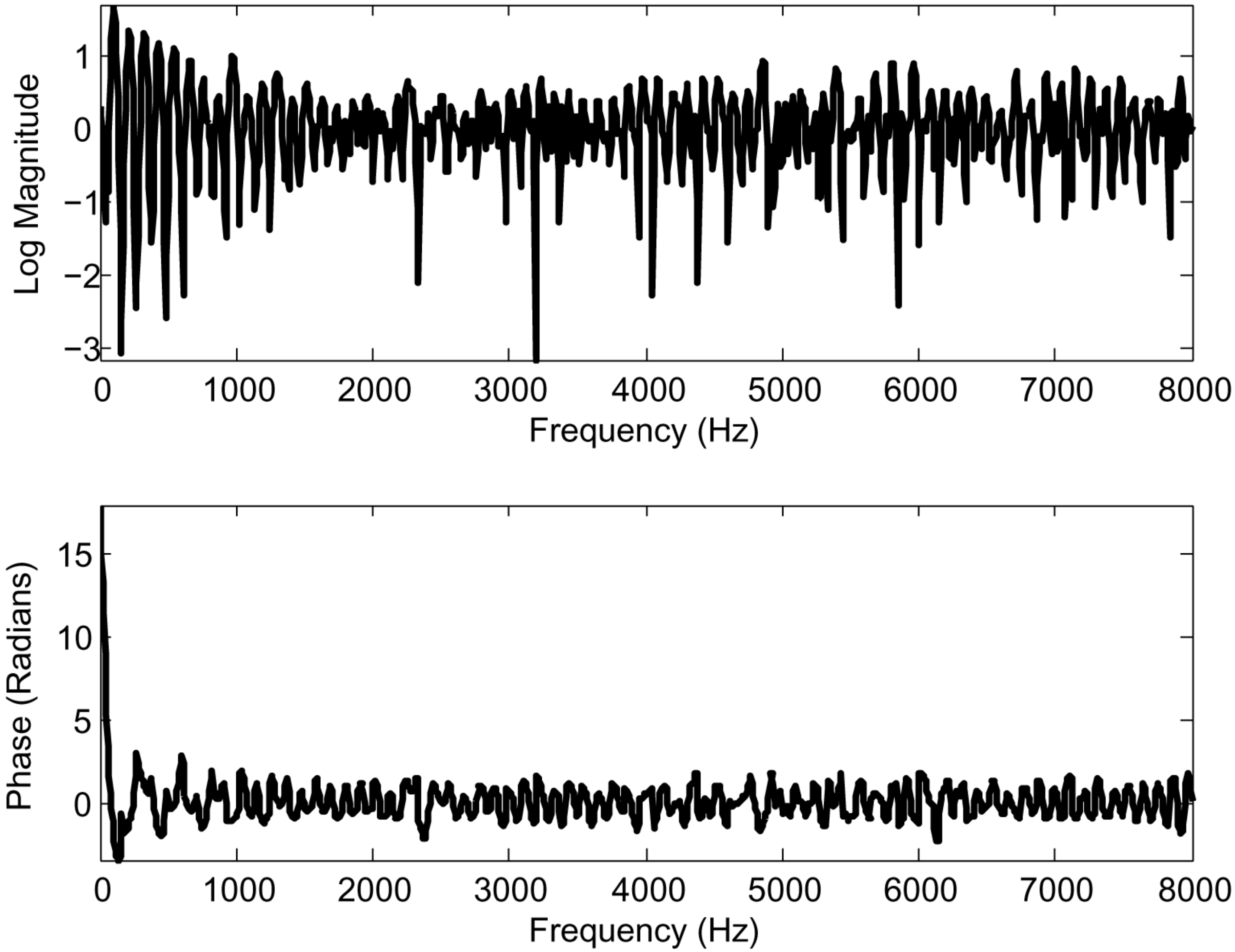
Cepstrally unwrapped phase, 50 quefrencies cutoff

Voiced Speech Example



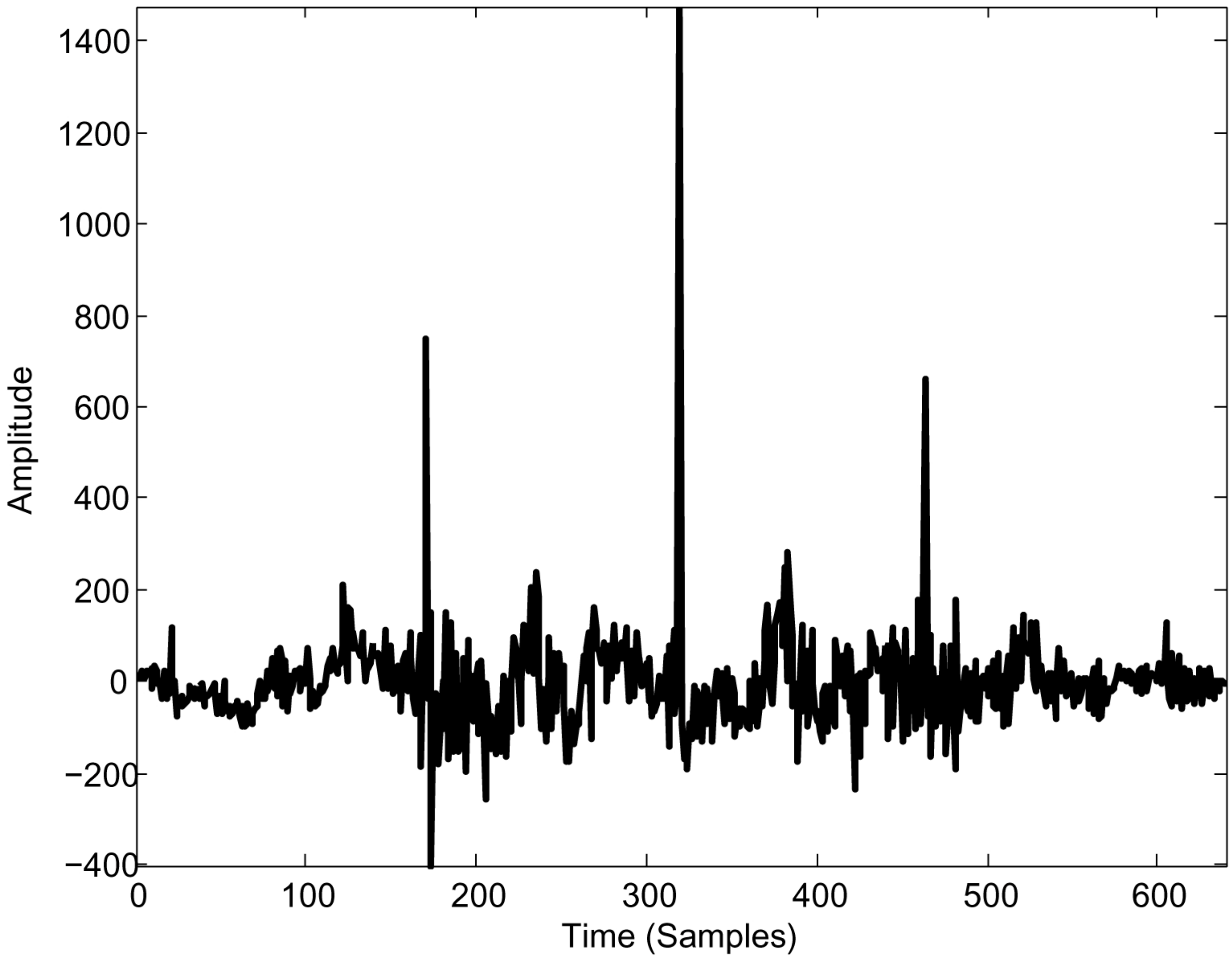
Combined impulse response of glottal pulse, vocal tract system, and radiation system

Voiced Speech Example



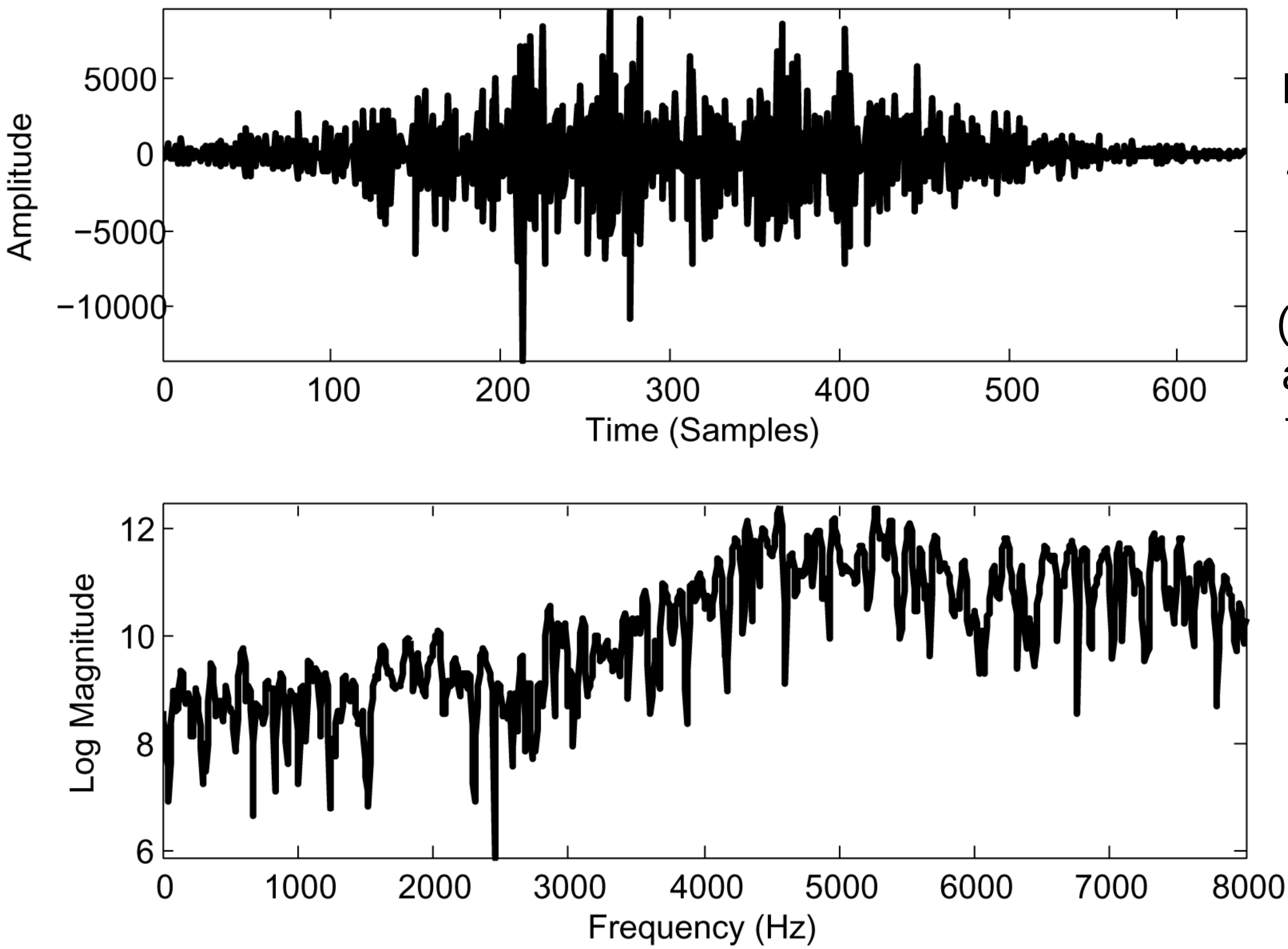
High quefrency liftering; cutoff quefrency=50; log magnitude and unwrapped phase

Voiced Speech Example



Estimated excitation function for voiced speech (Hamming window weighted)

Unvoiced Speech Example

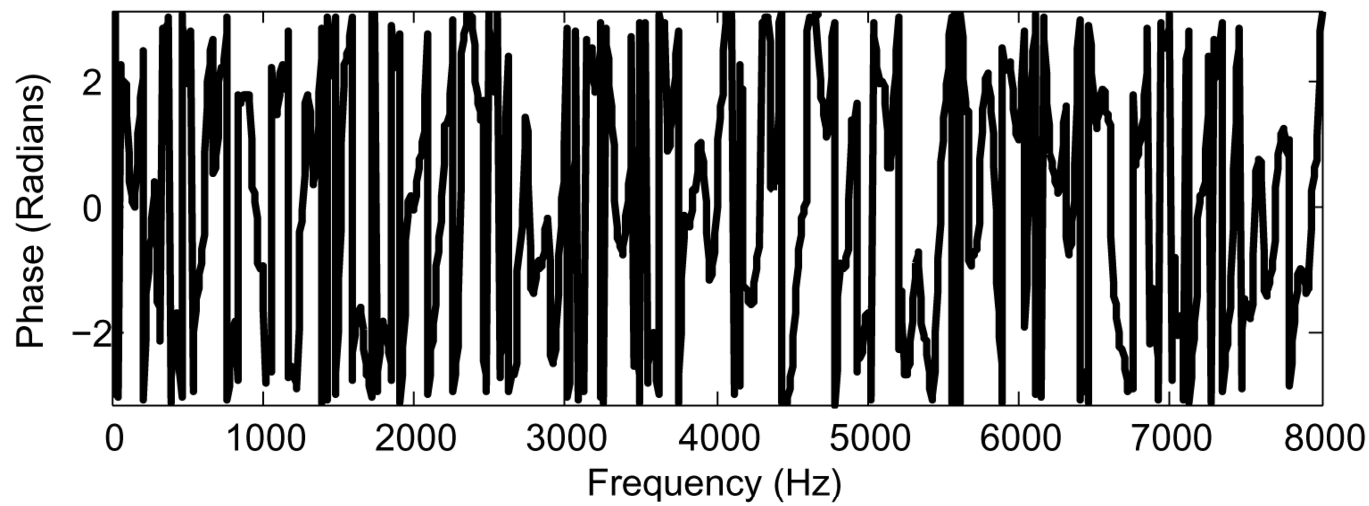


Hamming window

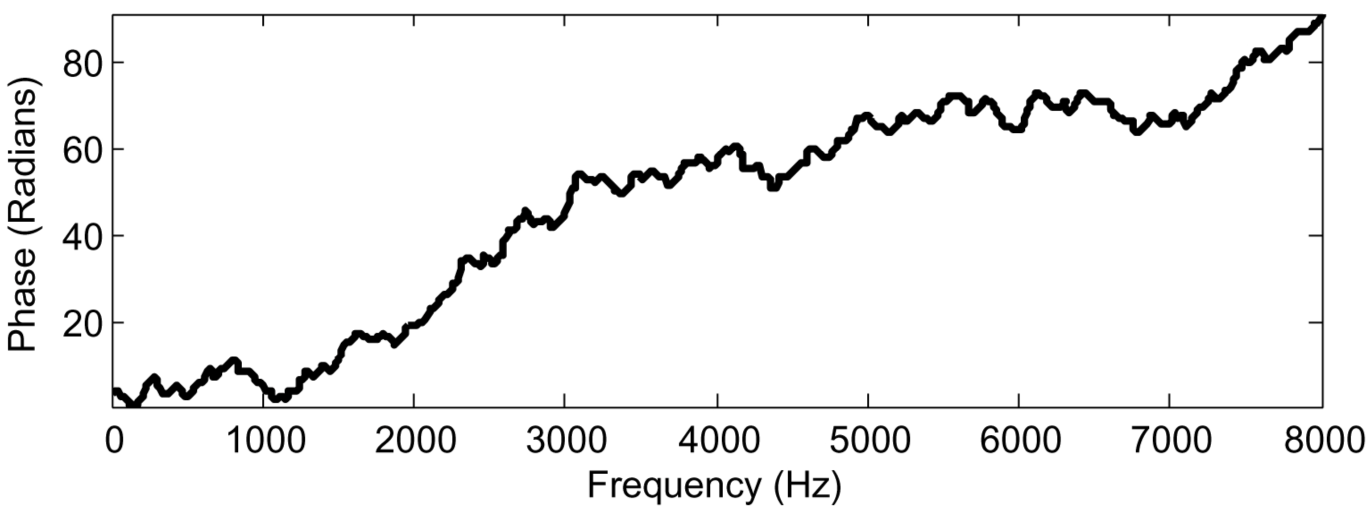
40 msec duration

(section beginning at sample 3200 in file test_16k.wav)

Unvoiced Speech Example

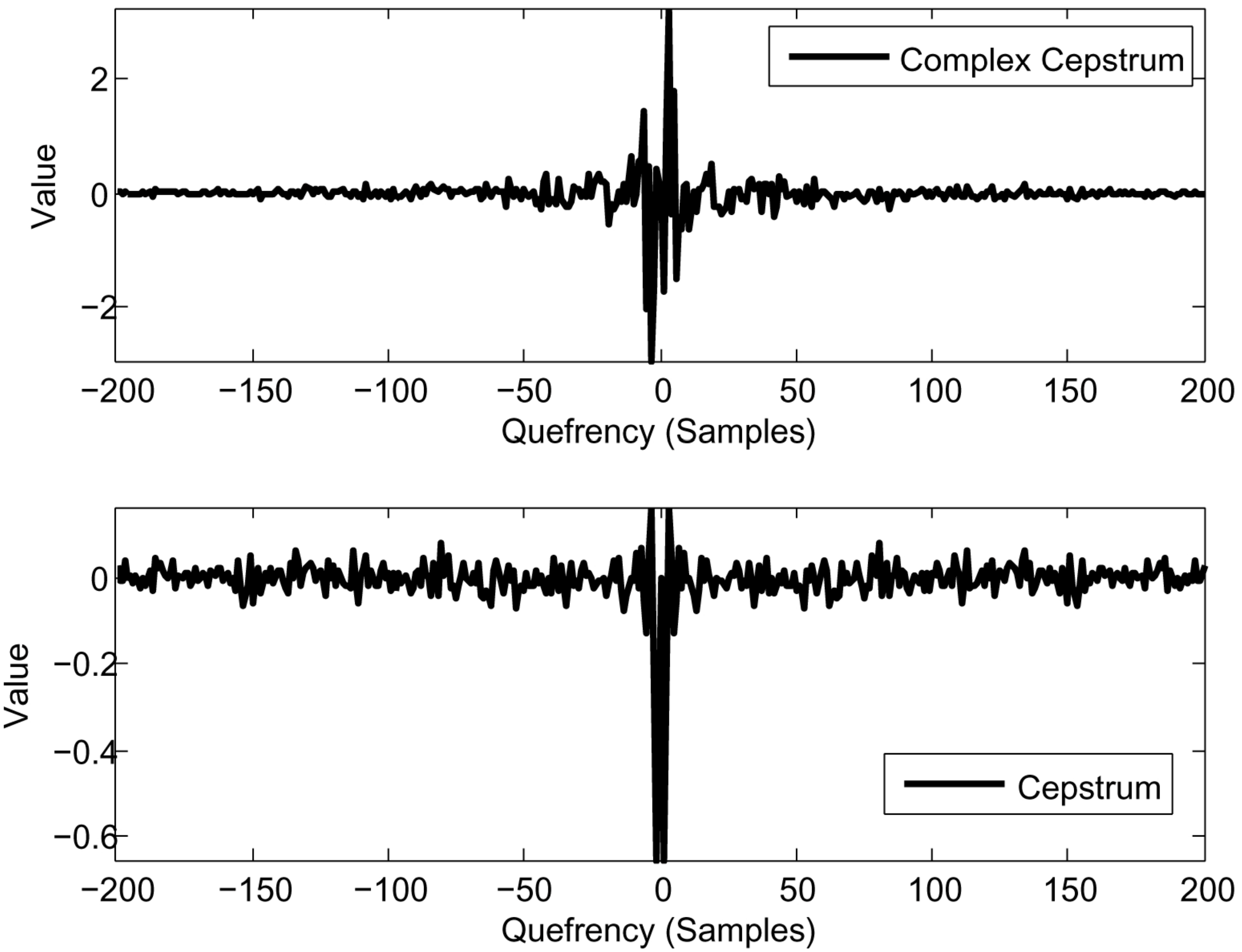


wrapped phase

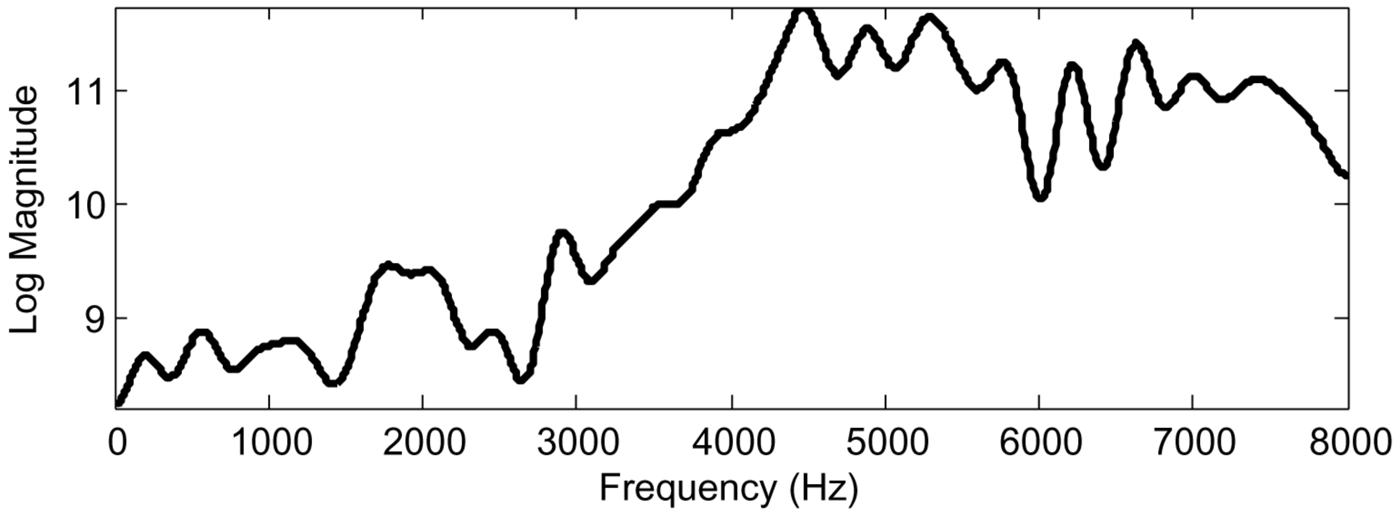


unwrapped phase

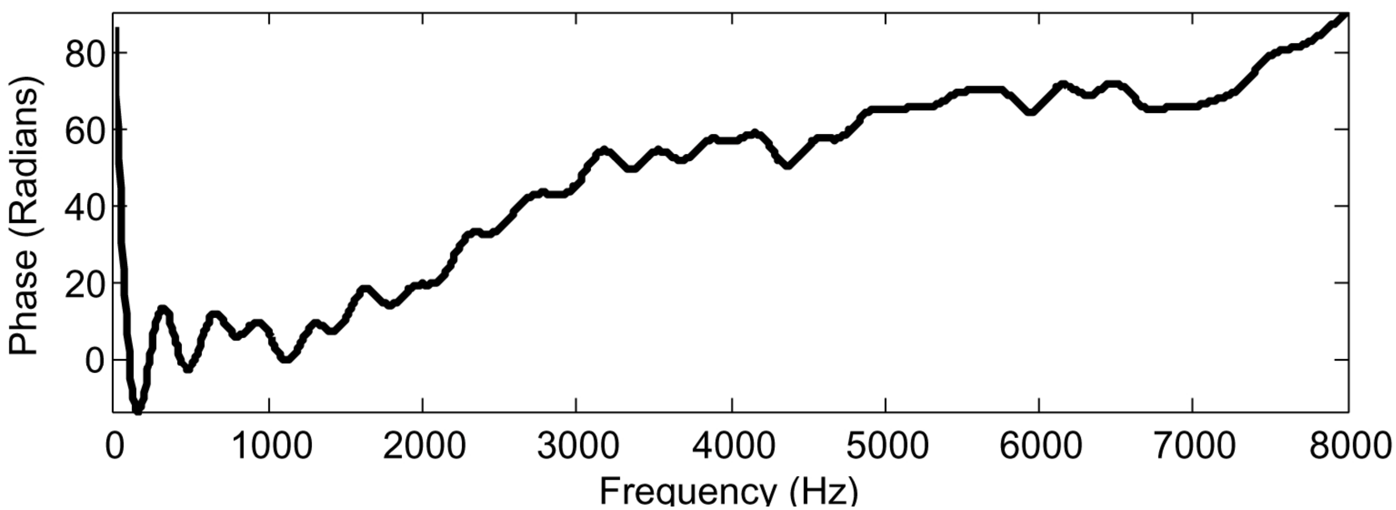
Unvoiced Speech Example



Unvoiced Speech Example

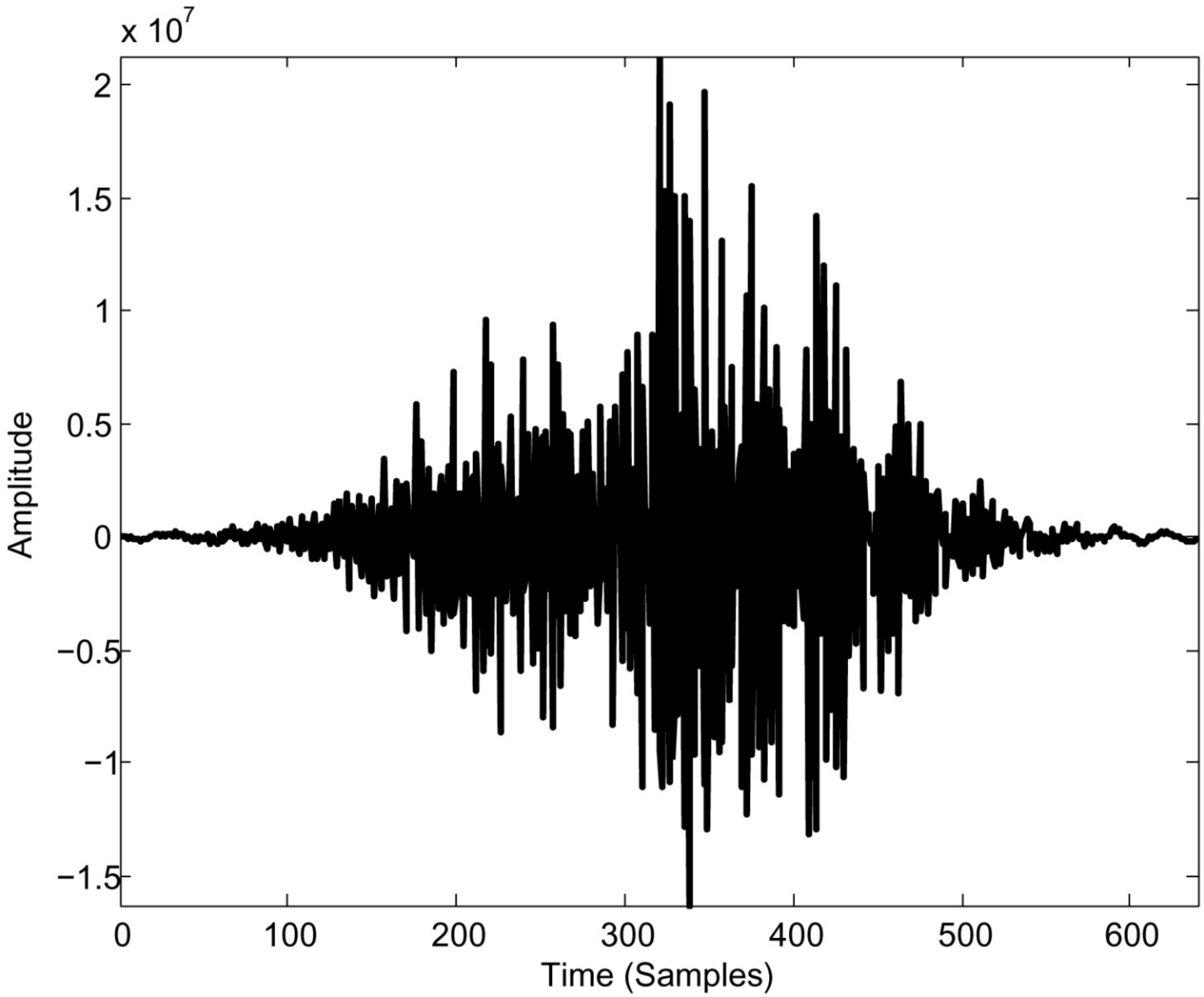


Cepstrally smoothed log magnitude, 50 quefrencies cutoff



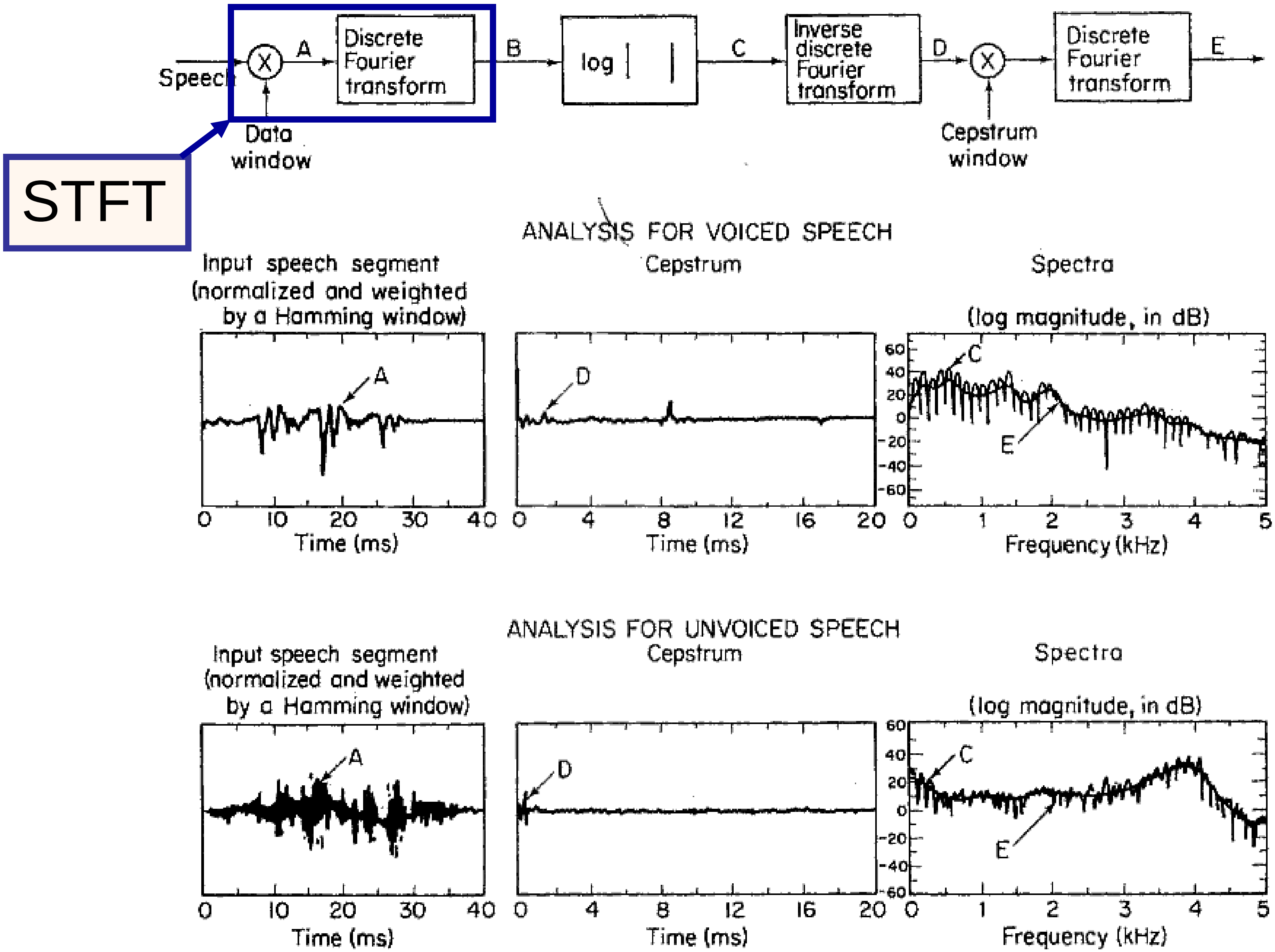
Cepstrally unwrapped phase, 50 quefrencies cutoff

Unvoiced Speech Example



Estimated excitation source for unvoiced speech section (Hamming window weighted)

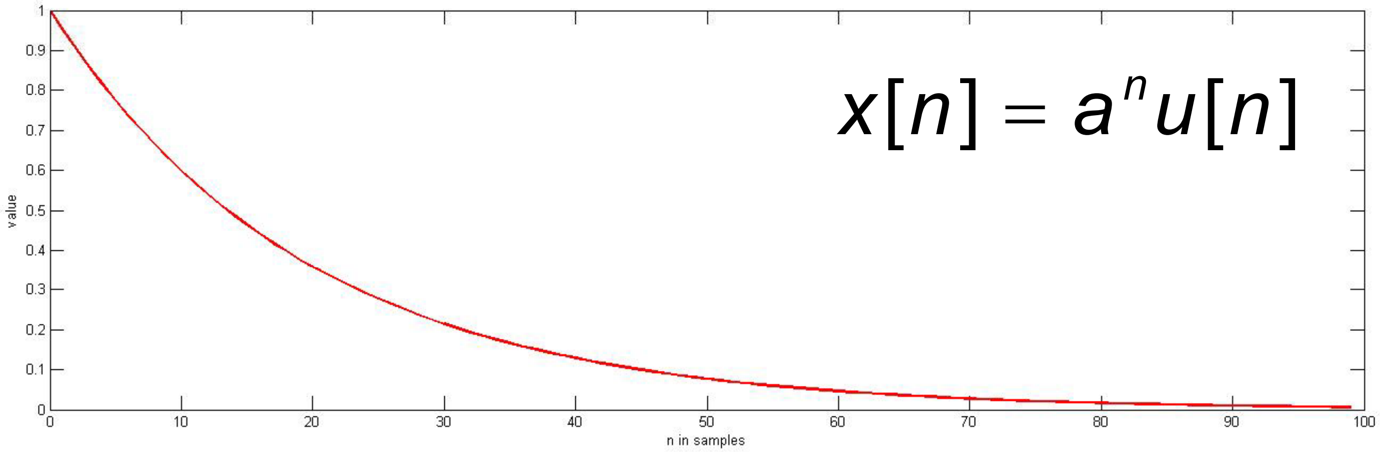
Short-Time Homomorphic Analysis

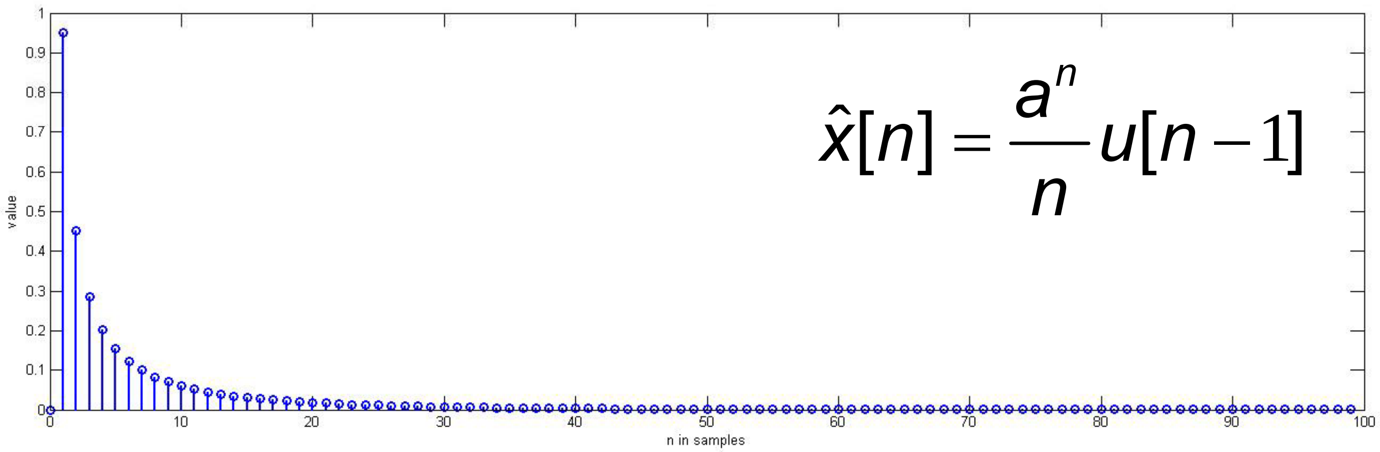


Review of Cepstral Calculation

- 3 potential methods for computing cepstral coefficients, $\hat{x}[n]$, of sequence *x[n]*
  - analytical method; assuming *X(z)* is a rational function; find poles and zeros and expand using log power series
  - recursion method; assuming *X(z)* is either a minimum phase (all poles and zeros inside unit circle) or maximum phase (all poles and zeros outside unit circle) sequence
  - DFT implementation; using windows, with phase unwrapping (for complex cepstra)

Example 1—single pole sequence (computed using all 3 methods)





Cepstral Computation Aliasing

- Effect of quefrency aliasing via a simple example

$$x[n] = \delta[n] + \alpha\delta[n - N_p]$$

- with discrete-time Fourier transform

$$X\left(e^{j\omega}\right) = 1 + \alpha e^{-j\omega N_p}$$

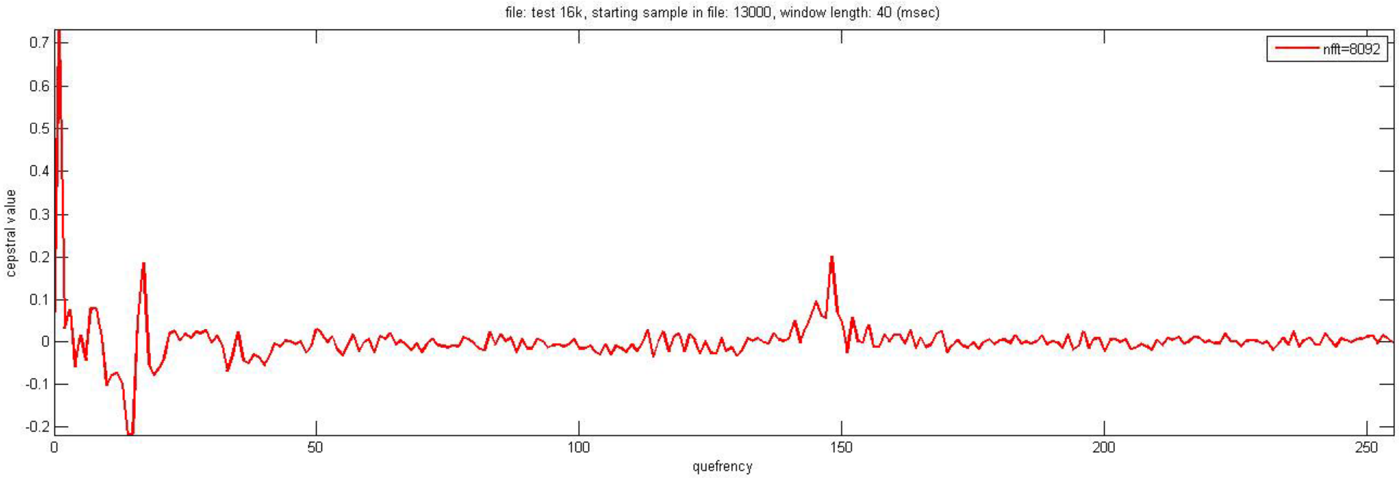
- We can express the complex logarithm as

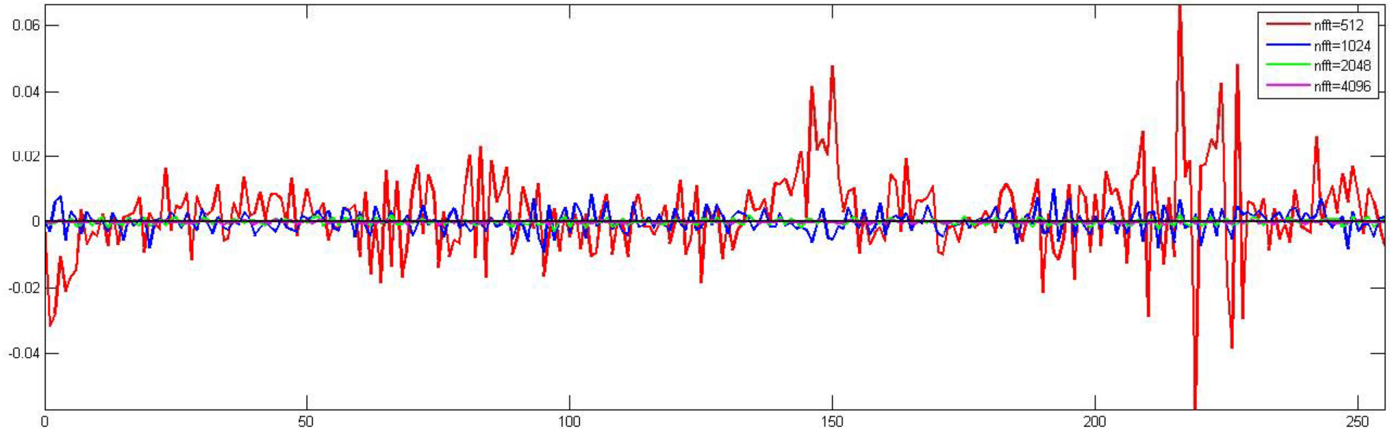
$$\hat{X}(e^{j\omega}) = \log\{1 + \alpha e^{-j\omega N_p}\} = \sum_{m=1}^{\infty}\left(\frac{(-1)^{m+1}\alpha^m}{m}\right)e^{-j\omega m N_p}$$

- giving a complex cepstrum in the form

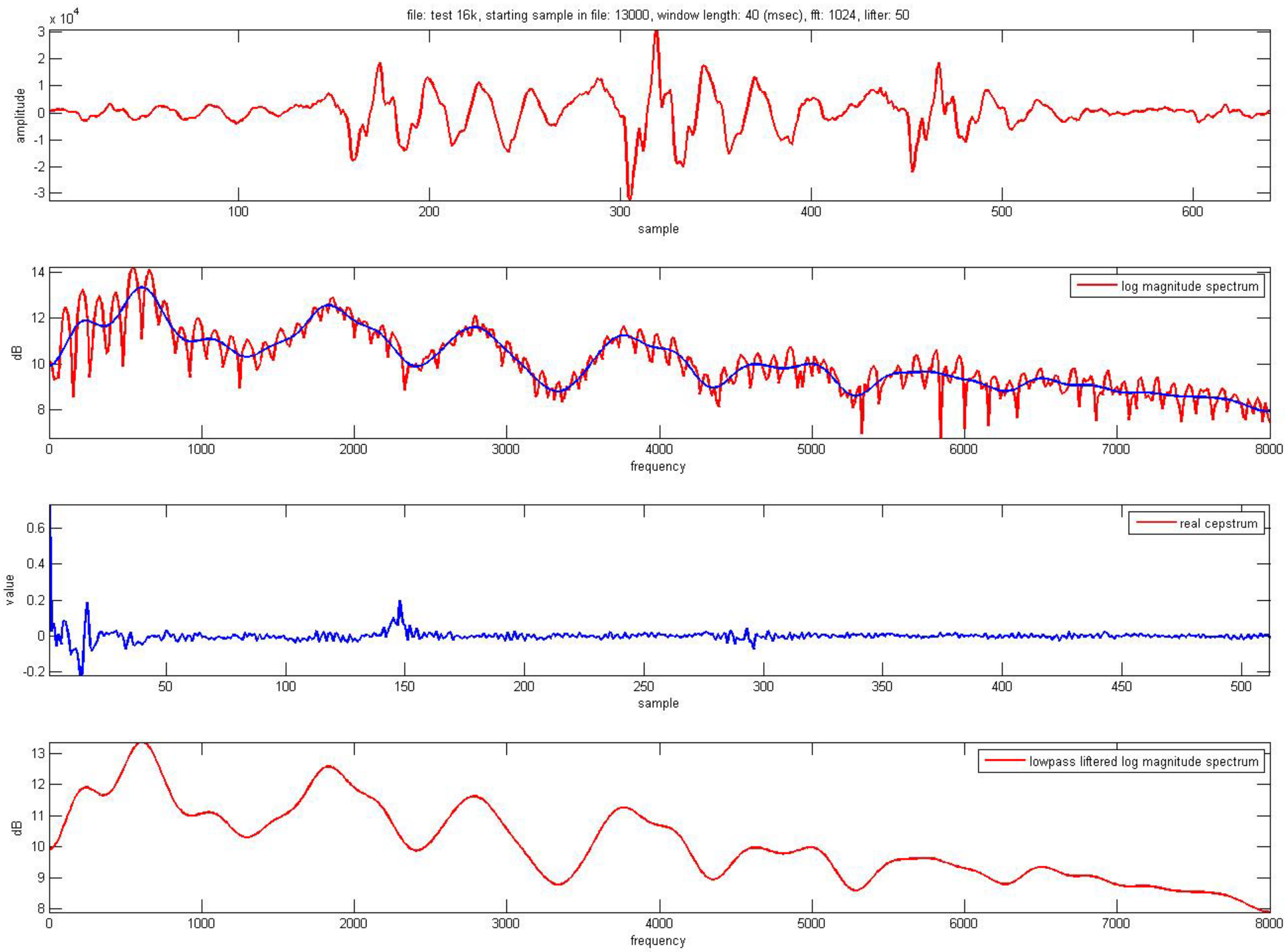
$$\hat{x}[n] = \sum_{m=1}^{\infty}\left(\frac{(-1)^{m+1}\alpha^m}{m}\right)\delta[n - mN_p]$$

Example 2—voiced speech frame

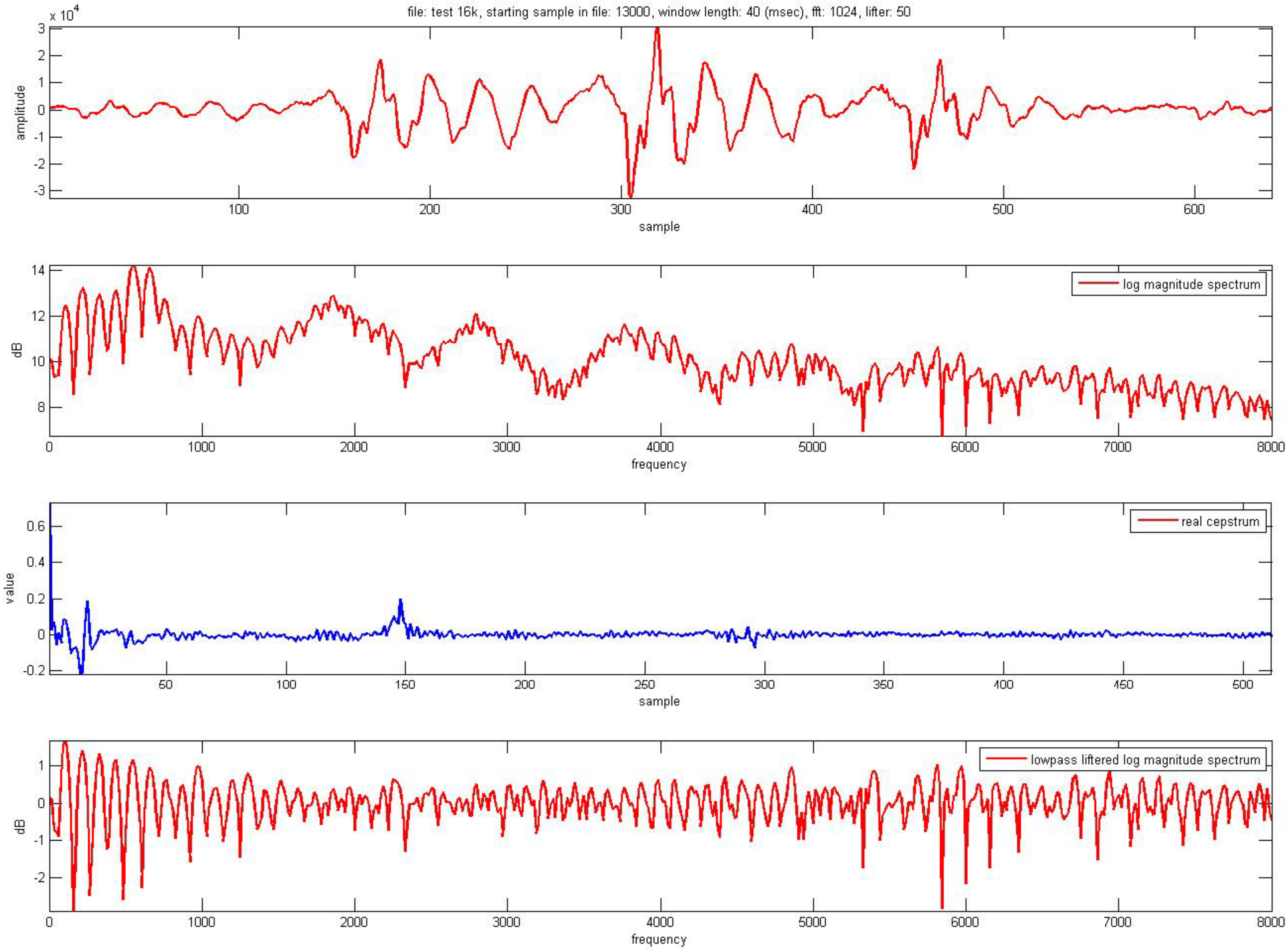




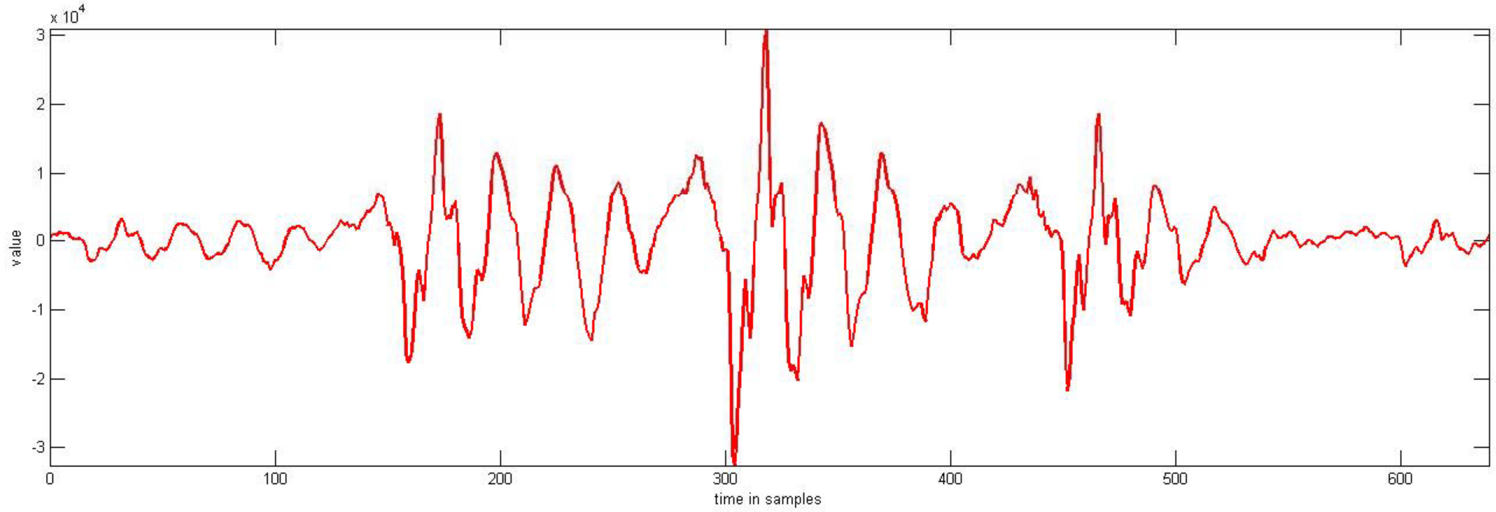
Example 3—low quefrency liftering

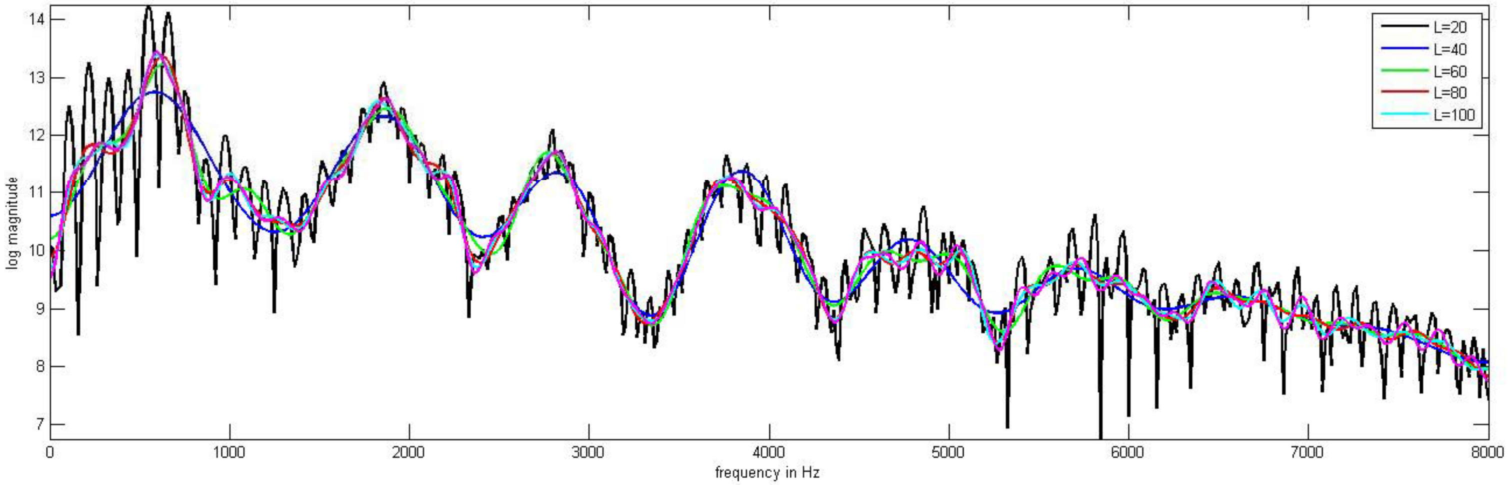


Example 3—high quefrency liftering

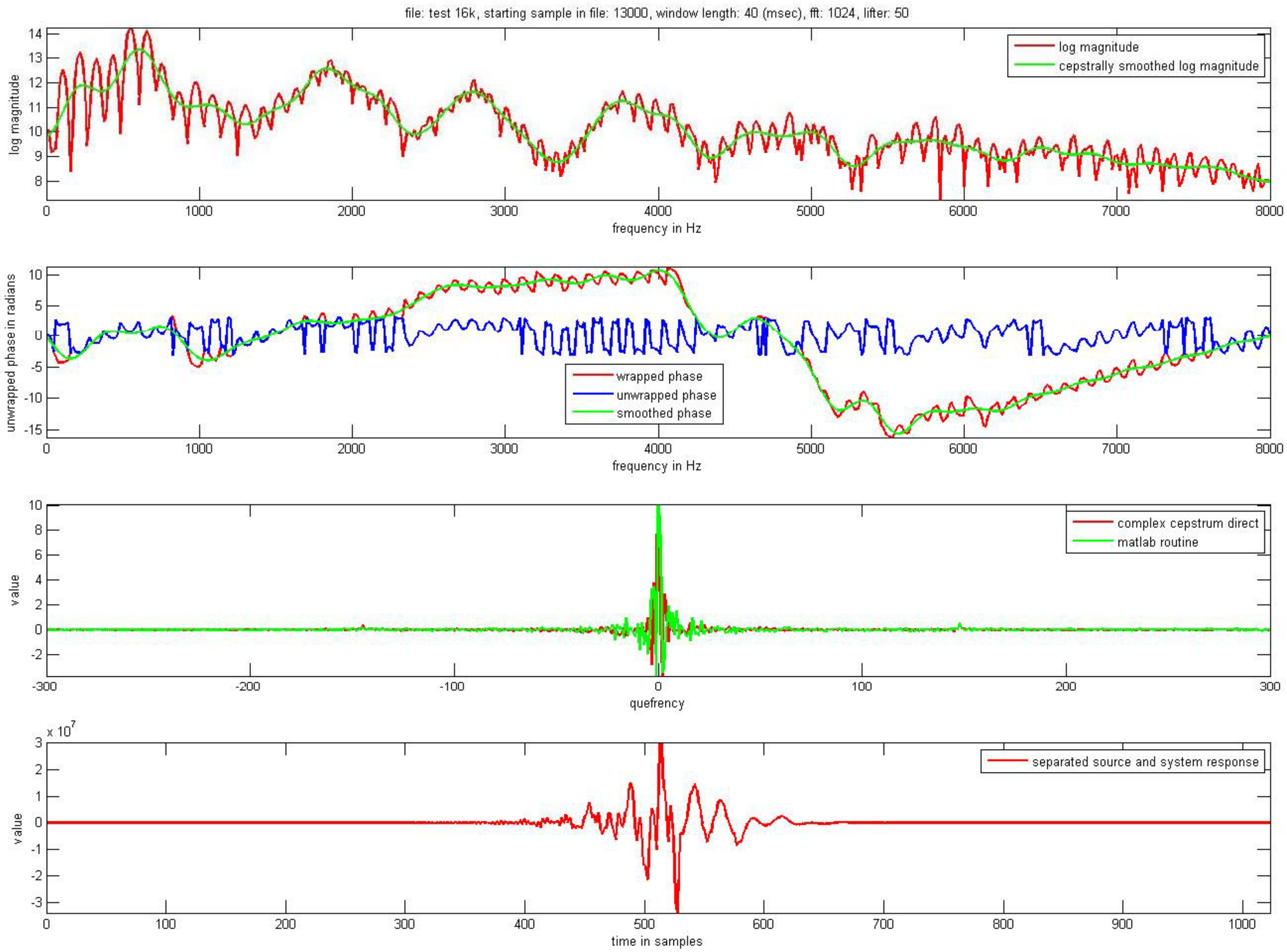


Example 4—effects of low quefrency lifter

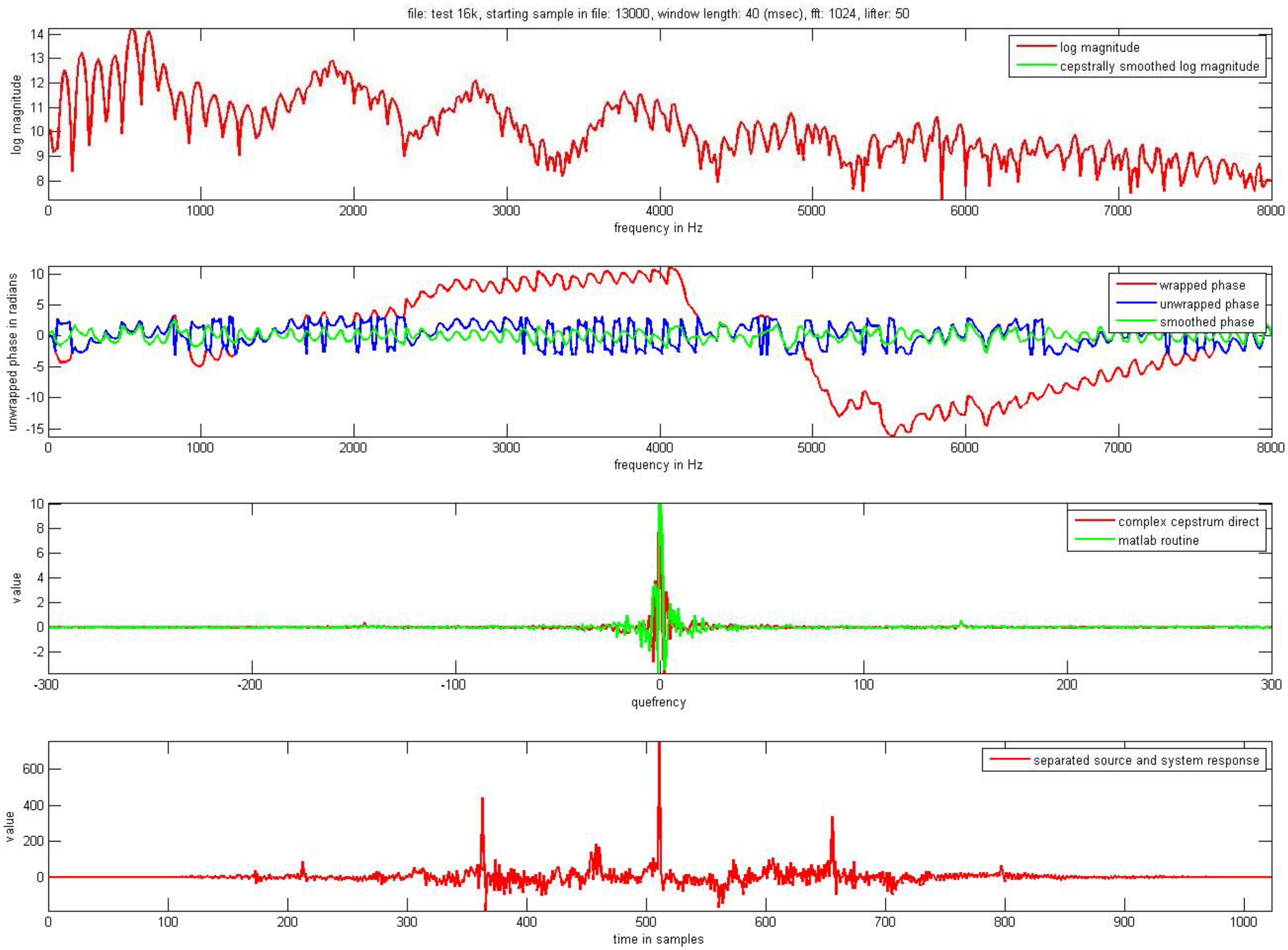




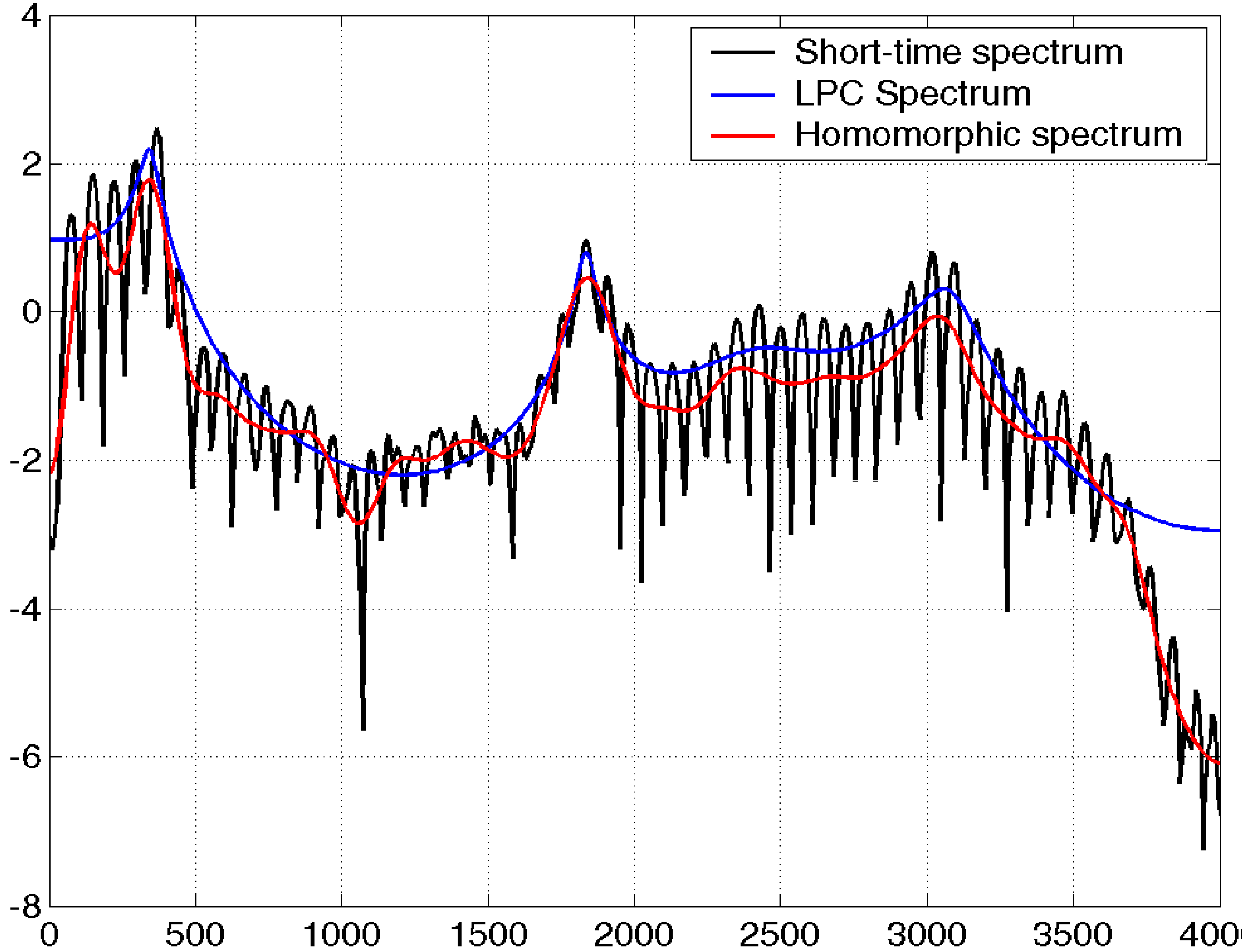
Example 5—phase unwrapping



Example 6—phase unwrapping

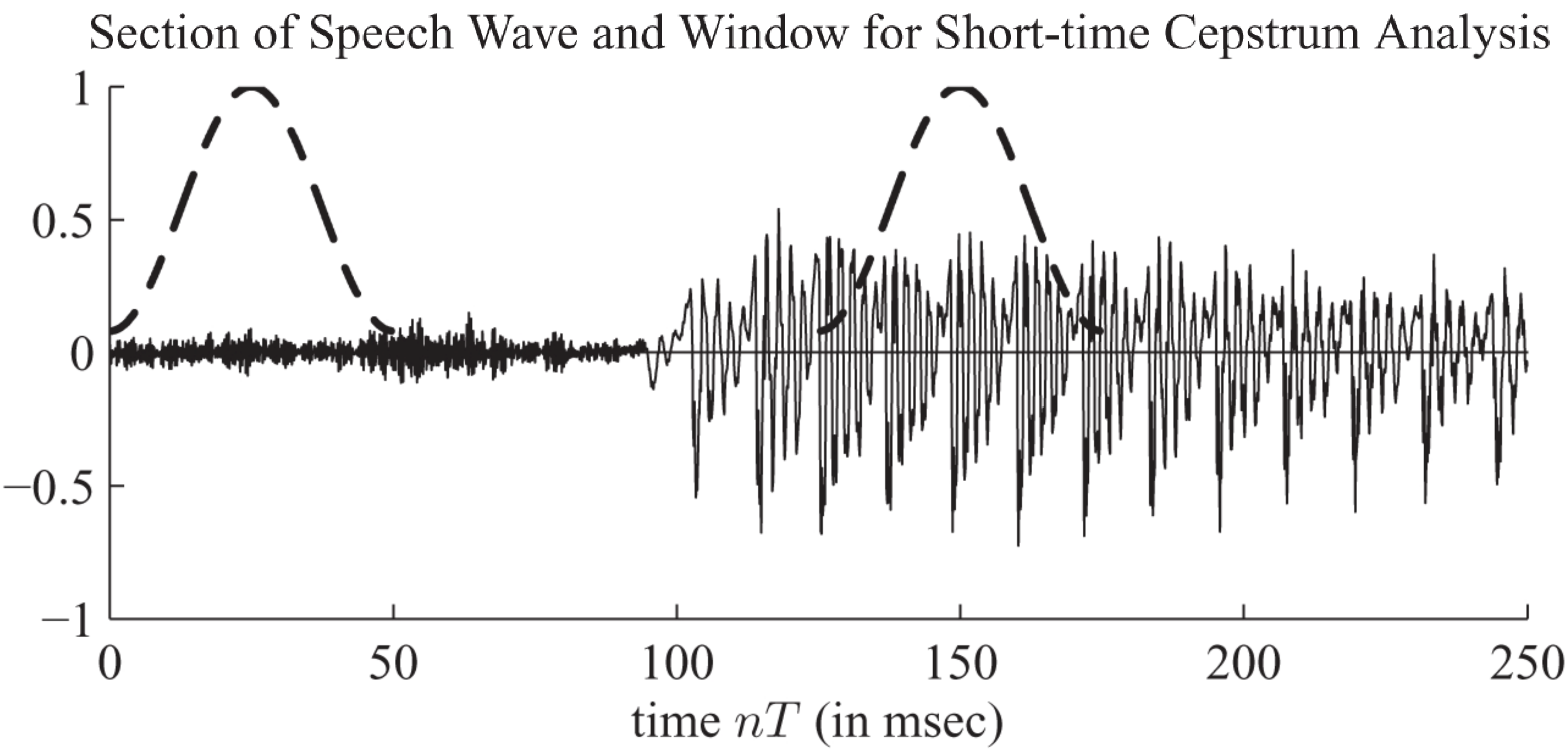


Homomorphic Spectrum Smoothing

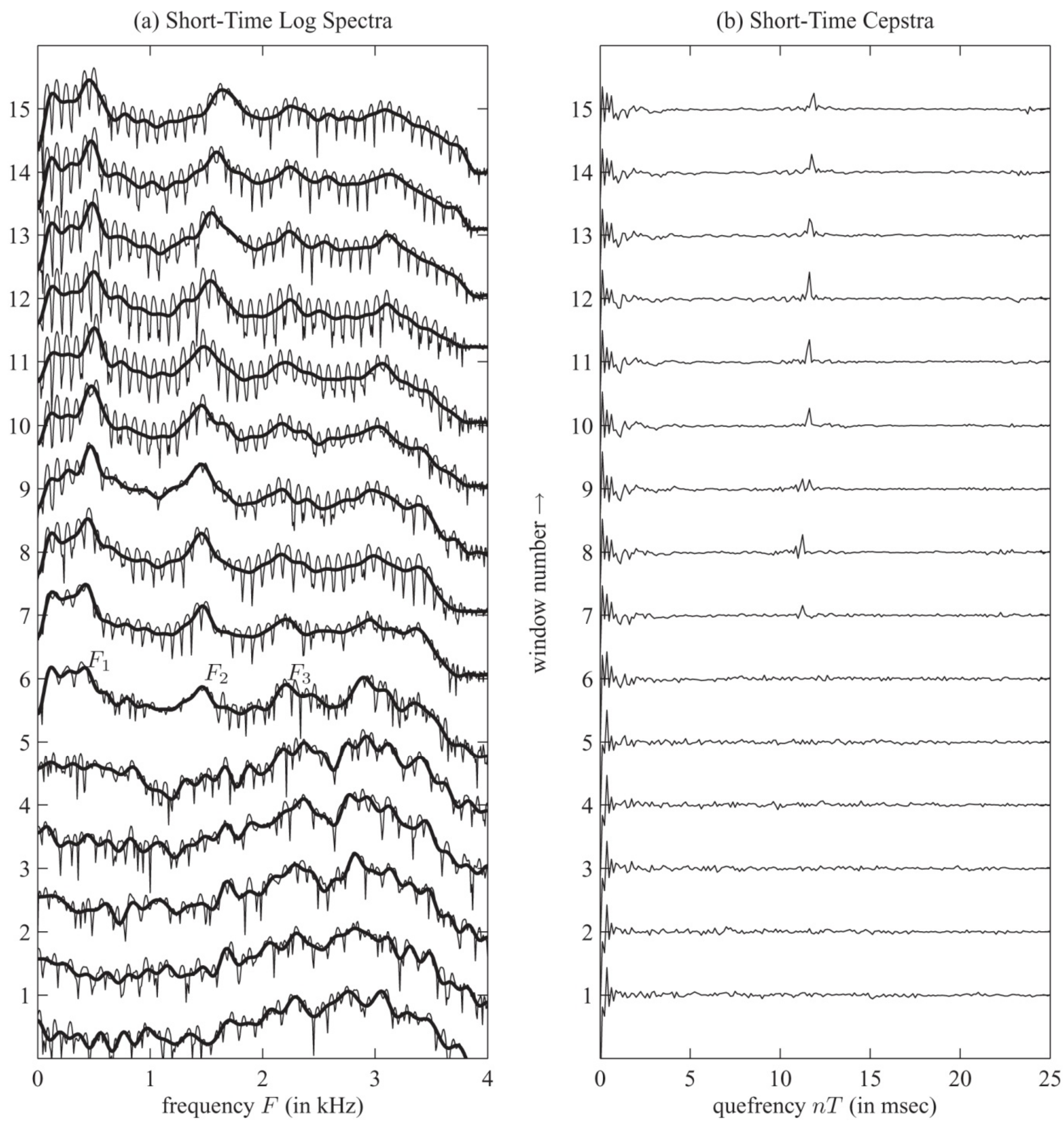


Running Cepstrum

Running Cepstrum



Running Cepstrums



Cepstrum Applications

Cepstrum Distance Measures

- The cepstrum forms a natural basis for comparing patterns in speech recognition or vector quantization because of its stable mathematical characterization for speech signals
- A typical "cepstral distance measure" is of the form:

$$D = \sum_{n=1}^{n_{co}} (c[n] - \overline{c}[n])^2$$

where $c[n]$ and $\overline{c}[n]$ are cepstral sequences corresponding to frames of signal, and D is the cepstral distance between the pair of sequences.

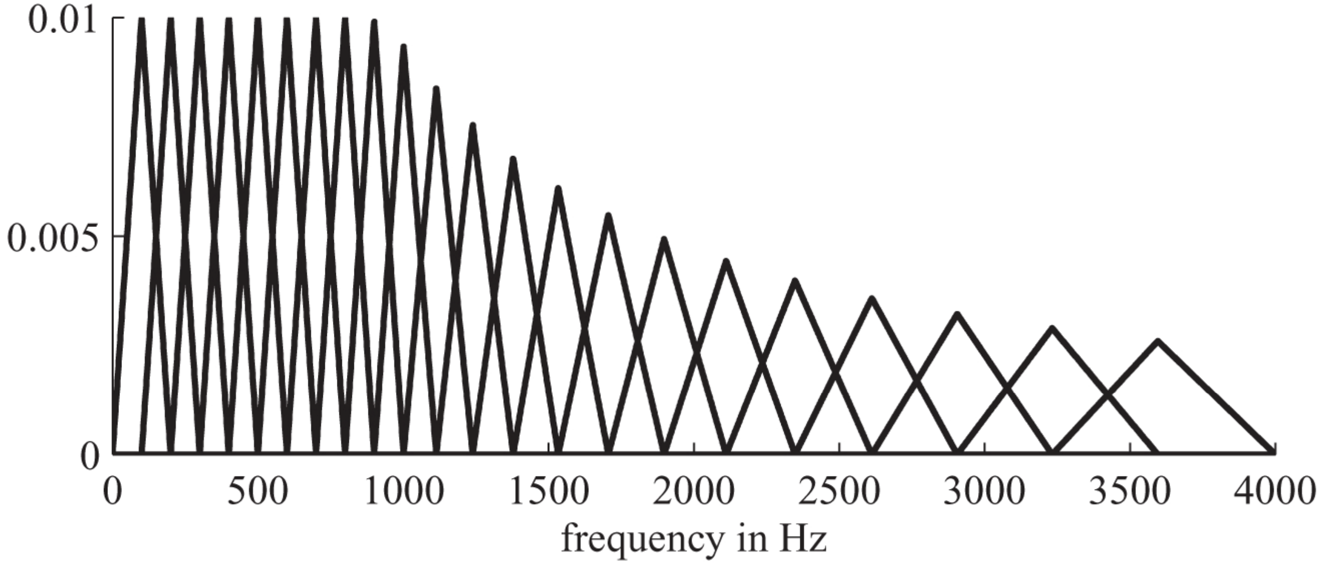
- Using Parseval's theorem, we can express the cepstral distance in the frequency domain as:

$$D = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\log |H(e^{j\omega})| - \log |\overline{H}(e^{j\omega})|)^2 d\omega$$

- Thus we see that the cepstral distance is actually a log magnitude spectral distance

Mel Frequency Cepstral Coefficients

- Basic idea is to compute a frequency analysis based on a filter bank with approximately critical band spacing of the filters and bandwidths. For 4 kHz bandwidth, approximately 20 filters are used.
- First perform a short-time Fourier analysis, giving $X_m[k],\ k = 0,1,\ldots, NF/2$ where m is the frame number and k is the frequency index (1 to half the size of the FFT)
- Next the DFT values are grouped together in critical bands and weighted by triangular weighting functions.



Mel Frequency Cepstral Coefficients

- The mel-spectrum of the m^{th} frame for the r^{th} filter $(r = 1, 2, \ldots, R)$ is defined as:

$$\text{MF}_m[r] = \frac{1}{A_r} \sum_{k=L_r}^{U_r} |V_r[k] X_m[k]|^2$$

where $V_r[k]$ is the weighting function for the r^{th} filter, ranging from DFT index L_r to U_r, and

$$A_r = \sum_{k=L_r}^{U_r} |V_r[k]|^2$$

is the normalizing factor for the r^{th} mel-filter. (Normalization guarantees that if the input spectrum is flat, the mel-spectrum is flat).

- A discrete cosine transform of the log magnitude of the filter outputs is computed to form the function mfcc$[n]$ as:

$$\text{mfcc}_m[n] = \frac{1}{R} \sum_{r=1}^{R} \log(\text{MF}_m[r]) \cos\left[\frac{2\pi}{R}\left(r + \frac{1}{2}\right) n\right], \quad n = 1, 2, \ldots, N_{\text{mfcc}}$$

- Typically $N_{\text{mfcc}} = 13$ and $R = 24$ for 4 kHz bandwidth speech signals.

Delta Cepstrum

- The set of mel frequency cepstral coefficients provide perceptually meaningful and smooth estimates of speech spectra, over time
- Since speech is inherently a dynamic signal, it is reasonable to seek a representation that includes some aspect of the dynamic nature of the time derivatives (both first and second order derivatives) of the short-term cepstrum
- The resulting parameter sets are called the delta cepstrum (first derivative) and the delta-delta cepstrum (second derivative).
- The simplest method of computing delta cepstrum parameters is a first difference of cepstral vectors, of the form:

$$\Delta \text{mfcc}_m[n] = \text{mfcc}_m[n] - \text{mfcc}_{m-1}[n]$$

- The simple difference is a poor approximation to the first derivative and is not generally used. Instead a least-squares approximation to the local slope (over a region around the current sample) is used, and is of the form:

$$\Delta \text{mfcc}_m[n] = \frac{\sum_{k=-M}^{M} k(\text{mfcc}_{m+k}[n])}{\sum_{k=-M}^{M} k^2}$$

where the region is M frames before and after the current frame

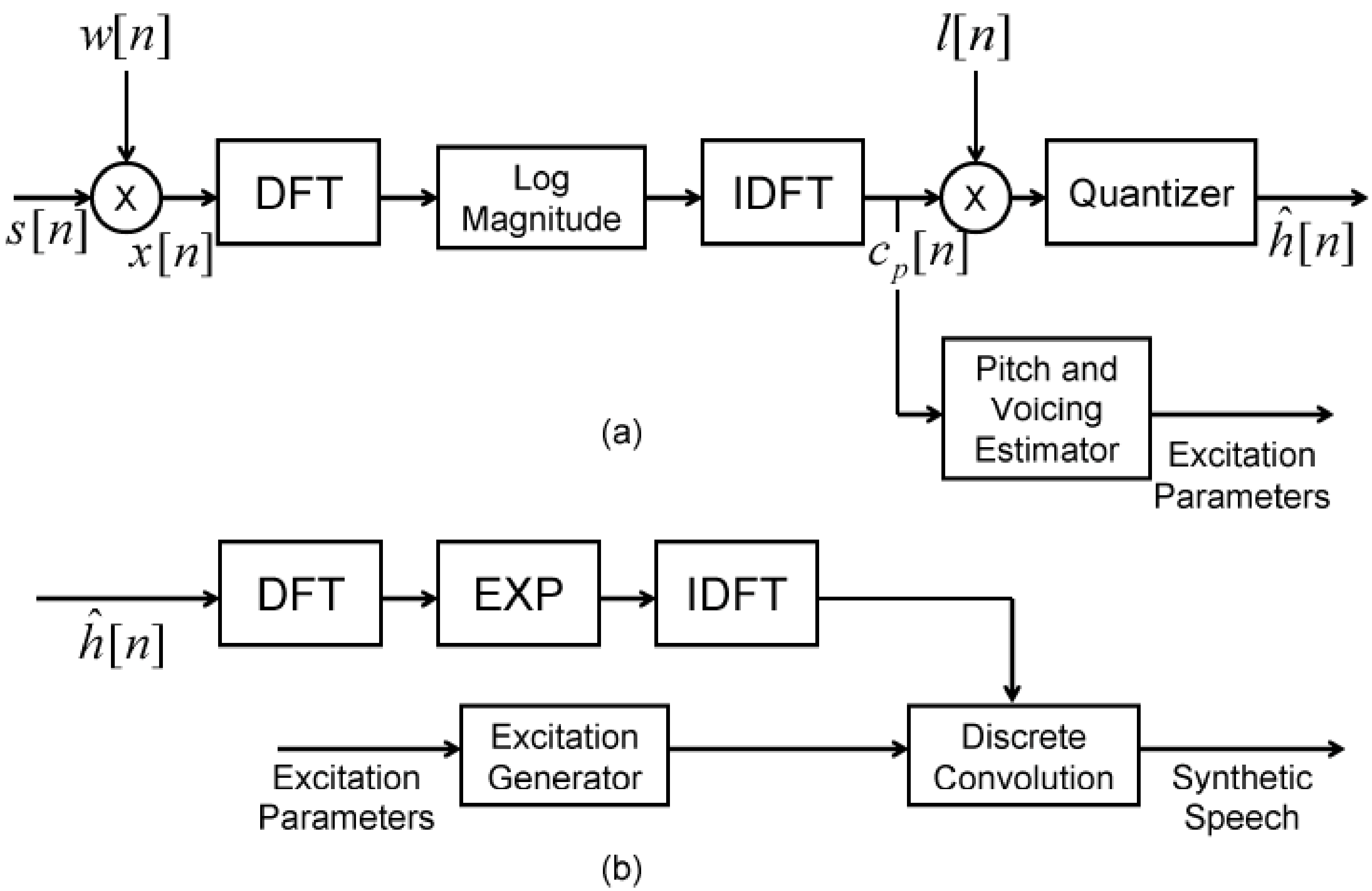
Homomorphic Vocoder

- time-dependent complex cepstrum retains all the information of the time-dependent Fourier transform => exact representation of speech
- time dependent real cepstrum loses phase information -> not an exact representation of speech
- quantization of cepstral parameters also loses information
- cepstrum gives good estimates of pitch, voicing, formants => can build homomorphic vocoder

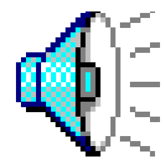
Homomorphic Vocoder

1. compute cepstrum every 10-20 msec
2. estimate pitch period and voiced/unvoiced decision
3. quantize and encode low-time cepstral values
4. at synthesizer-get approximation to $h_v(n)$ or $h_u(n)$ from low time quantized cepstral values
5. convolve $h_v(n)$ or $h_u(n)$ with excitation created from pitch, voiced/unvoiced, and amplitude information

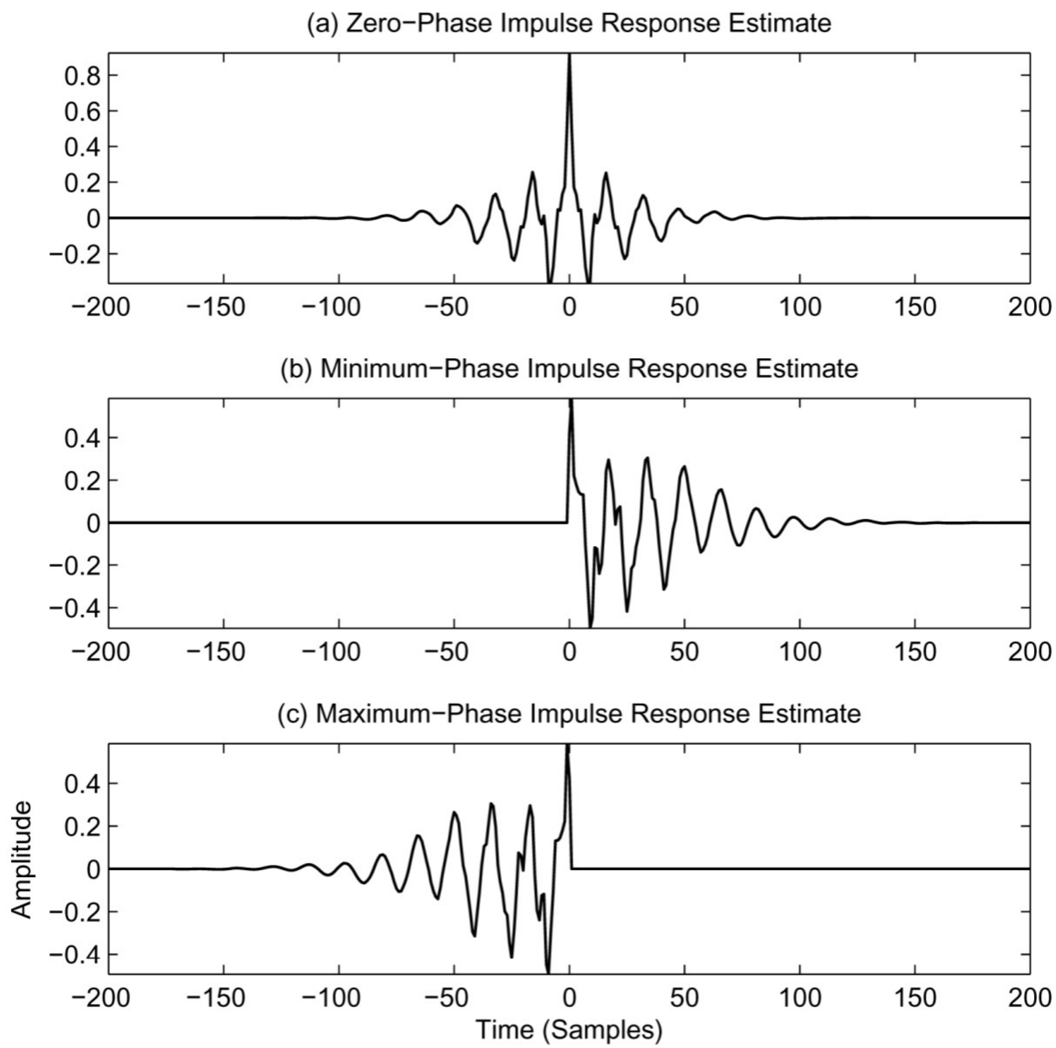
Homomorphic Vocoder



• *l(n)* is cepstrum window that selects low-time values and is of length 26 samples


homomorphic vocoder

Homomorphic Vocoder Impulse Responses



Summary

- Introduced the concept of the cepstrum of a signal, defined as the inverse Fourier transform of the log of the signal spectrum

$$\hat{x}[n] = F^{-1}\left[\log X(e^{j\omega})\right]$$

- Showed cepstrum reflected properties of both the excitation (high quefrency) and the vocal tract (low quefrency)
  - short quefrency window filters out excitation; long quefrency window filters out vocal tract
- Mel-scale cepstral coefficients used as feature set for speech recognition
- Delta and delta-delta cepstral coefficients used as indicators of spectral change over time