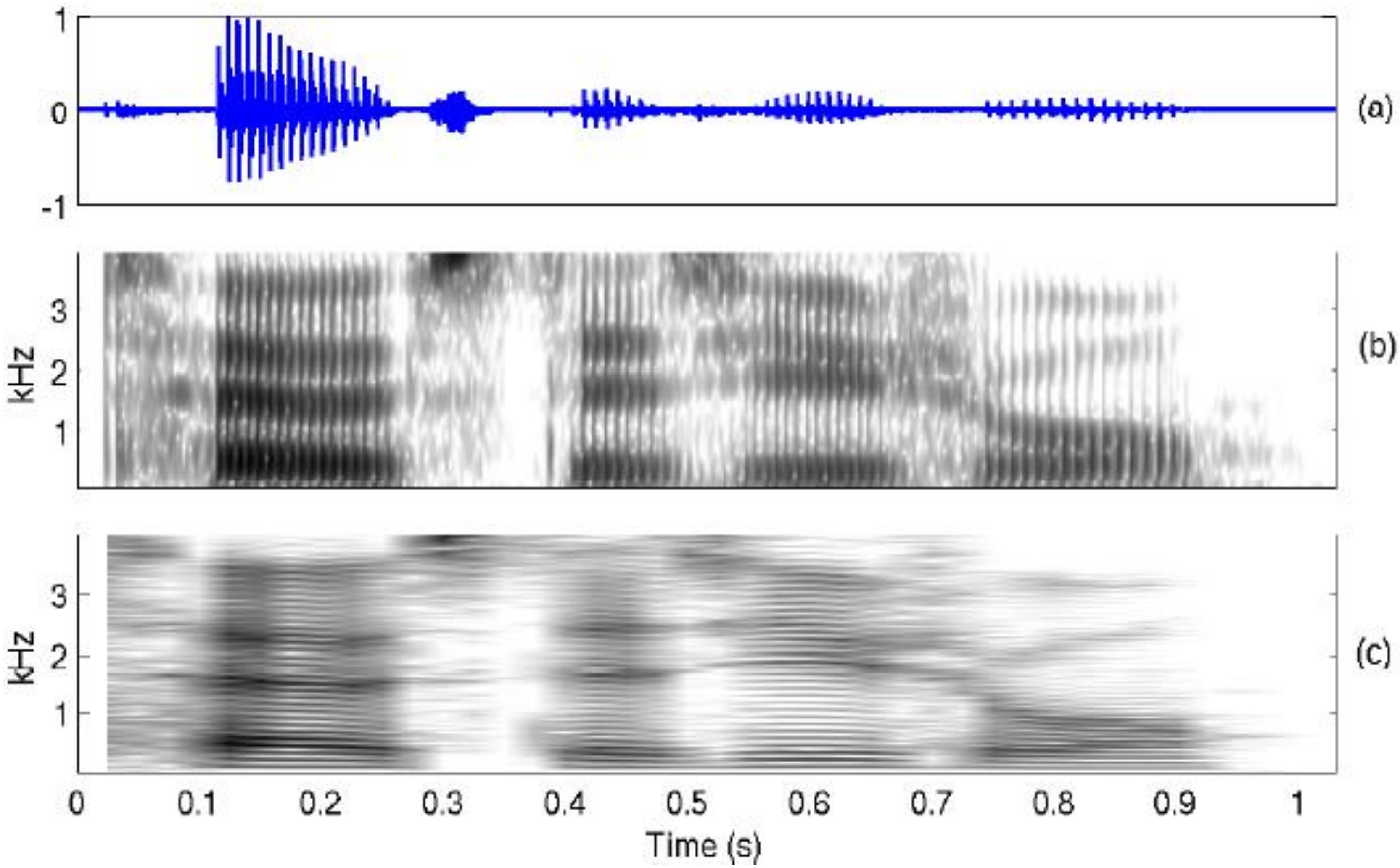
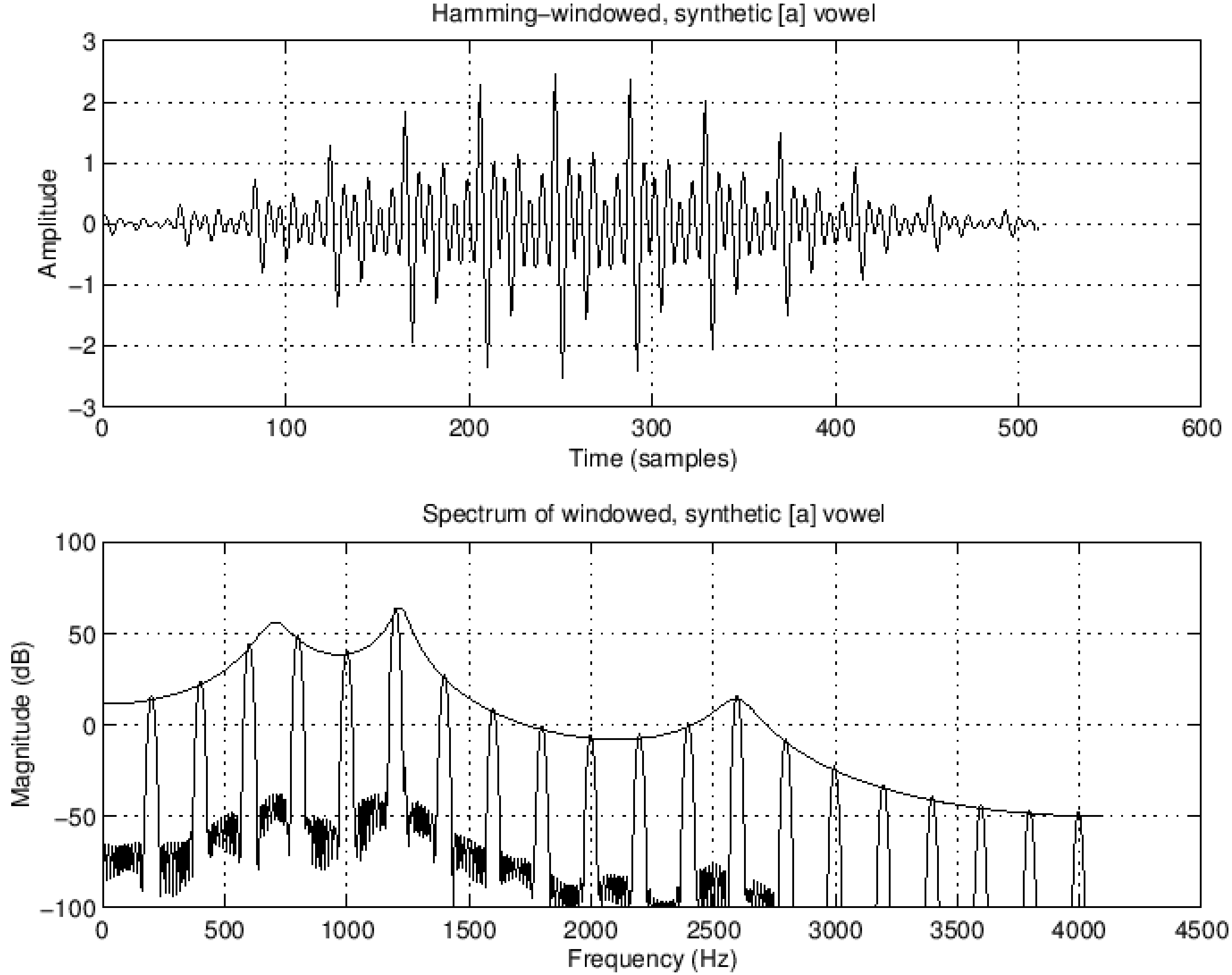


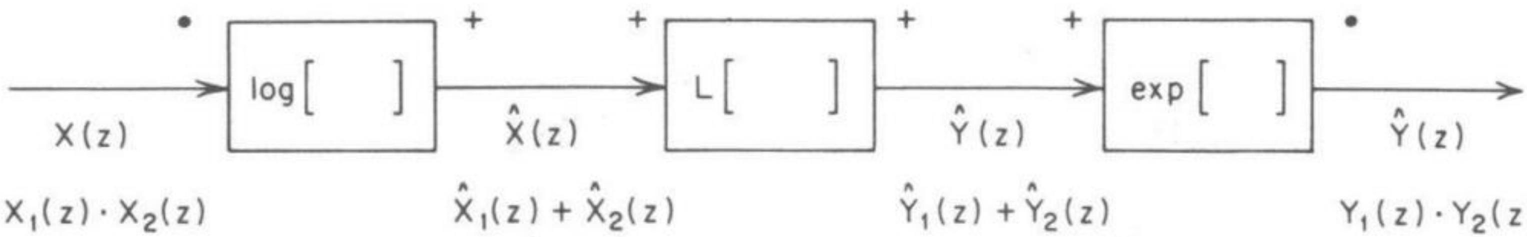
1
0
-1
(a)
kHz
3
2
1
(b)
kHz
3
2
1
(c)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Time (s)

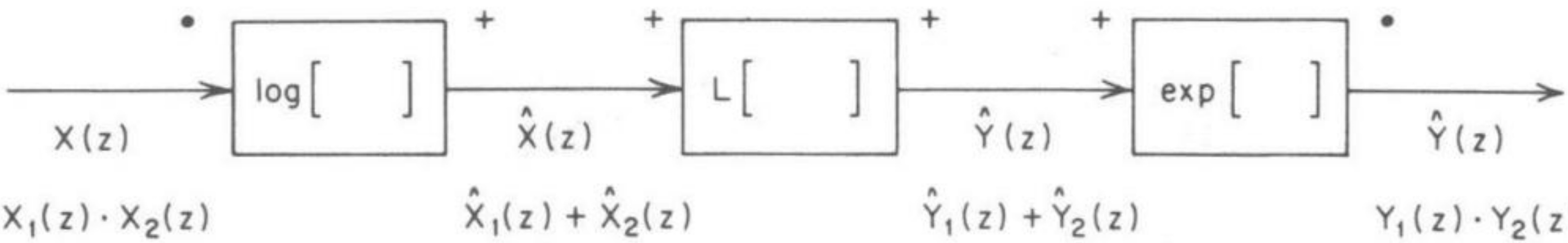


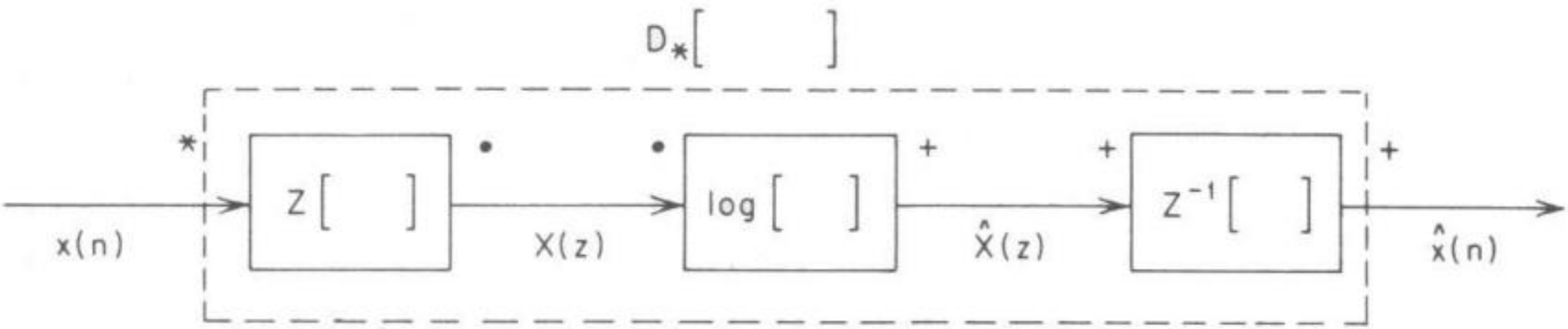
Can we separate into excitation and vocal tract response – deconvolution.

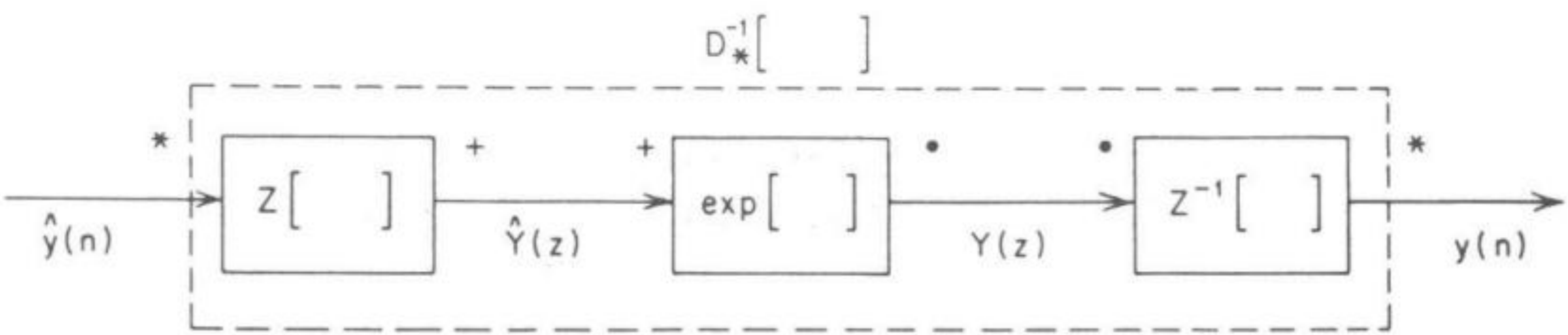
https://ccrma.stanford.edu/~jos/SpecEnv/LPC_Envelope_Example_Speech.html

Thus, the frequency domain representation of a homomorphic system for deconvolution can be represented as



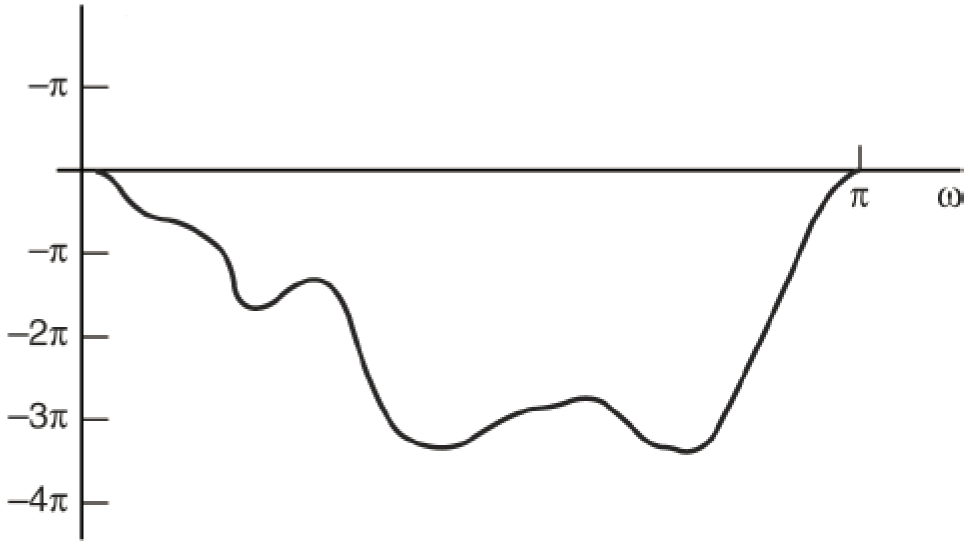

log[ ]
L[ ]
exp[ ]
X(z)
X̂(z)
Ŷ(z)
Ŷ(z)
$X_1(z) \cdot X_2(z)$
$\hat{X}_1(z) + \hat{X}_2(z)$
$\hat{Y}_1(z) + \hat{Y}_2(z)$
$Y_1(z) \cdot Y_2(z)$

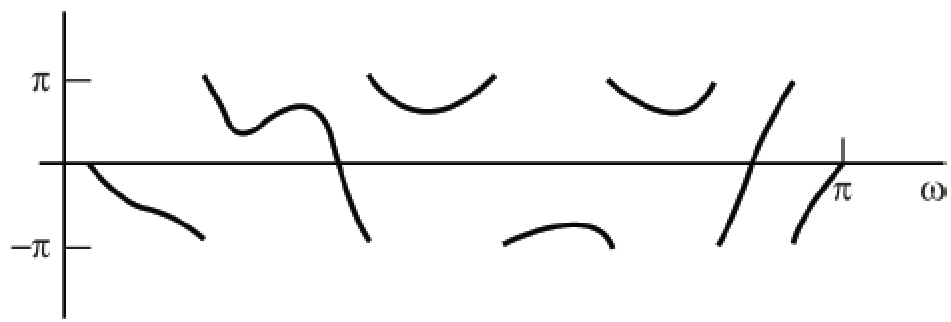

$D_*[\quad]$
Z[ ]
log[ ]
$Z^{-1}[\quad]$
x(n)
X(z)
X̂(z)
x̂(n)


$D_*^{-1}[\quad]$
Z[ ]
exp[ ]
$Z^{-1}[\quad]$
ŷ(n)
Ŷ(z)
Y(z)
y(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


−π
π
ω
−π
−2π
−3π
−4π


π
π
ω
−π

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

$$\hat{X}(e^{j\omega}) = \ln X(e^{j\omega}) = \ln |X(e^{j\omega})| + j \underbrace{\theta_x(\omega)}_{\text{not unique}}$$

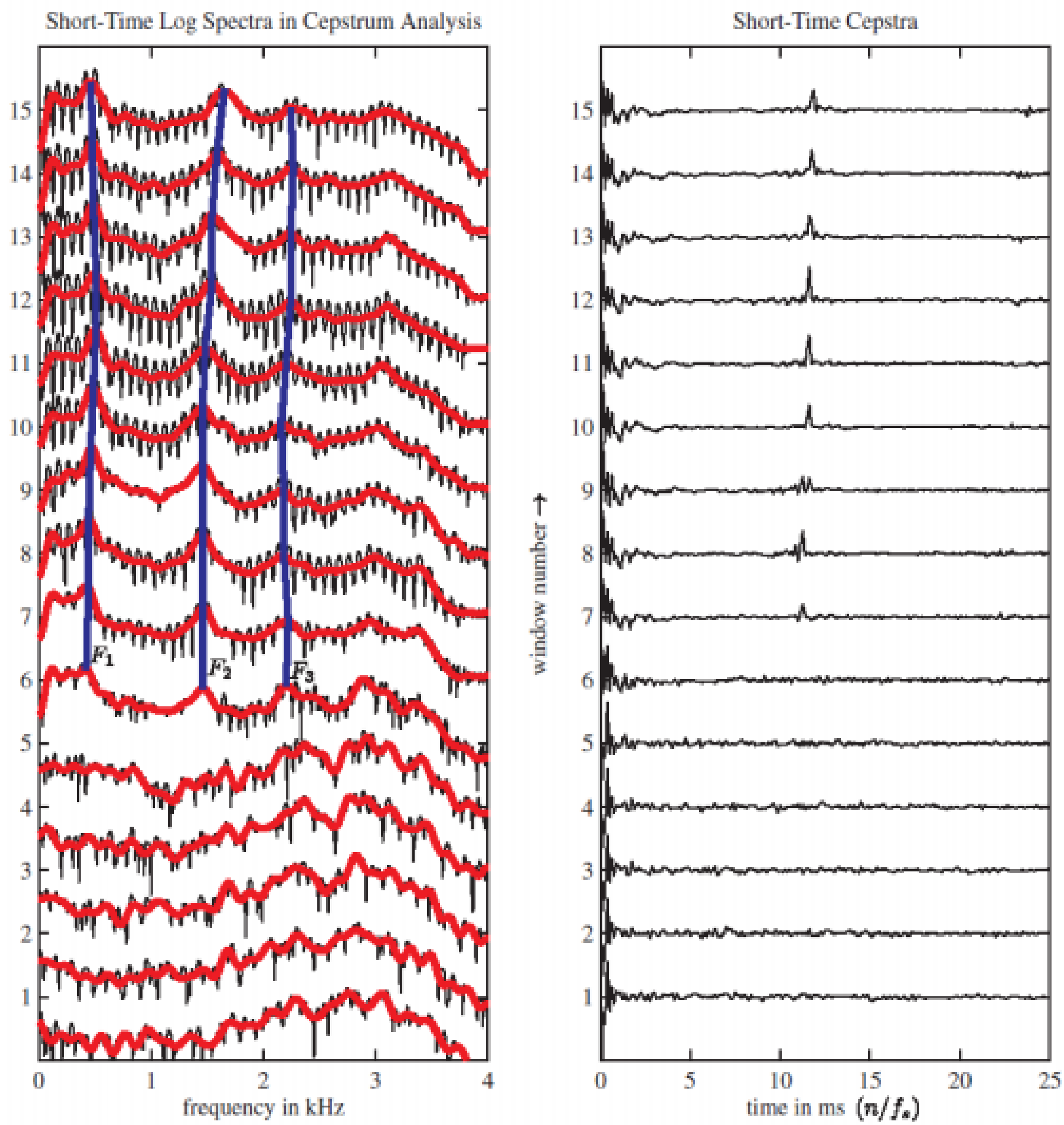
$$\hat{x}[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{X}(e^{j\omega}) e^{j\omega n} d\omega$$

$$= \underbrace{\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |X(e^{j\omega})| e^{j\omega n} d\omega}_{\rightarrow c_x[n]}$$

$$+ \underbrace{j \frac{1}{2\pi} \int_{-\pi}^{\pi} \theta_x(\omega) e^{j\omega n} d\omega}_{\rightarrow \hat{x}_{odd}[n]}$$

$\hat{x}[n]$ not unique

- one way to make unique is to use unwrapped phase. $\Phi^*_x(\omega)$



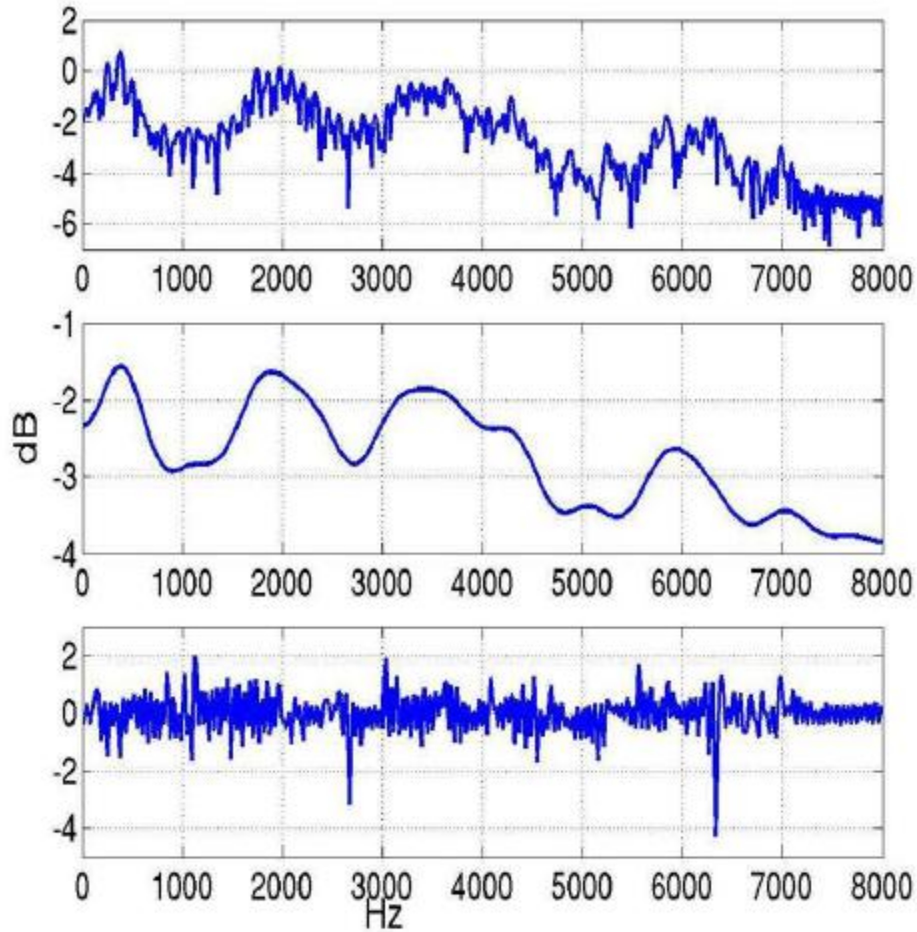
Short-Time Log Spectra in Cepstrum Analysis
Short-Time Cepstra
F_1
F_2
F_3
window number →
frequency in kHz
time in ms (n/f_s)

Spectral Envelope

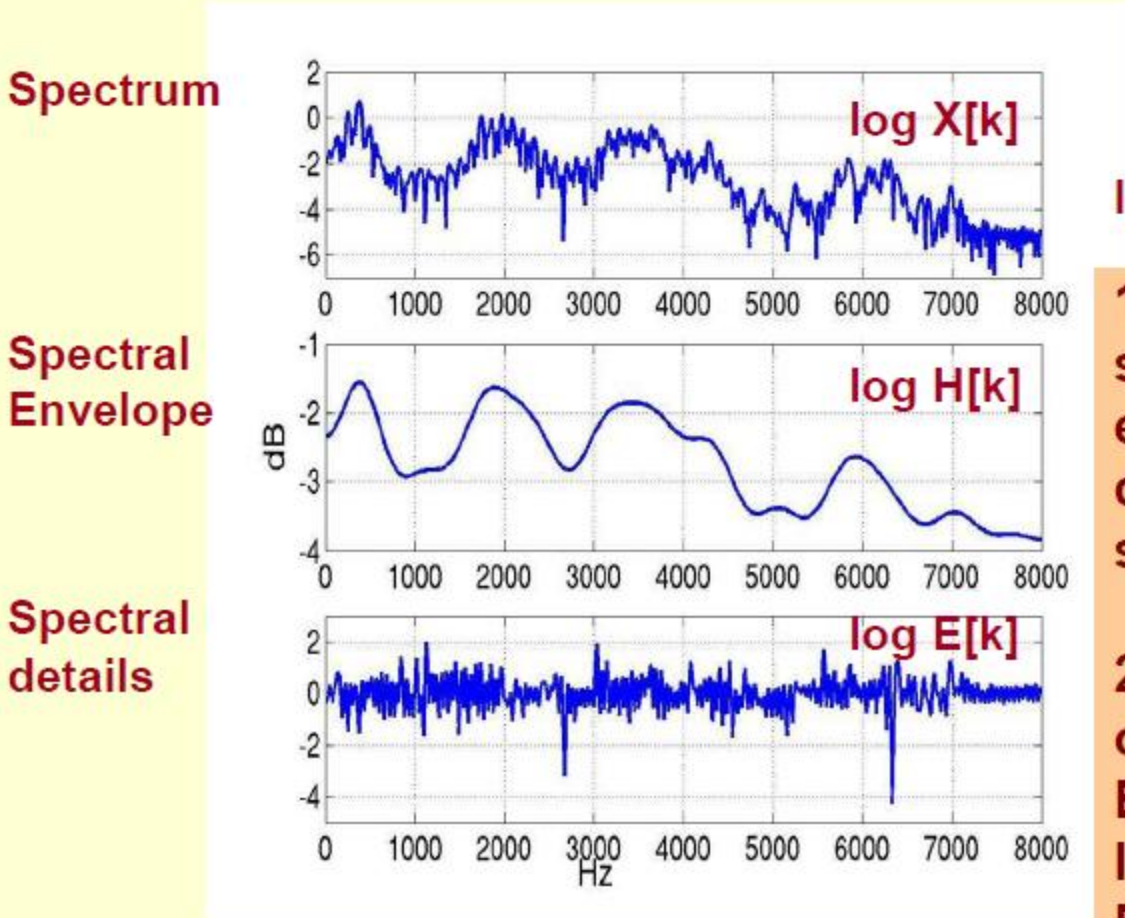
Spectrum

Spectral Envelope

Spectral details



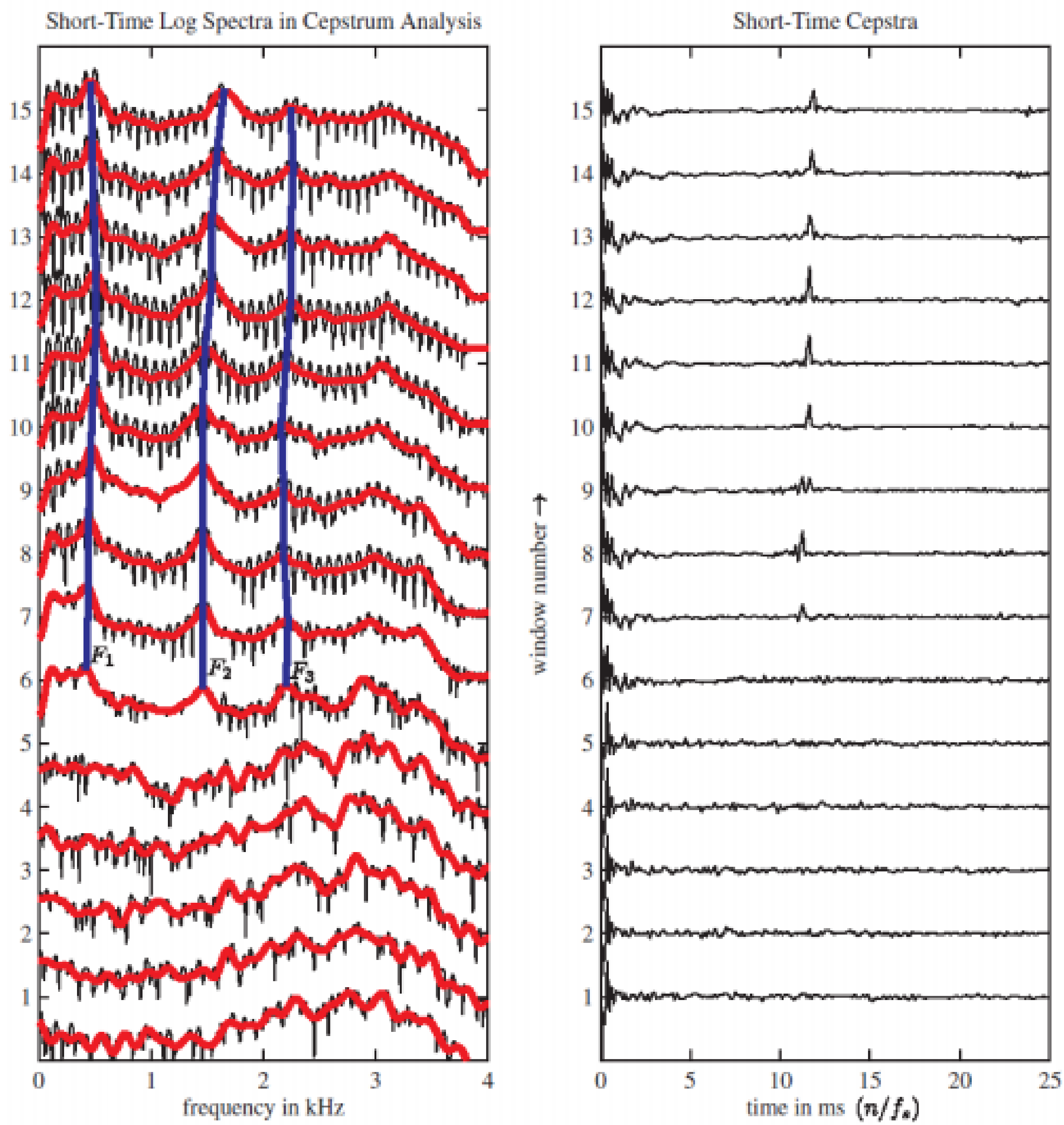
Spectral Envelope



log X[k] = log H[k] + log E[k]

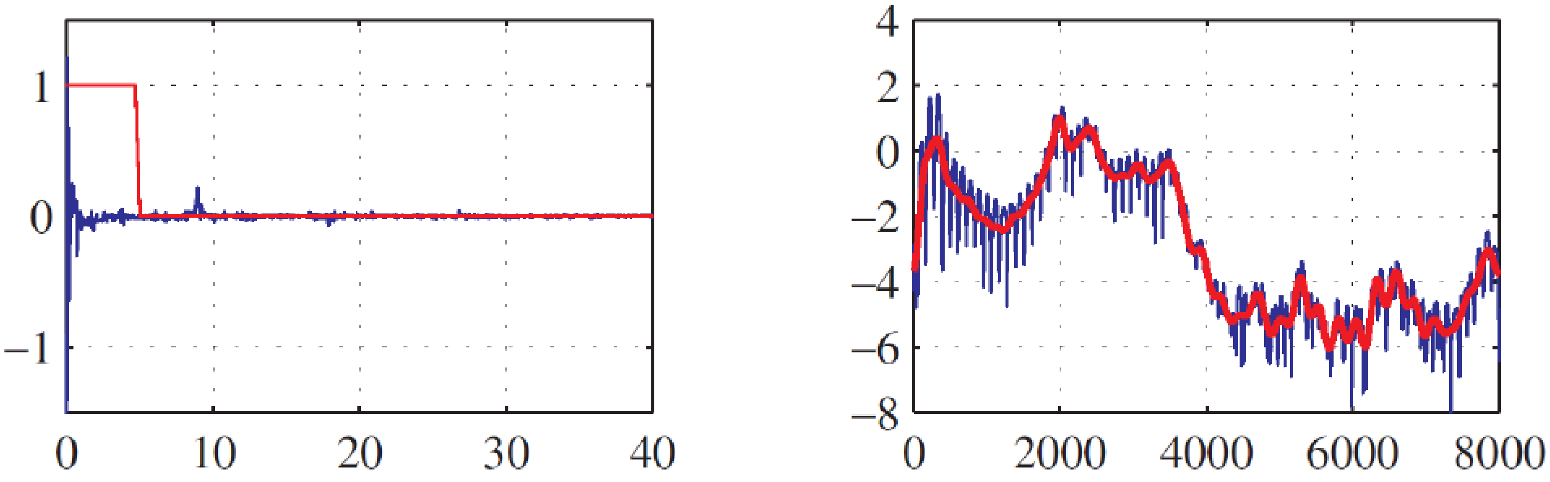
1. Our goal: We want to separate spectral envelope and spectral details from the spectrum.

2. i.e Given log X[k], obtain log H[k] and log E[k], such that log X[k] = log H[k] + log E[k]

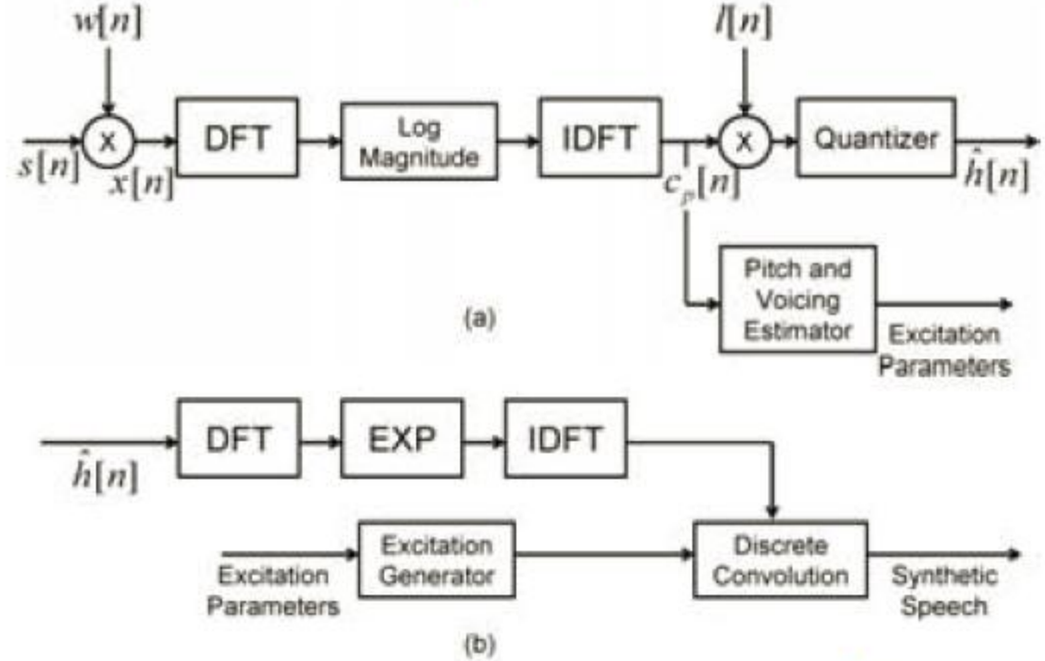


Short-Time Log Spectra in Cepstrum Analysis
Short-Time Cepstra
F_1
F_2
F_3
frequency in kHz
time in ms (n/f_s)
window number →

Liftering in the cepstral domain

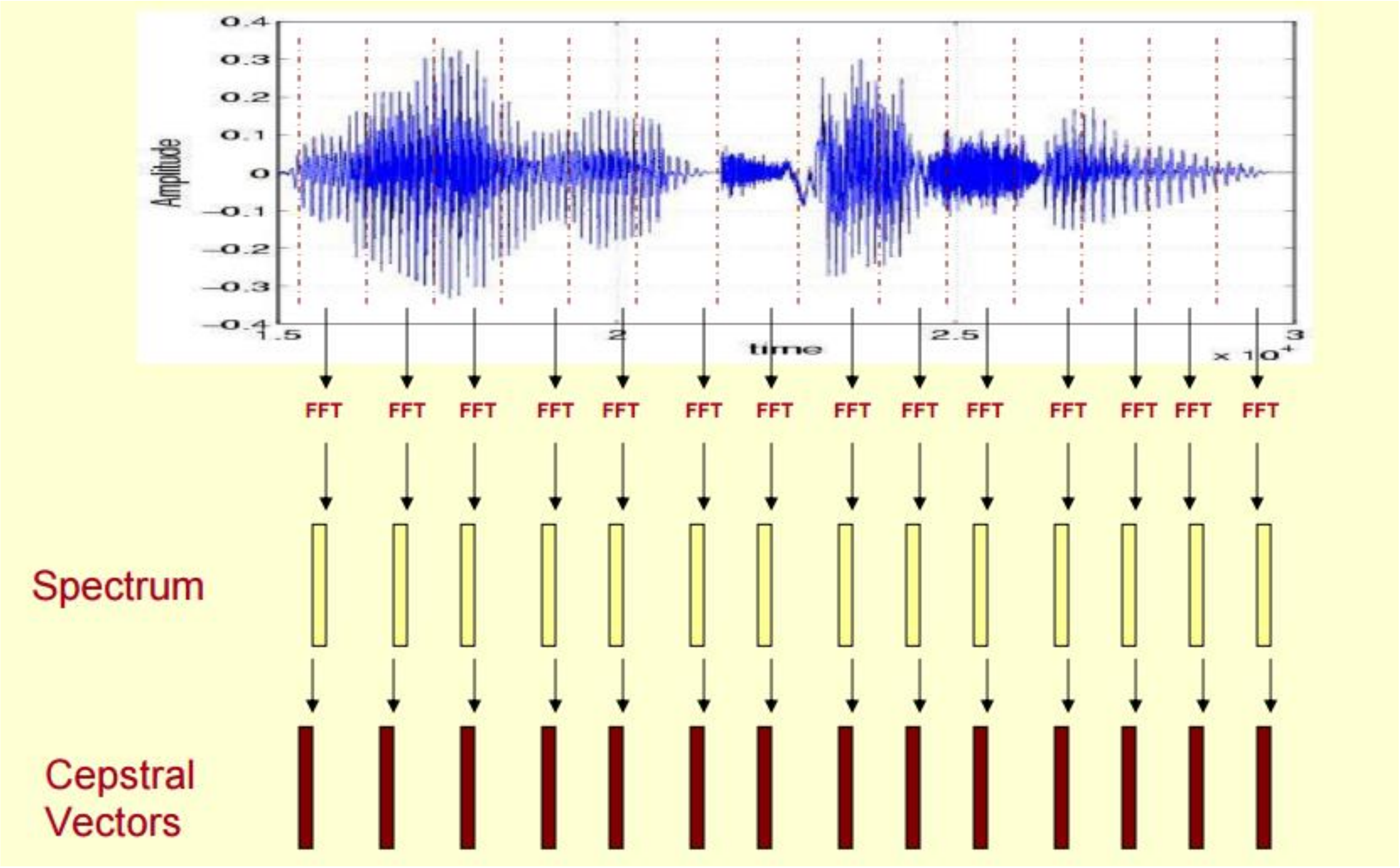


Homomorphic Vocoder

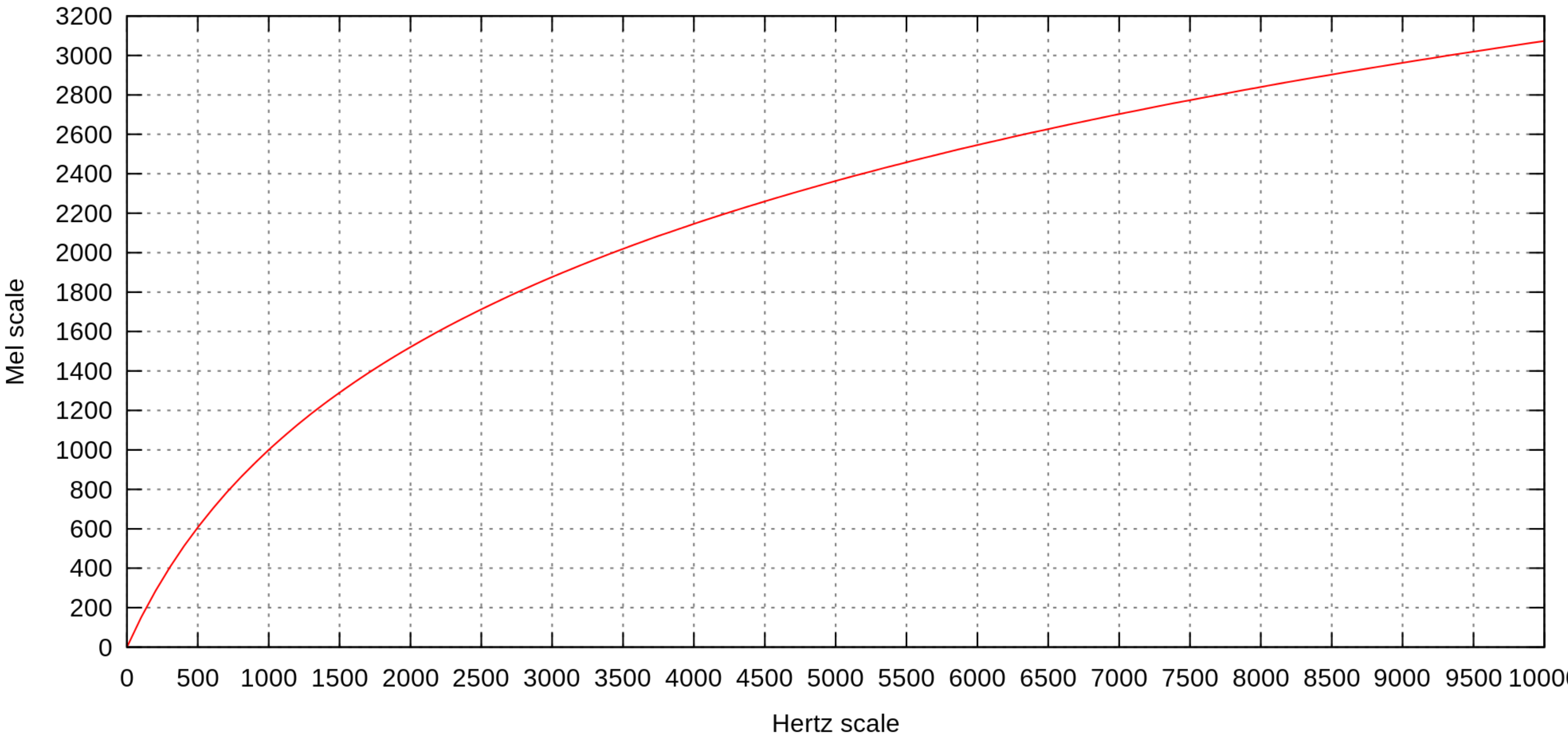


- *l(n)* is cepstrum window that selects low-time values and is of length 26 samples


homomorphic vocoder



0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
−0.4
Amplitude
1.5
2
2.5
3
time
x 10^4
FFT
Spectrum
Cepstral
Vectors

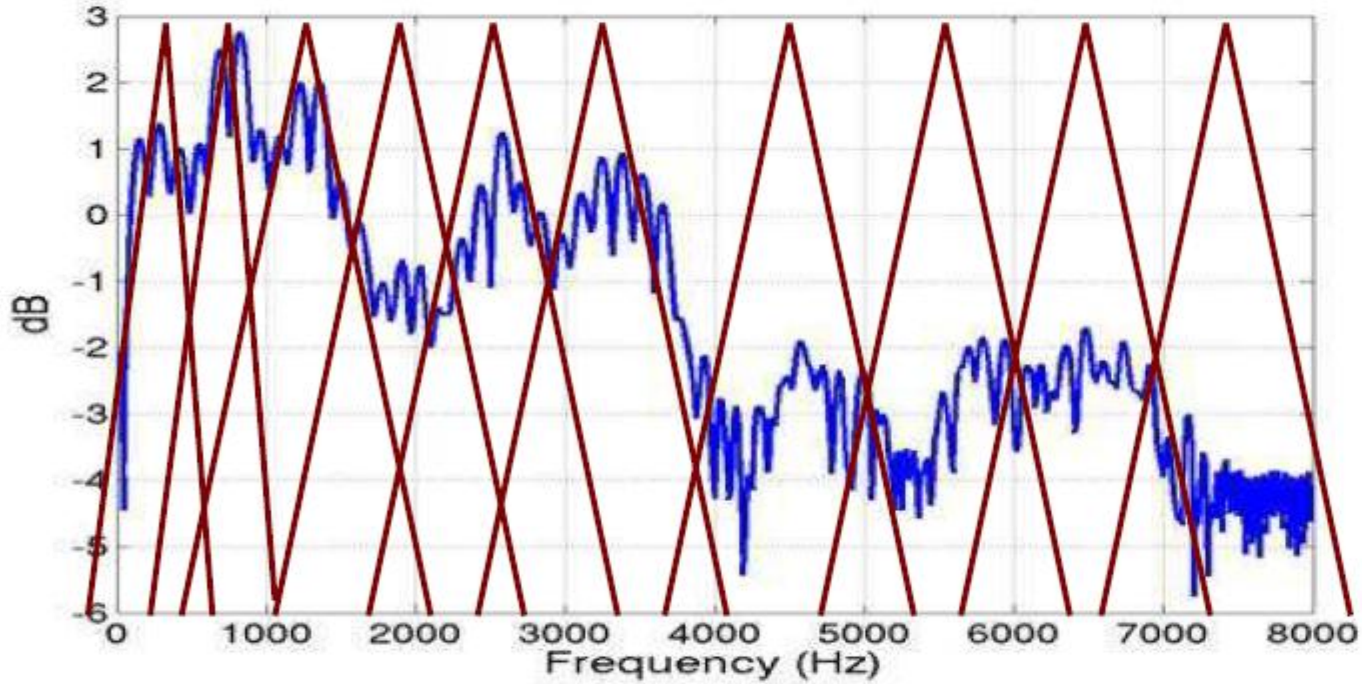


Mel scale
3200
3000
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
0 500 1000 1500 2000 2500 3000 3500 4000 4500 5000 5500 6000 6500 7000 7500 8000 8500 9000 9500 10000
Hertz scale

Mel-Frequency Filters

More no. of filters in low freq. region

Lesser no. of filters in high freq. region



- The mel-frequency spectrum at analysis time n is defined as

$$MF[r] = \frac{1}{A_r} \sum_{k=L_r}^{U_r} |V_r[k] X(n,k)|$$

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

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

  - which serves as a normalization factor for the r-th filter, so that a perfectly flat Fourier spectrum will also produce a flat Mel-spectrum

- For each frame, a discrete cosine transform (DCT) of the log-magnitude of the filter outputs is then computed to obtain the MFCCs

$$MFCC[m] = \frac{1}{R} \sum_{r=1}^{R} \log(MF[r]) \cos\left[\frac{2\pi}{R}\left(r + \frac{1}{2}\right) m\right]$$

- where typically $MFCC[m]$ is evaluated for a number of coefficients N_{MFCC} that is less than the number of mel-filters R
  - For $F_s = 8KHz$, typical values are $N_{MFCC} = 13$ and $R = 22$

Notes

– The MFCC is no longer a homomorphic transformation

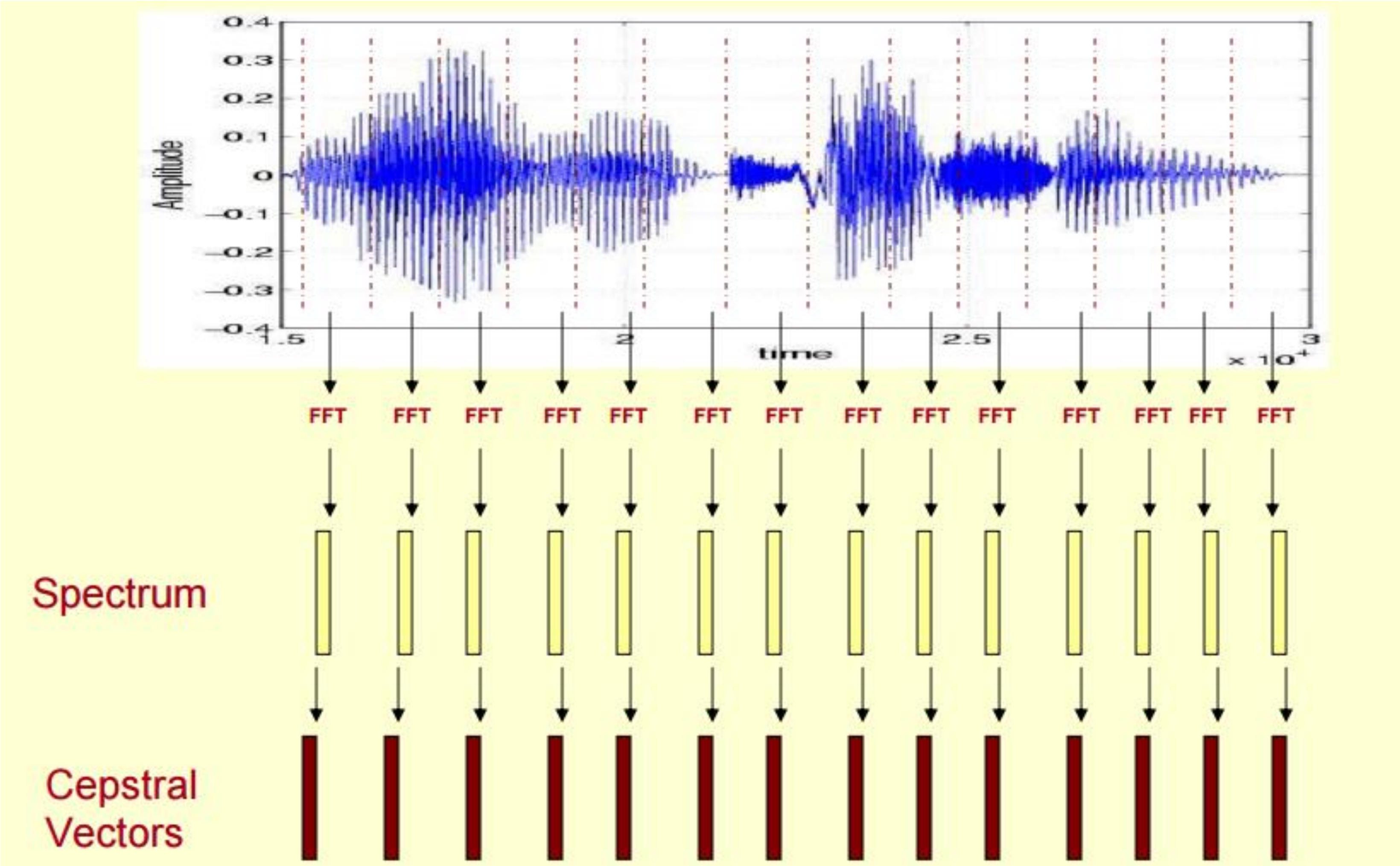
- It would be if the order of summation and logarithms were reversed, in other words if we computed

$$\frac{1}{A_r}\sum_{k=L_r}^{U_r} \log|V_r[k]X(n,k)|$$

- Instead of

$$\log\left(\frac{1}{A_r}\sum_{k=L_r}^{U_r} |V_r[k]X(n,k)|\right)$$

- In practice, however, the MFCC representation is approximately homomorphic for filters that have a smooth transfer function
- The advantages of the second summation above is that the filter energies are more robust to noise and spectral properties



0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
Amplitude
1.5
2
2.5
3
time
x 10^4
FFT
Spectrum
Cepstral
Vectors