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

$$\hat{X}(e^{j\omega}) = \ln X(e^{j\omega}) = \ln|X(e^{j\omega})| + j\,\theta_x(\omega)$$

$\theta_x(\omega)$ ← not unique

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

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

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

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

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

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

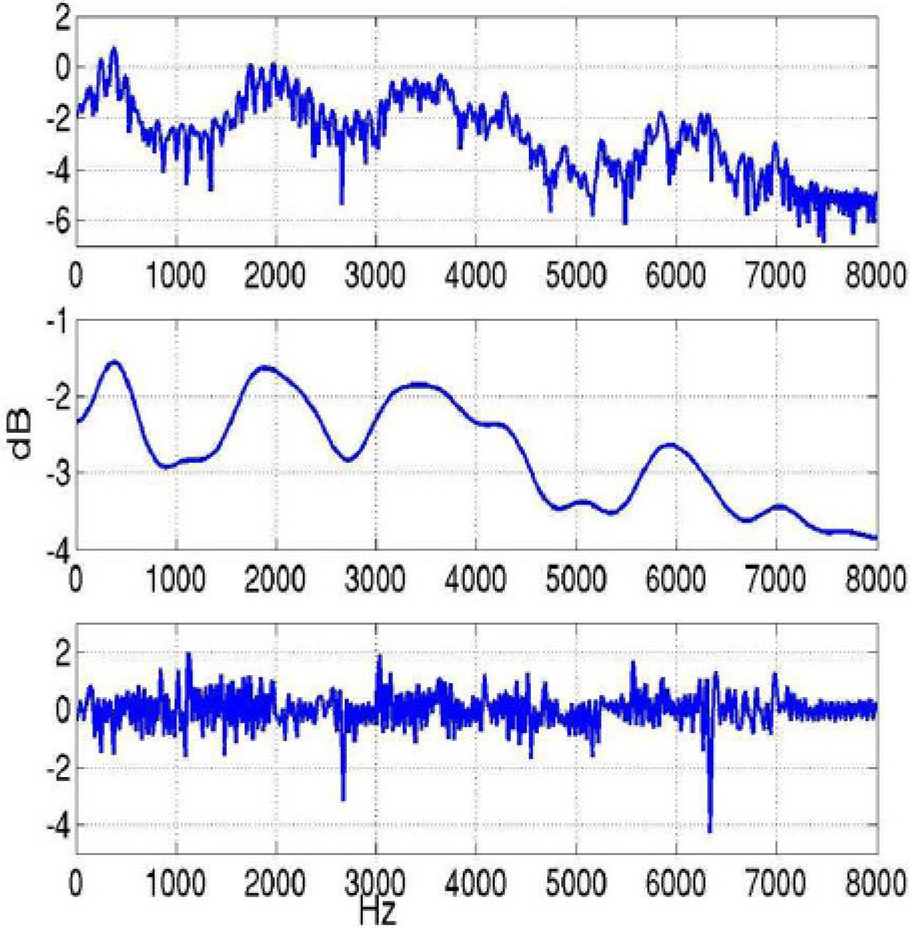
Example of a voiced spectrum from PRAAT

Spectral Envelope

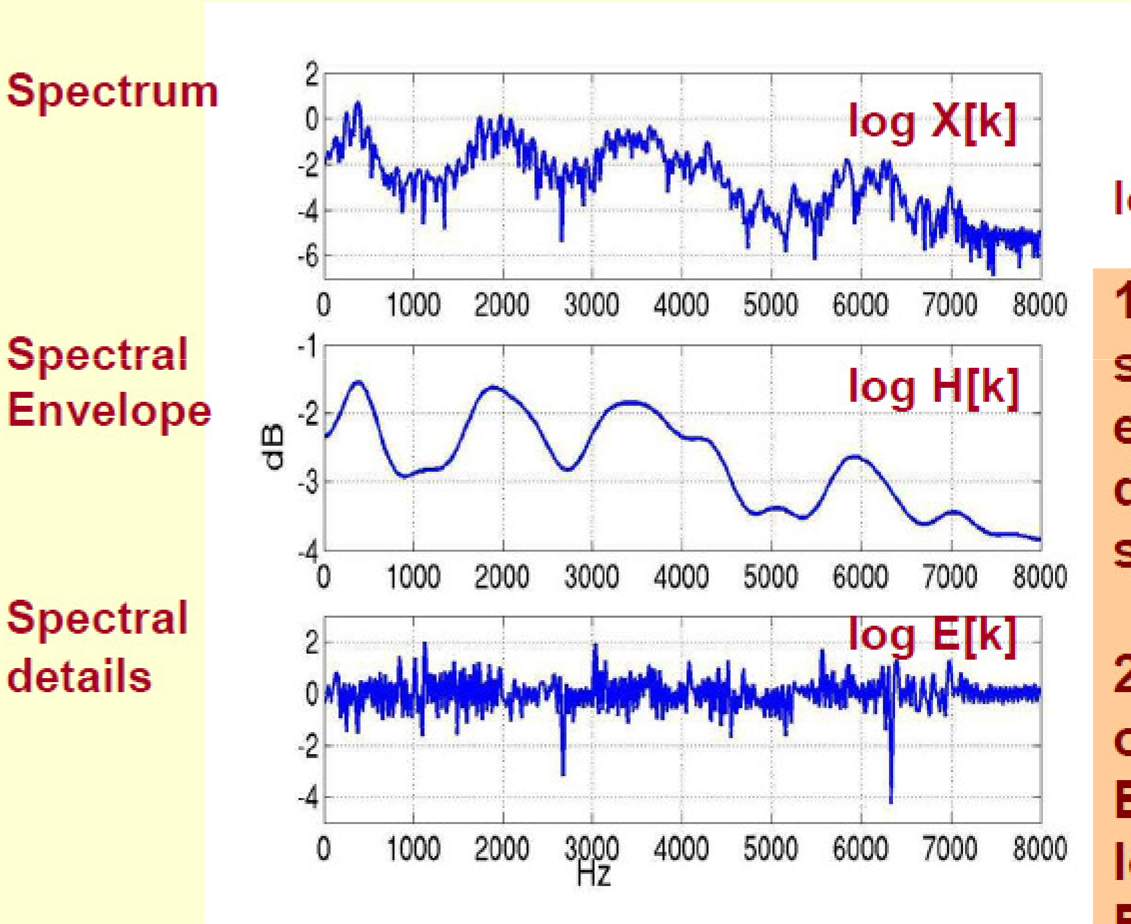
Spectrum

Spectral Envelope

Spectral details



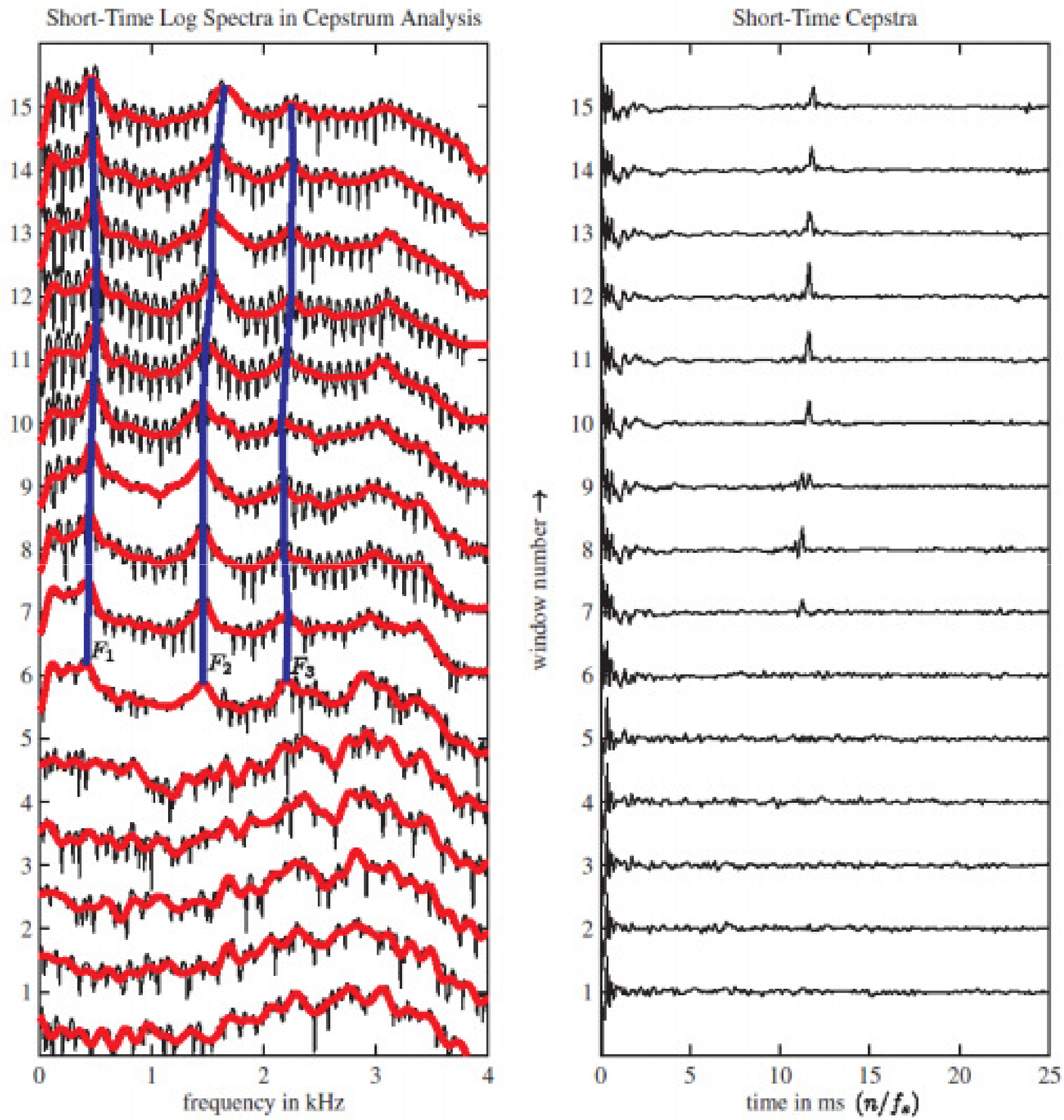
Spectral Envelope



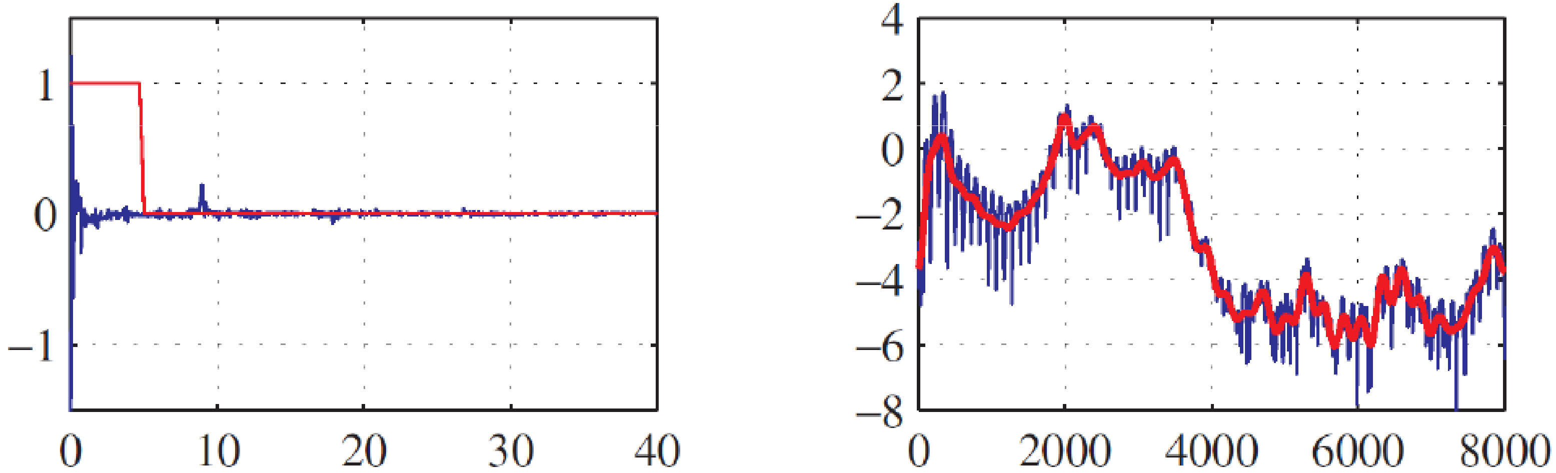
log X[k] = log H[k] + log E[k]

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

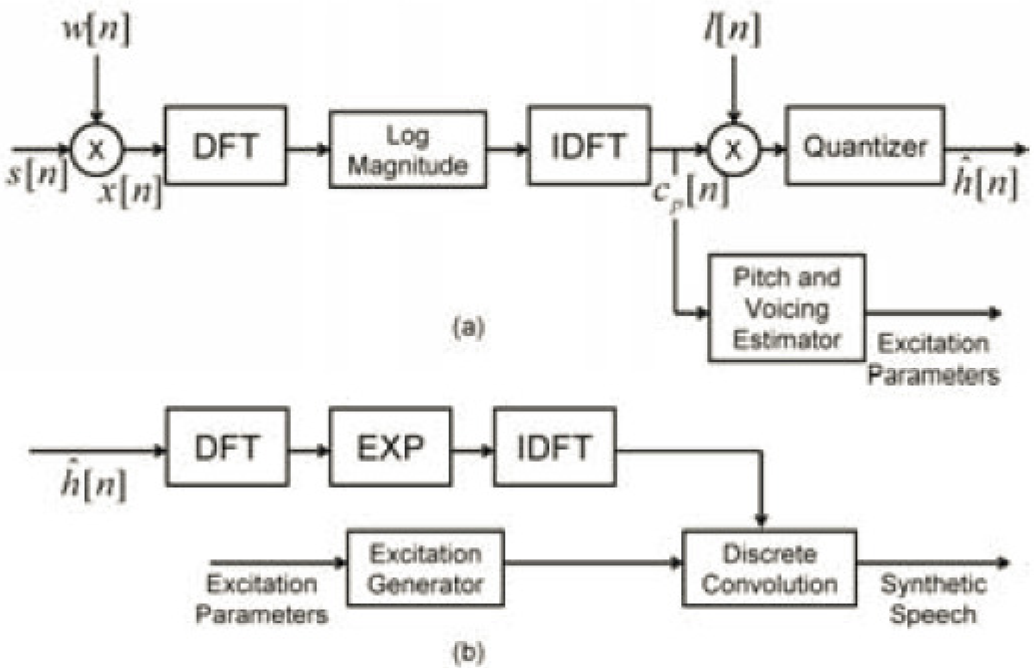
2. i.e Given log X[k], obtain log H[k] and log E[k], such that log X[k] = log H[k] + log E[k]


Short-Time Log Spectra in Cepstrum Analysis
Short-Time Cepstra
F1
F2
F3
frequency in kHz
window number →
time in ms (n/fs)

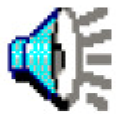
Liftering in the cepstral domain



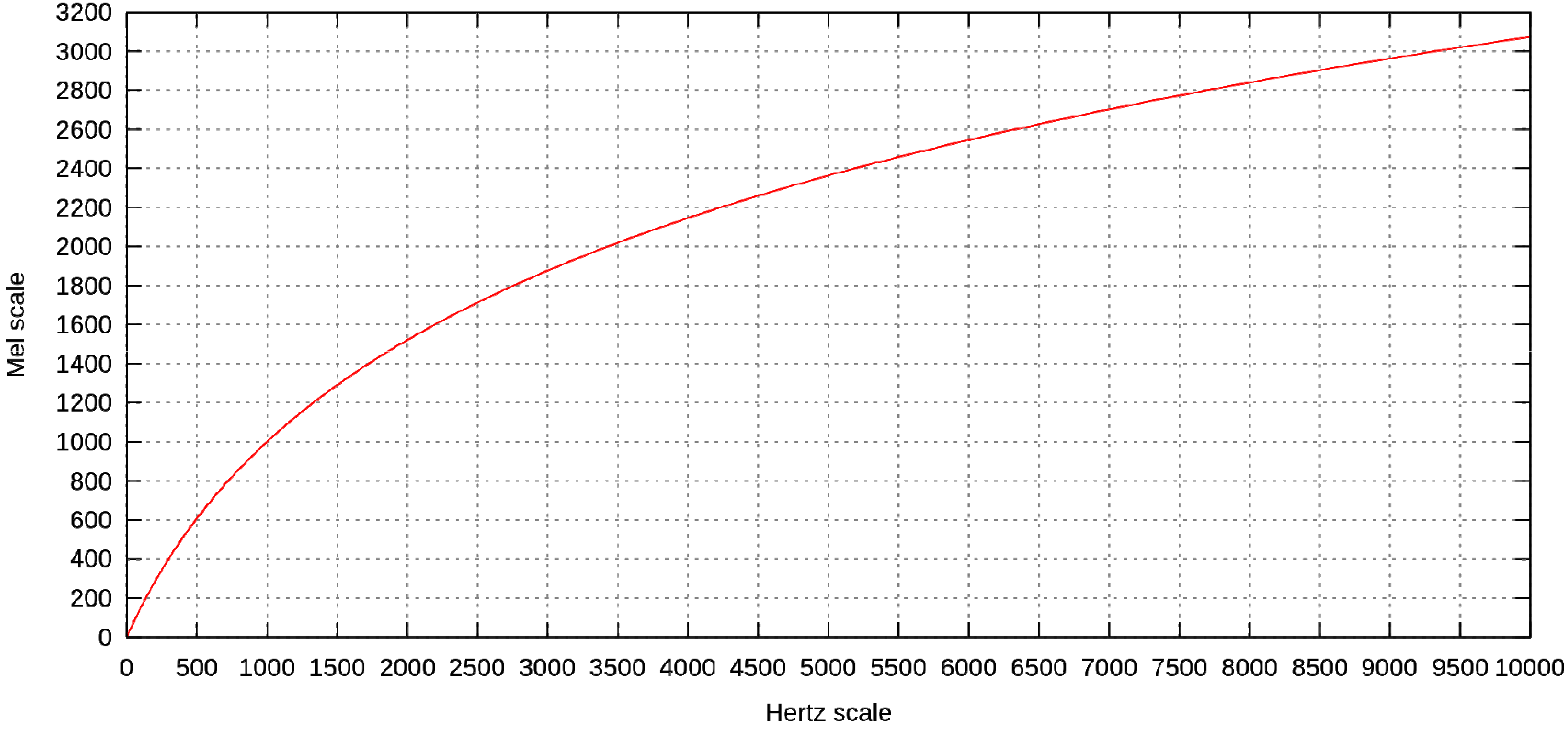
Homomorphic Vocoder



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



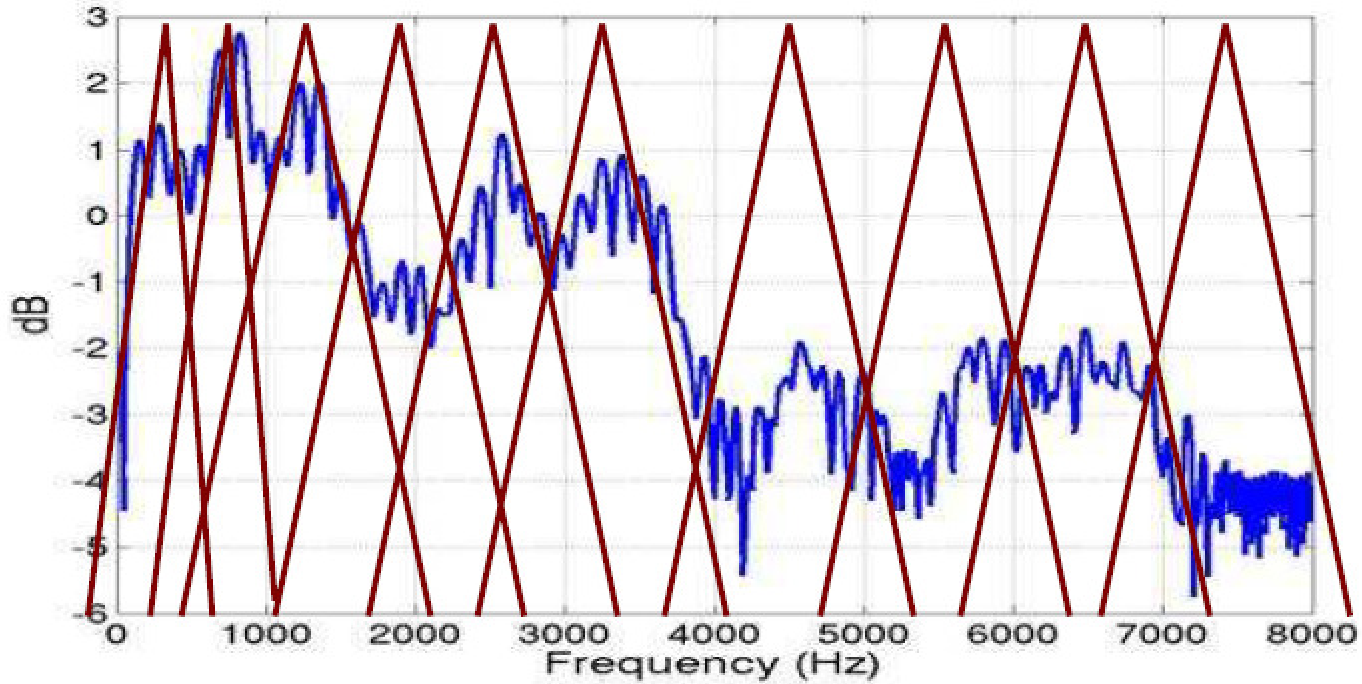
homomorphic vocoder


3200
3000
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
Mel scale
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

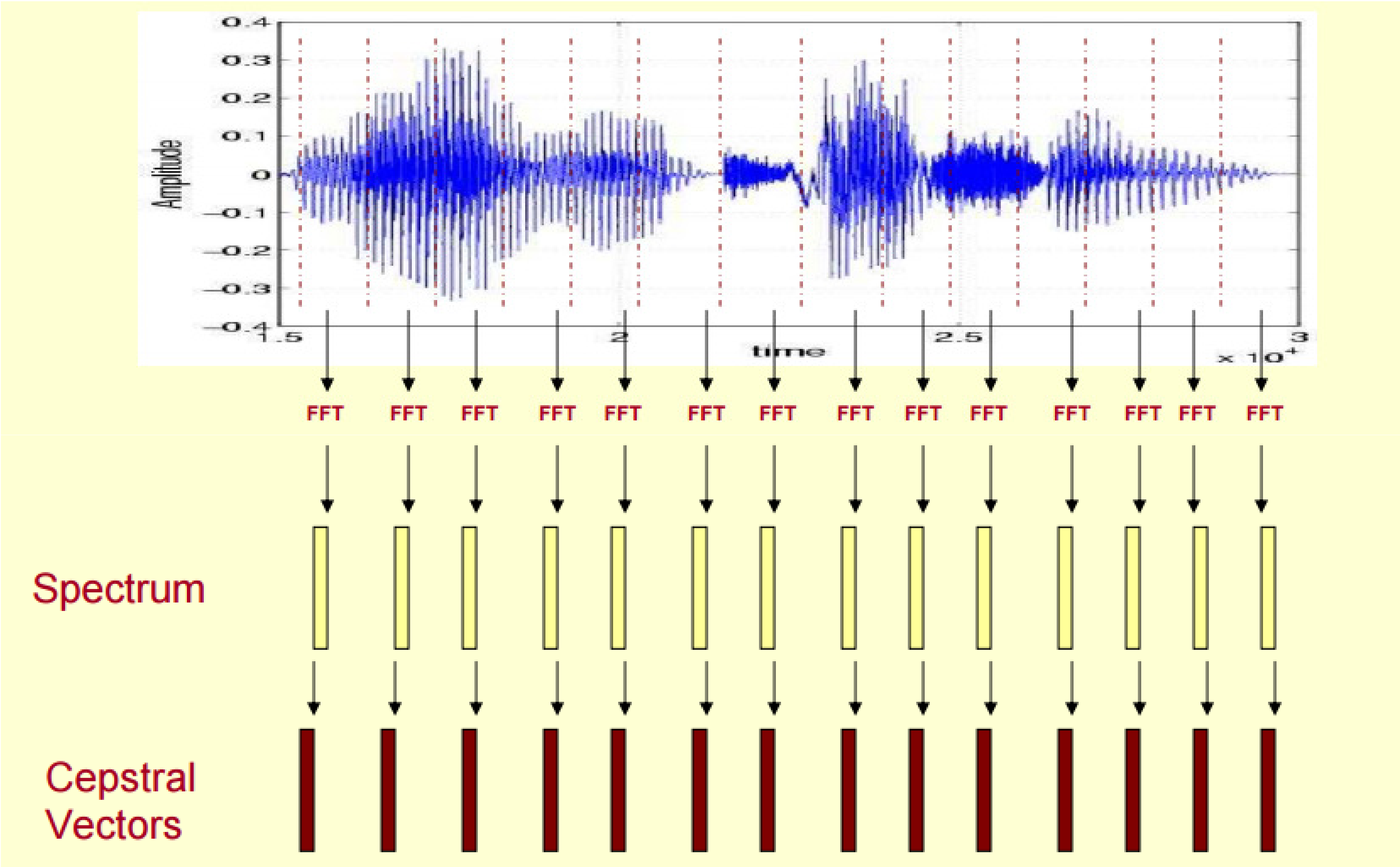
- The MFCC is no longer a homomorphic transformation
  - It would be if the order of summation and logarithms were reversed, in other words if we computed

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

  - 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



Amplitude
time
x 10⁴
FFT
Spectrum
Cepstral
Vectors