Spectral Distortion Measures

Distortion Measures : Mathematical Considerations

(a) $0 \leq d(x,y) < \infty$ — Positive Definiteness

(b) $d(x,y) = d(y,x)$ — Symmetry

(c) $d(x,y) \leq d(x,z) + d(y,z)$ — Triangle inequality

(d) $d(x+z, y+z) = d(x,y)$ — Invariant

Distance Vs Distortion

Distortion Measure

In Speech domain: Distance $\Longleftrightarrow$ Measure of Dissimilarity

Distortion Measures : Perceptual Considerations

Subjective Judgement of Sound difference (Phonetic Relevance)

Phonetically Relevant

- Formant Locations ($F_1, F_2, \ldots$)
- Formant Bandwidths
- Fundamental frequency (F_0)
- Overall Intensity

Phonetically Not Relevant

- Spectral Tilt
- Highpass filtering
- Lowpass filtering
- Bandpass / Notch filtering

Log-Spectral Distortion Measure

Spectral Distortion Measures

Log-Spectral Distance: $V(\omega) = \log S(\omega) - \log S'(\omega)$

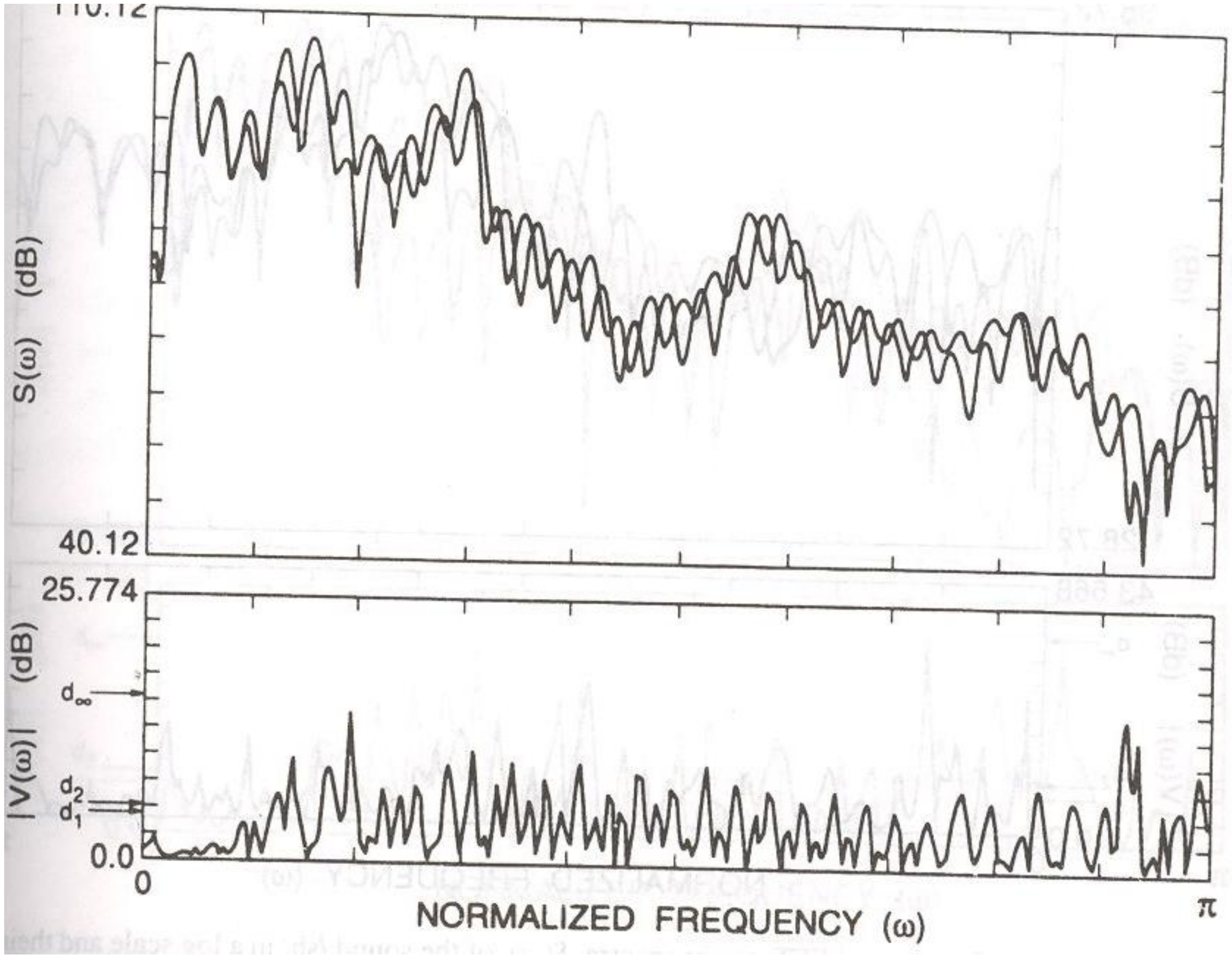
Lp Norm: $$d(S, S')^p = (d_p)^p = \int_{-\pi}^{\pi} |V(\omega)|^p \frac{d\omega}{2\pi}$$

$p = 1 \Rightarrow$ Absolute log spectral distortion

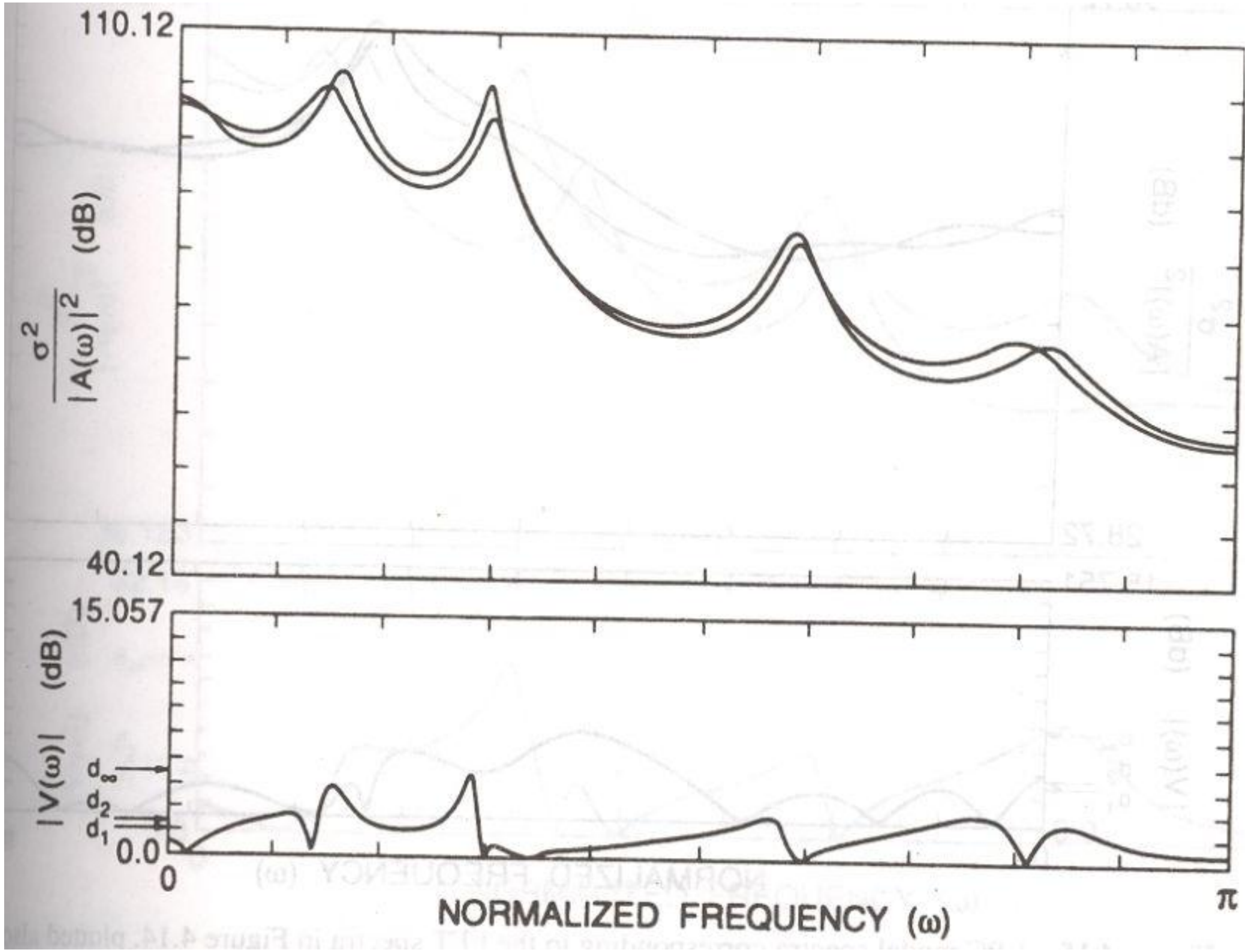
$p = 2 \Rightarrow$ RMS log spectral distortion

$p = \infty \Rightarrow$ peak log spectral distortion

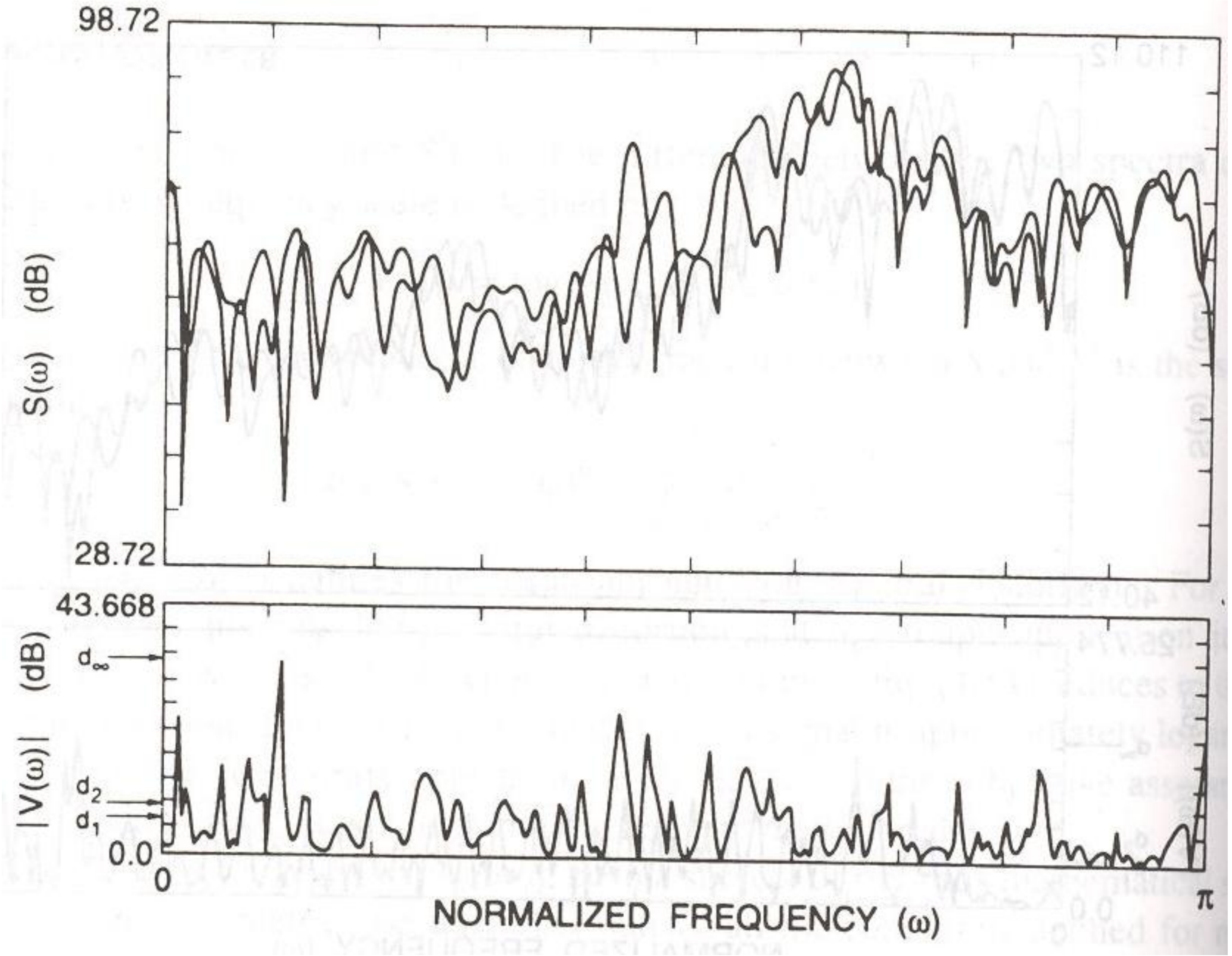
Spectral difference using FFT power spectra for the sound \e\



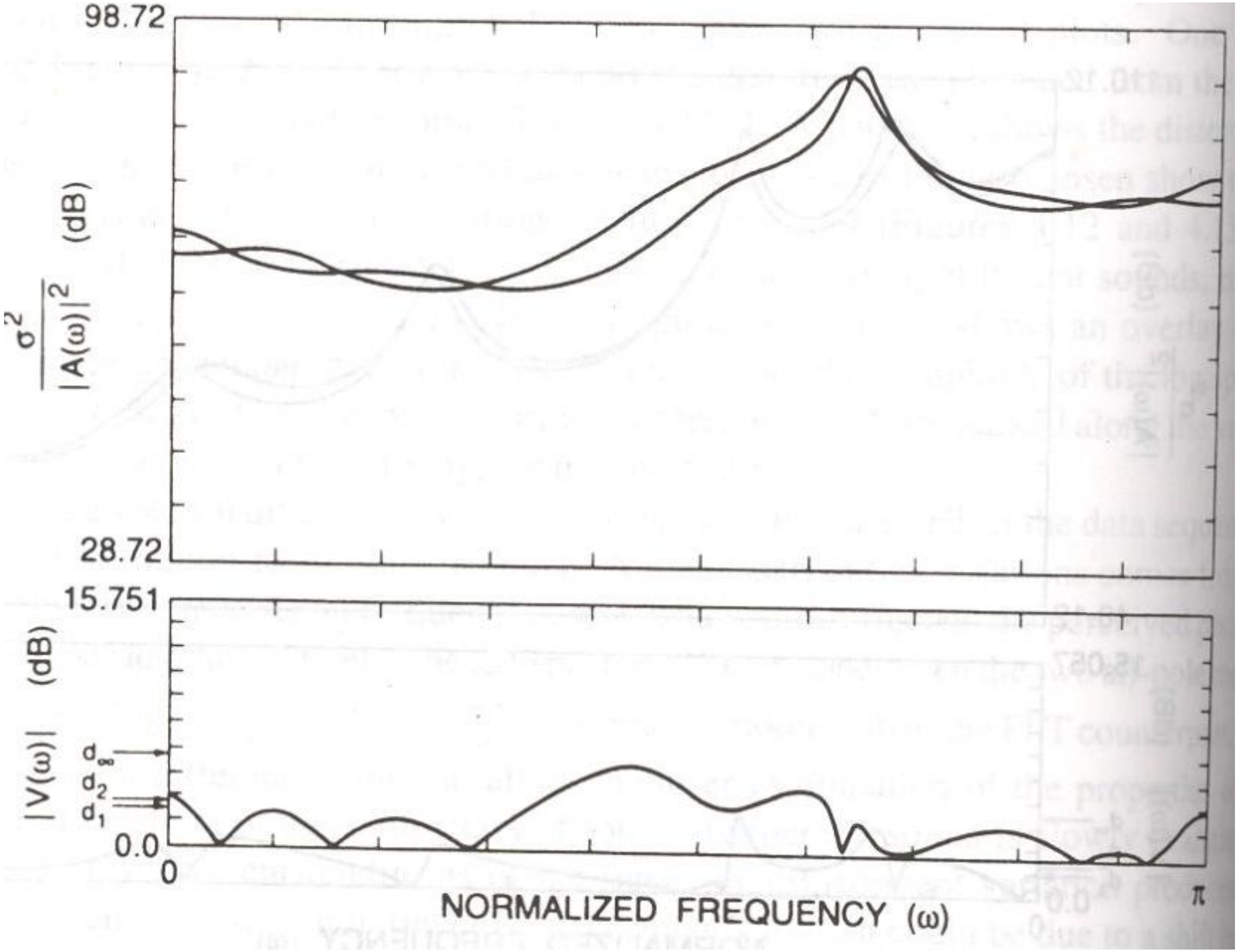
Spectral difference using LP power spectra for the sound \e\



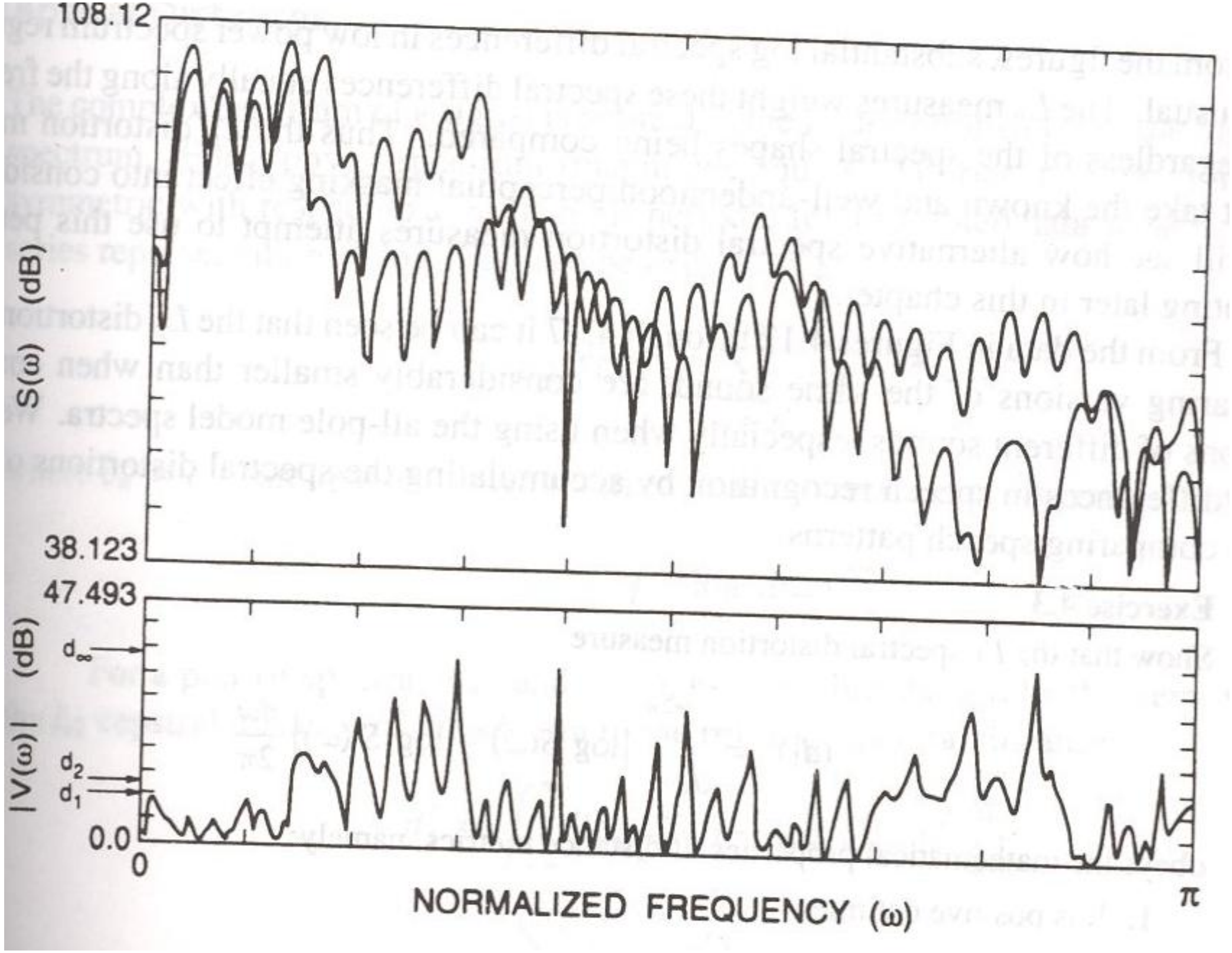
Spectral difference using FFT power spectra for the sound \sh\



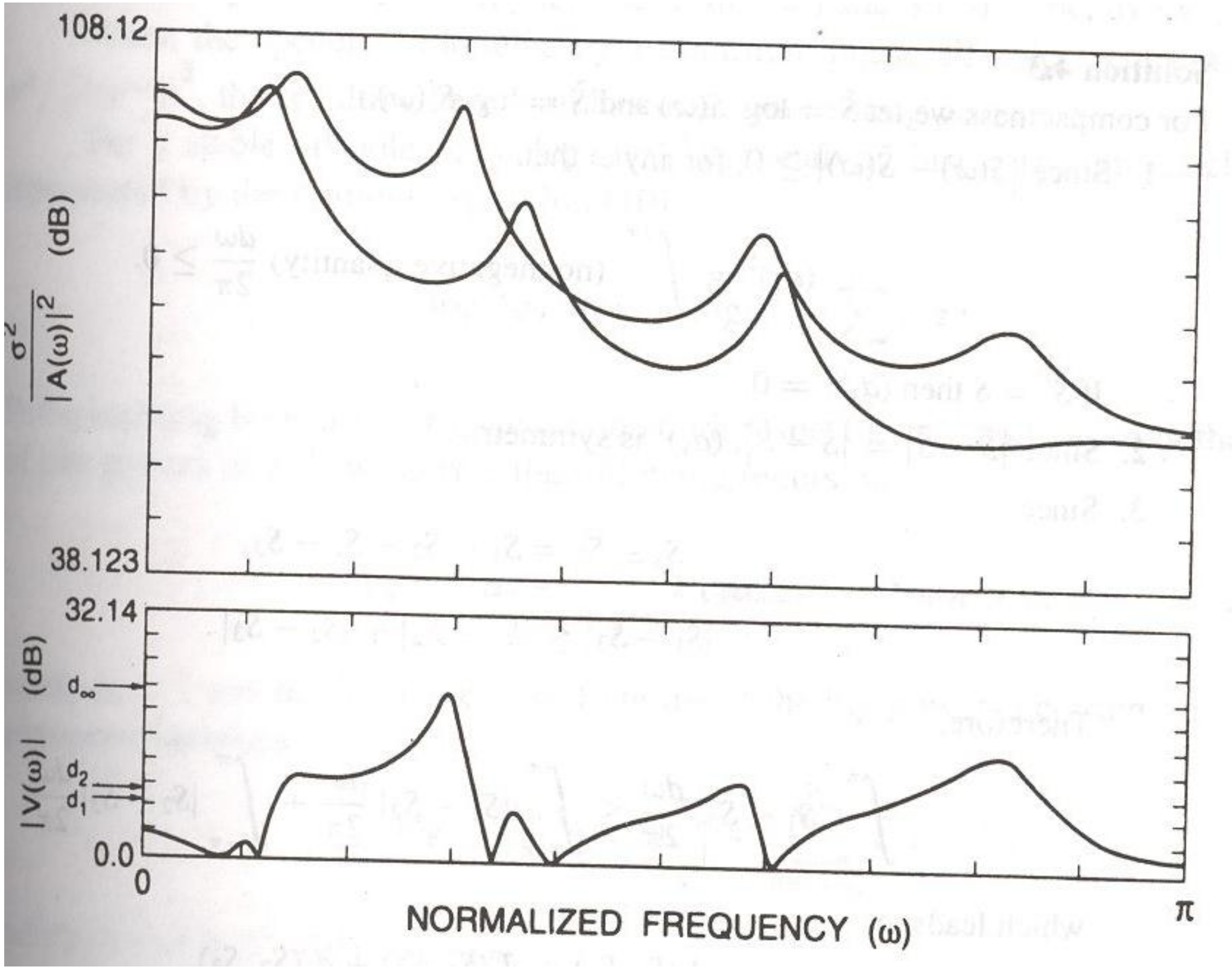
Spectral difference using LP power spectra for the sound \sh\



Spectral difference using FFT power spectra for the sounds \e\ & \i\



Spectral difference using LP power spectra for the sounds \e\ & \i\



Cepstral Distance

$$\log S(\omega) = \sum_{m=-\infty}^{\infty} c_m e^{-jm\omega}$$

$$d_2^2 = \int_{-\pi}^{\pi} \left| \log S(\omega) - \log S'(\omega) \right|^2 \frac{d\omega}{2\pi}$$

$$= \sum_{m=-\infty}^{\infty} (c_m - c_m')^2$$

Truncated cepstral distance

$$d_c^2(L) = \sum_{m=1}^{L} (c_m - c_m')^2$$

Truncated Cepstral Distance

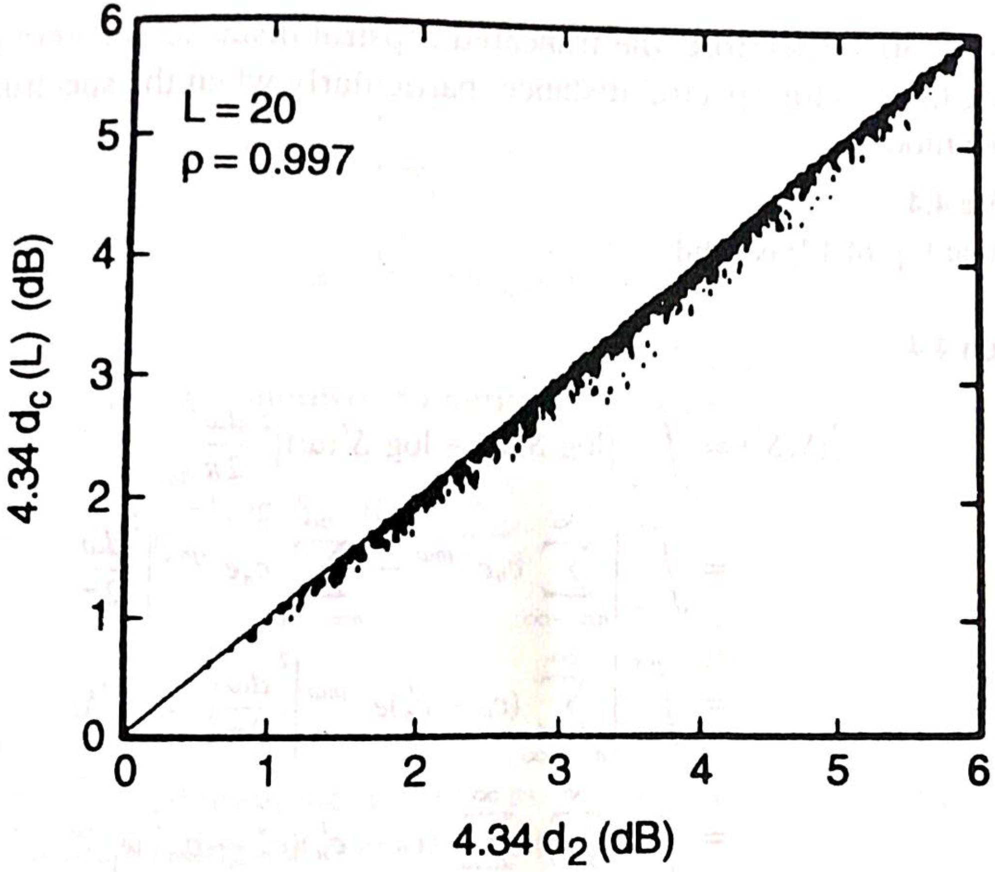


Figure 4.18 Scatter plot of d_2^2, the cepstral distance, versus $2d_c^2(L)$, the truncated cepstral distance (multiplied by 2), for 800 pairs of all-pole model spectra; the truncation is at $L = 20$ (after Gray and Markel [9]).

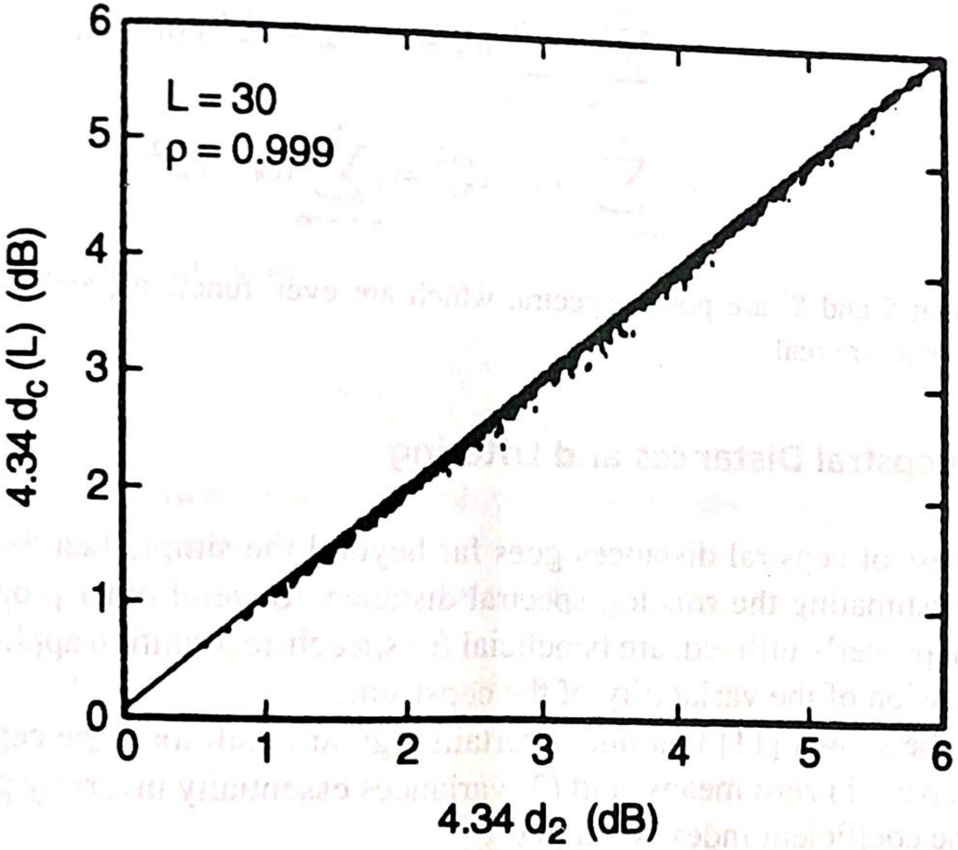


Figure 4.19 Scatter plot of d_2^2, the cepstral distance, versus $2d_c^2(L)$, the truncated cepstral distance (multiplied by 2), for 800 pairs of all-pole model spectra; the truncation is at $L = 30$ (after Gray and Markel [9]).

Weighted Cepstral Distance

$$E(c_m^2) \sim \frac{1}{m^2}$$

$$d_2^2 = \sum (c_m - c'_m)^2 \Big/ \frac{1}{m^2}$$

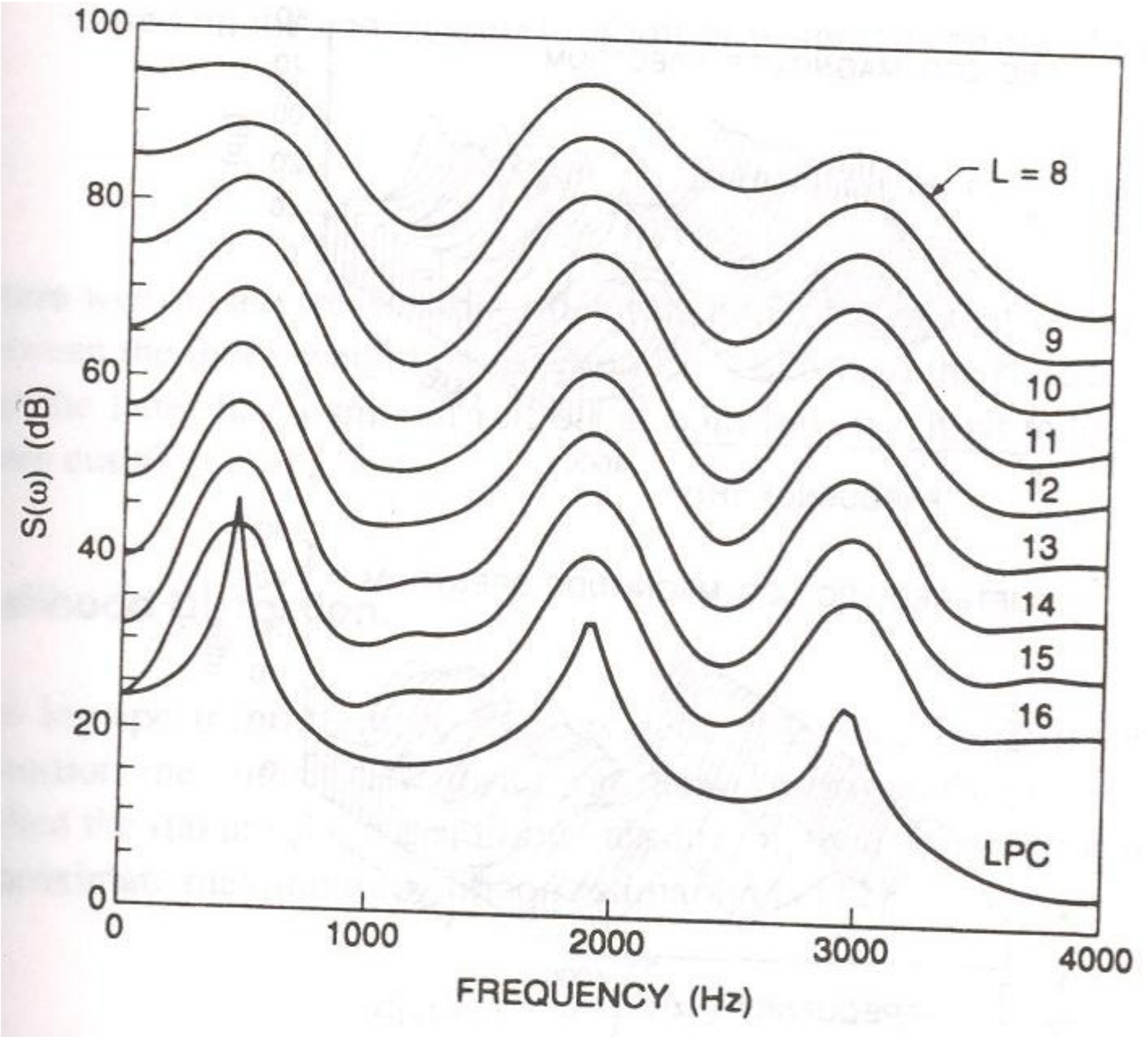
$$= \sum (m c_m - m c'_m)^2$$

Variability of Low Cepstral coeff: Transmission / Speaker char / Vocal effort

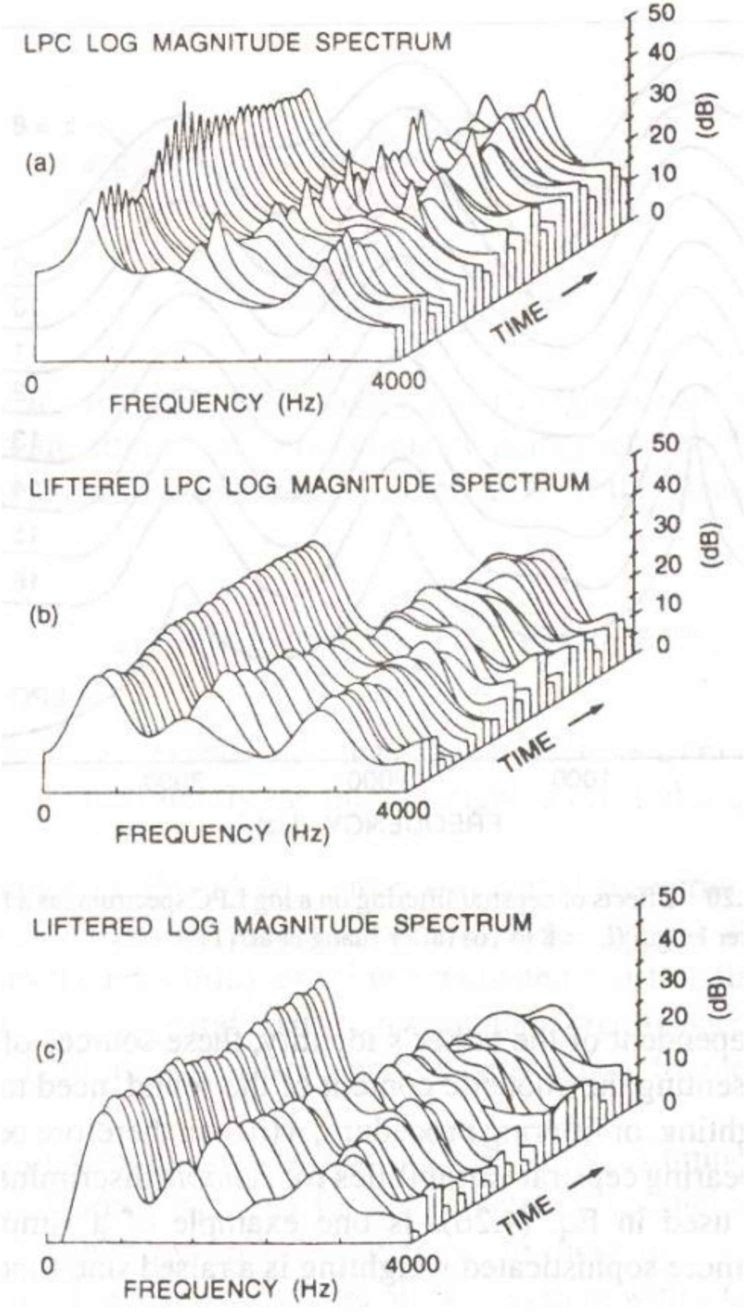
Variability of High Cepstral coeff: — Artifacts of LPC analysis (all-pole constraint, analysis window position, interaction between source & system etc..)

$$w(m) = \begin{cases} 1 + h \sin\left(\frac{m\pi}{L}\right); & m = 1, 2, \ldots L \\ 0; & m \le 0,\ m > L \end{cases}$$

Cepstrum Liftering (Weighted Cepstrum)



Comparison of Original and Liftered Magnitude Spectra



Other Distortion Measures

- Likelihood Distortion Measures : Itakura-Saito Distortion Measure
  - ✓Assymetric
- Cosh Distortion Measure
- KL Distortion Measure
- Cosine Distance