Why LPC is so popular (widely used)

- LPC provides good approximation to vocal tract spectral shape for voiced sounds
  less effective for unvoiced & transient regions.

- LPC modeling can be used to separate source & tract separation.

- LPC can be analytically tractable model, mathematically precise. Computationally less complex compared to filter bank-analysis

- LPC model works well for speech recognition

LPC Model

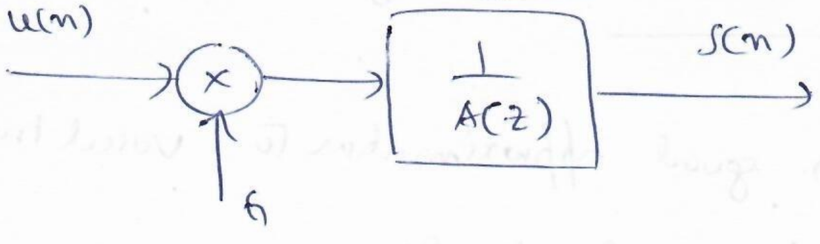
$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) + \cdots + a_p s(n-p)$$

$$= \sum_{k=1}^{p} a_k s(n-k) + G u(n)$$

$$S(z) = \sum a_k z^{-k} S(z) + G U(z)$$

$$H(z) = \frac{S(z)}{G U(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \frac{1}{A(z)}$$

generating filter



Synthesis model

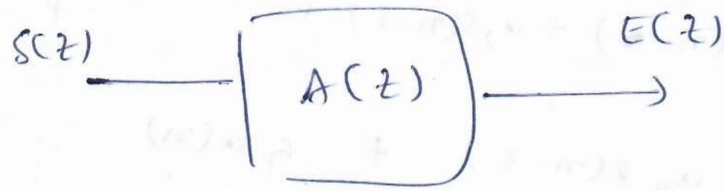
$$e(n) = s(n) - \hat{s}(n)$$

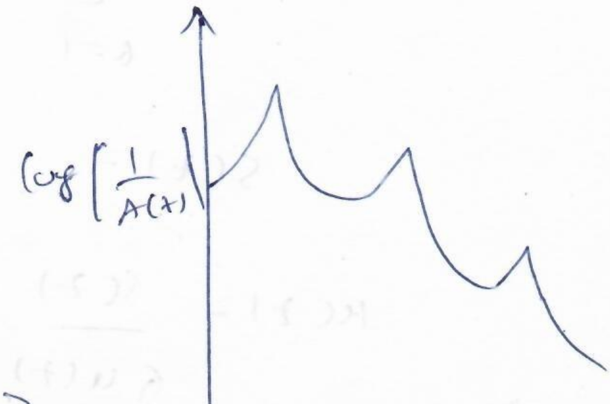
$$= s(n) - \sum_{k=1}^{p} a_k s(n-k)$$

$$E(z) = S(z)\left[1 - \sum_{k=1}^{p} a_k z^{-k}\right]$$

Analysis model (inverse filter formulation)

$$\frac{E(z)}{S(z)} = 1 - \sum_{k=1}^{p} a_k z^{-k} = A(z)$$





LP analysis of speech

$$\hat{s}(n) = \sum_{k=1}^{p} a_k s(n-k)$$

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^{p} a_k s(n-k)$$

$$E = \sum_n e^2(n) = \sum \left[s(n) - \sum_{k=1}^{p} a_k s(n-k) \right]^2$$

optimum values of a_k's

$$\frac{\partial E}{\partial a_k} = 0 \qquad k = 1, 2, \ldots p$$

$$\sum_k a_k \sum_m s_{m-k} s_{m-i} = -\sum s_m s_{m-i}$$

$1 \le i \le p$

Autocorrelation Method

$$\sum_k a_k R(i-k) = R(i)$$

$i = 1, 2, \ldots p$

$$\begin{bmatrix} R(0) & R(1) & R(2) & \cdots & R(p-1) \\ R(1) & R(0) & R(1) & \cdots & R(p-2) \\ \vdots & & & & \\ R(p-1) & R(p-2) & R(p-3) & \cdots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix}$$

Co-variance Method

$$\phi(i,0) = \sum_{k=1}^{p} a_k \phi(i,k)$$

$$\begin{bmatrix} \phi(1,1) & \phi(1,2) & \phi(1,3) & \cdots & \phi(1,p) \\ \phi(2,1) & \phi(2,2) & \phi(2,3) & \cdots & \phi(2,p) \\ \vdots & & & & \\ \phi(p,1) & \phi(p,2), & \phi(p,3) & \cdots & \phi(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \phi(1,0) \\ \phi(2,0) \\ \vdots \\ \phi(p,0) \end{bmatrix}$$

$$\phi(i,k) = \sum_{n} s(n-i)\, s(n-k)$$

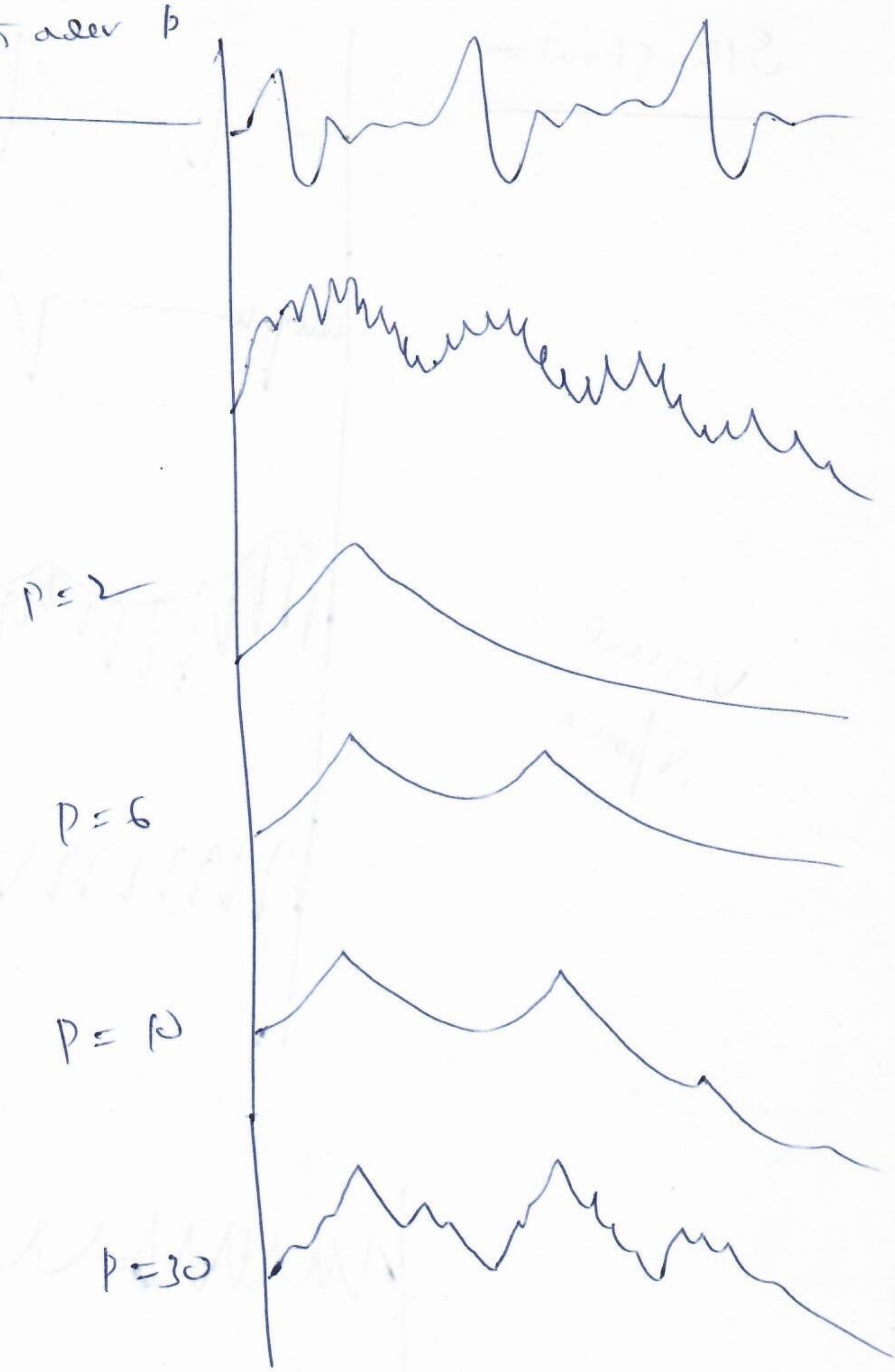
Total error (E_{min})

$$E = \sum_{n} e^2(n) = \sum_{n} \left[s(n) - \sum_{k=1}^{p} a_k s(n-k) \right]^2$$

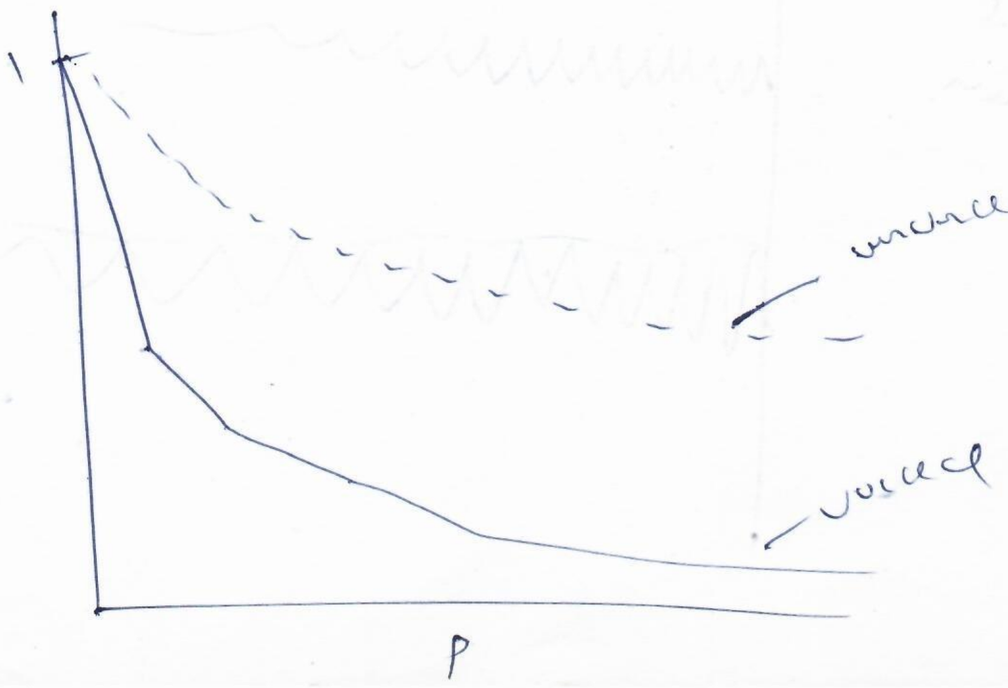
$$= \sum_{n} s^2(n) - \sum_{k=1}^{p} a_k \sum_{n} s(n)\, s(n-k)$$

$$= R(0) - \sum_{k=1}^{p} a_k R(k)$$

LP spectrum with order p

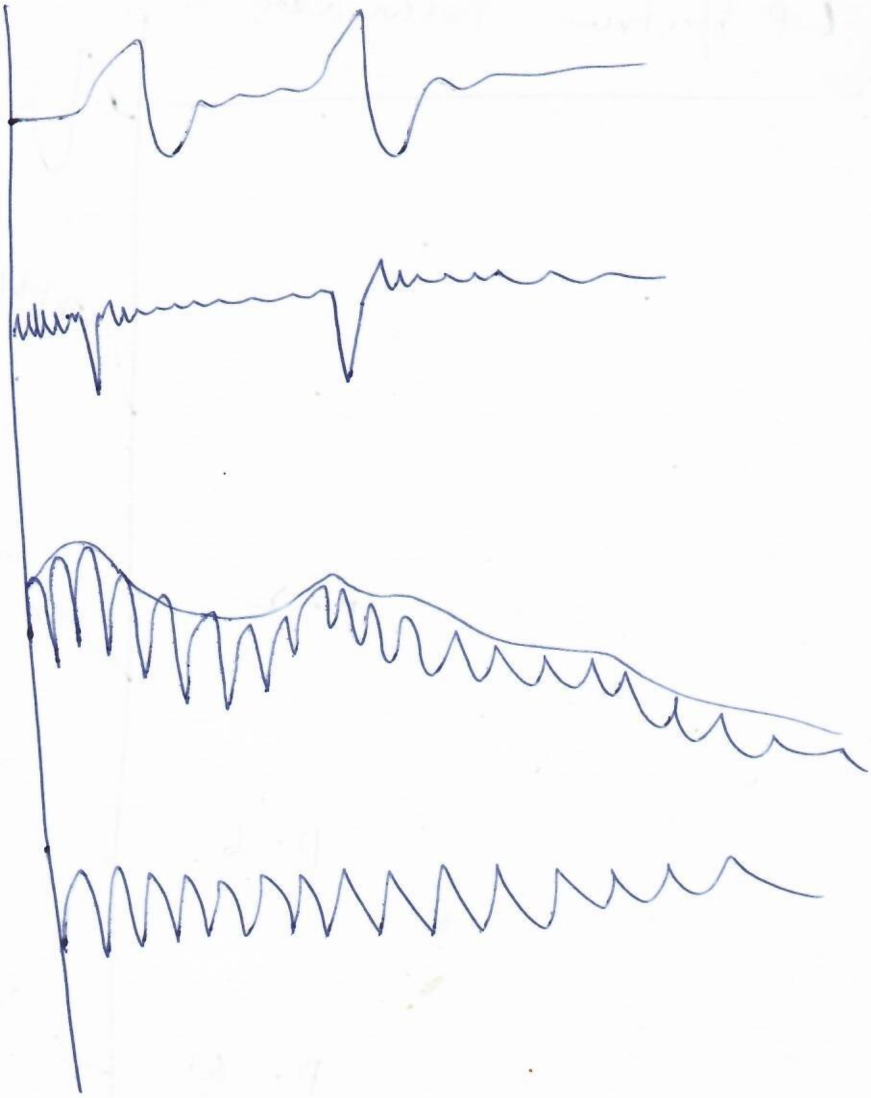


Normalized error curve

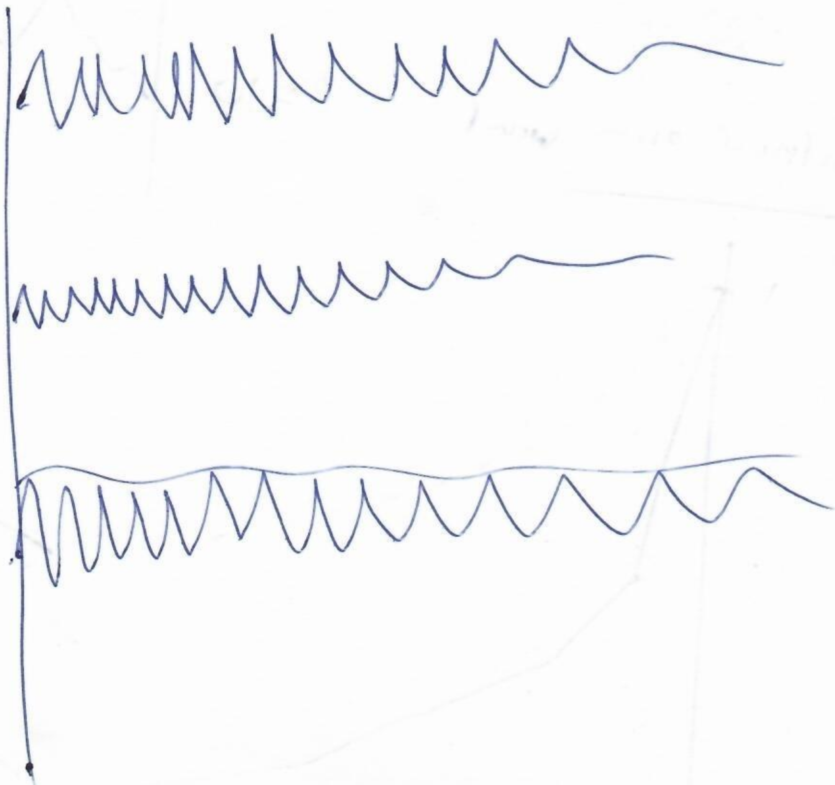


Illustrations

Voiced Speech



unvoiced Speech



Frequency domain formulation

$$E(z) = \left[1 + \sum_{k=1}^{p} a_k z^{-k}\right] S(z) = A(z)\, S(z)$$

Parseval's theorem

$$E = \sum_{-\infty}^{\infty} e^2(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|E(e^{j\omega})\right|^2$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|A(e^{j\omega})\right|^2 \left|S(e^{j\omega})\right|^2$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} A(e^{j\omega})\, A(e^{-j\omega}) \,.\, P(\omega)$$

minimize E w.r.t a_k $\Rightarrow$ $\frac{\partial E}{\partial a_k} = 0$

$$R(i) = \sum_{k=1}^{p} a_k R(i-k) \qquad i = 1, 2, \dots p$$

$$R(i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P(\omega)\, e^{j\omega i}\, d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} P(\omega) \cos(\omega i)\, d\omega$$

LP Spectral Matching

Signal Spectrum $= P(\omega)$

All-pole model spectrum $= \hat{P}(\omega)$

$$\hat{P}(\omega) = |H(e^{j\omega})|^2 = \frac{\sigma^2}{|A(e^{j\omega})|^2} = \frac{\sigma^2}{|1+\sum a_k e^{-j\omega k}|^2}$$

$$P(\omega) = \frac{|E(e^{j\omega})|^2}{|A(e^{j\omega})|^2}$$

$$P(\omega) = \hat{P}(\omega) \Rightarrow |E(e^{j\omega})|^2 = \sigma^2$$

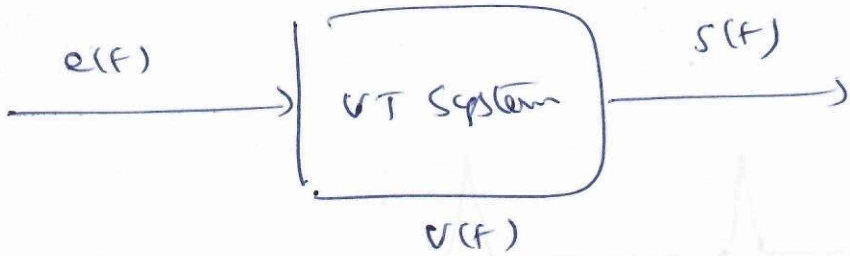
Filter $A(z) \to$ whitening filter

$$E = \frac{1}{2\pi}\int_{-\pi}^{\pi} |E(e^{j\omega})|^2 = \frac{1}{2\pi}\int |A(e^{j\omega})|^2 |S(e^{j\omega})|^2$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{P(\omega)}{\hat{P}(\omega)} \sigma^2 d\omega = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{P(\omega)}{\hat{P}(\omega)} d\omega$$

σ can be obtained by $\hat{R}(0) = R(0)$

Cepstrum analysis



$$s(t) = e(t) * v(t)$$

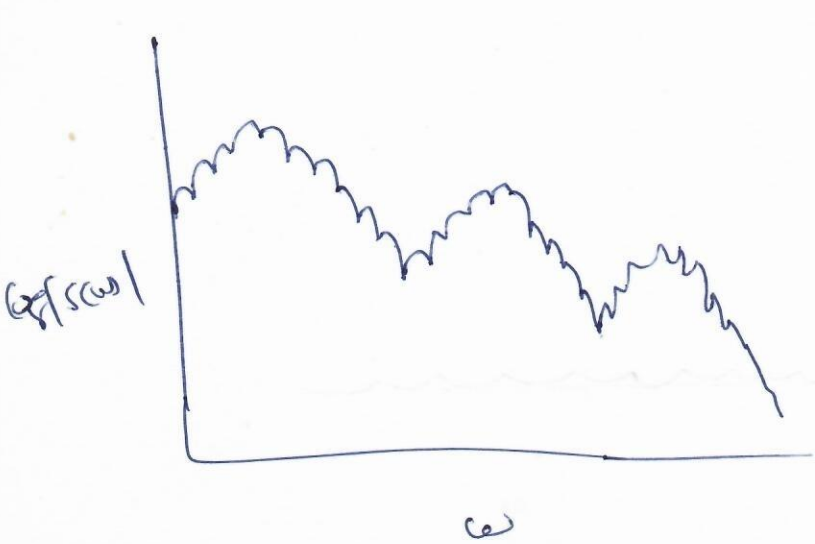
$$S(\omega) = E(\omega) \cdot V(\omega)$$

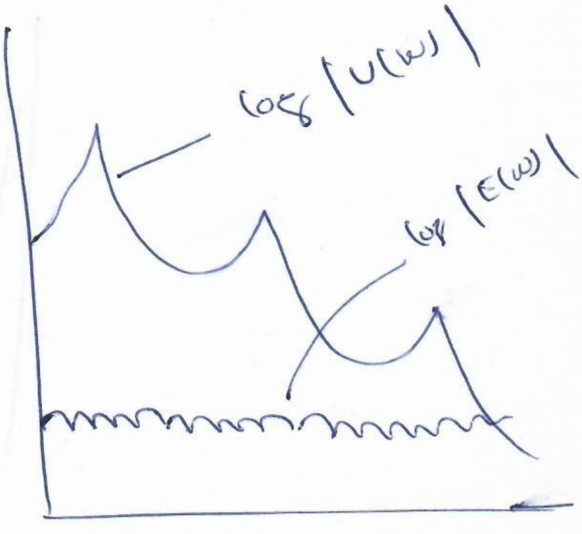
$$\log(S(\omega)) = \log(E(\omega)) + \log(V(\omega))$$

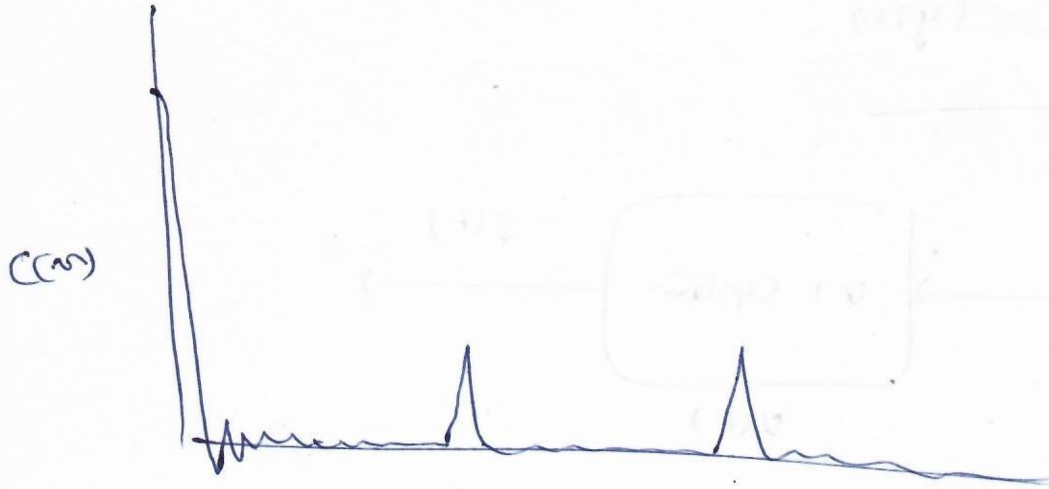
$$S(\omega) = |S(\omega)| \, e^{j\theta(\omega)}$$

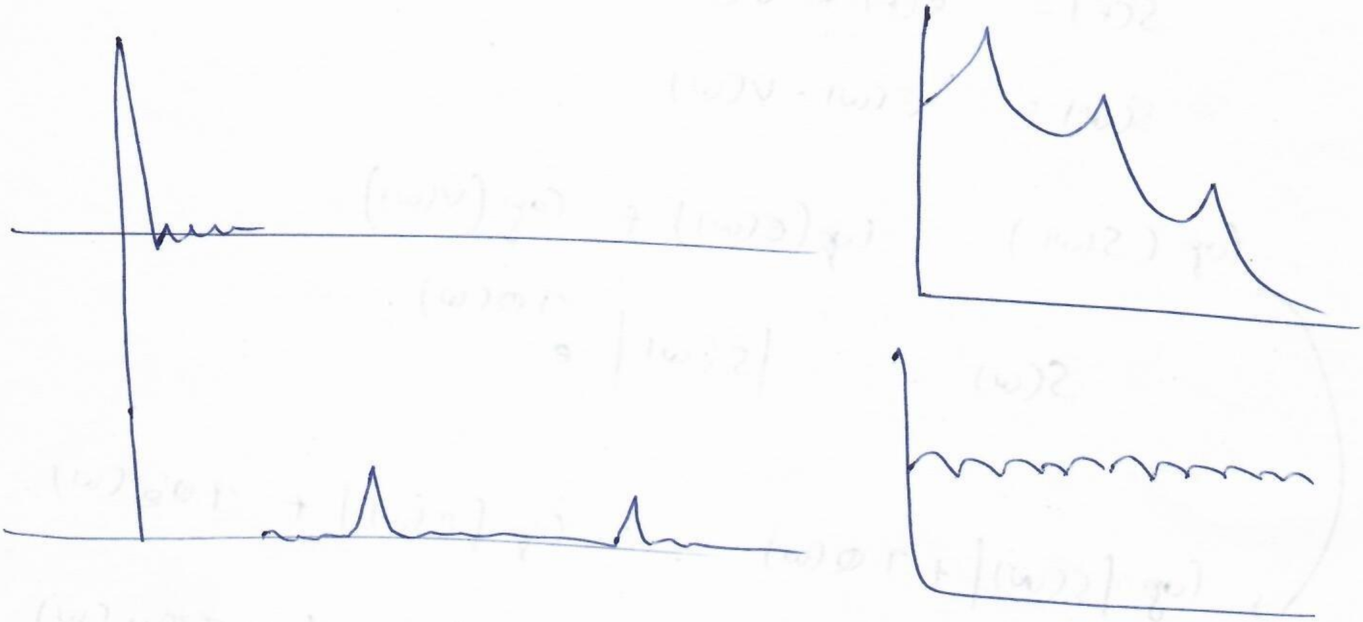
$$\log|S(\omega)| + j\theta(\omega) = \log|E(\omega)| + j\theta_e(\omega) + \log|V(\omega)| + j\theta_v(\omega)$$

$$\log|S(\omega)| = \log|V(\omega)| + \log|E(\omega)|$$

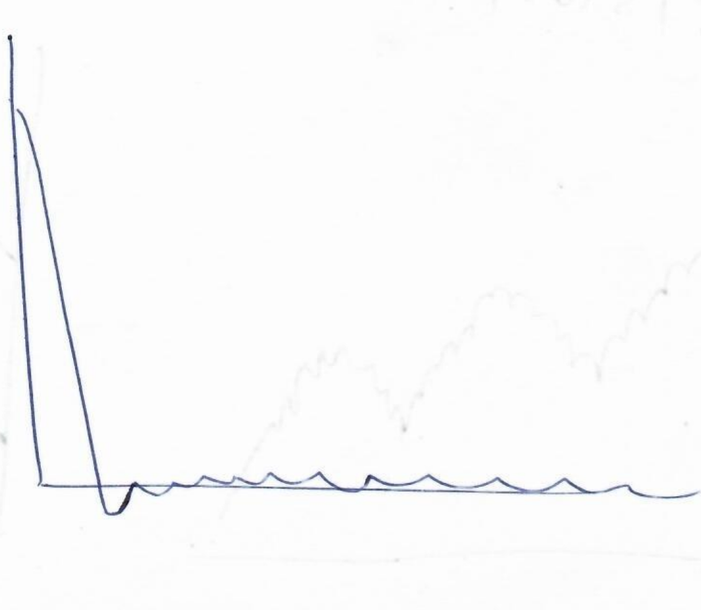








unvoiced cepstrum



Steps in Deriving the LPCCs from Speech

Preemphasis

Frame blocking

windowing

Autocorrelation analysis

LPC analysis

Cepstral analysis

cepstral weighting

Cepstral derivative

Steps in Computation of PLPs

FFT (Power Spectrum)

Critical band integration & resampling

- Trapezoidal filters (approximate to critical band masking curve)

Equal loudness curve - weighting

Power law of hearing (cube root on spectral content)

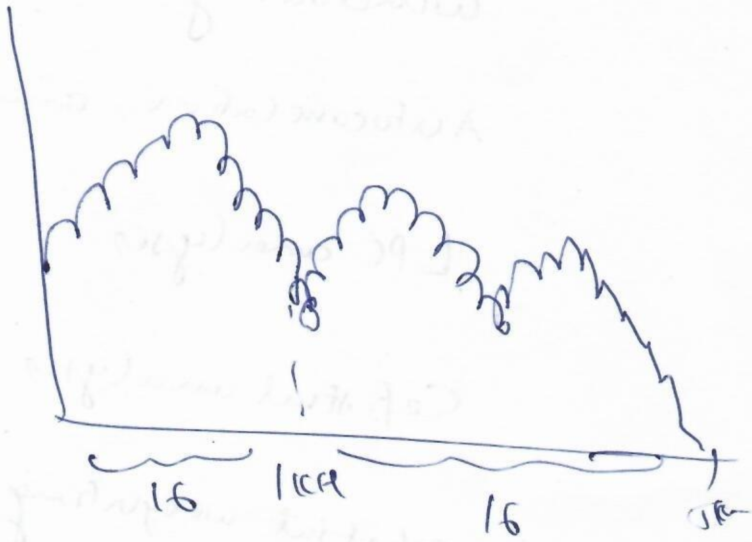
IDFT (Mel linear auto correlation coefficients)

Spectral smoothing

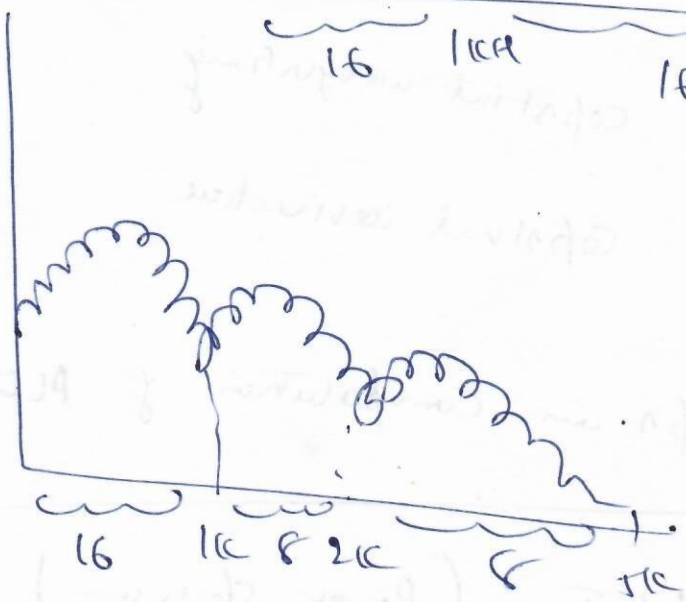
Solving the set of linear equations

cepstral recursion

Spectral coefficient



Mel-spectral



Mel-cepstral

IFT [Mel-spectral coefficients]

Different formulations of LP analysis

Mean Square error

Wiener filter

Autocorrelation matching

Autocovariance

Spectrum matching

Maximum entropy

Maximum likelihood

Lattice filter

$$E_{i+1} = \left(1 - k_i^2\right) E_i \qquad i = 1, 2, \ldots p$$

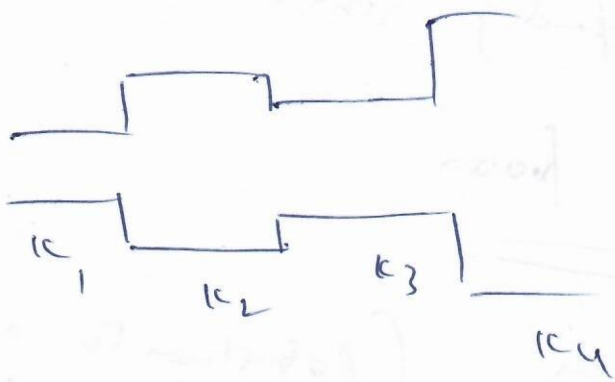
$$E_p = \prod_{i=1}^{p} \left(1 - k_i^2\right)$$

k_i — reflection coefficients

PARCOR, Autocorrel coeff

LPCs

Log area coefficients



$$g_m = \log\left(\frac{1 - k_i}{1 + k_i}\right)$$

Dynamic features

$$\Delta C_i(n) = \frac{\sum_{k=-N}^{N} k\, C_i(n+k)}{\sum_{k=-N}^{N} k^2}$$

RASTA-PLP

CMS: Cepstral Mean Subtraction

Robustness to Convolutional error.

Spectral Analysis

Bank of Compressing Static Nonlinearities — log

Bank of Linear BPF

Bank of expanding Static Nonlinearities — exponentiated

PLP process

Spectral Subtraction (Robustness to additive Noise)