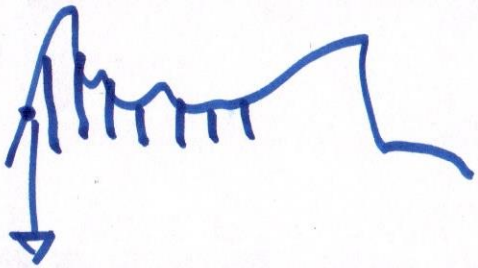


16-bit PCM

1 16-bit level

Speech/Hearing properties to exploit to reduce bit rate:

1. Correl^n between samples
2. Periodicity (due to vocal cords vibration)
3. Residual (source excit^n) can be modeled simply.
4. VT & Fo vary relatively slowly
5. Ear is rel. phase-insensitive (in the short term)
6. Masking makes noise less audible in strong-freq regions
7. Non-unif. freq resol^n of the ear.

2.4 kbps LPC Vocoder

Bit Allocation / Frame (20 ms)

CDEEP
IIT Bombay

10 LSFs : 4 x 4 bits + 6 x 3 bits = 34 bits

(4 → lower LSFs)

Gain : 6 bits

Pitch : 8 bits

V/UV : 1 bit

Sync : 1 bit

---

50 bits / 20 ms

= 2.4 kbps , MOS = 2.5

Waveform Coder

16k 128 kbps

Model - based

2.4 kbps

Hybrid Codec

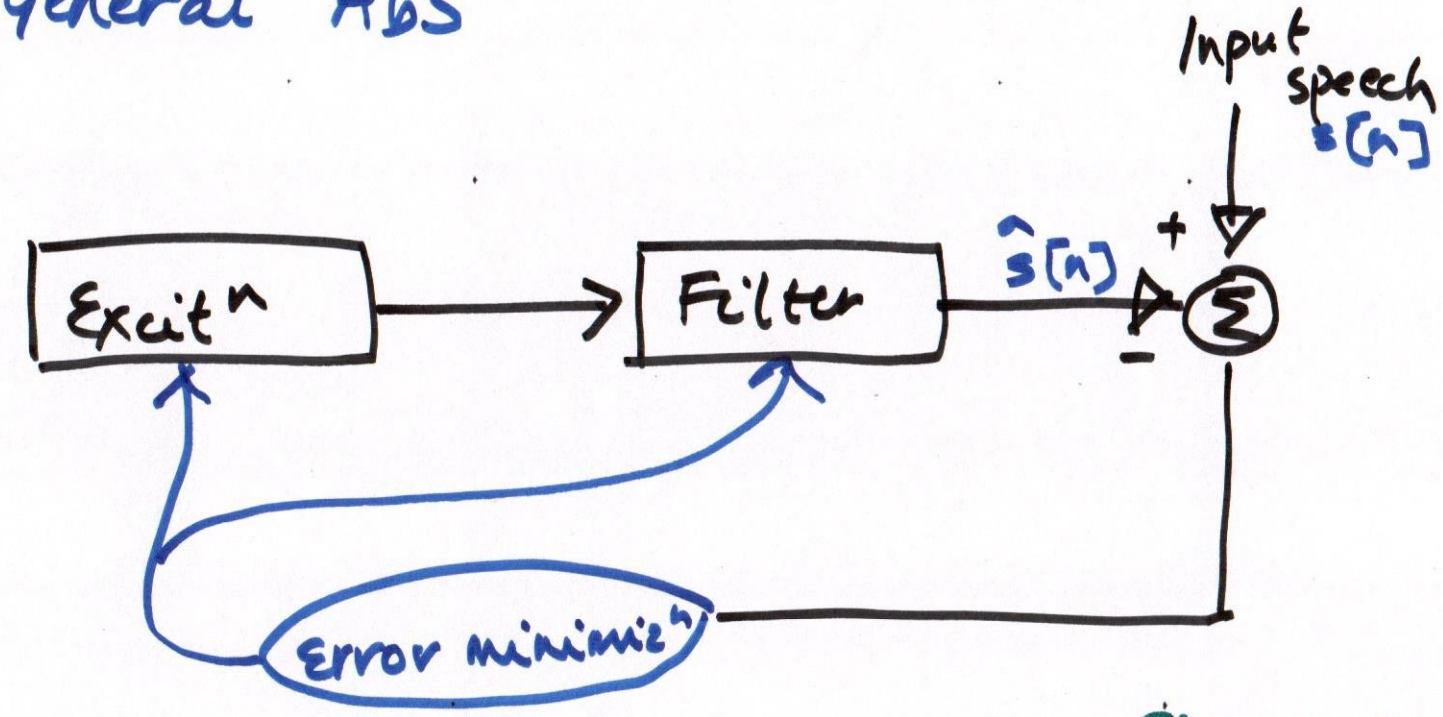
Analysis-by-Synthesis (AbS)

CDEEP
IIT Bombay

Does not follow the approach of finding model parameters & then quantising them.

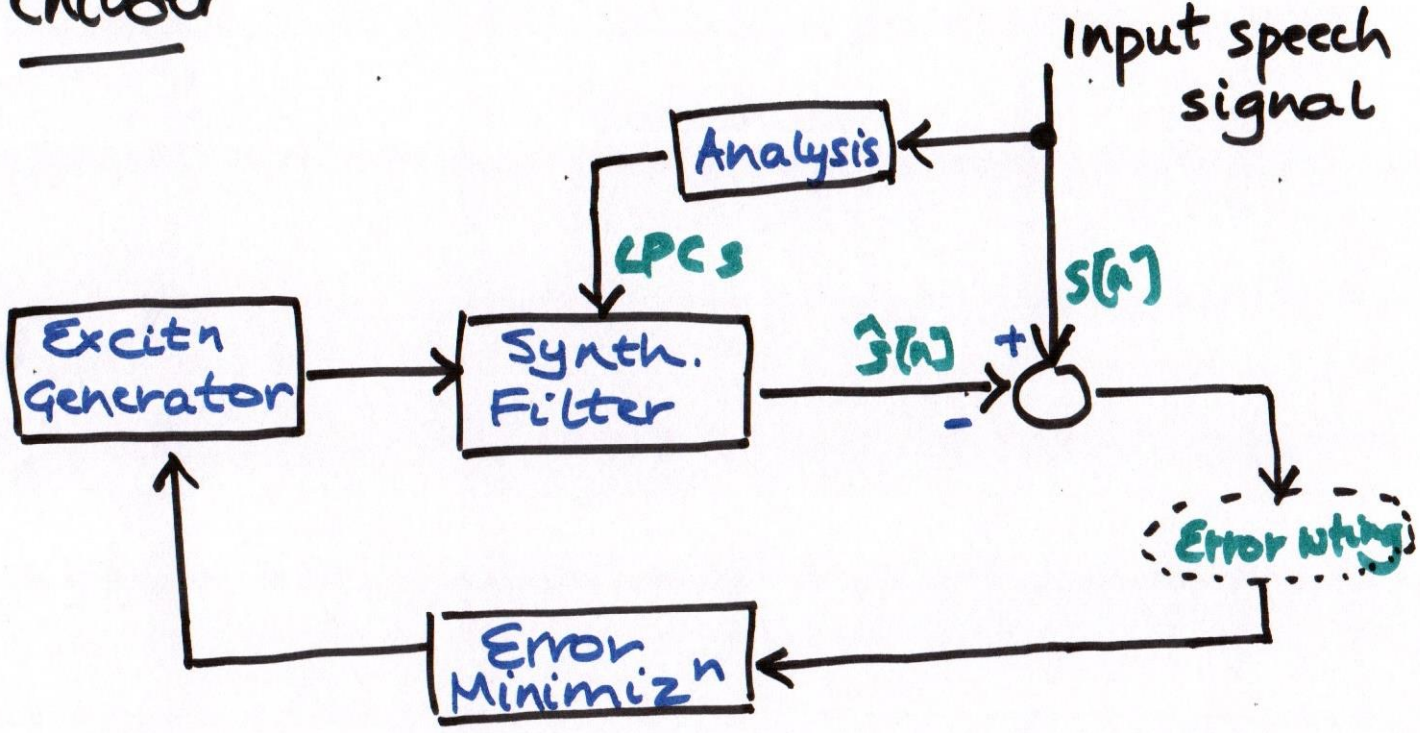
But rather gets quantised parameters by the synthesis of candidate output speech signals using all possible of a set of model parameters and identifying those that jointly minimise the error between the synthesized speech and the input speech.

General Abs



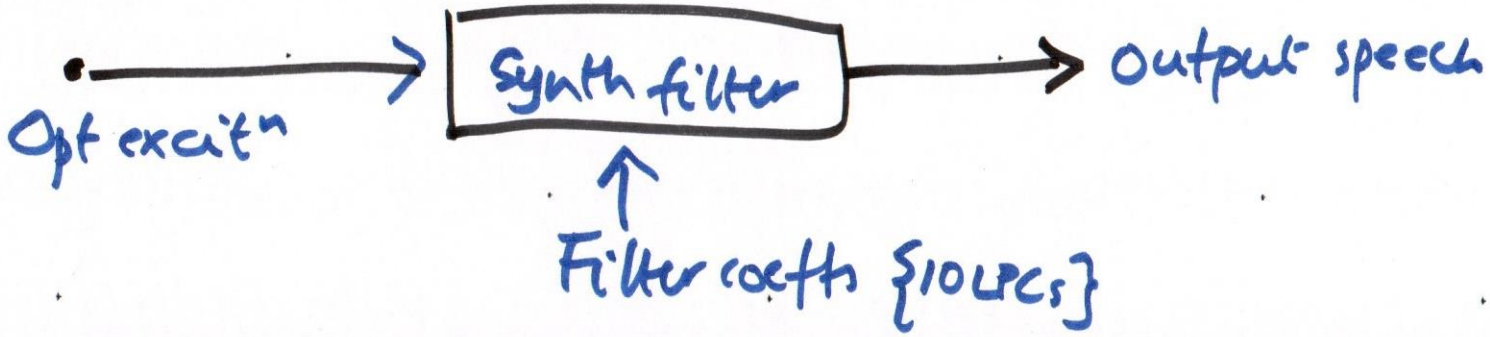
Closed-loop optimiz^n

Encoder

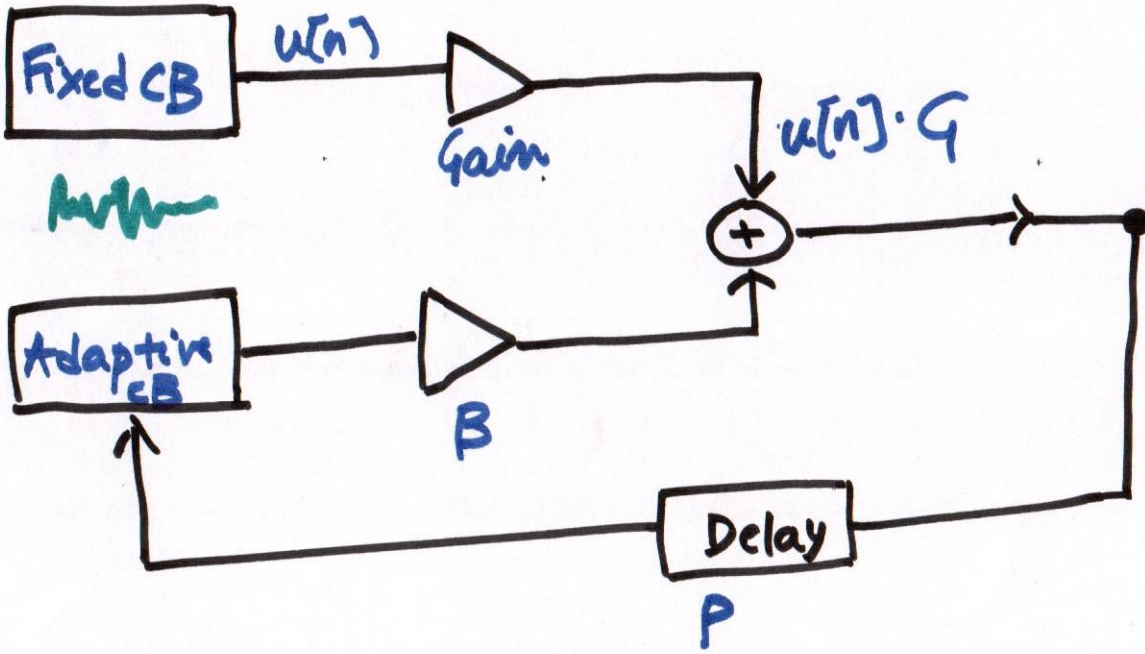


Transmitted info
- Filter coeffs
- Excitn params

Decoder



Excitation Generator



$$y[n] = G u[n] + B\, y[n-P]$$

Error minimizn

$$\sum_{\text{frame}} e^2[n] = \int |E(\omega)|^2 d\omega$$

$$= \int |S(\omega) - \hat{S}(\omega)|^2 d\omega$$

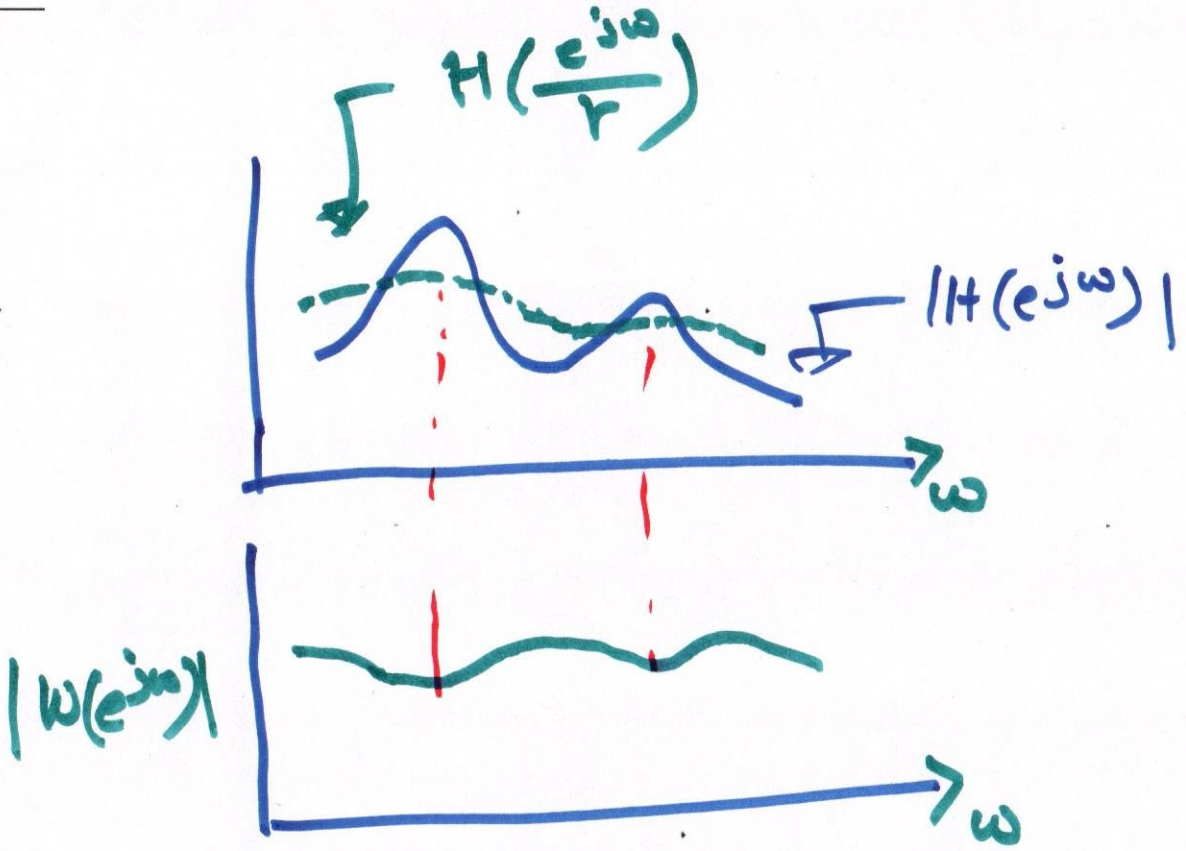
CDEEP
IIT Bombay

Error wting Filter:

$$W(z) = \frac{A(z)}{A(z/r)} = \frac{H(\frac{z}{r})}{H(z)}, \quad r<1$$

If $z = r z_0$ is a pole of $H(z)$, then

$z = r z_0$ " " " of $H(\frac{z}{r})$



$\Rightarrow$ we're minimizing $\int |W(e^{j\omega}) \cdot E(e^{j\omega})|^2 \, d\omega$

CELP (Code-excited LP): FS1016

Uses 30ms frame size

Bit allocation:

10 LSFs : $4 \times 4 + 6 \times 3 = 34$

Pitch : delay : 4×7

gain : 4×5

Codebook : index : 4×9

gain : 4×5

---

138 bit / 30ms $\rightarrow$ 4.8 kbps