LINEAR PREDICTIVE CODING for speech (vocoders)

LPC → specific, used for speech

PCM → can be used for anything other than speech also

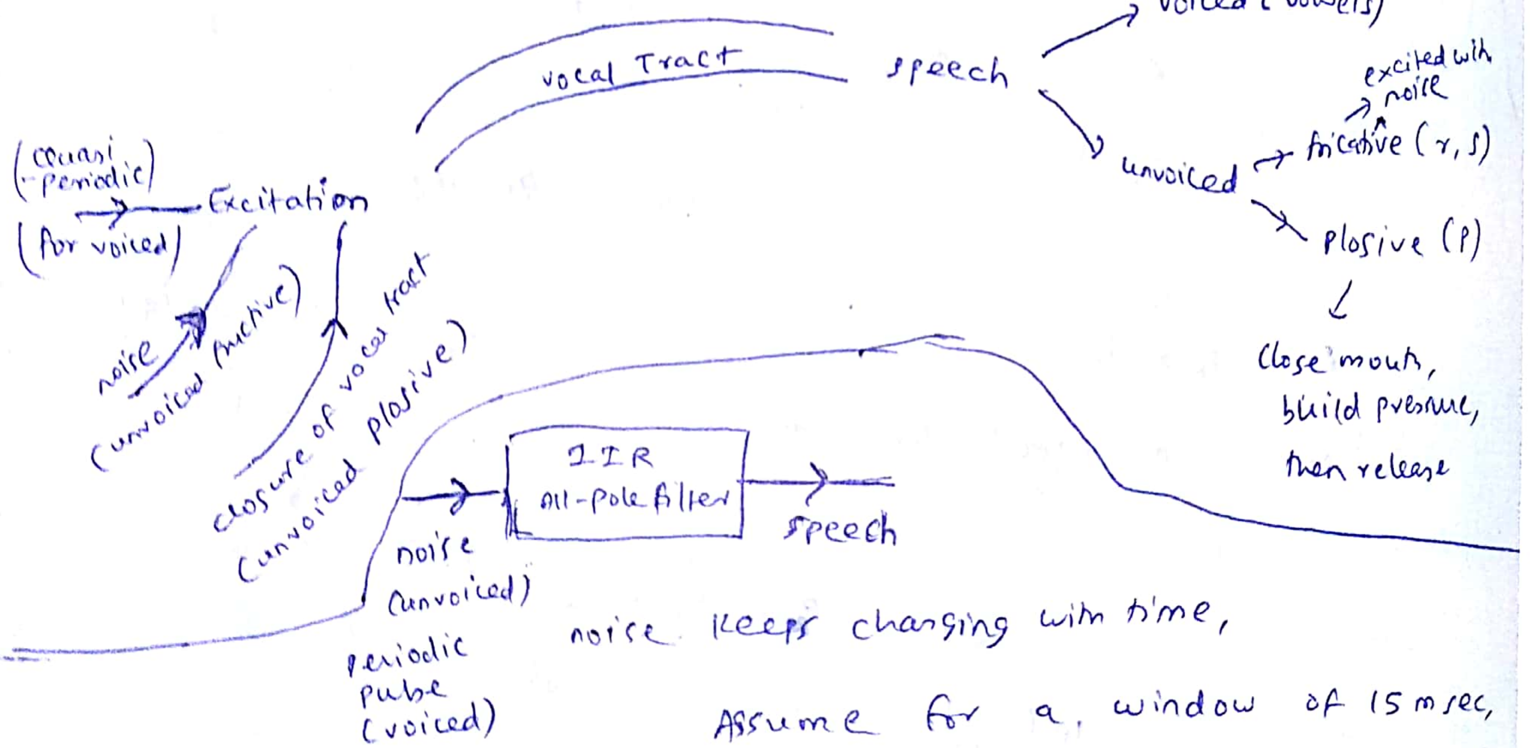
because it models human vocal tract.

**Type of Source Coding**

CELP

Code ~~Exce~~ Excited LPC

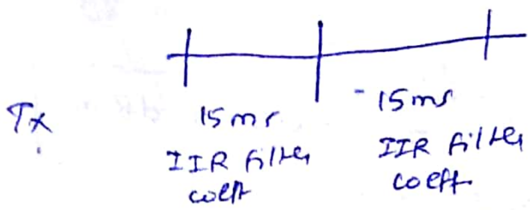
IS 95 CDMA




noise keeps changing with time,

Assume for a window of 15 msec, speech characteristics will be same.

From Tx, IIR filter coefficients are only transmitted



$\cancel{y(n) = a_1 y(n-1)}$

For PCM, we directly see encoded bits

For DPCM, we transmit the prediction error.

But here, " " " IIR filter coeffts.

$$y(n) = a_1 y(n-1) + a_2 y(n-2) + \cdots a_M y(n-M) + G x(n)$$

$$\frac{Y(z)}{X(z)} = \underset{\text{SYNTHESIS FILTER}}{H(z)} = \frac{G}{(1 + a_1 z^{-1} + \cdots a_M z^{-M})}$$

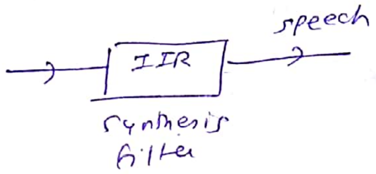
IIR filter $\quad H(z) = \dfrac{G}{1 + \sum\limits_{k=1}^{M} a_k z^{-k}} \qquad z = e^{j\omega}$

We have to transmit $\quad Tx(\text{Gain}, \underbrace{a_1, a_2, \cdots a_M}_{\text{IIR filter coeffts}})$

suppose order $M = 10$

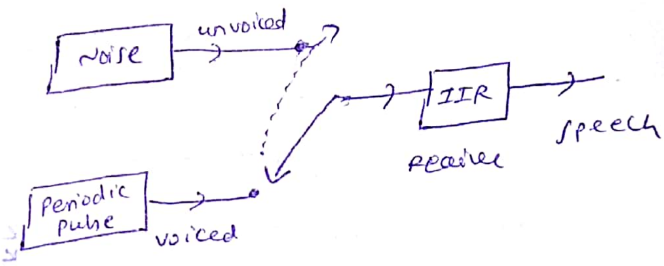
$Tx(G, a_1, a_2, \cdots a_{10})$

Receiver




Voiced / Unvoiced =)

Pitch =)




Each filter coeff =) 6 bit / coeff.

order 10 =) 10 coeffs =) $6 \times 10 = 60$ bits for IIR filter

Gain = 6 bits $\qquad$ pitch = 5 bits

voiced/unvoiced = 1 bit

So, for every 15 msec, we are transmitting

$60 + 6 + 5 + 1 = 72$ bits for 15 msec

Bit rate = 4.8 kbps (much lower than PCM (64 kbps) and DPCM (48 kbps))

but provides same quality.

## Computation at the ~~Receiver~~ Transmitter:

Determine voiced / unvoiced

(Energy of voiced) >> Energy of unvoiced +

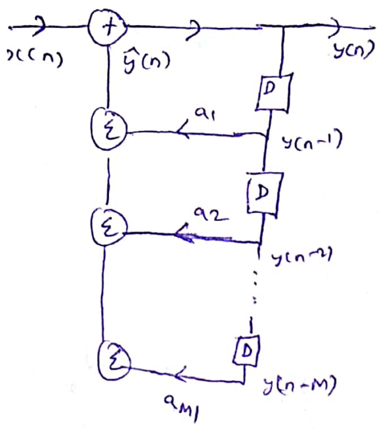
(zero crossing of unvoiced) >> (zero crossing of voiced)

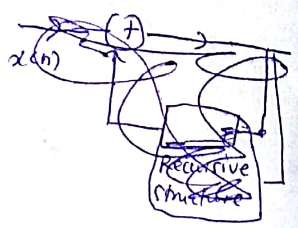
(Auto-correlation coefficient) voiced >> Auto-correlation coefficient of unvoiced

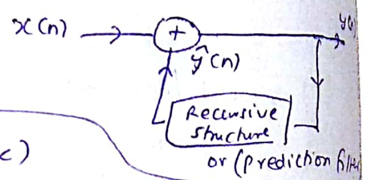
This is how Tx decides whether it is voiced or unvoiced.

occurs at Tx ← **Determination of IIR filter coefficient**

(Synthesis filter)




$$y(n) = x(n) + \sum_k a_k y(n-k)$$

$$y_n = x(n) + \sum_k a_k y(n-k)$$

$$-x(n) = \underbrace{\sum_k a_k y(n-k)}_{\text{Prediction of } y(n) \text{ based on past sample}} - y(n) = \hat{y}(n) - y(n) = e$$

$e_n =$ (Prediction Error)

**Job of Tx:** For every 15 msec, find $a_1, a_2, \dots a_M$ that minimises $e_n^2$ (Mean squared error)

Assume in 15 msec, we get N samples.

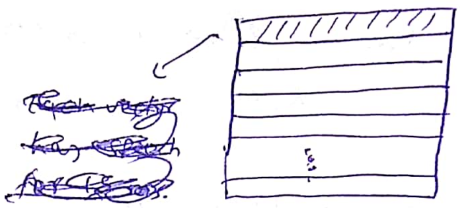
$$E = \sum_{n=1}^{N} e_n^2$$

Find $\{a_1, a_2, \dots a_M\}$ min(E)

**Minimize E:** $\frac{\partial E}{\partial a_1} = 0$ $\quad \frac{\partial E}{\partial a_2} = 0$ and so on...

## CELP


Code book



1000 code vectors

For every 15 ms, the ~~speech~~ ~~voice~~ best code word is chosen and that code vector is transmitted.

So, Tx transmits → 10 × 6 = 60 coeff. bits

6 bits each code vector