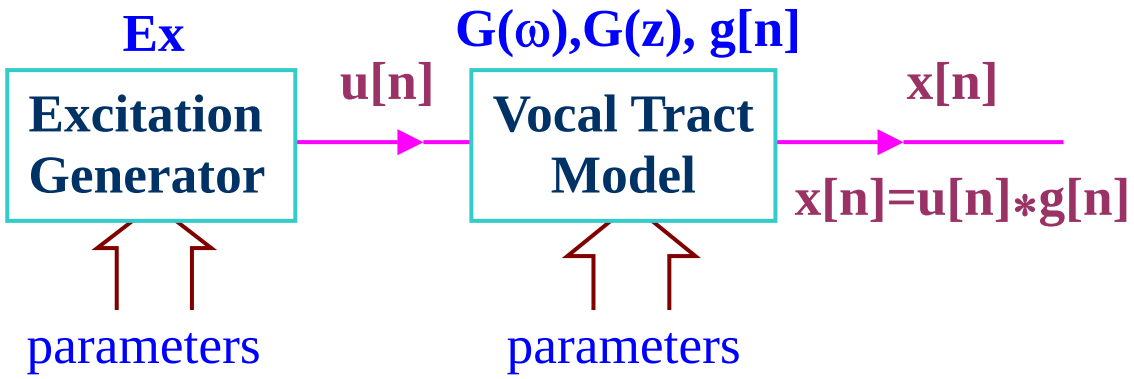
2.5 Linear Prediction Coding (LPC) of Speech Signals

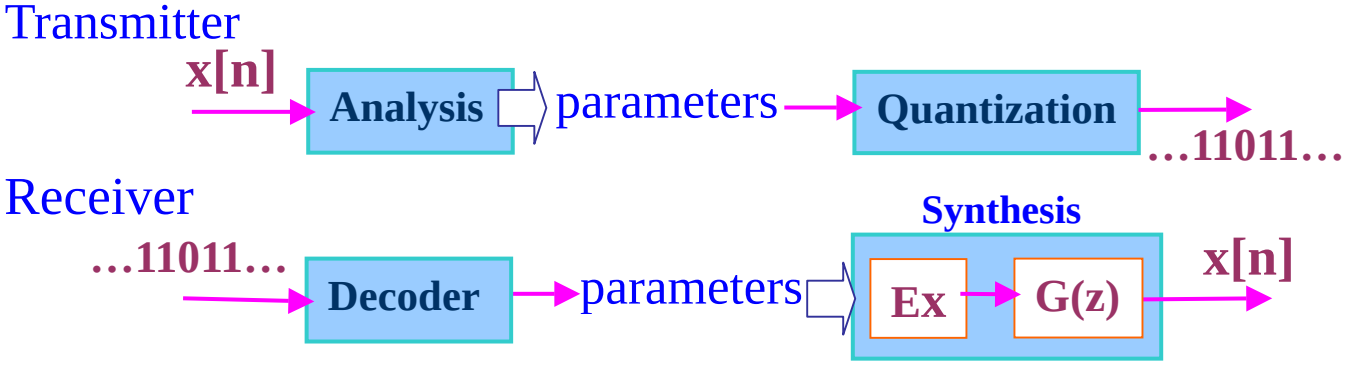
Speech Source Model and Source Coding

- **Concept**



  - digitization and transmission of the parameters will be adequate
  - at receiver the parameters can produce x[n] with the model
  - much less parameters with much slower variation in time lead to much less bits required
  - the key for low bit rate speech coding

- **Analysis and Synthesis**



- High computation requirements are the price for low bit rate

Speech Source Model and Source Coding

- **Z-transform Representation of the Vocal Tract Model**
  - Z-transform

$$x[n] \overset{Z}{\longleftrightarrow} X(z) = \sum_{n=-\infty}^{\infty} x[n]\, z^{-n}$$

similar to

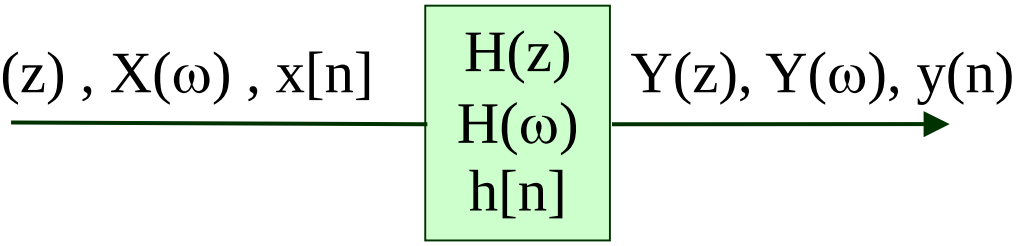
$$X(\omega) = \sum_{n=-\infty}^{\infty} x[n]\, e^{-j\omega n}, \quad z = e^{j\omega}$$

$$x[n-k] \overset{Z}{\longleftrightarrow} X(z)\, z^{-k}$$

$z = e^{j\omega}$ for discrete-time

$z = e^{j\omega T_s}$ for continuous-time

$$\sum_{k=1}^{P} a_k x[n-k] \overset{Z}{\longleftrightarrow} X(z) \left[\sum_{k=1}^{P} a_k z^{-k} \right]$$



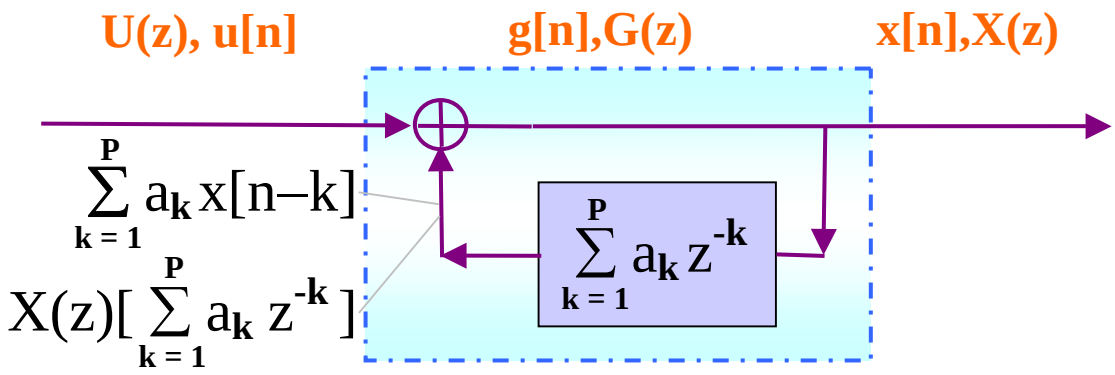
$y[n] = x[n] * h[n]$

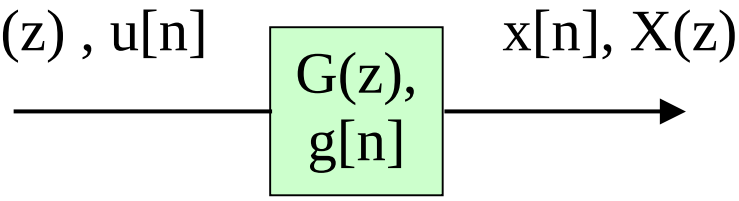
$Y(z) = X(z)H(z)$

Speech Source Model and Source Coding

- Vocal Tract Model

$$u[n] + \sum_{k=1}^{P} a_k\, x[n-k] = x[n]$$





$$G(z) = \frac{1}{1 - \sum_{k=1}^{P} a_k\, z^{-k}} = \frac{X(z)}{U(z)},$$

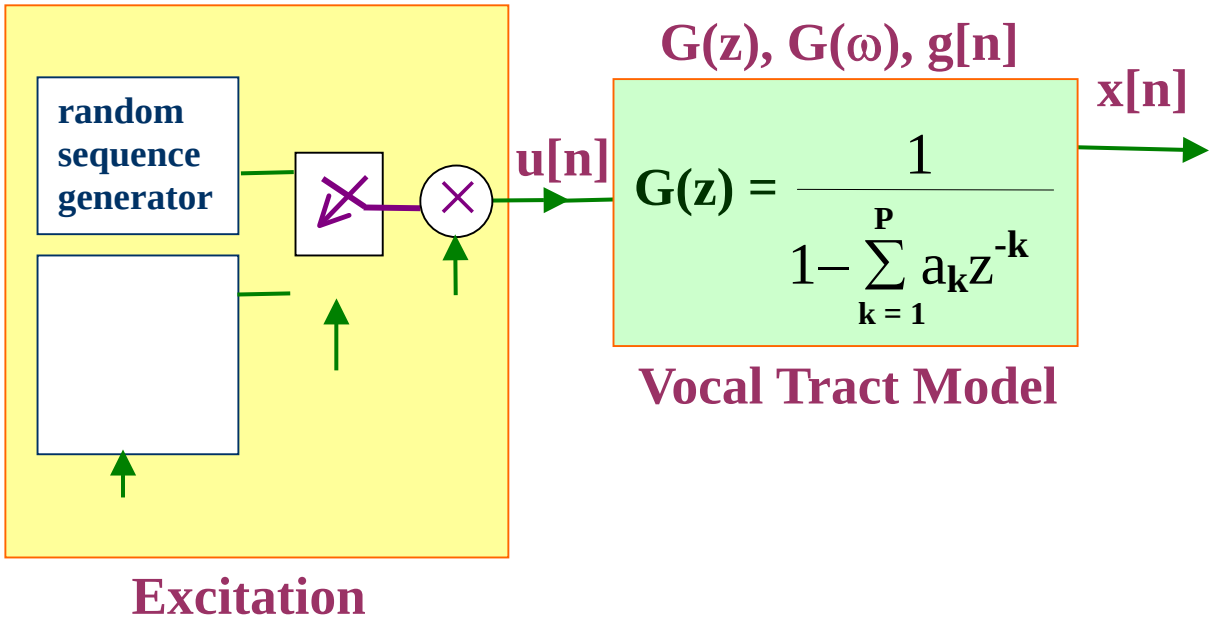
$G(z) \rightarrow G(\omega)$ with peaks at several ω_k called formant frequencies

- This turns out to be a reasonable model for the frequency response of the vocal tract

$\{a_k\}$: Linear Prediction Coding (LPC) coefficients

Speech Source Model and Source Coding

- **Speech Source Model**



- Excitation parameters

  v/u : voiced/ unvoiced

  N : pitch for voiced

  G : signal gain

  → excitation signal u[n]

- Vocal Tract parameters

  $\{a_k\}$: LPC coefficients

  →formant structure of speech signals

- A good approximation, though not precise enough

LPC Analysis of Speech

- **Vocal Track Model**

$$u[n] + \sum_{k=1}^{P} a_k\, x[n-k] = x[n]$$

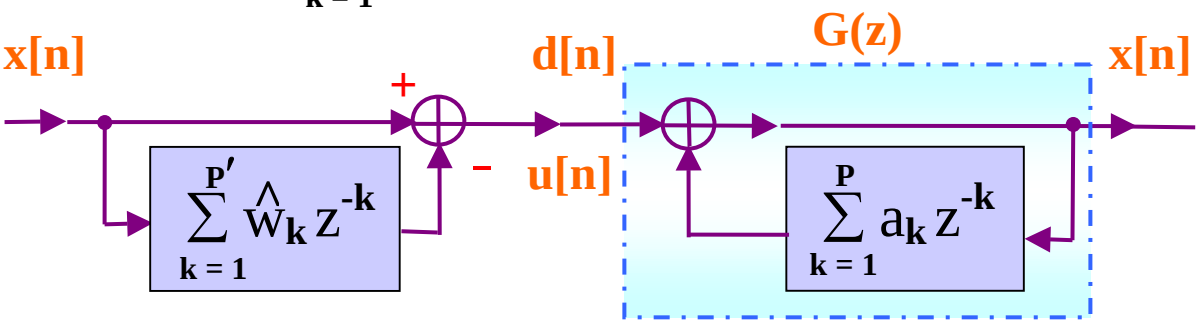
- **Linear Prediction**

$$d[n] + \sum_{k=1}^{P'} \hat{w}_k\, x[n-k] = x[n]$$

uncorrelated unpredictable

correlated with previous samples statistically

$$u[n] + \sum_{k=1}^{P} a_k\, x[n-k] = x[n]$$



→ if G(z) is a good model and if P′ = P

$\hat{w}_k$ previously obtained will be good estimates of a_k

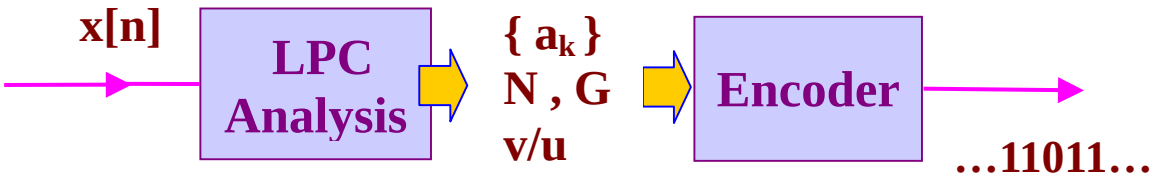
$\bar{a} = \bar{R}^{-1}\bar{r}, \quad \bar{a} = [a_1, a_2, \ldots, a_P]^t$

r_k and a_k are time-varying, obtained from local statistics

different LPC algorithms exist for computing $\{a_k\}$, d[n] can be an approximation of u[n]

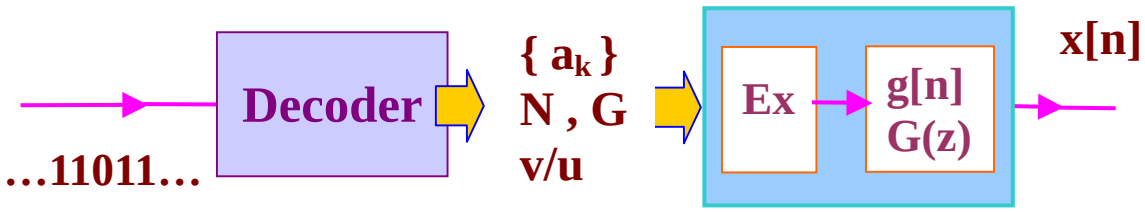
LPC Vocoder(Voice Coder)

transmitter



N by pitch detection
v/u by voicing detection

receiver



{a_k} can be non-uniform or vector quantized to reduce bit rate further

Ref : Chap 21, 31.8 of Gold

3.3 (3.3.1 up to 3.3.9) of Rabiner and Juang, "Fundamentals of Speech Recognition", Prentice Hall, 1993