Consider $e[n]$ (prediction error signal)

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

$$\Rightarrow E(z) = S(z)\, A(z)$$

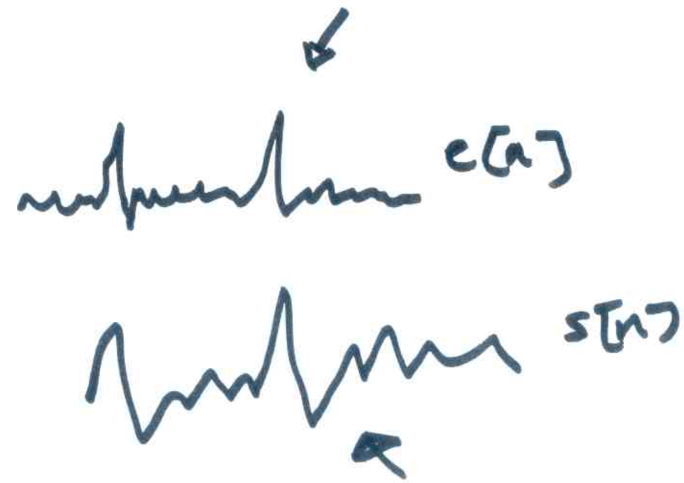
$$s[n] \longrightarrow \boxed{A(z)} \longrightarrow e[n]$$

$$\frac{G}{H(z)}$$

"inverse filtering"

Nature of $e[n]$:

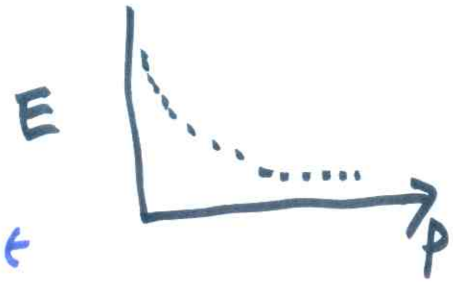
- In voiced speech, $e[n]$ will have large values around glottal pulses
- In uv speech, $e[n]$ is expected to have a flat spectrum (white noise)



Choosing LP analysis parameters :

i) Choice of LP order p :

The order needs to be large enough to represent each formant using a pair of complex conjug poles.

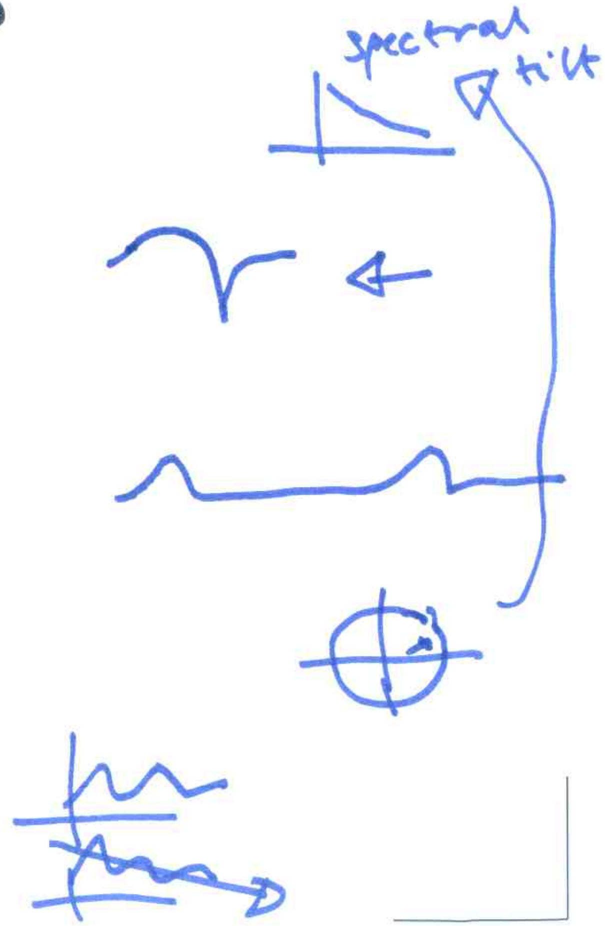


Also, we need ~2 poles to represent or capture the spectral shaping due to glottal wave shape + radn

$$p \approx 2 \times \# \text{formants} + 2$$

\# formants : 1 per 1000 Hz (male)
1 per 1200 Hz (female)

$\Rightarrow \boxed{p = \frac{F_s}{1000} + 2}$ = 10 for 8 kHz sampling



Modeling zeros in speech spectrum:

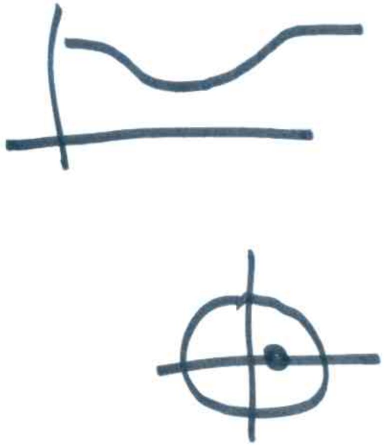
The all-pole model can handle zeros indirectly.

e.g. $1 - az^{-1} = \dfrac{1}{\dfrac{1}{1-az^{-1}}}$

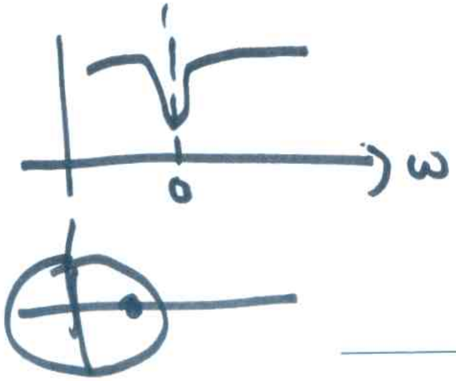
$|a| < 1$

$$= \frac{1}{1 + \sum_{n=1}^{\infty} (az^{-1})^n}$$

We can drop terms where $a^n \ll 1$



We generally assume that another 2-4 poles can handle spectral zeros, if any.



Choice of N (window length)

$$r[k] = \sum x[n]\, x[n+k]$$ ⇐ accurate estimate requires long window

⇒ trade-off

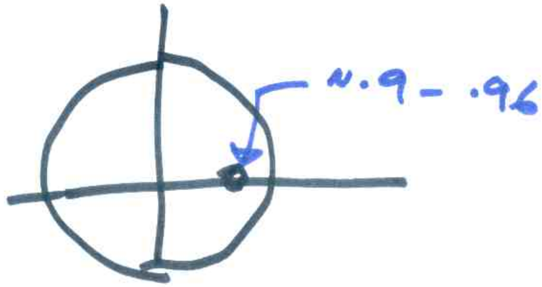
$N \sim 20$ ms

~ 200 samples at 8 kHz sampling.

Pre-processing by pre-emphasis:

For voiced speech, the glottal "roll-off" (spectral tilt) appears like a 6 dB fall-off in the spectral envelope.

A "pre-emphasis" filter $(1 - \alpha z^{-1})$ counter-acts the spectral roll-off



$s[n] \to$ $\boxed{\text{pre-emph.}}$ $\xrightarrow{s_p[n]}$ $\boxed{\text{LP Analysis}}$ $\to$ $\{a_n\}$, G, $e[n]$

$\alpha = 0.95$