5

ECE 438 lecture 4 April 2023

Announcements:

Office Hours 3:30p EDT today

HW #8 due tonight at 11:59p EDT via gradescope

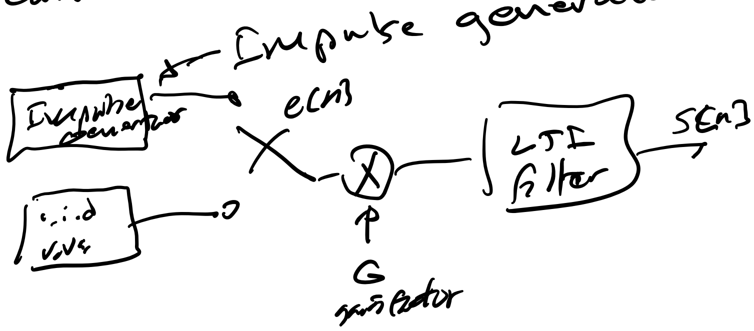
Linear predictive coding (LPC) speech

look wikipedia for a nice history of the development of LPC speech

Module 4.3 — handwritten

Notes are prepared by Prof. Michael Zoltowski

Recall DT model for speech:



LTI filter characterizes the vocal tract response for a single phoneme — call it h[n] since it is a filter

Assume that h[n] is all poles

pitch period $P = NT_s = N\frac{1}{F_s}$

has many — T_s sampling — F_s sampling frequency

different interval
meanings

All pole model:

$$V(z) = H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$ p-poles

So the model parameters are:

1. pitch period N
2. gain G
3. coefficients a_k, $k = 1, \dots, p$
4. p - number of poles

All-pole model is well-suited to resonances i.e. formants, in frequency response

Nasal sounds are better characterized by zeros

But we can model them by adding more poles

What are we trying to do?

Idea is to transmit these parameters rather than the samples of the speech waveform

Usual practice is 13-14 poles for speech sampled at 10kHz - toll quality speech

Have $S(z) = V(z) E(z)$

$$V(z) = \frac{G}{1 - \sum a_k z^{-k}}$$

$1 - \sum_{k=1}$

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

$$s[n] = G e[n] + \sum_{k=1}^{p} a_k s[n-k]$$ — model for speech waveform

Choose our predictor to be

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

known — known

Choose unknown parameter α_k, $k = 1, \ldots, p$ to minimize prediction error

$$f(n) = s[n] - \hat{s}[n] = s[n] - \sum_{k=1}^{p} \alpha_k s[n-k]$$

More precisely, we want to minimize the total squared error:

$$E = \sum_n f^2[n]$$

What is the relation between the a_k's and the α_k's?

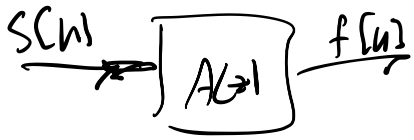
Assume we are "lucky" and choose

$$\alpha_k = a_k, \quad k = 1, \dots, p$$

What does this imply?

let $A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k}$

Consider the following system:

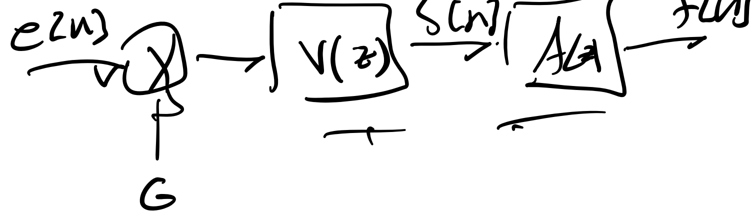


Hypothesis: $F(z) = A(z)S(z)$

$$F(z) = S(z) - \sum_{k=1}^{p} \alpha_k z^{-k} S(z)$$

$$f[n] = s[n] - \underbrace{\sum_{k=1}^{p} \alpha_k s[n-k]}_{\hat{s}[n]}$$

Replace s[n] by the model:



but $V(z) = \frac{1}{A(z)}$, $\alpha_k = a_k$, $k = 1, \dots, p$

$|H(e)|$

So $F(z) = G\,E(z)$ or $f[n] = G\,e[n]$

Thus the error can tell us what is the pitch period (repetition of pulses in $e[n]$)

Estimation of unknown coefficients α_k; $k=1,\dots,p$

New concept: frame

frame is just $\underline{\underline{N}}$ samples of the speech waveform

Define a window $w[m] \neq 0$ only for $0 \le m \le N-1$

N - window length

let $s_n[m] = w[m]\,s[m+n]$

frame location

shift speech waveform back in time by n samples

$$f_n[m] = s_n[m] - \hat{s}_n[m]$$

$$\hat{s}_n[m] = \sum_{k=1}^{p} \alpha_k\, s_n[m-k]$$

Note:

$s_n[m] \neq 0$ only for $0 \le m \le N-1$

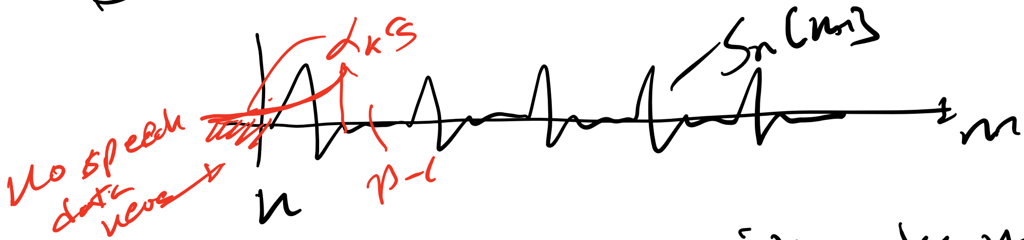
and $\hat{s}_n[m] \neq 0$ only for $0 \le m \le N+p-1$

thus $f_n[m] \neq 0$ only for $0 \le m \le N+p-1$

thus $$E_n = \sum_{m=-\infty}^{\infty} (f_n[m])^2 = \sum_{m=0}^{N+p-1} (f_n[m])^2$$

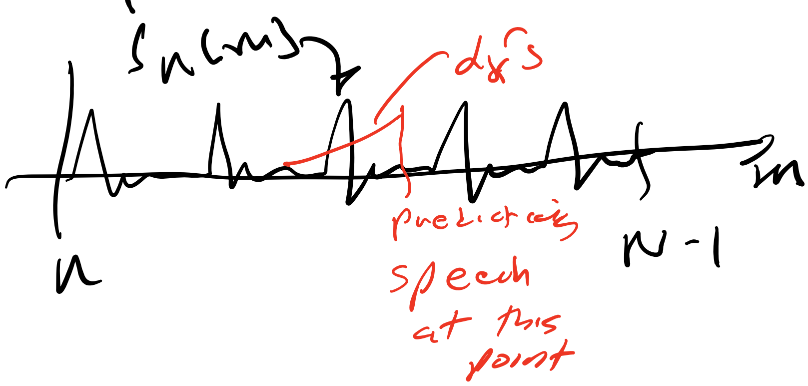
Epochs for prediction are:

(a) $0 \le m \le p-1$



trying to predict $s_n[m]$ with only partial data $\Rightarrow$ larger errors

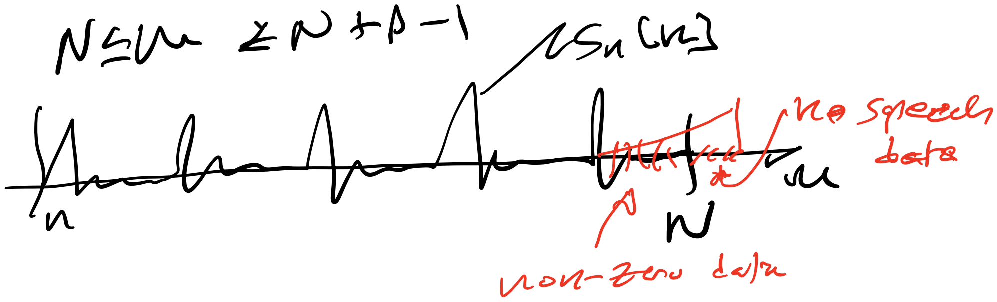
(b) $p \le m \le N-1$



prediction is based on a full set of data

$\Rightarrow$ it is as good as we can get

(c) $N \leq m \leq N+p-1$



trying to predict zero from nonzero data

$\Rightarrow$ larger errors