EEE 6110 Speech Processing.

Dr. Ciira Maina
ciira.maina@dkut.ac.ke

13th March, 2019

Speech Model

- The sampled speech signal is modelled as the output of a linear filter
- The properties of the linear filter are slowly varying
- The system is excited by either quasi-periodic pulses or random noise

Linear Prediction

- The excitation is denoted $e[n]$
- The output is the speech signal denoted $s[n]$
- We have

$$s[n] = \sum_{k=1}^{p} a_k s[n-k] + Ge[n] \tag{1}$$

- Taking the z-transform we obtain the transfer function of the linear system

$$H(z) = \frac{S(z)}{E(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} \tag{2}$$

- The parameters of the model are the gain parameter G and the coefficients a_k

Linear Prediction

- Let

$$\tilde{s}[n] = \sum_{k=1}^{p} \alpha_k s[n-k] \tag{3}$$

- The prediction error or residual is given by

$$d[n] = s[n] - \tilde{s}[n] = s[n] - \sum_{k=1}^{p} \alpha_k s[n-k] \tag{4}$$

- We have

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} = \frac{D(z)}{S(z)} \tag{5}$$

- If the speech signal obeys the model and $a_k = \alpha_k$, $A(z)$, the prediction error filter, is the inverse filter of the system.

Linear Prediction

- We estimate the coefficients using minimum mean-square prediction error
- For a short segment $\hat{n}$ such that $s_{\hat{n}}[m] = s[m + \hat{n}]$ we have

$$E_{\hat{n}} = \sum_m (s_{\hat{n}}[m] - \tilde{s}_{\hat{n}}[m])^2 = \sum_m (s_{\hat{n}}[m] - \sum_{k=1}^{p} \alpha_k s_{\hat{n}}[m-k])^2 \quad (6)$$

- To estimate α_k we take the derivative of $E_{\hat{n}}$ and set it to zero

Linear Prediction

- We get
$$\sum_{m} d_{\hat{n}}[m] s_{\hat{n}}[m-i] = 0 \quad 1 \leq i \leq p \tag{7}$$
- This can be expressed as a set of p linear equations
- These equations are known as the Yule-Walker equations and can be solved using the Levinson Durbin algorithm

Spectral Analysis

- We have

$$H(e^{j\omega}) = H(z)\Big|_{z=e^{j\omega}} = \frac{G}{1 - \sum_{k=1}^{p} a_k e^{-j\omega k}} \tag{8}$$

- This is the frequency response of an all-pole filter
- Peaks occur at roots of the denominator

Cepstral Processing

- Homomorphic transformations transform convolutions into sums
- The cepstrum of a discrete time signal is defined as

$$c[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |X(e^{j\omega})| e^{j\omega n} d\omega \tag{9}$$

- The complex cepstrum is defined as

$$\hat{c}[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\{X(e^{j\omega})\} e^{j\omega n} d\omega \tag{10}$$

- If $x[n] = x_1[n] * x_2[n]$, then $\hat{c}[n] = \hat{c}_1[n] + \hat{c}_2[n]$
- It is approximated by computing the inverse DFT of the logarithm of the DFT of the signal.

Mel-Frequency Cepstral Coefficients

- Real cepstrum of a windowed short time signal
- Computed using the FFT
- Useful in pattern recognition applications
- Employs a filter bank whose center frequencies and bandwidths are based on critical bands.
- Thought to mimic human processing of speech

Readings

- HAH - Chapter 8
- RS - Chapter 7 and 9