Presentation 4 : Digital Models for the Speech Signal

Kocaeli University
Speech Processing
Digital Model for Speech Production

- Detailed model for speech production adapted from Quatieri’s book

![Diagram of a digital model for speech production]

- Impulse Train
- Random Noise
- Impulsive Input
- Linear/Non-linear Combiner
- Impulsive Input
- Random Noise
- Impulse Train
- Output: Speech
Digital Model for Speech Production

- **Source**
  - Periodic impulse train for the voiced sounds
  - Random noise for the fricative consonants
  - An impulse for the plosive consonants

- **Filter composed of three parts**
  - Glottal Pulse Model, $G(z)$
  - Vocal Tract Model, $V(z)$
  - Radiation Model, $R(z)$

- $A_v$, $A_n$ and $A_i$ represent gain coefficients
Digital Model for Speech Production

- Glottal Pulse Model, $G(z)$
  - An approximation
    $$G(z) = \frac{1}{(1 - \beta z)^2}$$
  - It represents two identical poles outside the unit circle when $\beta < 1$

- Radiation Model, $R(z)$
  - Can be approximated with a zero slightly inside the unit circle, $\alpha < 1$
    $$R(z) = 1 - \alpha z^{-1}$$
Vocal Tract Model, $V(z)$

- It can be approximated with an all-pole model when the losses in the system are omitted

$$V(z) = \frac{1}{\prod_{k=1}^{c_i} (1 - c_k z^{-1})(1 - c_k^* z^{-1})}$$

- We assume $c_i$ poles of the $V(z)$ lie inside the unit circle
Digital Model for Speech Production

- Simpler model for speech production adapted from Rabiner and Schafer’s book
Digital Model for Speech Production

- **Source**
  - Periodic impulse train for the voiced sounds
  - Random noise for the unvoiced sounds

- **Filter composed of three parts**
  - Glottal Pulse Model, $G(z)$
  - Vocal Tract Model, $V(z)$
  - Radiation Model, $R(z)$
In linear predictive analysis, $G(z)$, $V(z)$ and $R(z)$ are combined together.

They are represented as a single transfer function:

$$H(z) = G(z)V(z)R(z)$$

Simplified model for speech production:

- Impulse Train
- Random Noise
- Voiced/Unvoiced Switch
- Speech

$G(z)$ and $X$ are shown in the diagram.
Linear Predictive Analysis

- $H(z)$ is assumed to be an all-pole model

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

- Then speech samples $s[n]$ are related to the excitation $u[n]$ by the simple difference equation

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

$$s[n] = Gu[n] + \sum_{k=1}^{P} a_k s[n - k]$$
Questions?

Thank you!