Speech Signal Generation Modeling

Yangang Cao

April 11, 2019

Sound Source Power:

- **Periodic pulse:** produce vowels (voiced) – a
- **Random noise:** produce fricatives (unvoiced) – th
- **Impulse:** produce plosives (unvoiced) – h

However, the real satitution is complicated
such as:

- **z:** voiced fricatives
- **b:** unvoiced plosives
- **m:** nasals

and so on

Excitation Model:

$$g[n] = (\beta^{-n}u[-n]) * (\beta^{-n}u[-n])$$

corresponding z transformation:

$$G(z) = \frac{1}{(1-\beta z)^2}$$

Track Model:

$$V(z) = \frac{G}{1 - \sum_{k=1}^{N} a_k z^{-k}} = \prod_{i=1}^{M} \frac{a_i}{1 - b_i z^{-1} - c_i z^{-2}}$$

and

$$c_i = -e^{-2\pi B_i T}$$

$$b_i = 2e^{-\pi B_i T} \cos(2\pi F_i T)$$

$$a_i = 1 - b_i - c_i$$

$$G = a_1 \cdot a_2 \cdot a_3 \cdot \cdots \cdot a_M$$

Radiant Model:

$$
\begin{aligned}
R(z) &= 1 - \alpha z^{-1} \\
&= \frac{1}{\sum_{k=0}^{\infty} a^k z^{-k}} \\
&= \frac{1}{\prod_{k=0}^{\infty} (1 - b_k z^{-1})}
\end{aligned}
$$

All-Pole Model of Speech Signal

- **Periodic pulse:**

$$X_v(z) = A_v G(z) V(z) R(z)$$

- **Random noise:**

$$X_n(z) = A_n U(z) V(z) R(z)$$

- **Impulse:**

$$X_i(z) = A_i V(z) R(z)$$