

CECS401
Fundamentals of Spoken Language Processing

Note-4
Thursday 9/2/99

Topic-3 Digital Models of Speech signals

A. Acoustic-tube model of speech signal

(from Digital processing of speech signals, by L. R. Rabiner and R. W. Schafer, Prentice Hall, 1978)

- Digital models of speech serve as the mathematical basis for the analysis and synthesis of speech.
- A completely detailed acoustic theory incorporating all the effects in the vocal tract in speech production is not yet available (effects including time variation of vocal tract shape, losses due to heat conduction and viscous friction at the vocal tract walls, softness of the vocal tract walls, radiation of sounds at lips, nasal coupling, excitation of sounds in the vocal tract)
- The acoustic tube model is a simplified and useful physical model of speech production.

Acoustic tube model

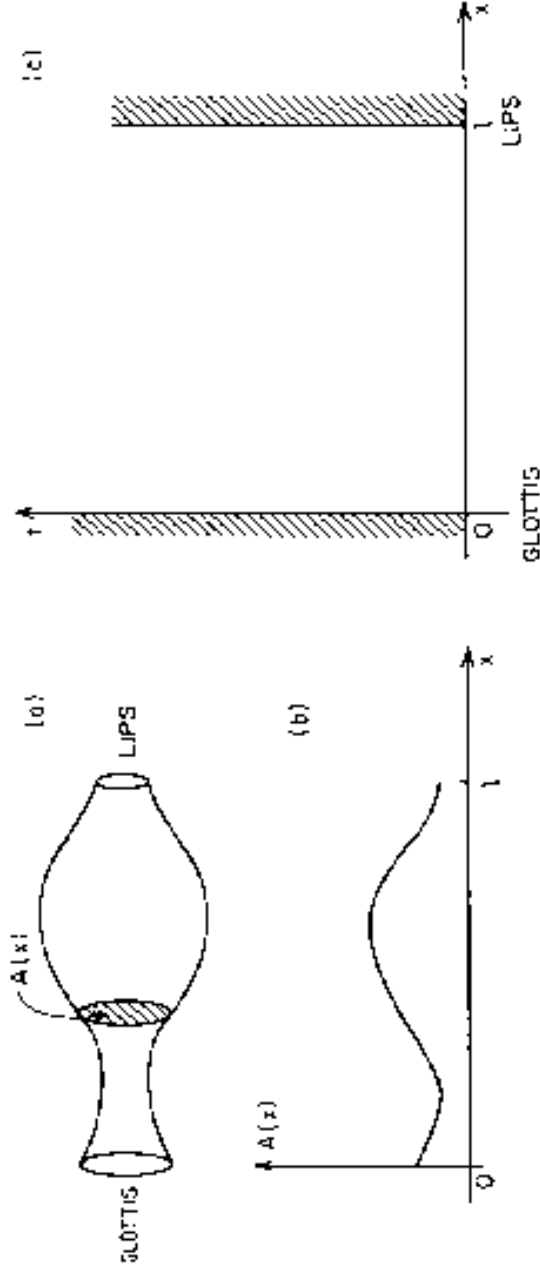


Fig. 3.13 (a) Schematic vocal tract; (b) corresponding area function; (c) $x-t$ plane for solution of wave equation.

$A(x,t)$: cross-sectional area of the tube as a function of the distance along the tube and time.

$u(x,t)$: volume velocity flow at position x and time t .

$p(x,t)$: sound pressure at position x and time t .

- A set of partial differential equation can be derived to describe the sound wave propagation in the tube.
- Closed form solutions are not available in general. Numerical solutions can be obtained.
- Boundary condition at each end of the tube:
 - glottal excitation
 - termination of the nasal and oral tracts

Effects of vocal tract on speech production

1. Radiation at the lips

The radiation effect at the lip opening causes energy losses at high frequencies

2. Vocal tract transfer function of vowels

The transfer function is characterized by a set of resonances, i.e., formants.

The formant frequencies depend on the area function, and the formant bandwidths depend on the losses in the tube.

Example

Computed frequency response ($U(l, \Omega)/U_g(\Omega)$) for a set of measured area functions.

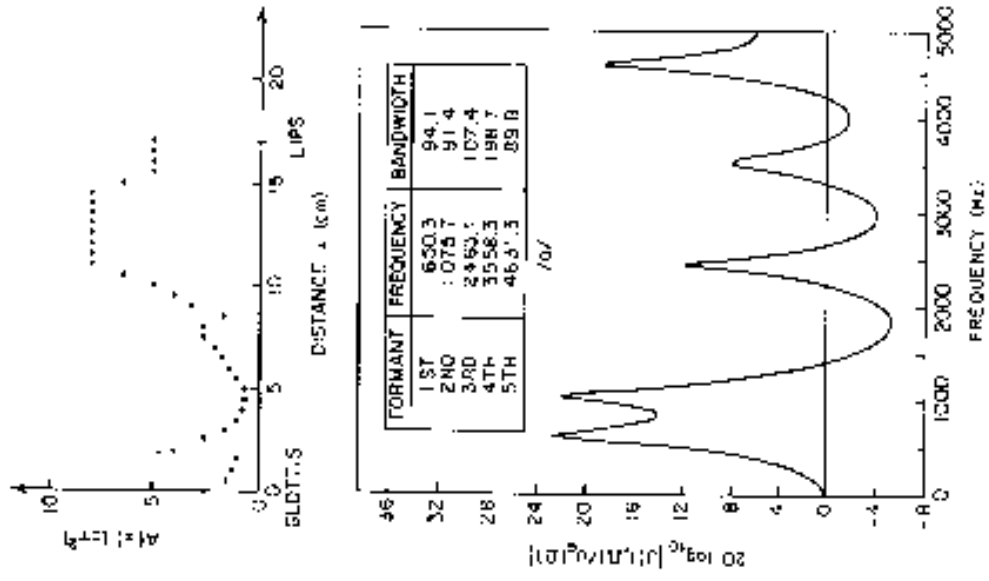
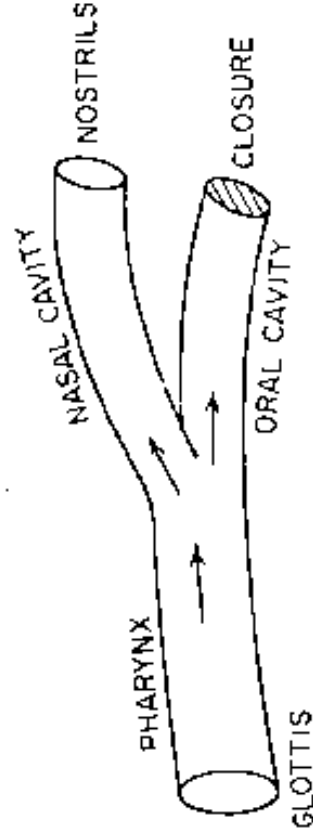


Fig. 3.23 Area function (after Fant [11]) and frequency response (after Fant and [18]) for the Danish vowel /a/.

3. Nasal coupling



The closed oral cavity can trap energy at certain frequencies, and the vocal system is therefore characterized by resonances and antiresonances (spectral zeros).

The formant bandwidths are broader than non-nasal sounds due to greater losses in the nasal tract.

4. Excitation

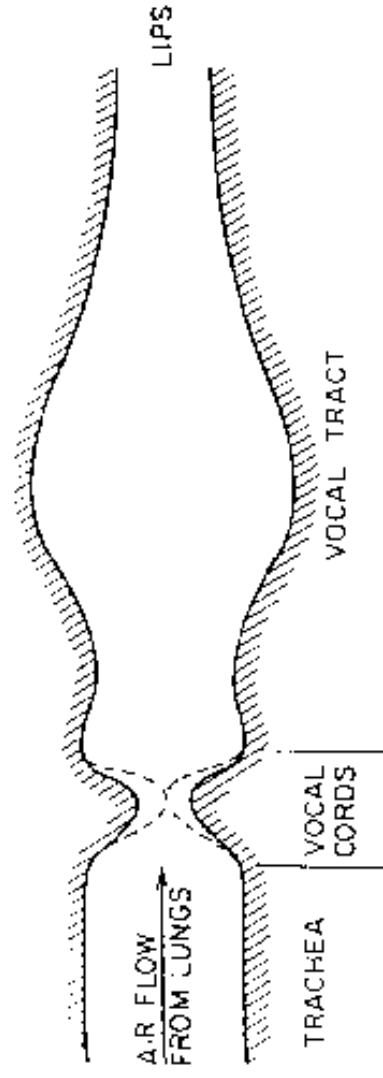


Fig. 3.28 Schematic representation of the vocal system.

Vibration of vocal cords — quasi-periodic pulse-like excitation.

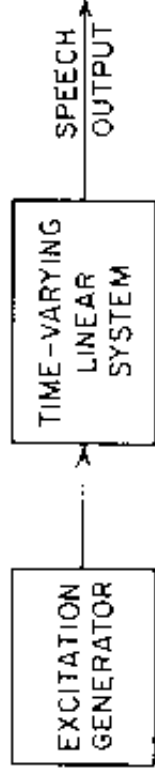
Constriction in the vocal tract — noise like excitation.

Closure in the vocal tract and then release — transient excitation.

B. Linear system model based upon the acoustic theory

The acoustic theory points the way to a simplified approach to modeling production of speech signals.

Linear system model separates the excitation features from the vocal tract and radiation features.



Lossless tube models

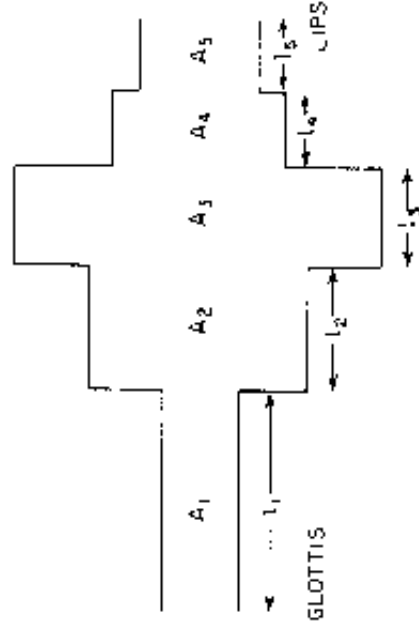


Fig. 3.32 Concatenation of 5 lossless acoustic tubes.

Lossless tube model represents the vocal tract by a concatenation of lossless uniform acoustic tubes.

A large number of short tubes can well approximate a continuously varying area function.

The model provides a convenient transition between continuous-

Equal-length lossless tube model

$$(l_1 = l_2 = \dots = l_N)$$

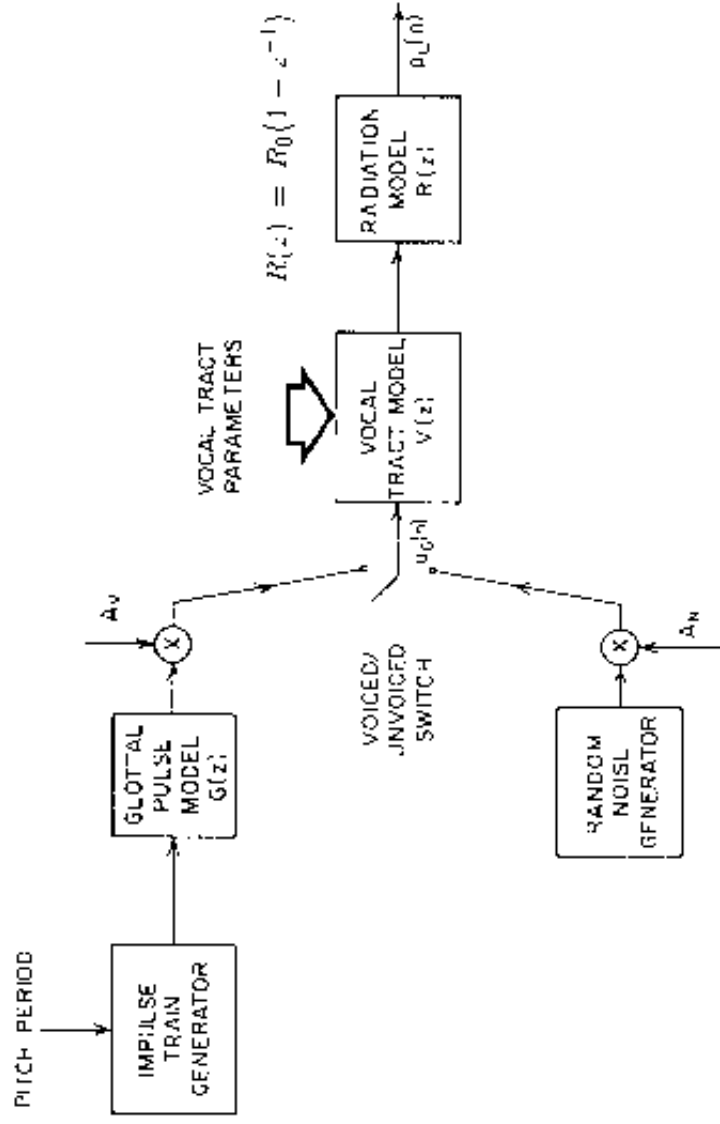
System transfer function is all pole

$$V(z) = \frac{U_l(z)}{U_g(z)} = \frac{G}{1 - \sum_{k=1}^N \alpha_k z^{-k}}$$

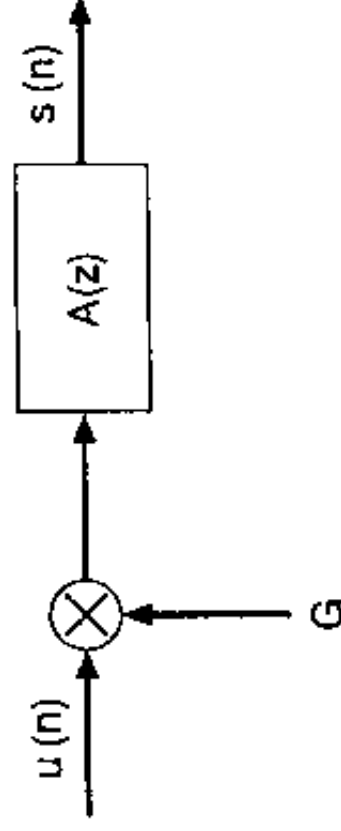
α_k 's are functions of reflection coefficients between adjacent cross-sectional areas and are time-varying.

The excitation generator provides input to the system in the form of either a train of glottal pulses or randomly varying noises.

Complete digital model of speech production



C. Linear predictive coding (LPC) of speech



$$s(n) = \sum_{i=1}^P a_i s(n-i) + Gu(n)$$

$s(n)$ is the speech sample, $u(n)$ is the normalized excitation, G is the gain.

Taking z-transform

$$S(z) = \sum_{n=0}^{\infty} s_n z^{-n}$$

$$z^{-i} S(z) = \sum_{n=0}^{\infty} s_{n-i} z^{-n}$$

$$S(z) = \sum_{i=1}^P a_i z^{-i} S(z) + GU(z)$$

leading to the system equation

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{i=1}^P a_i z^{-i}} = \frac{1}{A(z)}$$

LPC has wide applications in speech processing, including:

- speech coding
- speech synthesis
- formant and pitch analysis
- speech and speaker recognition