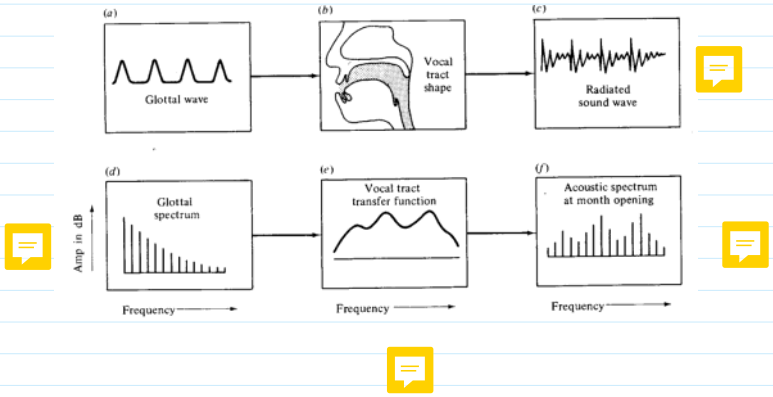
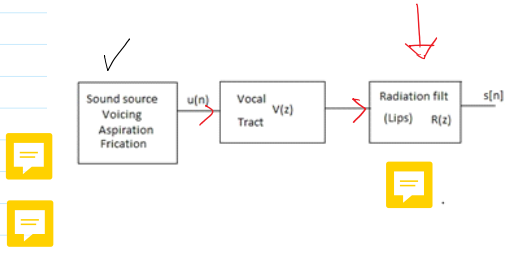
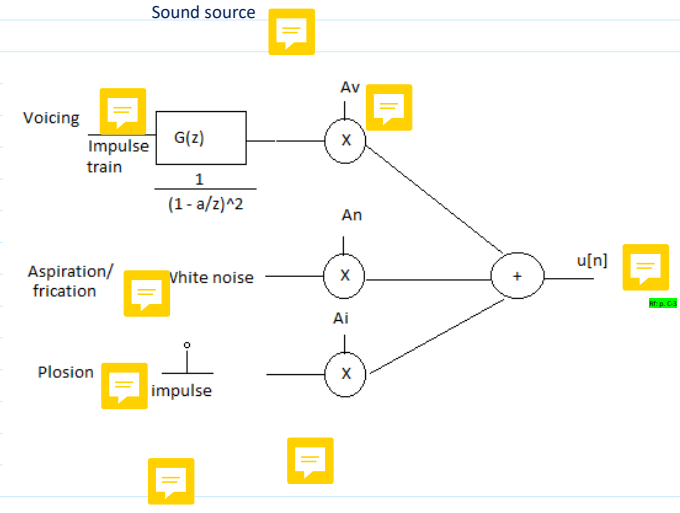
We model the speech signal as the result of a spectrally rich excitation signal that is filtered by a (time-varying or quasi-stationary) LTI system representing the vocal tract.



The source-filter model of speech production

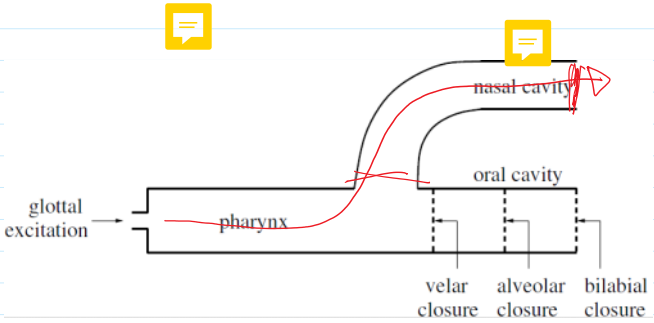


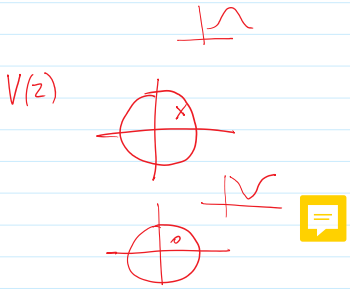


Vocal tract: V(z)

- Vowels
- Nasals
- Fricatives
- Glides, semivowels
- Plosives

Nasal consonants and nasalized vowels

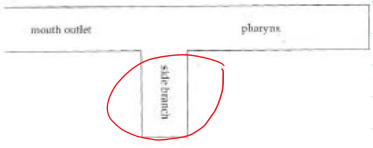




Helmholtz resonator theory: Resonances of the closed side-branch are "absorbed" or trapped and are missing in the sound out of the main branch.

Semi-vowels /l/ /r/

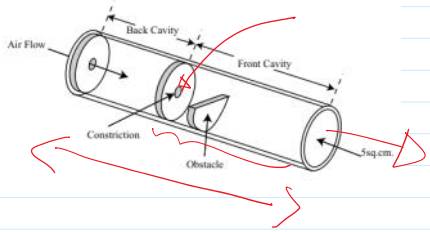
/w/ /y/



- The pocket (side branch) is about 4 cm. This gives an anti-formant around 2100Hz (so between F2 and F3).

Fricatives /f/, /v/, /s/, /sh/, /ksh/ This, /θ/

Figure 2-4.2 Physical Model of Fricative Production (after Shadle)

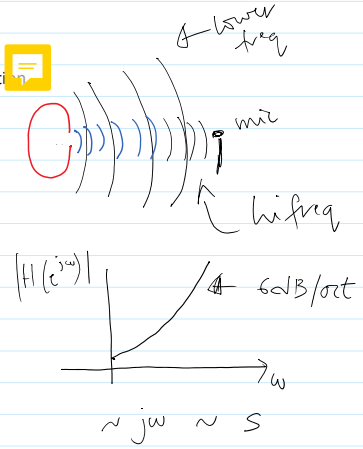


Plosives

Closing phase ---> Closure ---> Release ---> Aspiration ---> Opening

upper

Lip radiation



$R(z) = 1 - \alpha z^{-1}$

$\alpha \approx 0.95 - 0.98$

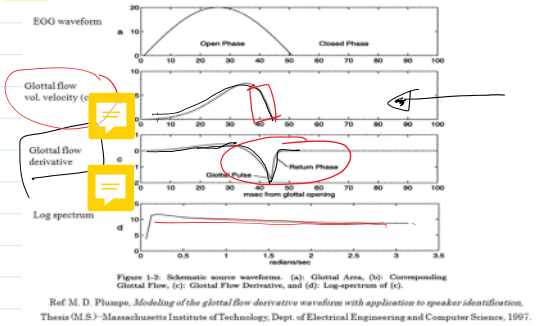


Figure 1-2: Schematic source waveforms. (a): Glottal Area, (b): Corresponding Glottal Flow, (c): Glottal Flow Derivative, and (d): Log-spectrum of (c).

Ref: M. D. Plumpe, *Modeling of the glottal flow derivative waveform with application to speaker identification*, Thesis (M.S.)--Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 1997.

- Voicing
- MoA
- PoA

Variants of normal speech

- Whisper
- Shouting
- Singing — formant tuning

/z/ /s/

/p/ /ʃ/