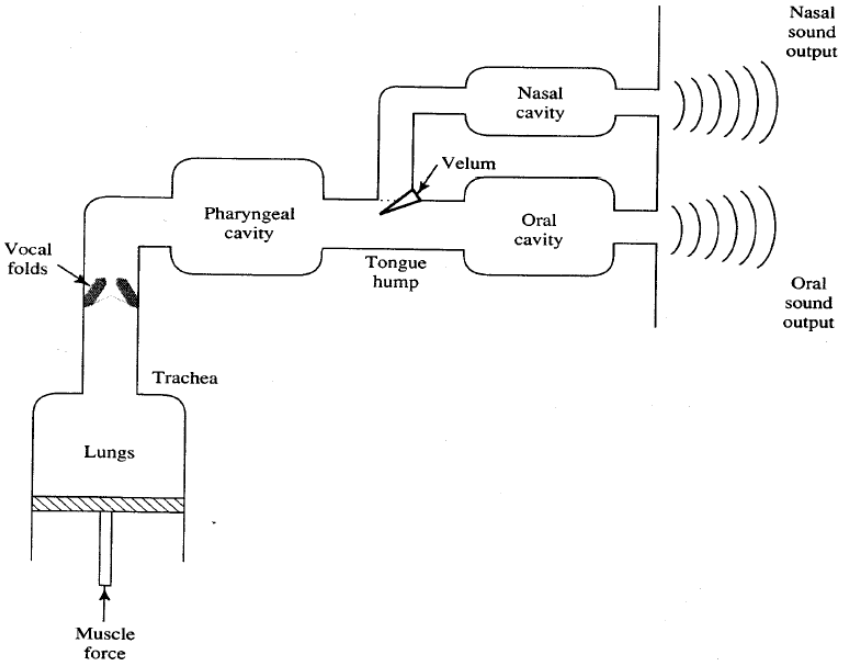
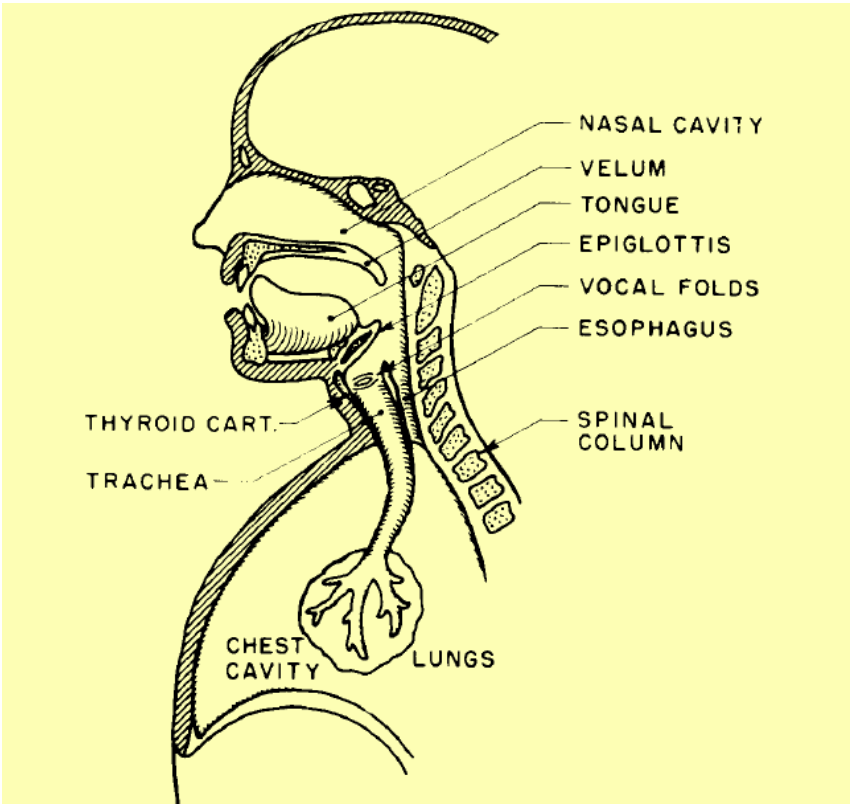
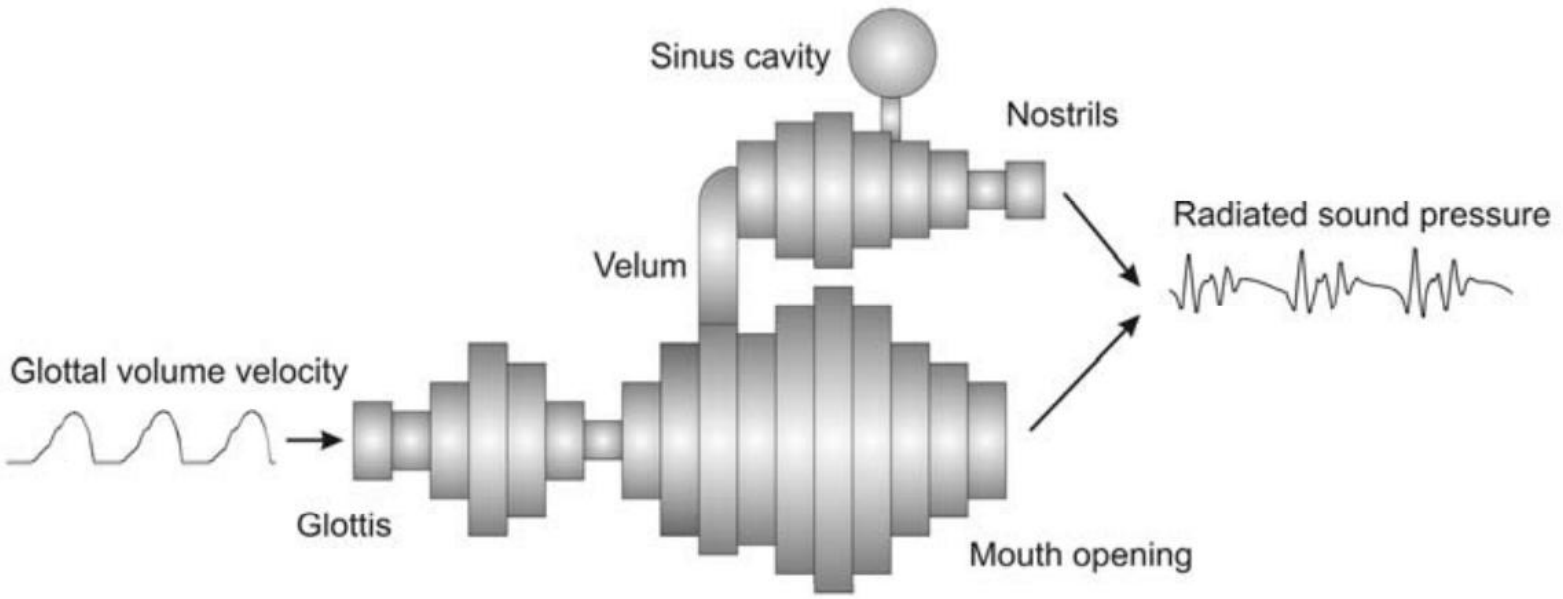
Pipeline Model





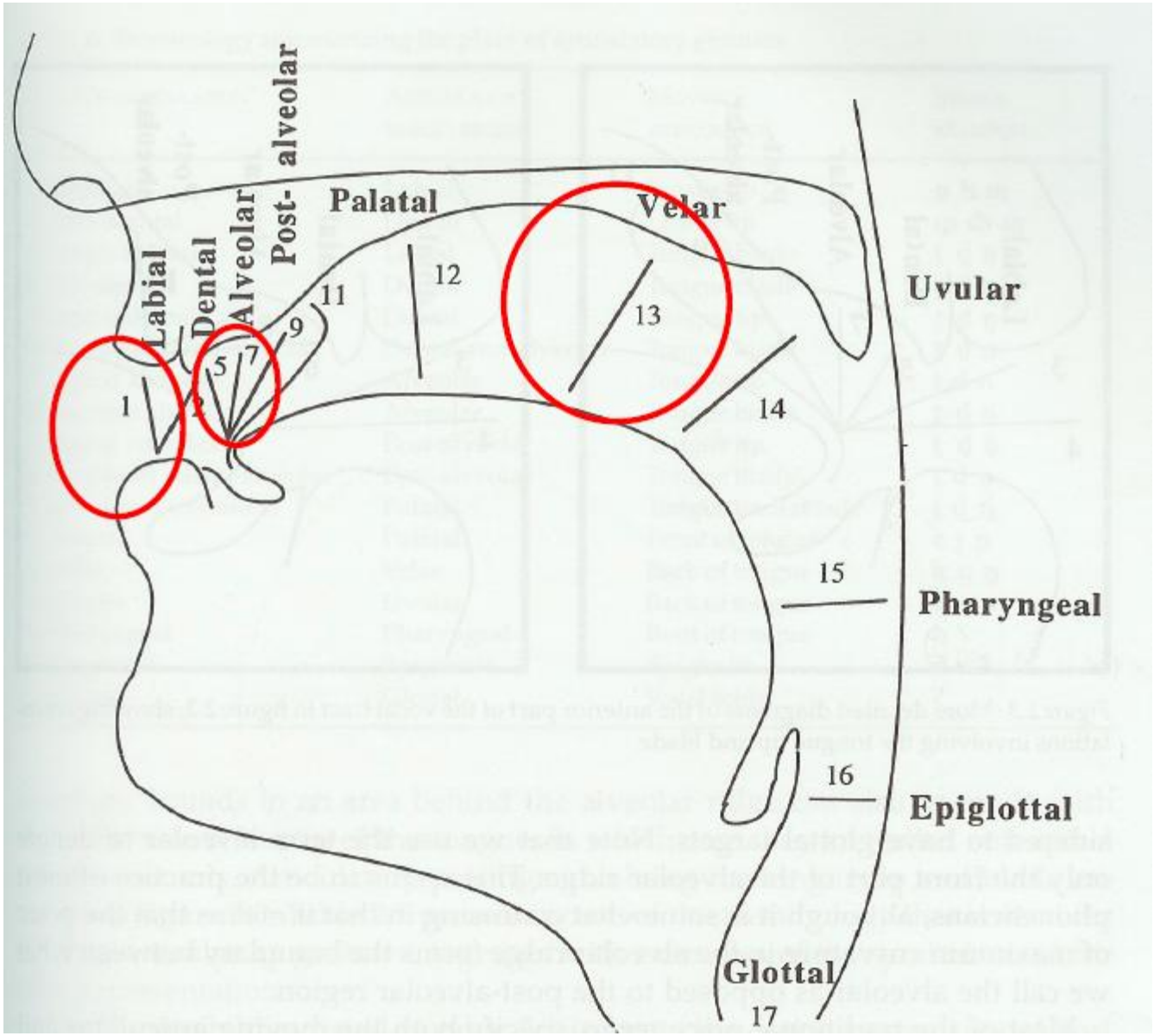
N-tube model

N-tube model

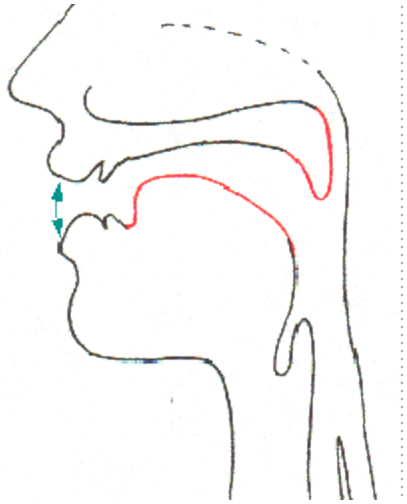


http://clas.mq.edu.au/speech/acoustics/frequency/vocal_tract_resonance.html

Atal, Bishnu S., et al. "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique." *The Journal of the Acoustical Society of America* 63.5 (1978): 1535-1555.

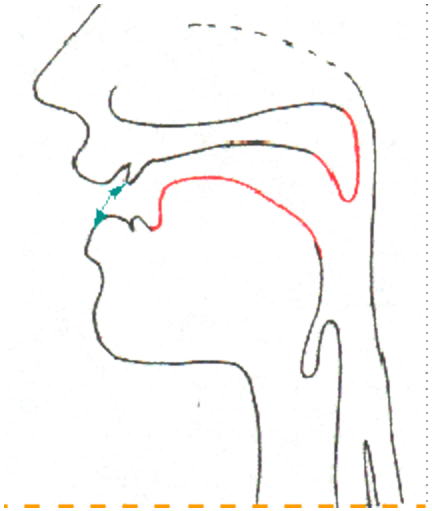


http://www.phonetics.ucla.edu/index/sounds.html



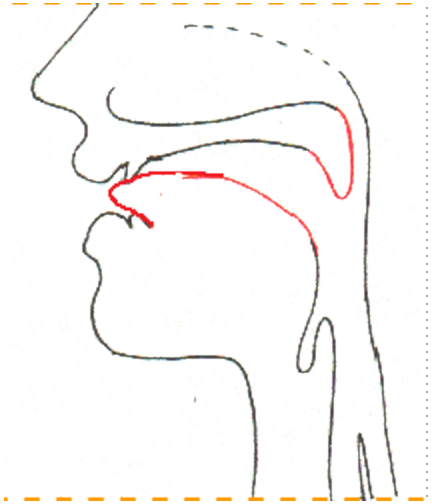
Labial

b, p, m



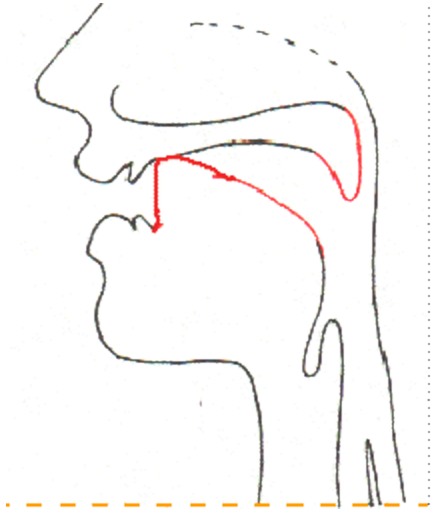
Labio-dental

f, v



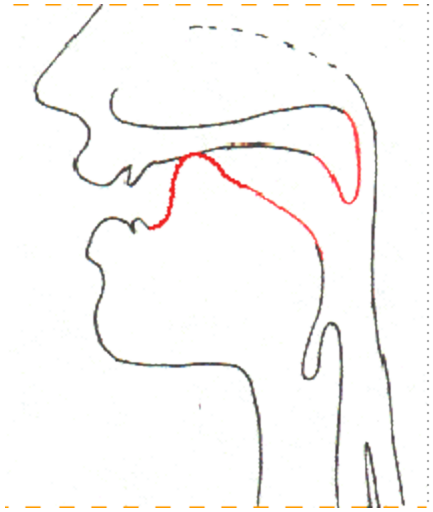
Interdental

θ, ð



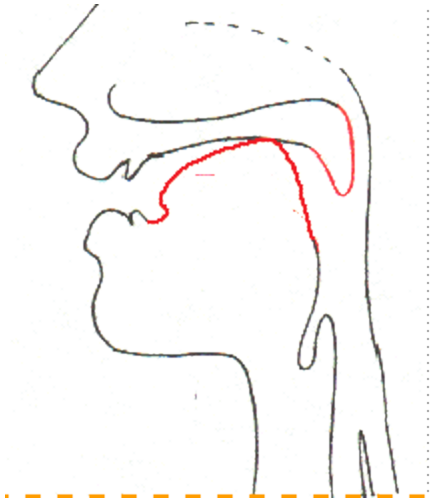
Alveolar

d, t, s, z, n



Palatal

š, ž

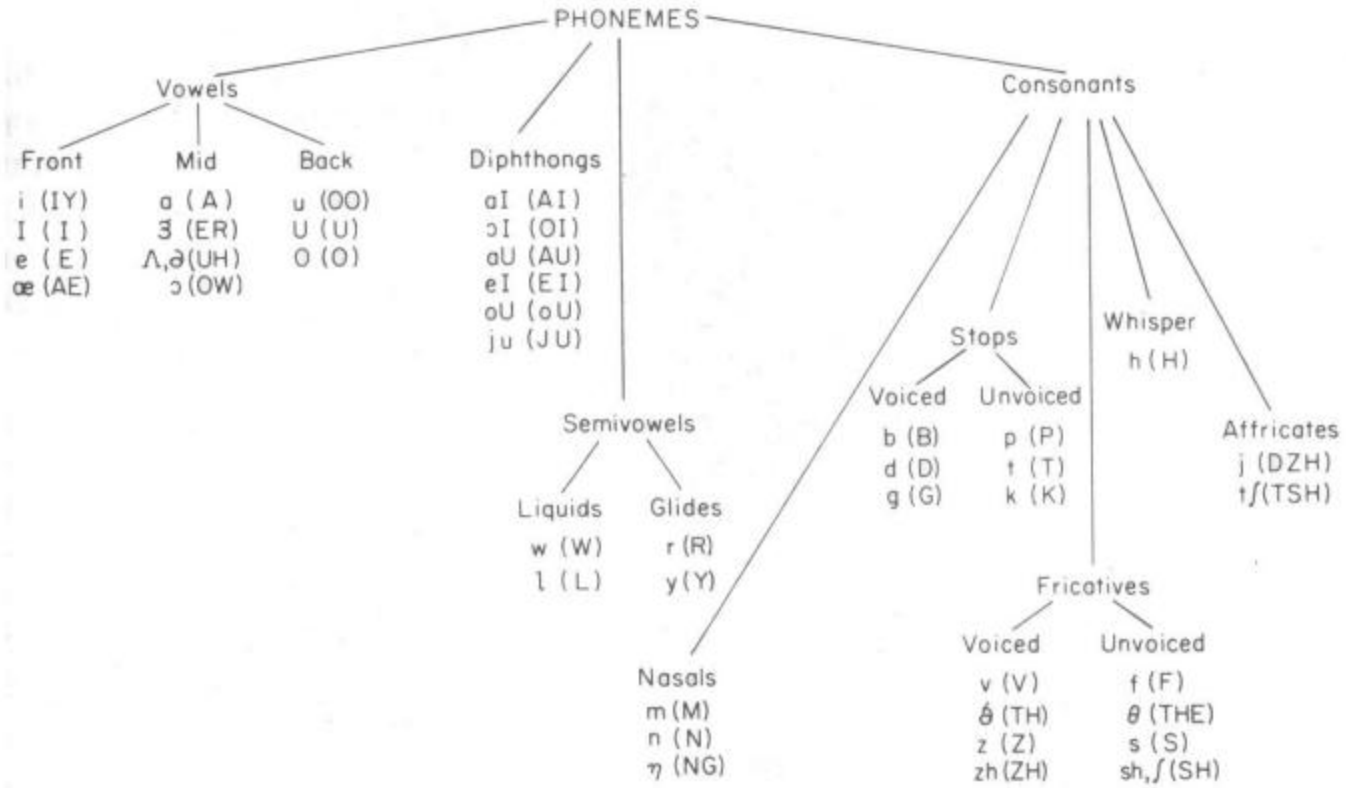


Velar

g, k, ŋ

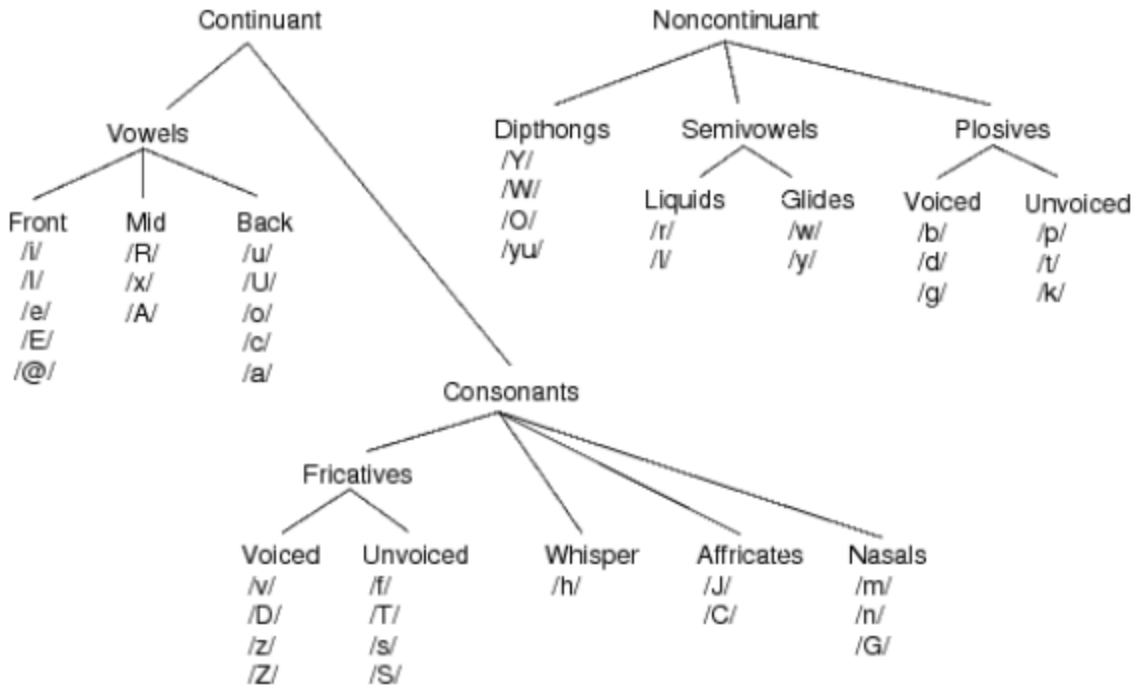
Four general classes of sounds in American English

- Vowels, diphthongs, semivowels, and consonants
- Each can be further divided according to articulators (manner, place)

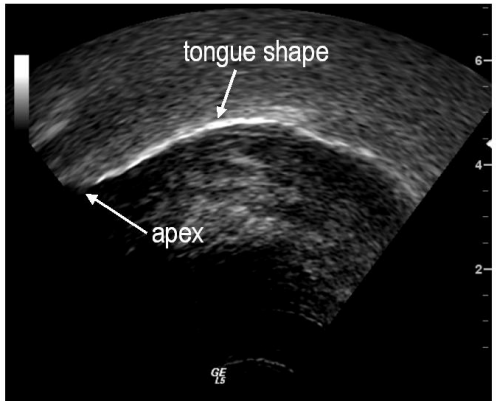


Alternatively, phoneme classes can be divided into

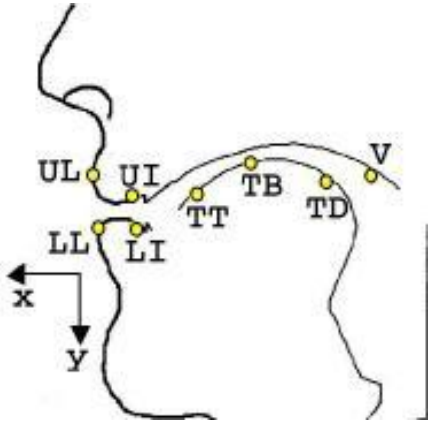
- Continuants: produced by a fixed vocal tract configuration
  - Includes vowels, fricatives, and nasals
- Non-continuants: vocal tract configuration changes over time
  - Diphthongs, semivowels, stops and affricatives



Ultrasound

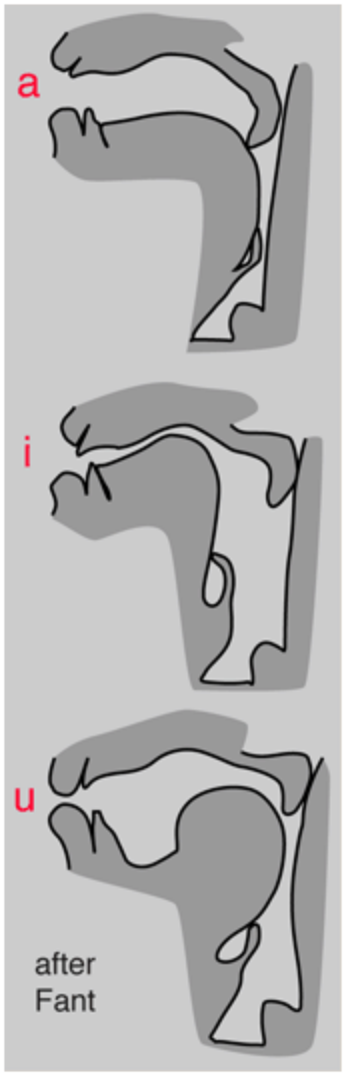


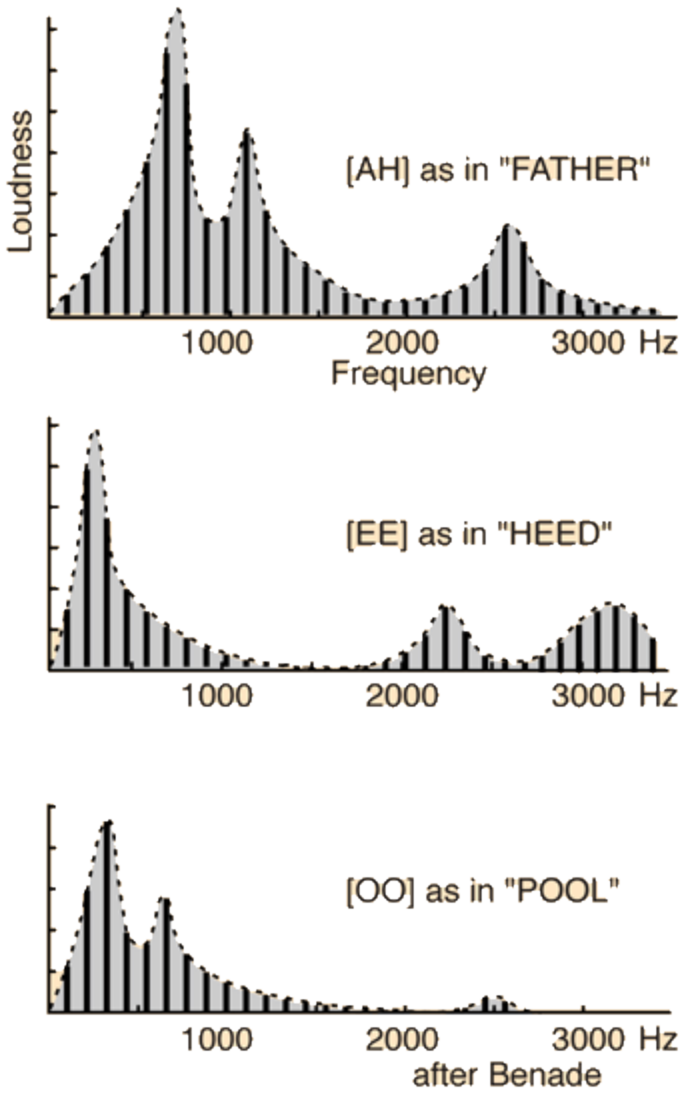
Articulograph



rtMRI

Articulation to acoustic





For vowels

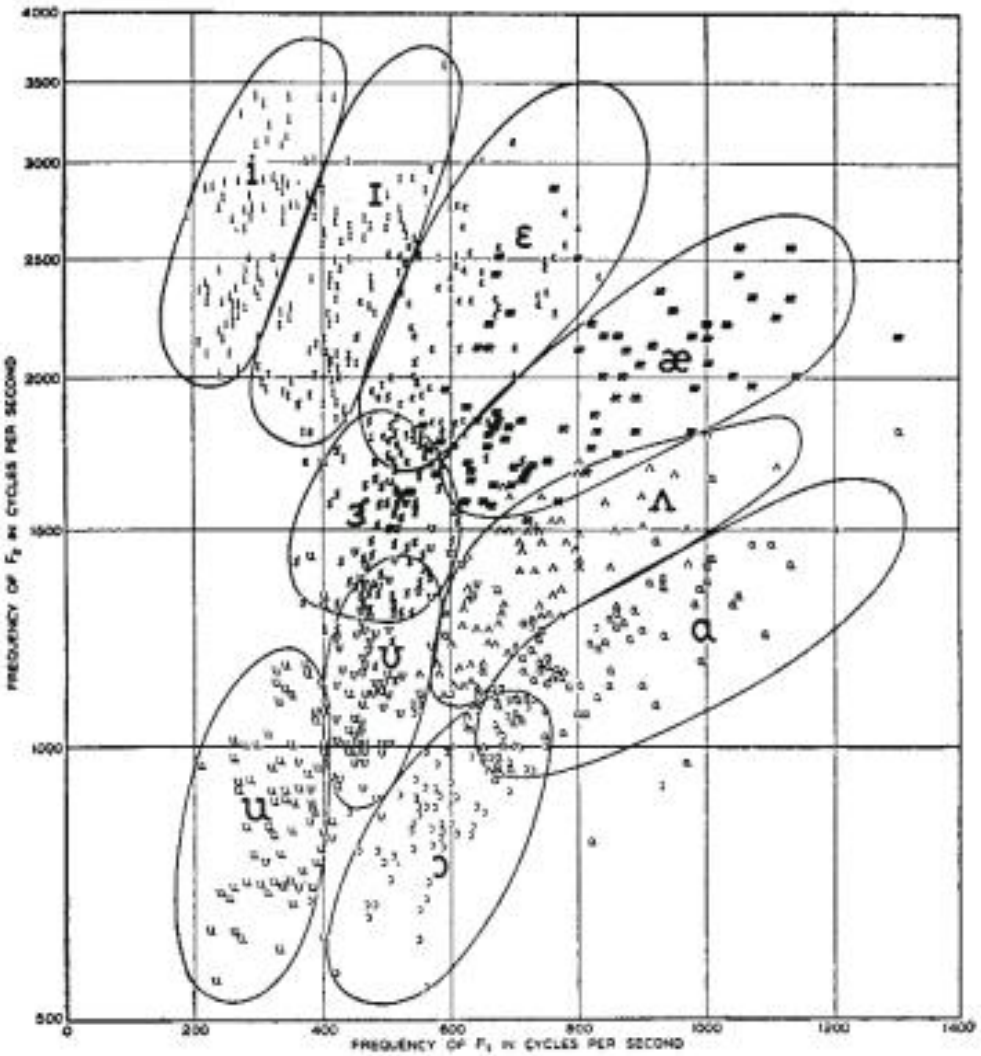
Formant

Formants are frequency peaks which have, in the spectrum, a high degree of energy. They are especially prominent in vowels. Each formant corresponds to a resonance in the vocal tract (roughly speaking, the spectrum has a formant every 1000 Hz). First three formant for few vowels (with example word and IPA symbol) are:

	Vowel	F1(Hz)	F2(Hz)	F3(Hz)
heed	iː	280	2620	3380
hid	ɪ	360	2220	2960
head	e	600	2060	2840
had	æ	800	1760	2500
hudd	ʌ	760	1320	2500
hard	ɑː	740	1180	2640
hod	ɒ	560	920	2560
hoard	ɔː	480	760	2620
hood	ʊ	380	940	2300
Who'd	uː	320	920	2200
heard	ɜː	560	1480	2520

Adult male formant frequencies in Hertz collected by J.C.Wells around 1960. Note how F1 and F2 vary more than F3.

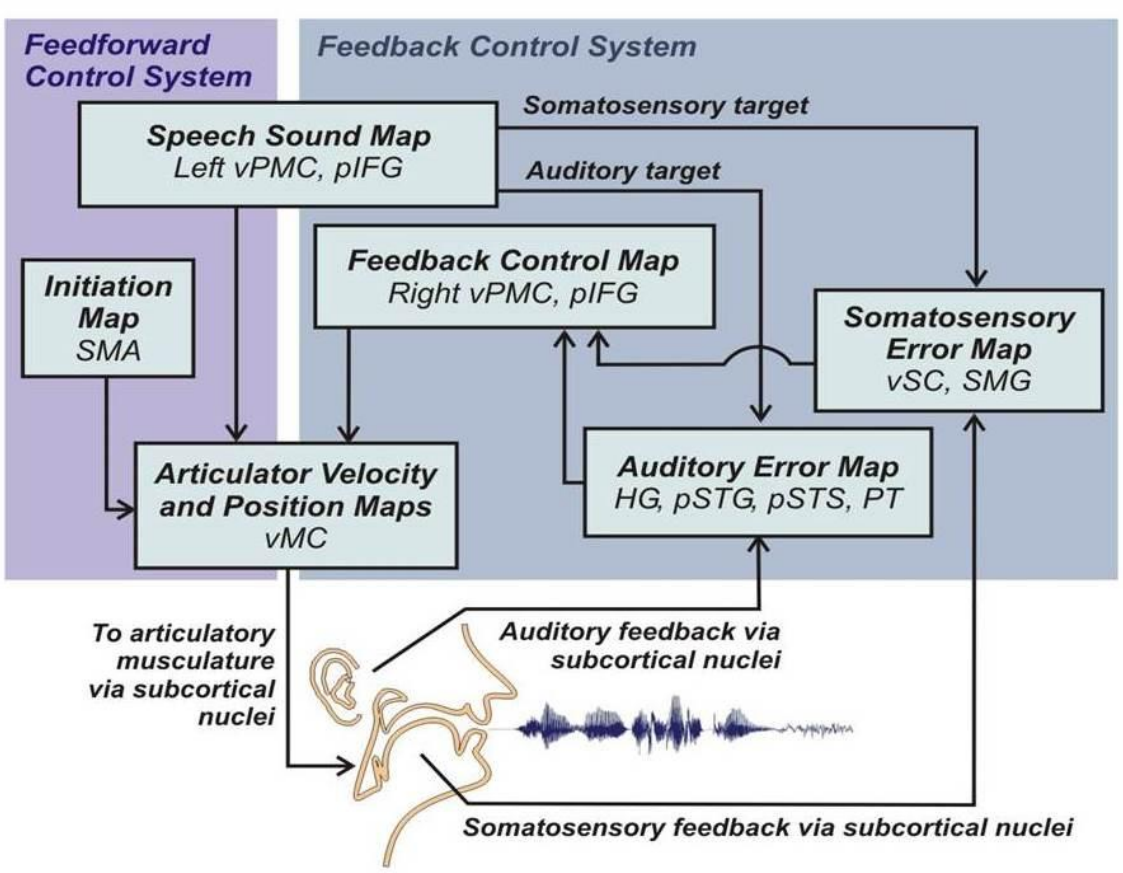
Formant



Frequency of second formant *versus* frequency of first formant for ten vowels by 76 speakers.

Speech production models

DIVA Model

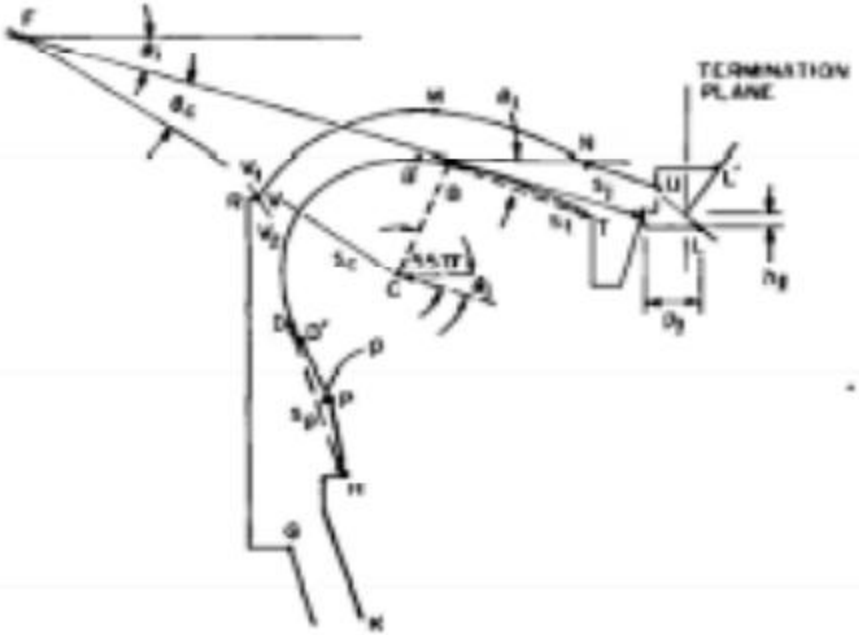


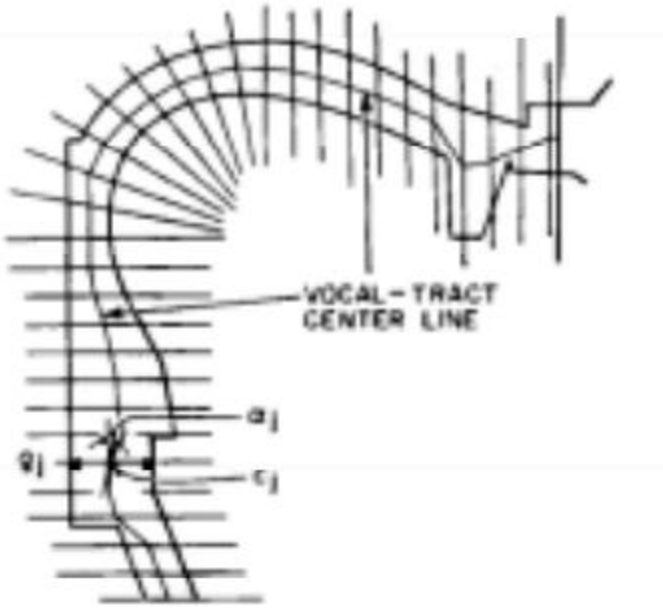
Guenther, Ghosh, and Tourville (2006) *Brain and Language*

http://www.bu.edu/speechlab/research/the-diva-model/

Speech production models

Articulatory Model





Mermelstein, Paul. "Articulatory model for the study of speech production."The Journal of the Acoustical Society of America 53.4 (1973): 1070-1082.

Speech production models

TaDA Model

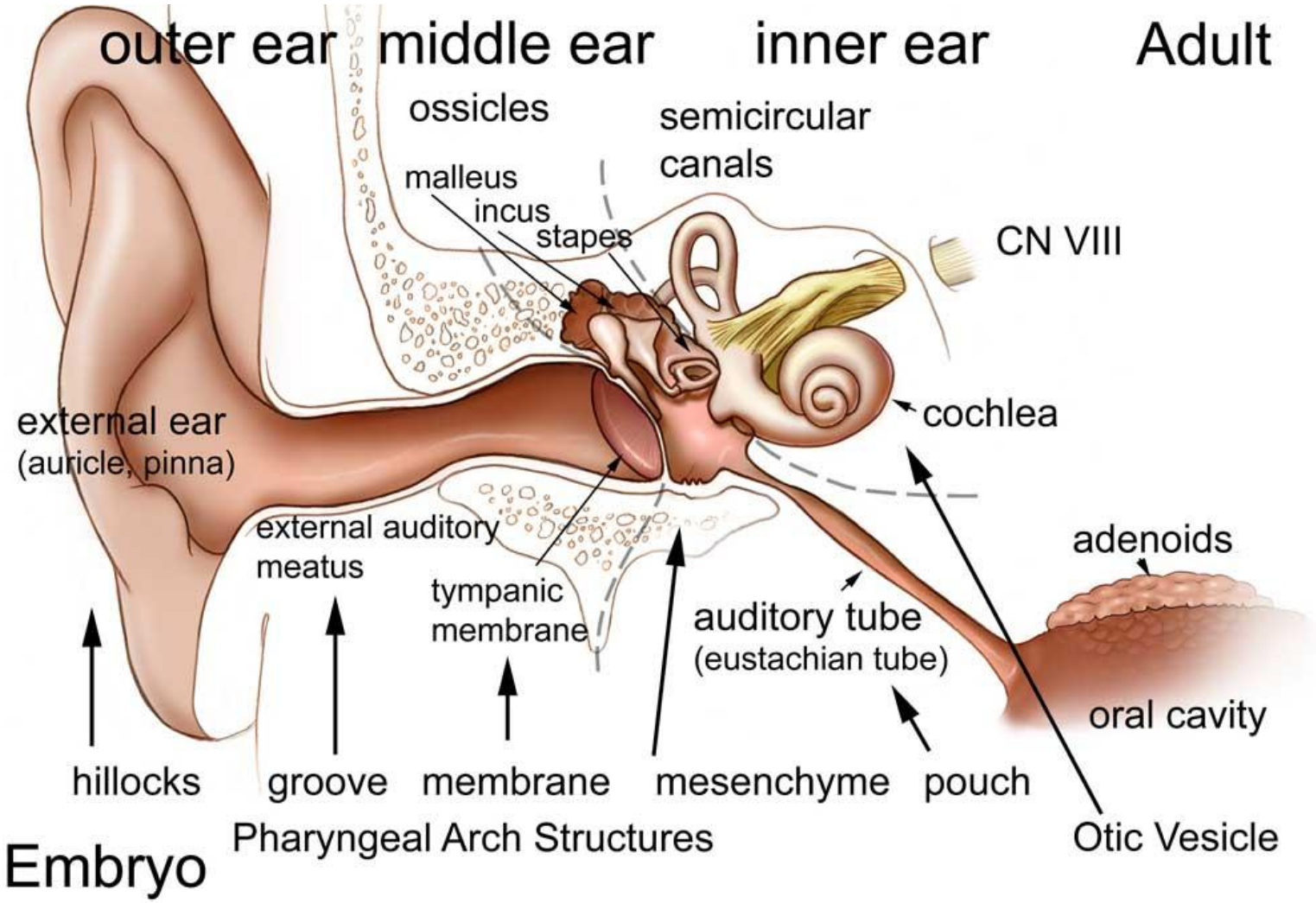
•Saltzman, Elliot. "Task dynamic coordination of the speech articulators: A preliminary model." AVAILABLE FROM US Department of Commerce, National Technical information Service, 5285 Port Royal Rd., Springfield, VA 22151. PUB TYPE Reports-Research/Technical (143) EDRS PRICE MF01/PC11 Plus Postage. (1986): 9.

•Nam, Hosung, et al. "TADA: An enhanced, portable Task Dynamics model in MATLAB." The Journal of the Acoustical Society of America 115.5 (2004): 2430-2430.

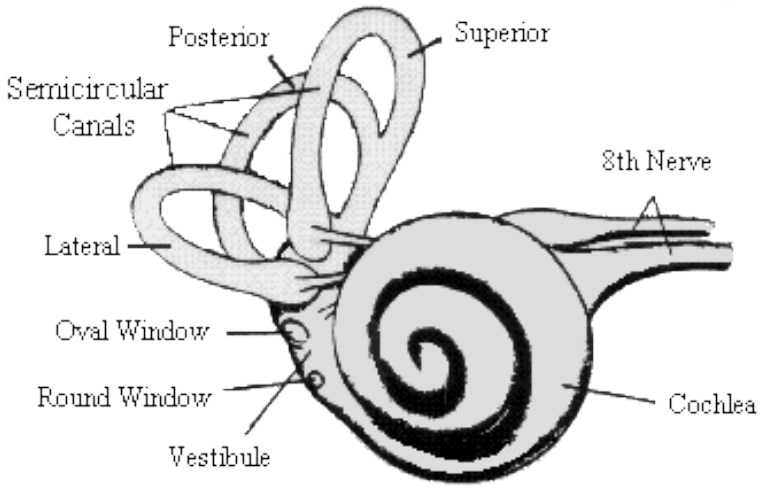
Forward Model

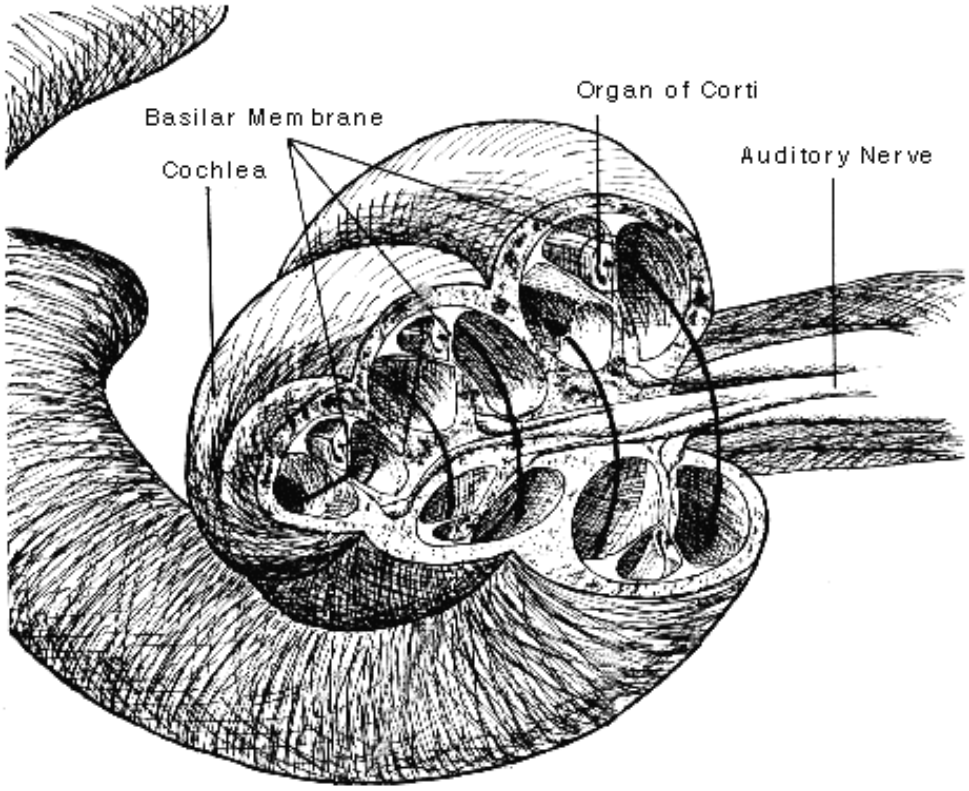
Heinks-Maldonado, Theda H., Srikantan S. Nagarajan, and John F. Houde. "Magnetoencephalographic evidence for a precise forward model in speech production." Neuroreport 17.13 (2006): 1375-1379.

The human ear

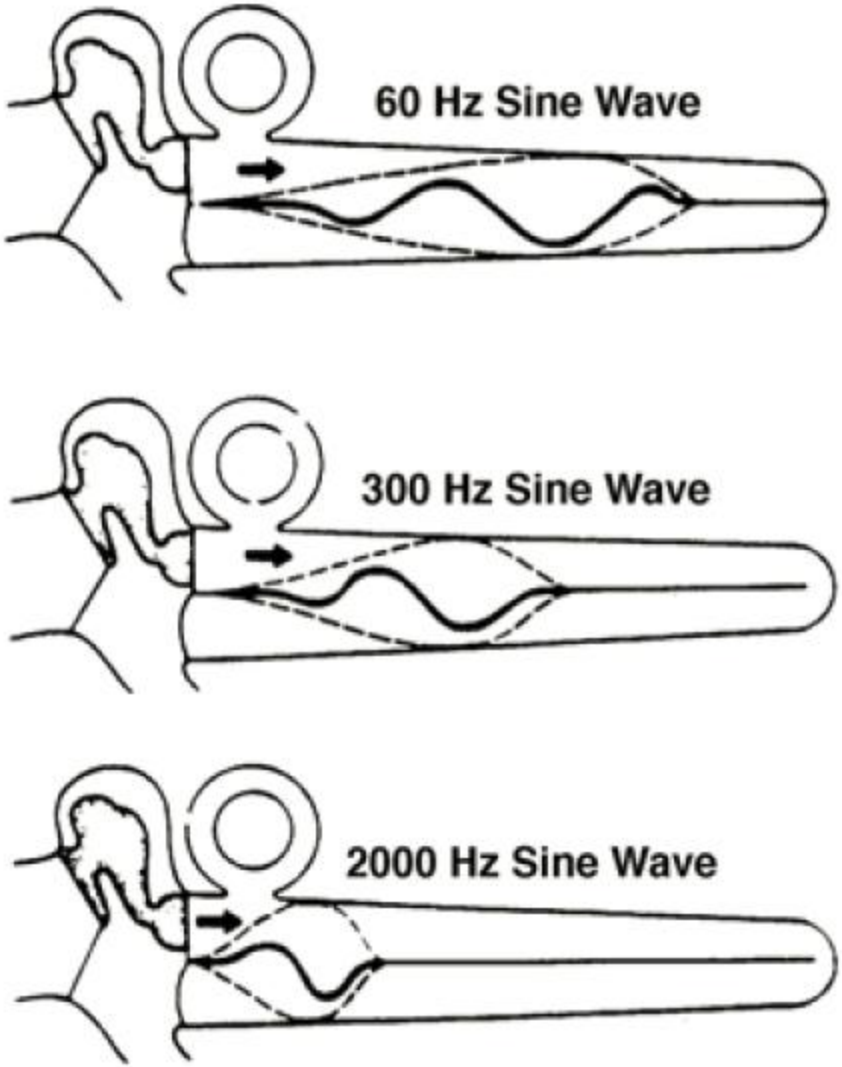


Cochlea – organ of hearing

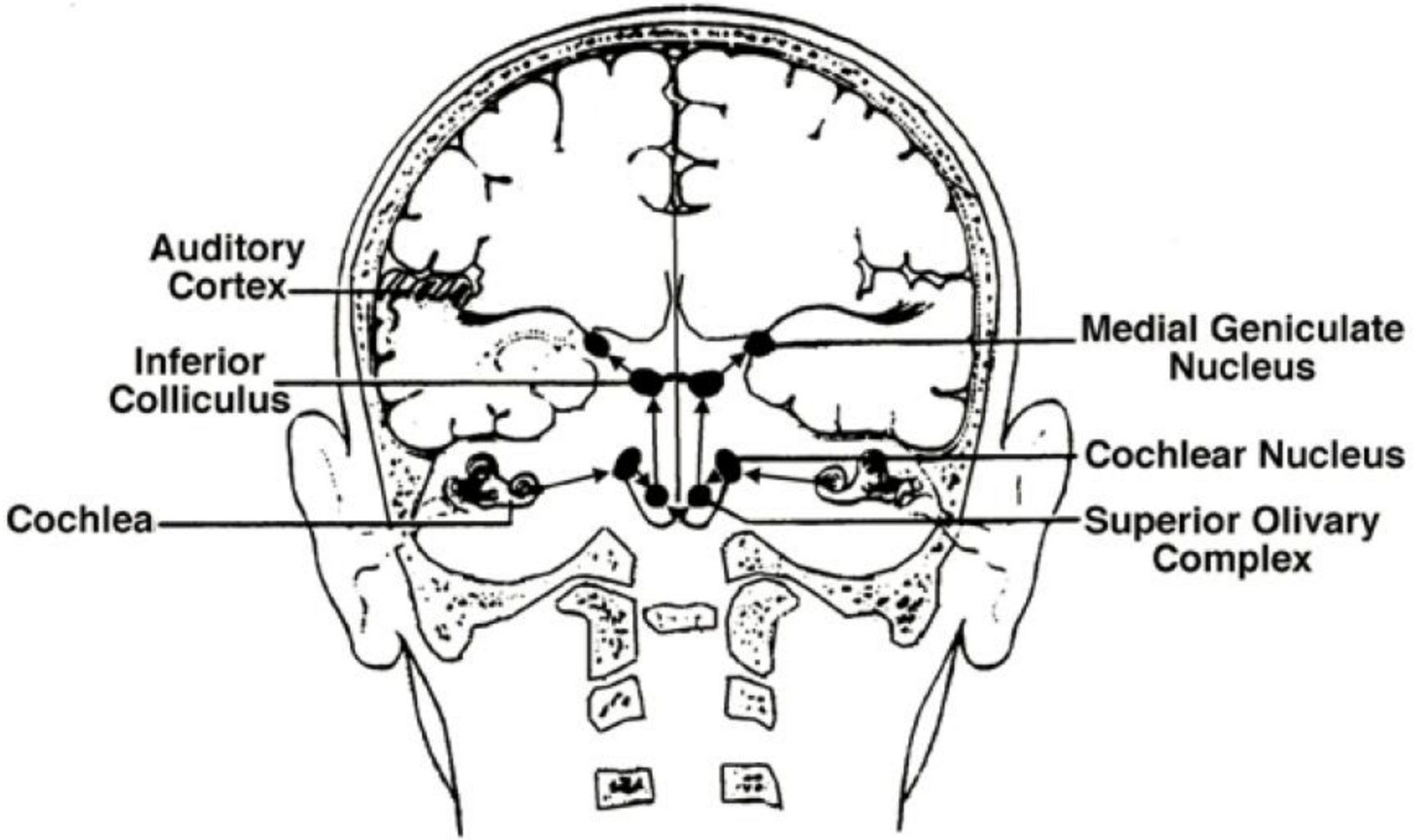




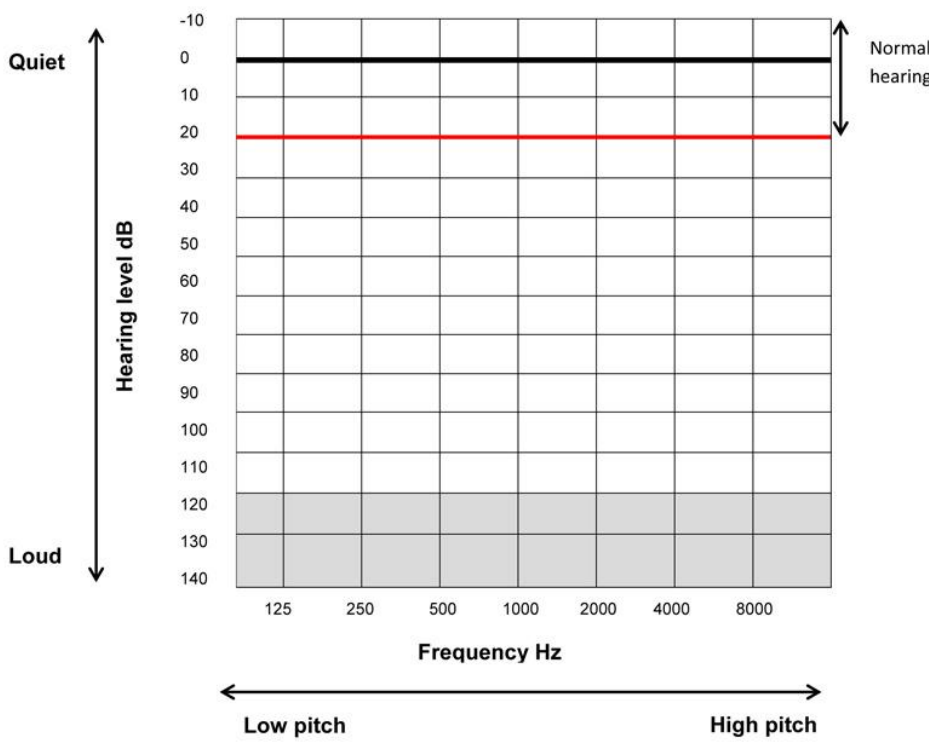
Tonotopic Mapping

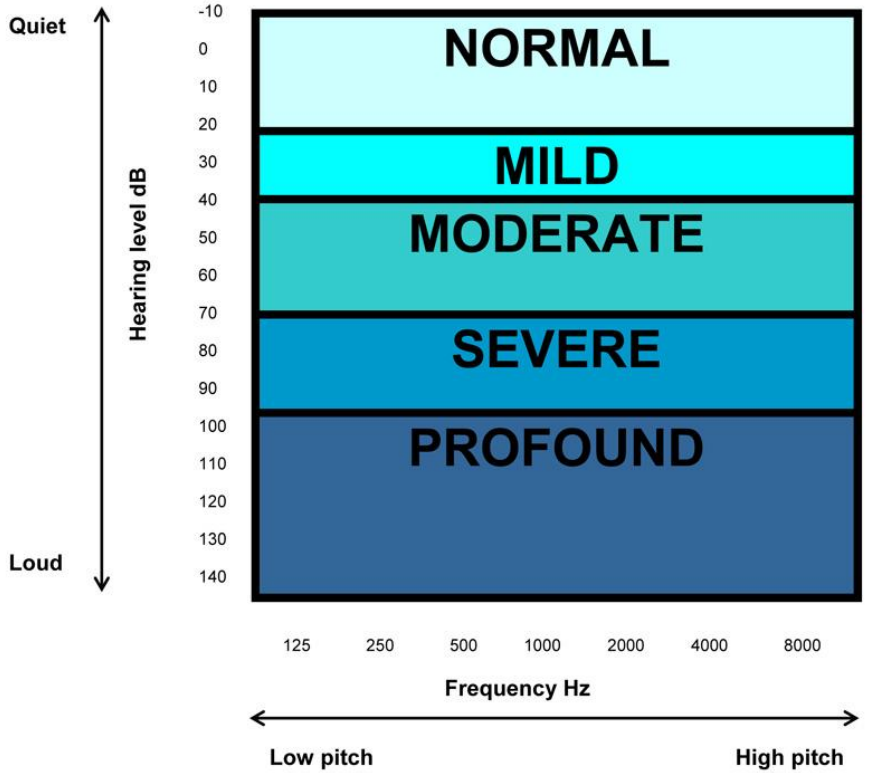


Central Auditory system



Audiogram


Quiet
Loud
Hearing level dB
-10
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
125
250
500
1000
2000
4000
8000
Frequency Hz
Low pitch
High pitch
Normal hearing


Quiet
Loud
Hearing level dB
-10
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
NORMAL
MILD
MODERATE
SEVERE
PROFOUND
125
250
500
1000
2000
4000
8000
Frequency Hz
Low pitch
High pitch