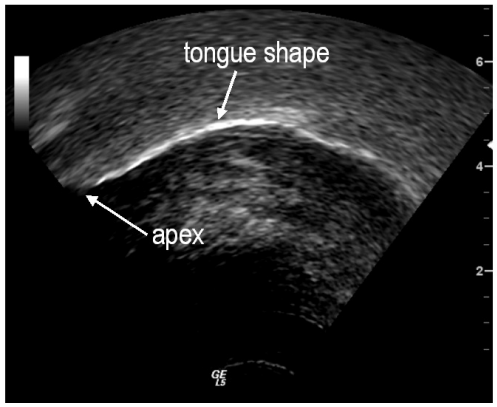
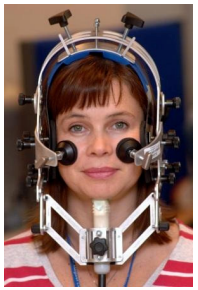
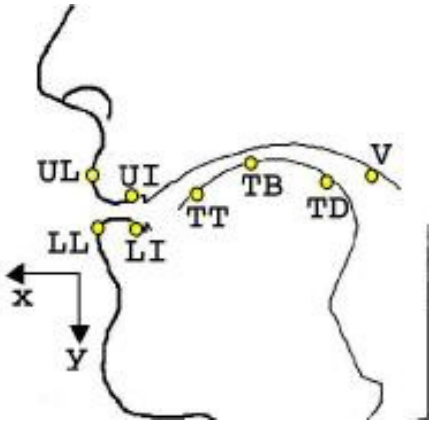
Ultrasound



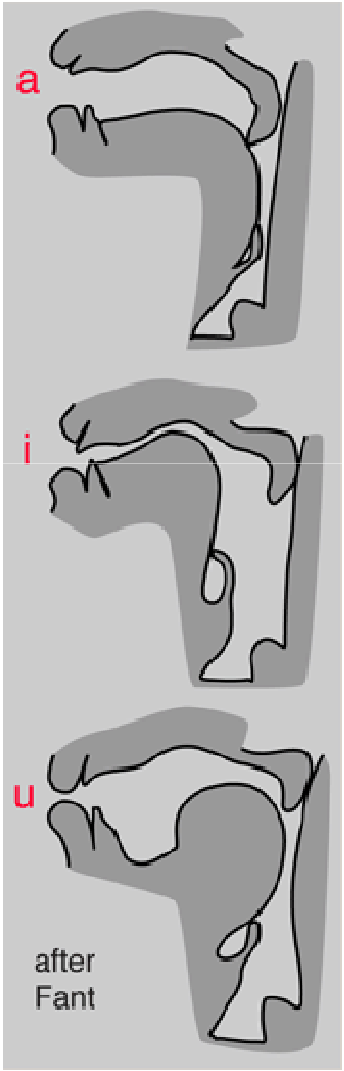


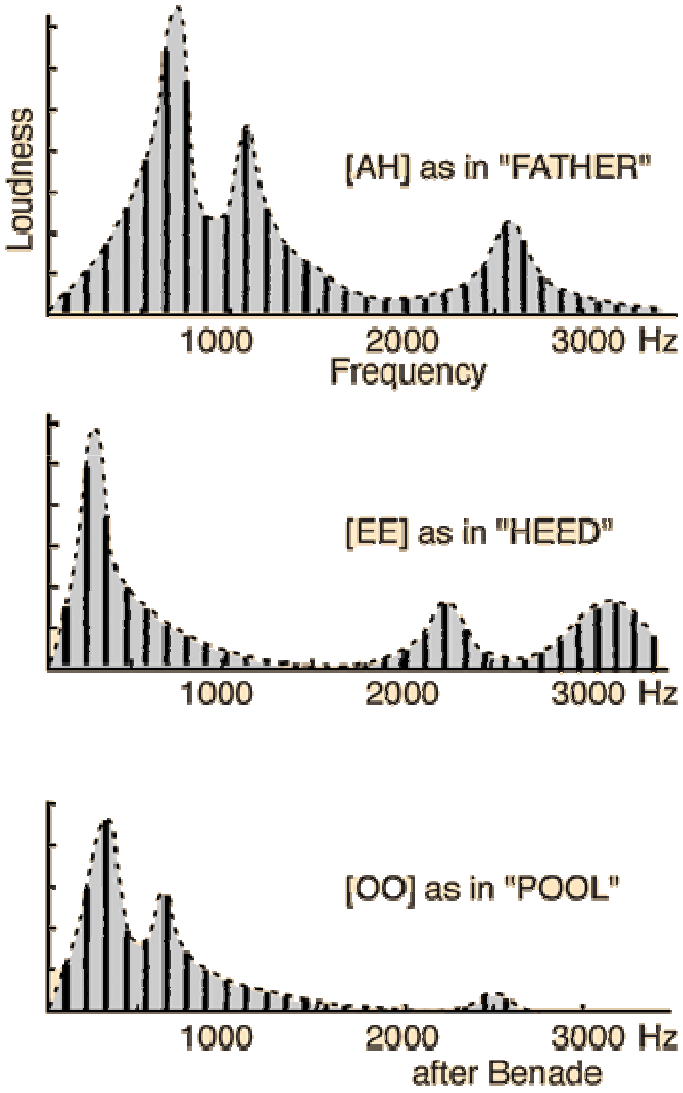
Articulograph



rtMRI

Articulation to acoustic





For vowels

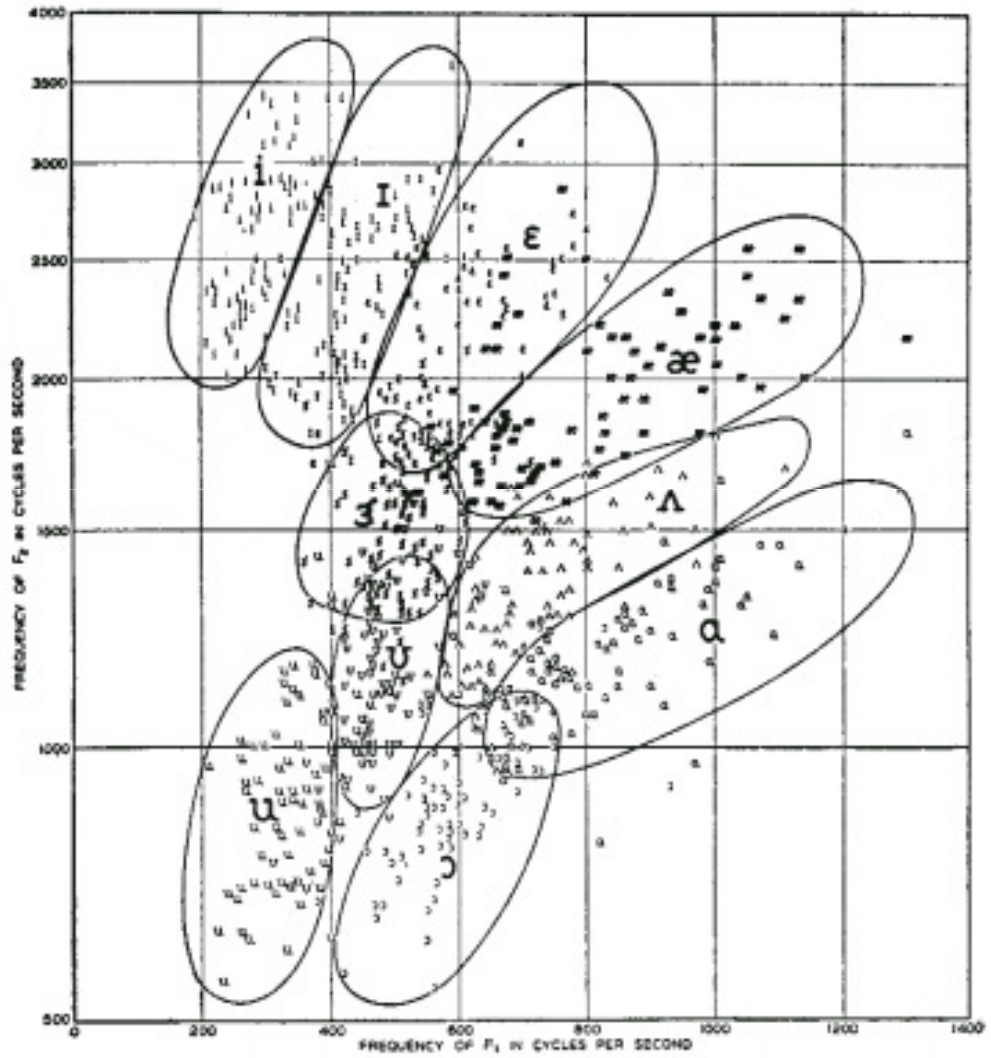
Formant

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.

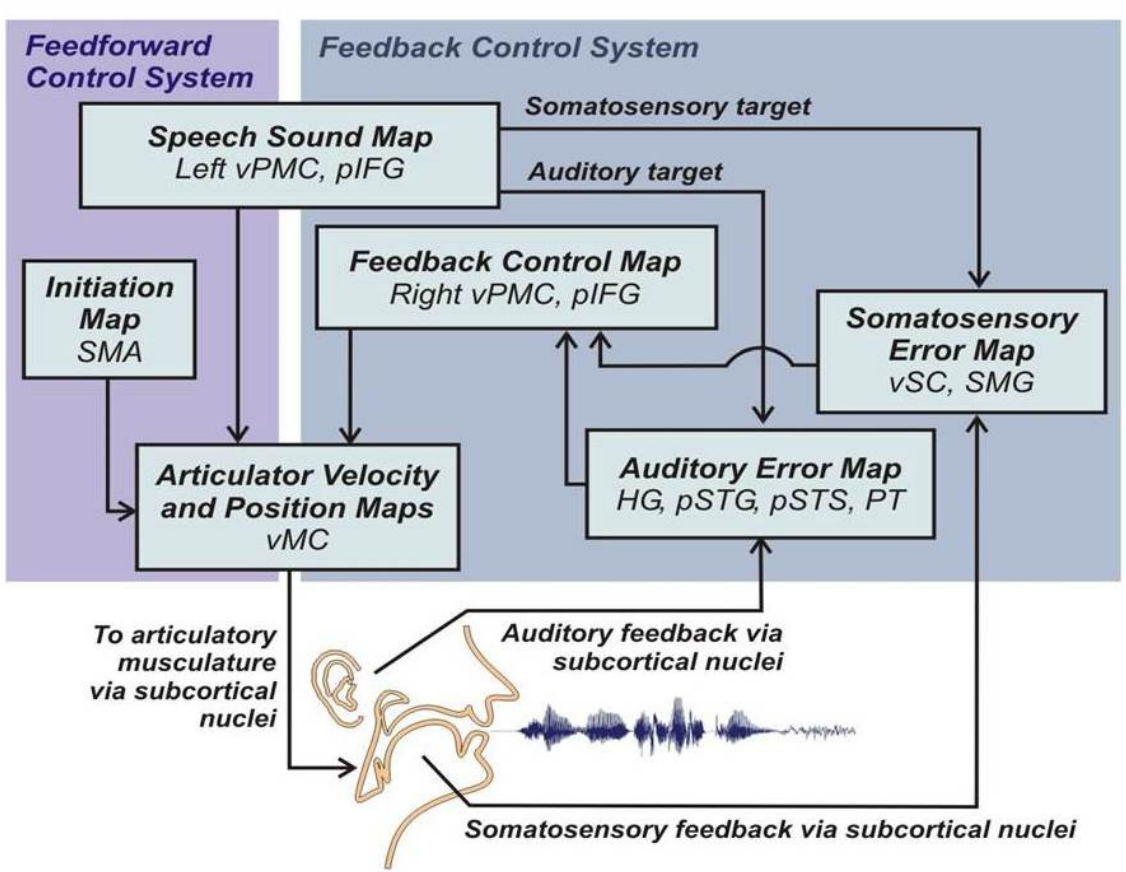
Formant



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

Speech production models

DIVA Model

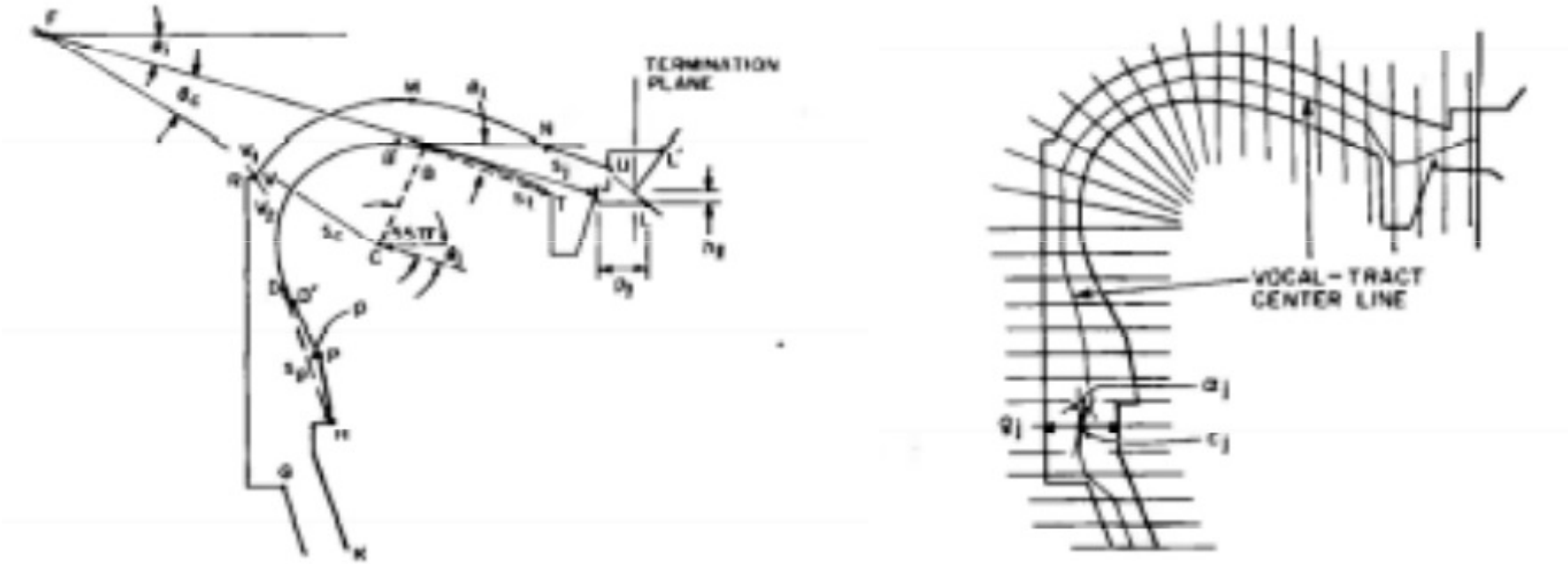


Guenther, Ghosh, and Tourville (2006) *Brain and Language*

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

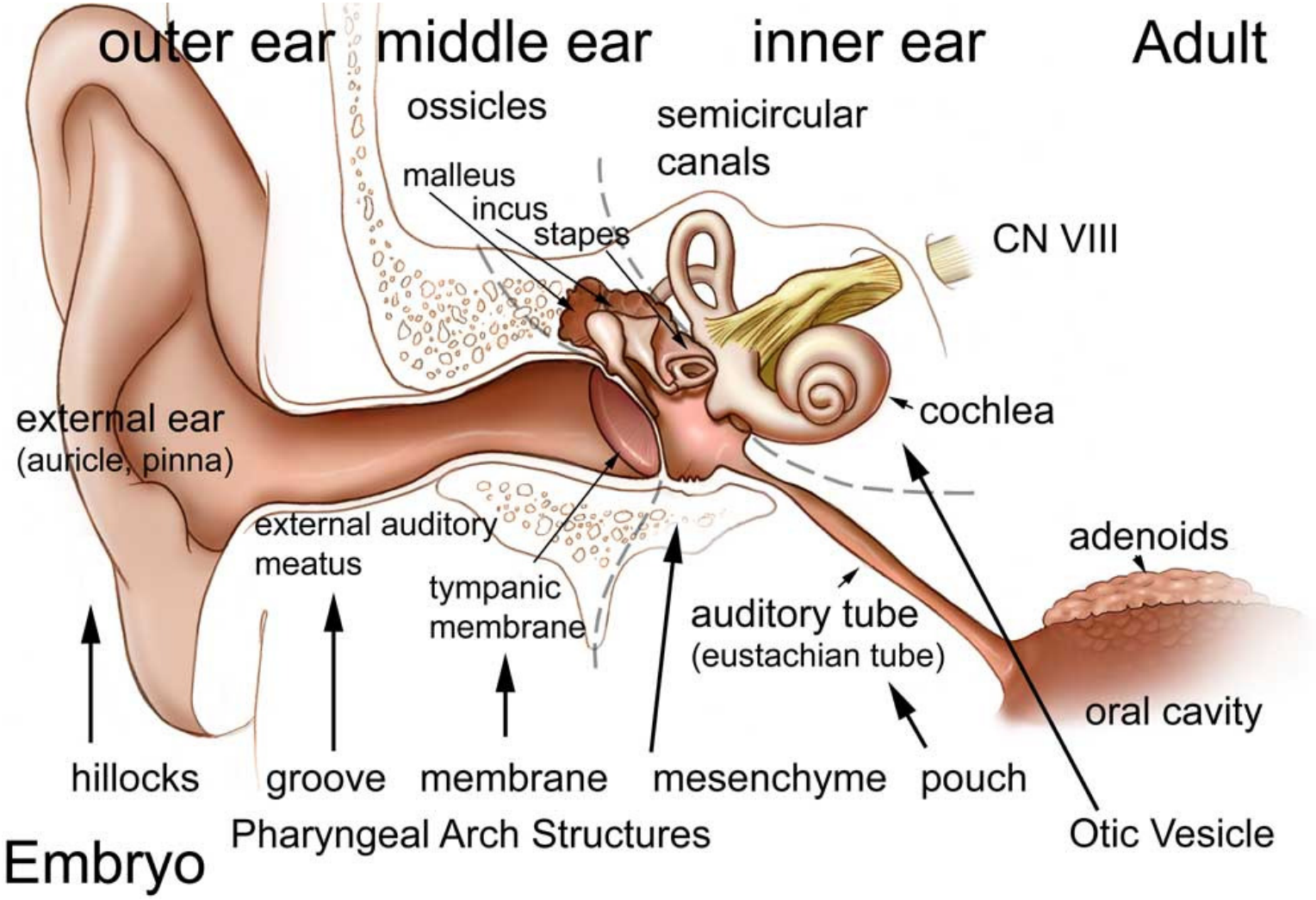
TaDA Model

•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.

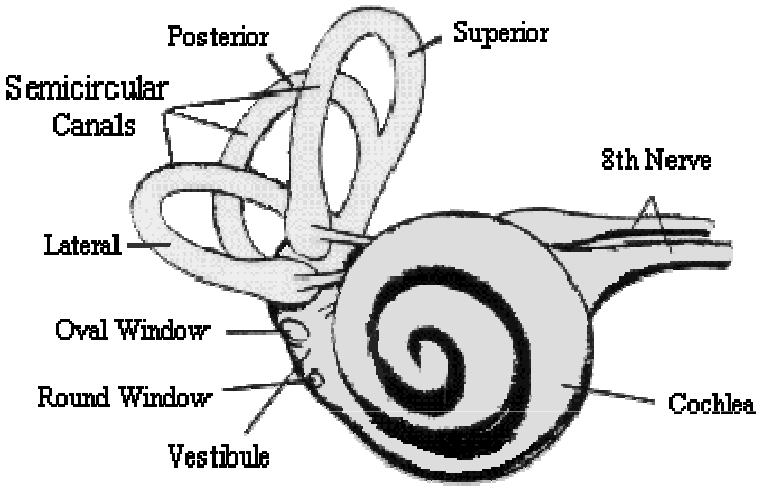
Forward Model

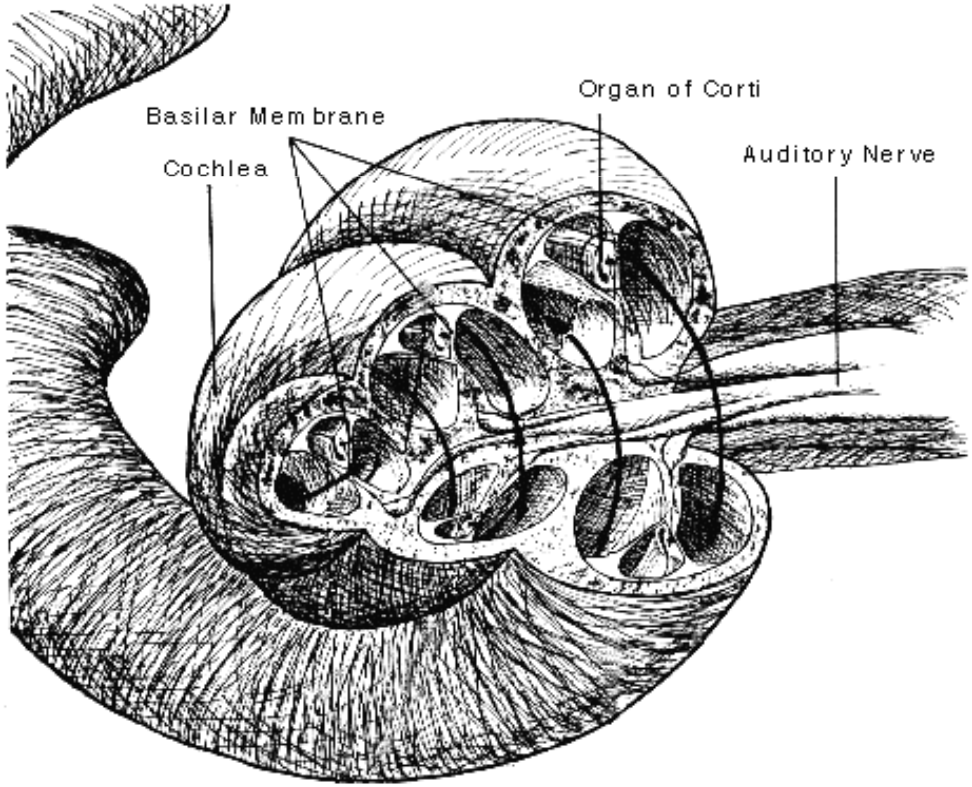
Heinks-Maldonado, Theda H., Srikantan S. Nagarajan, and John F. Houde. "Magnetoencephalographic evidence for a precise forward model in speech production." Neuroreport 17.13 (2006): 1375-1379.

The human ear

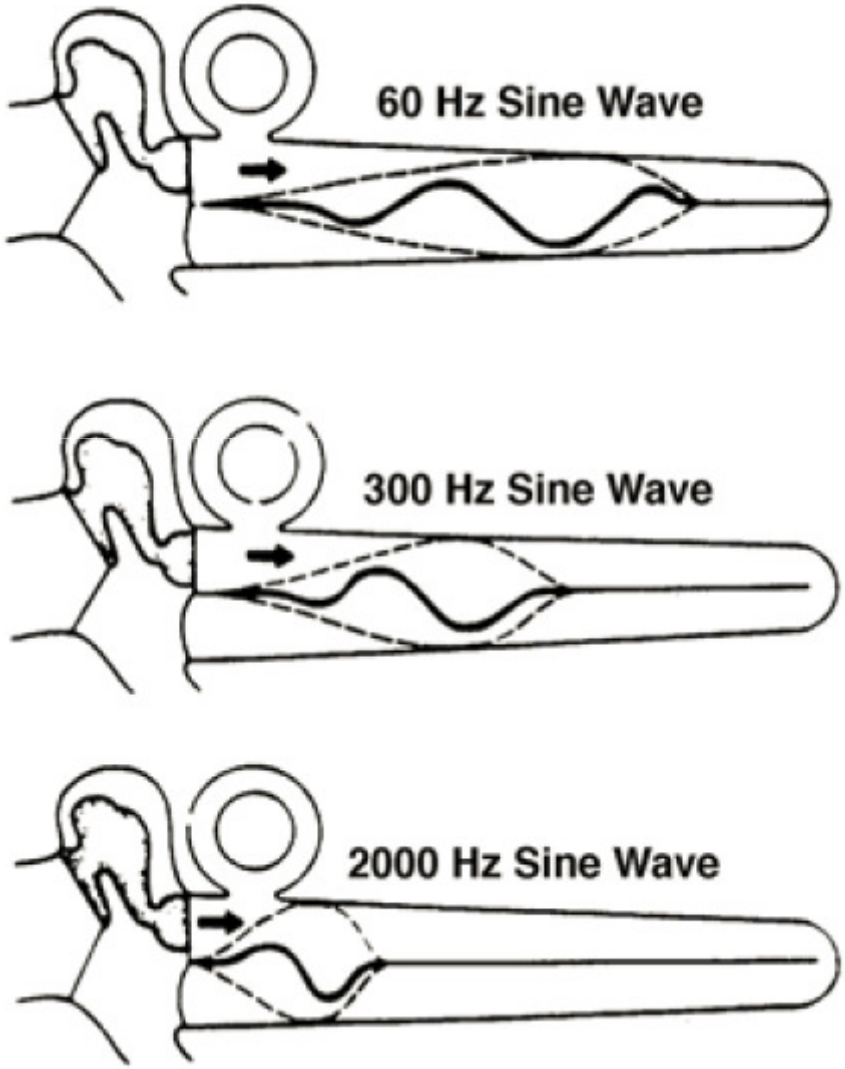


Cochlea – organ of hearing

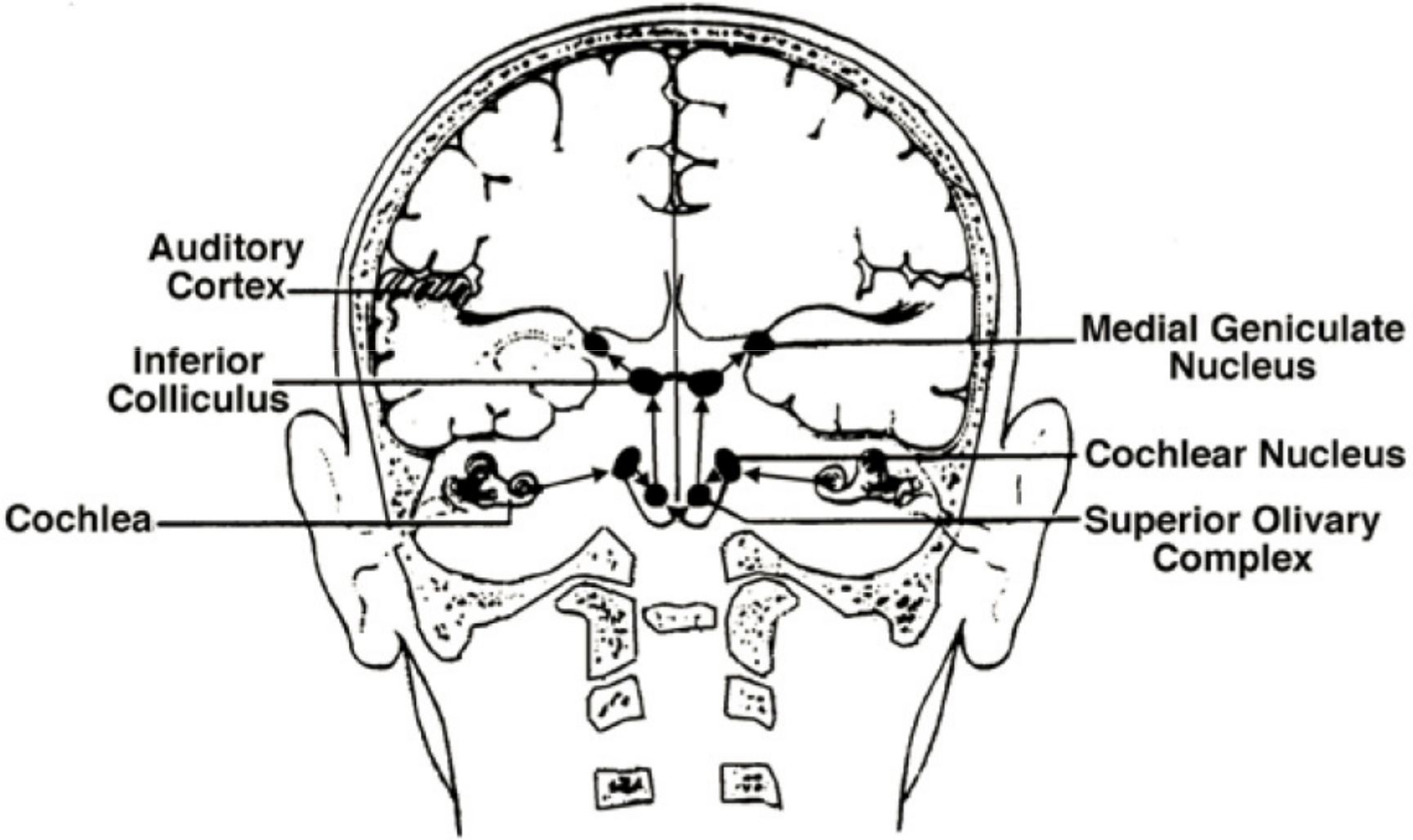




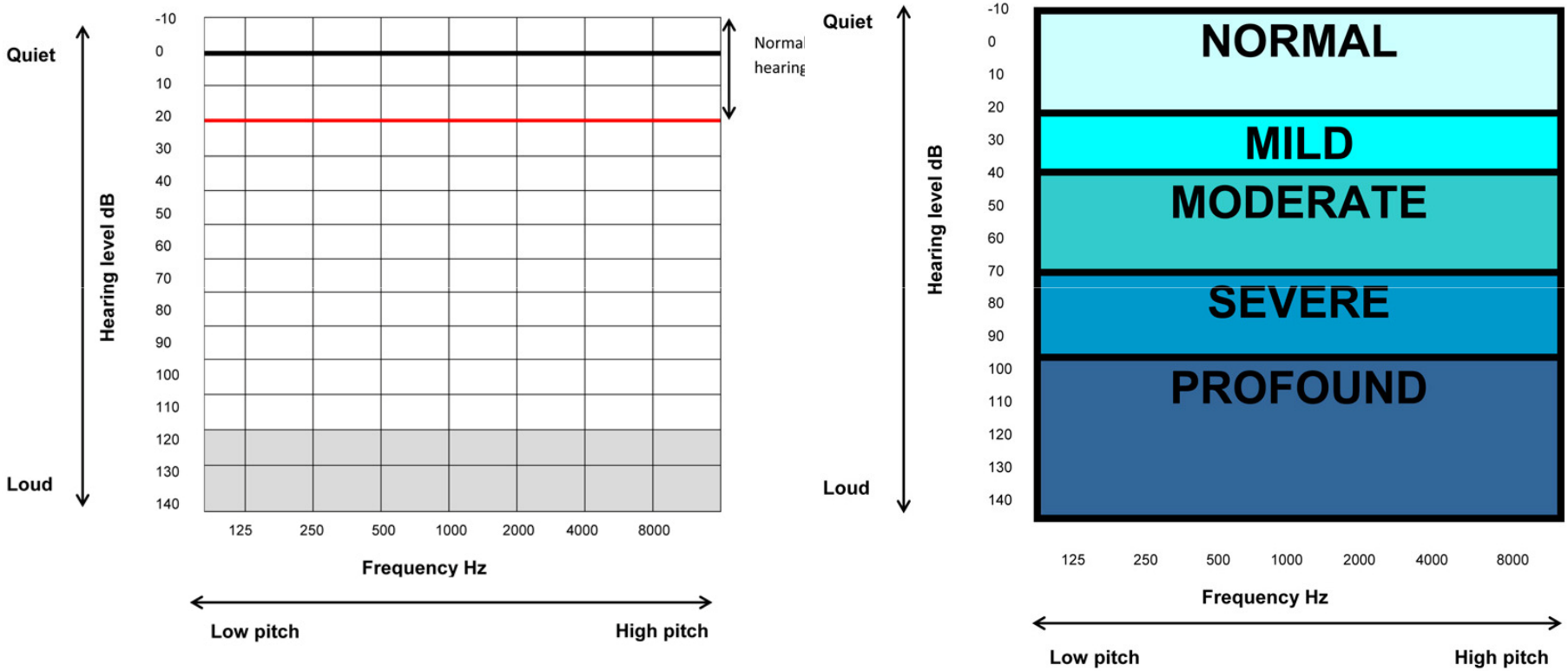
Tonotopic Mapping



Central Auditory system



Audiogram

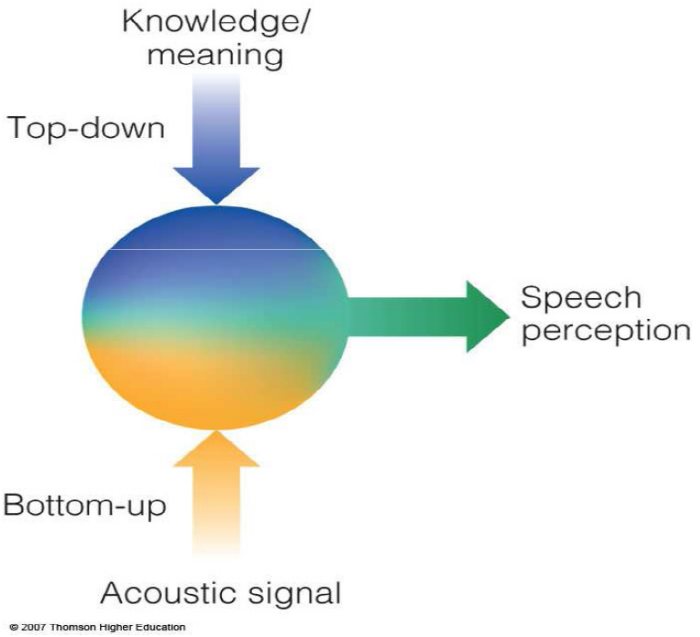

Quiet
Loud
Hearing level dB
-10
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
125
250
500
1000
2000
4000
8000
Frequency Hz
Low pitch
High pitch
Normal hearing
Quiet
Loud
Hearing level dB
NORMAL
MILD
MODERATE
SEVERE
PROFOUND
Frequency Hz
Low pitch
High pitch

Theories of Speech Perception

- **Active theories** suggests that speech perception and production are closely related
  - Listener knowledge of how sounds are produced facilitates recognition of sounds
- **Passive theories** emphasizes the sensory aspects of speech perception
  - Listeners utilize internal filtering mechanisms
  - Knowledge of vocal tract characteristics plays a minor role, for example when listening in noise conditions

Bottom up Top Down

- Top-down processing works with knowledge a listener has about a language, context, experience, etc.
  - Listeners use stored information about language and the world to make sense of the speech
- Bottom-up processing works in the absence of a knowledge base providing top-down information
  - listeners receive auditory information, convert it into a neural signal and process the phonetic feature information

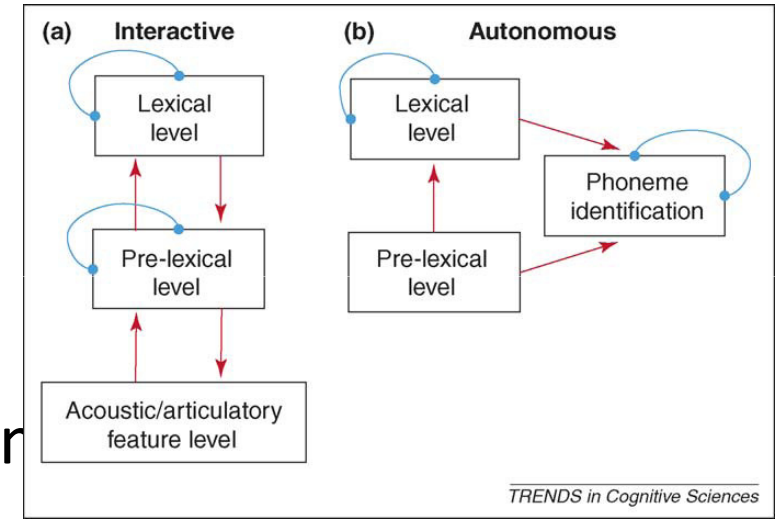


Knowledge driven top-down approach are less resistant to Additive Noise in case of non-sense syllable, nonsense words, incoherent sentence, short utterance, ungrammatical sentence

Bottom-up approach results in error propagation upto top

Autonomous vs. Interactive

- **Autonomous theories** posit feed-forward processing with lexical influence restricted to post-perceptual decision processes (uni-directional)
- **Interactive theories** posit information and knowledge fror many sources available to the listener a re involved at any or all stages of the processing of the signal (bi-directional)



Theories of Speech Perception

Marslen-Wilson's Cohort Model

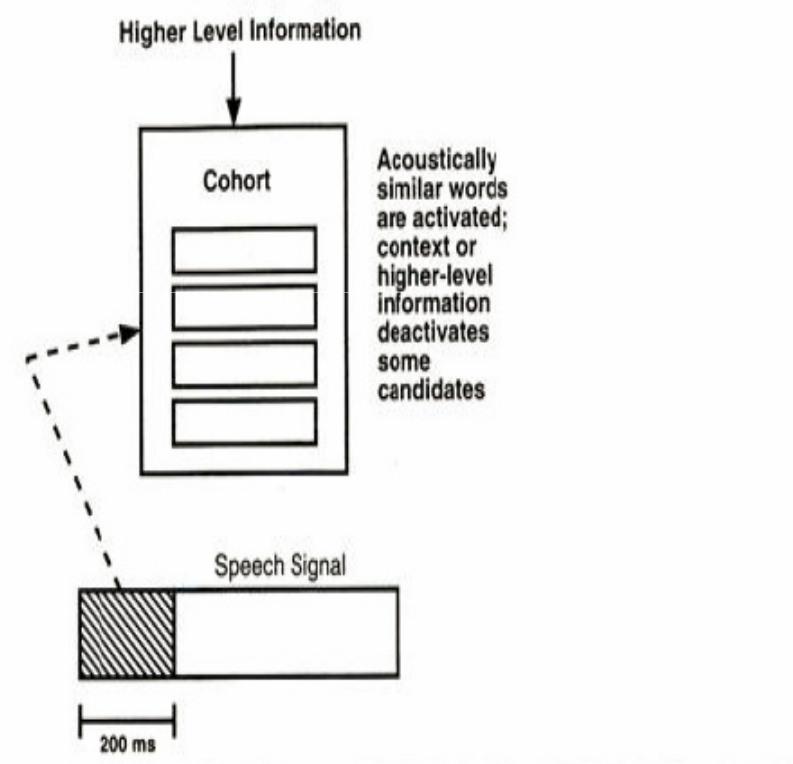


FIGURE 6-11 The Cohort Model of Word Recognition. *(Reprinted with permission by Singular Publishing Group, Inc. Kent, R. [1997].* The Speech Sciences. *San Diego: Singular, 388.)*

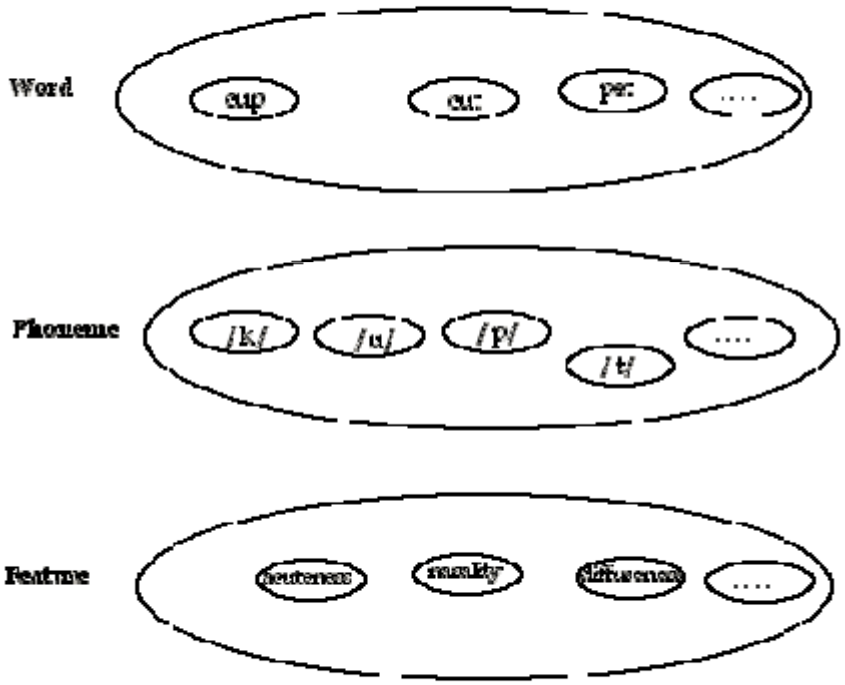
- Mental representations of words activated (in parallel) on the basis of bottom-up input (sounds)
- Can be de-activated by subsequent input
  - bottom-up (phonological)
  - top-down (contextual)

(Marslen-Wilson, 1980)

Theories of Speech Perception

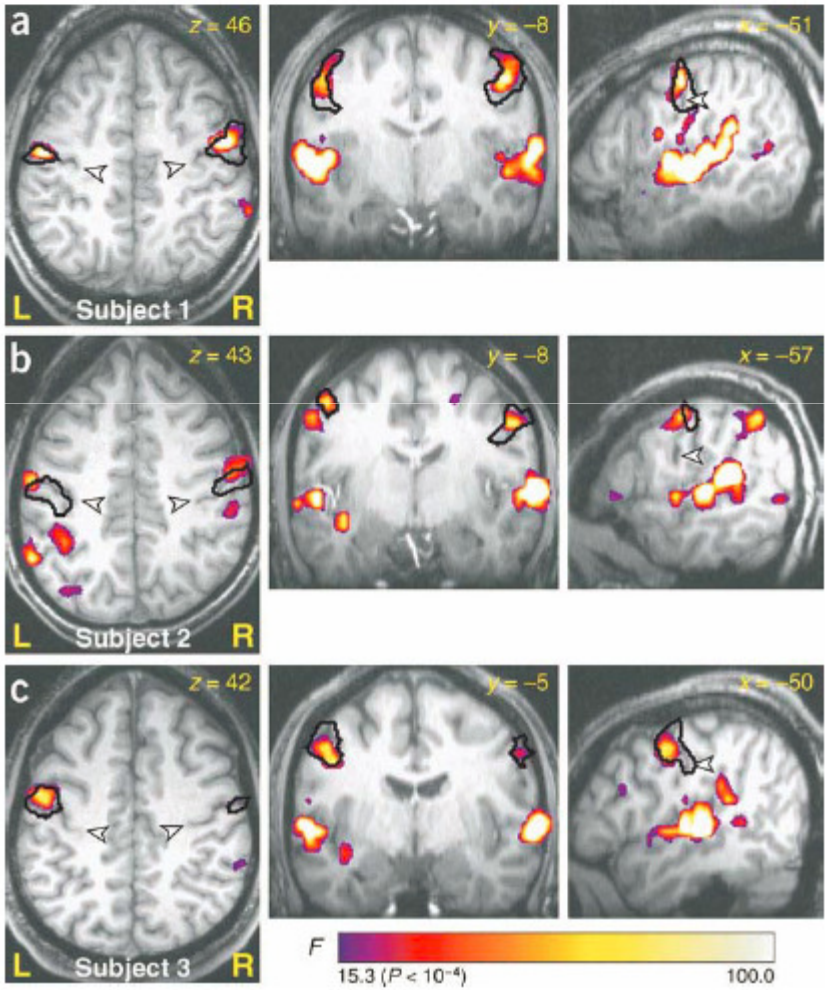
TRACE Model

- Like the interactive-activation model of printed word recognition, TRACE has three sets of interconnected detectors
  - Feature detectors
  - Phoneme detectors
  - Word detectors
- These detectors span different stretches of the input (feature detector span small parts, word detectors span larger parts)
- The input is divided into "time slices" which are processed sequentially.



(McClelland & Elman, 1986)

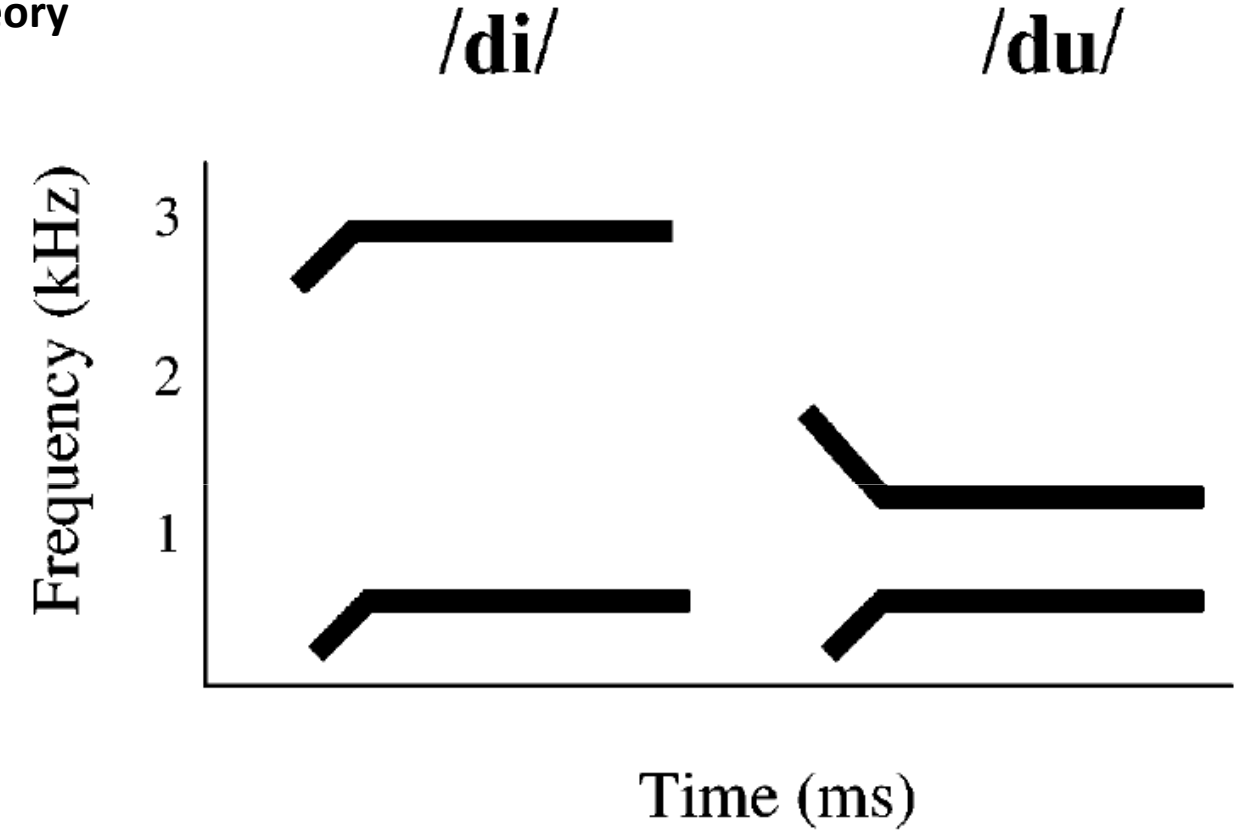
Wilson et al., 2004



- Black areas are premotor and primary motor cortex activated when subjects produced the syllables
- White arrows indicate central sulcus
- Orange represents areas activated by listening to speech
- Extensive activation in superior temporal gyrus
- Activation in motor areas involved in speech production (!)

Theories of Speech Perception

Motor Theory



Motor theory postulates that speech is perceived by reference to how it is produced; that is, when perceiving speech, listeners access their own knowledge of how phonemes are articulated. Articulatory gestures such as rounding or pressing the lips together are units of perception that directly provide the listener with phonetic information.

(Liberman, et al., 1967; Liberman & Mattingly, 1985)

Theories of Speech Perception

Analysis by Synthesis (Stevens & Halle, 1960)

- In this model, speech perception is based on auditory matching mediated through speech production.

When a listener hears a speech signal, he or she analyzes it by mentally modeling the articulation (in other words, the listener tries to synthesize the speech his or herself). If the 'auditory' result of the mental synthesis matches the incoming acoustic signal, the hypothesized perception is interpreted as correct.

Theories of Speech Perception

Direct Realist Theory (Fowler, 1986)

- Direct realism postulates that speech perception is direct (i.e., happens through the perception of articulatory gestures), but it is not special. All perception involves direct recovery of the distal source of the event being perceived (Gibson).

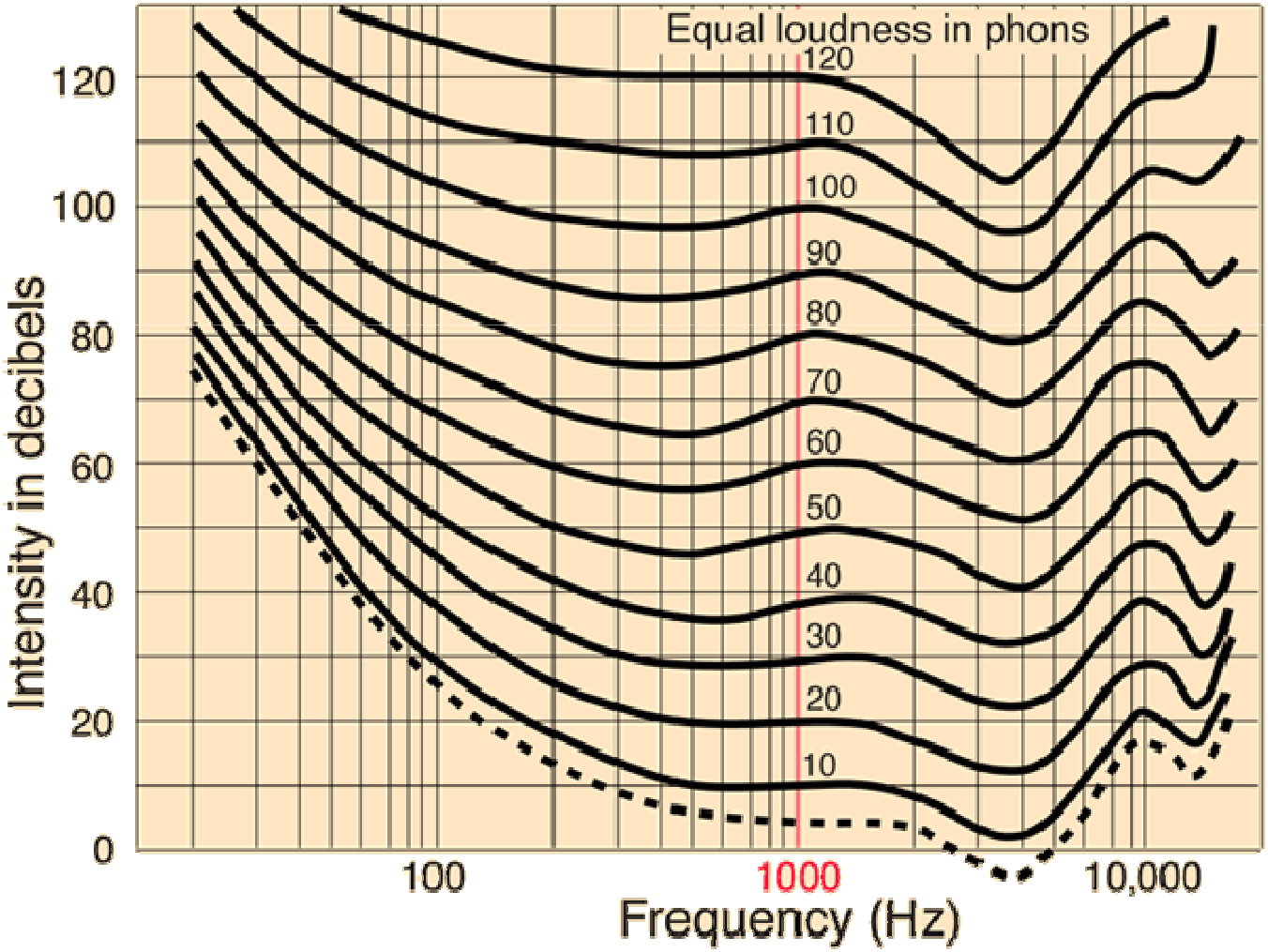
  In vision, you perceive *objects* (e.g., trees, cars, etc.). Likewise with smell you perceive e.g., cookies, roses, etc. Why not in the auditory perception of speech?

  - So, listeners perceive tongues and lips.

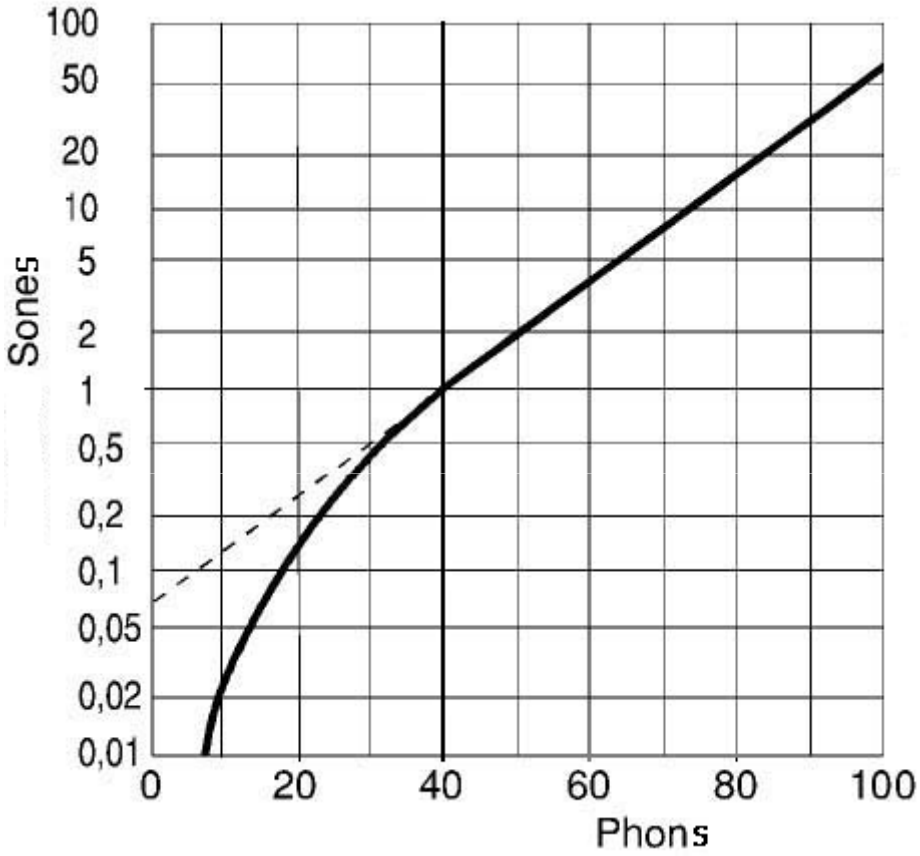
The articulatory gestures that are the objects of speech perception are not *intended* gestures (as in Motor Theory). Rather, they are the *actual* gestures.

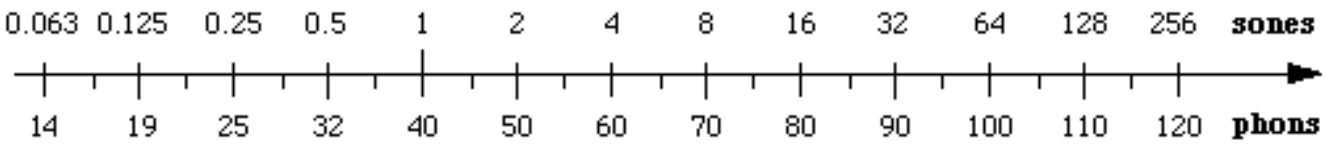
Psycho-acoustic Experiments

Fletcher Munson Curve – loudness as a function of frequency and intensity



Sone vs Phon


Sones
100
50
20
10
5
2
1
0,5
0,2
0,1
0,05
0,02
0,01
0
20
40
60
80
100
Phons


0.063 0.125 0.25 0.5 1 2 4 8 16 32 64 128 256 sones
14 19 25 32 40 50 60 70 80 90 100 110 120 phons