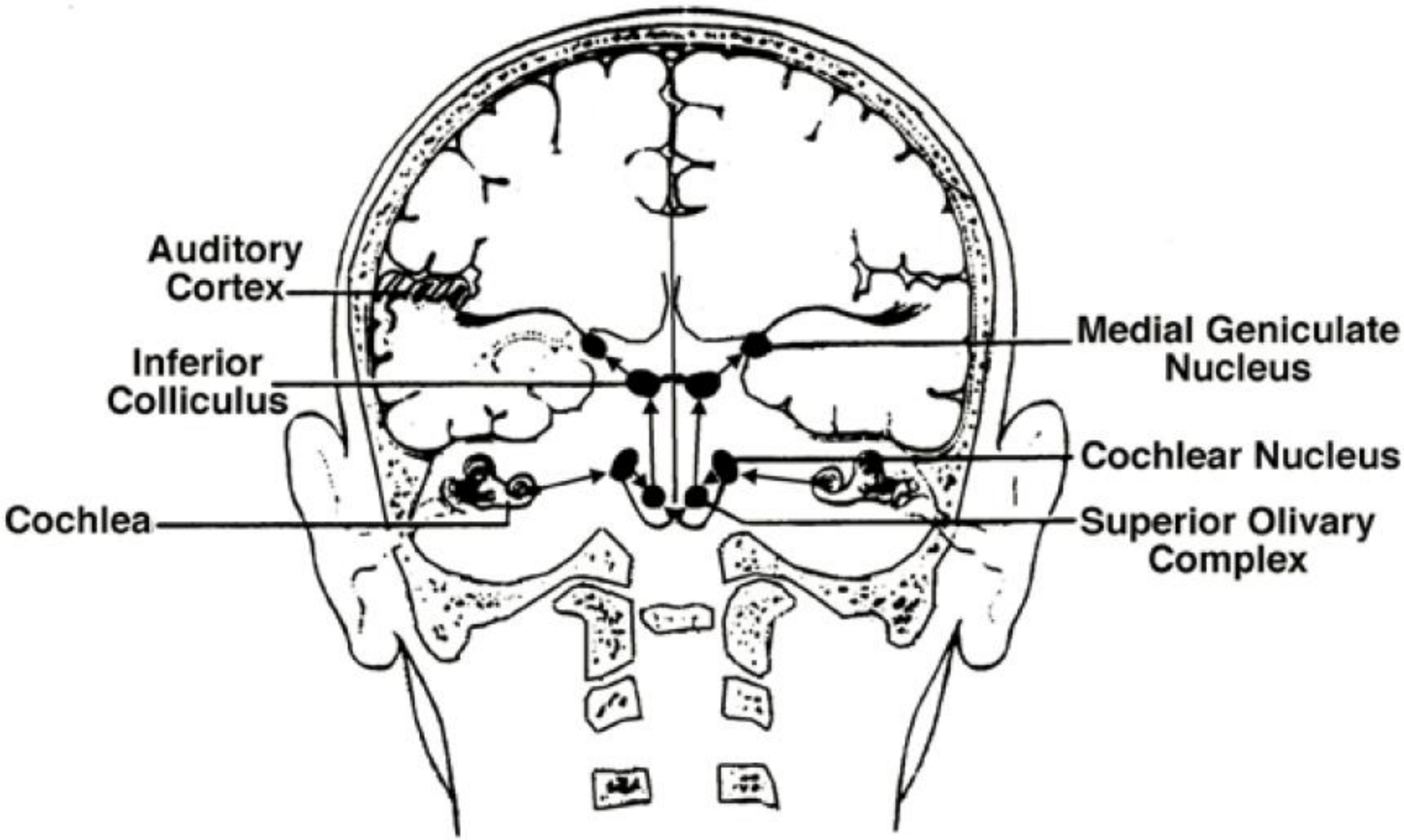
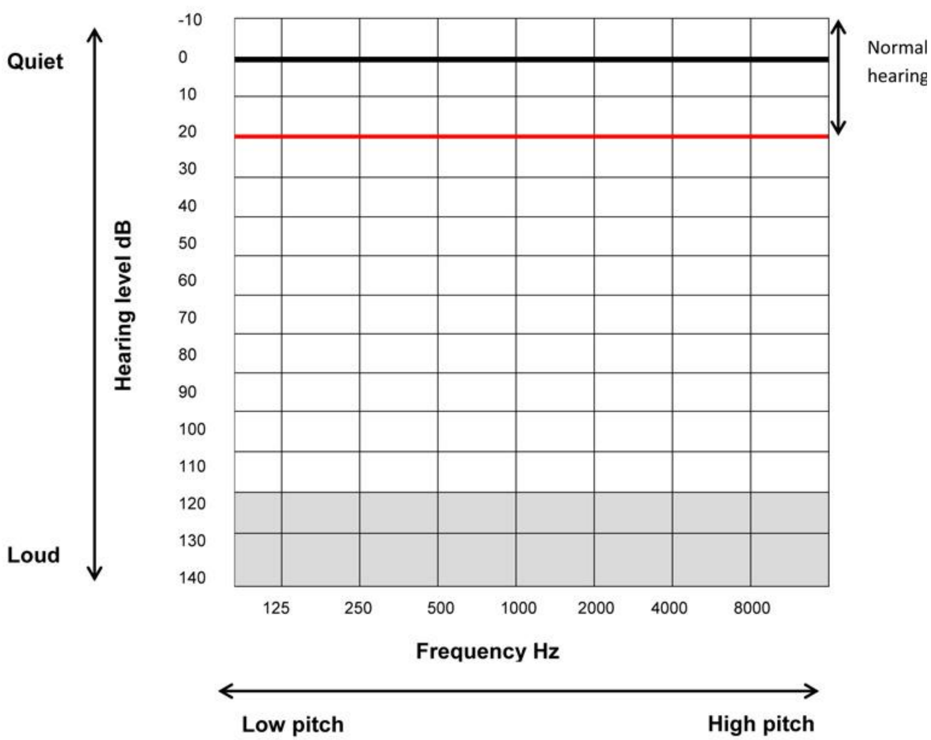
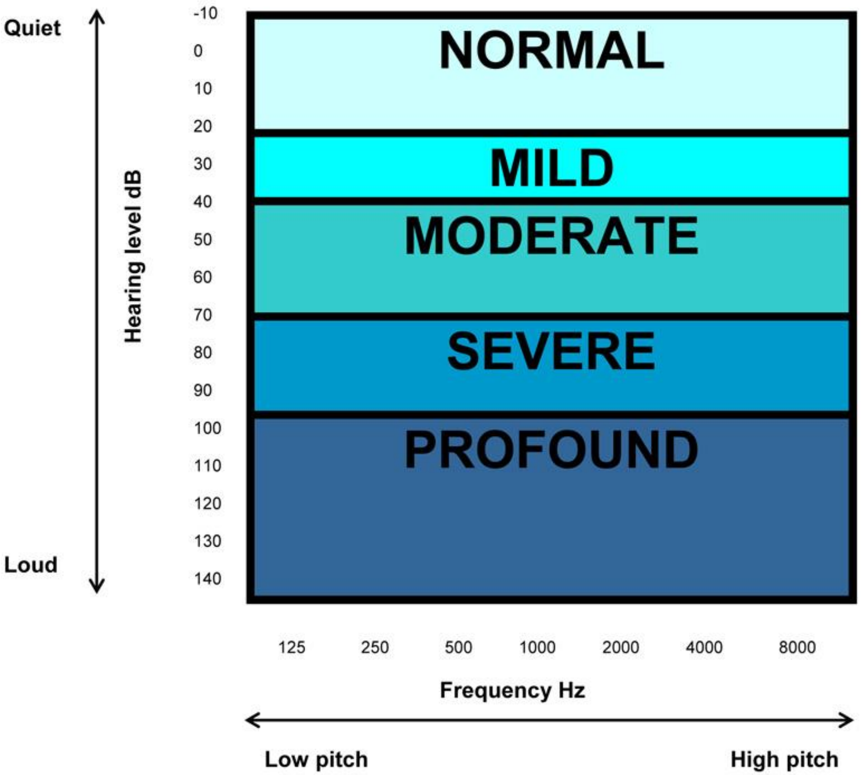
Central Auditory system



Audiogram


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

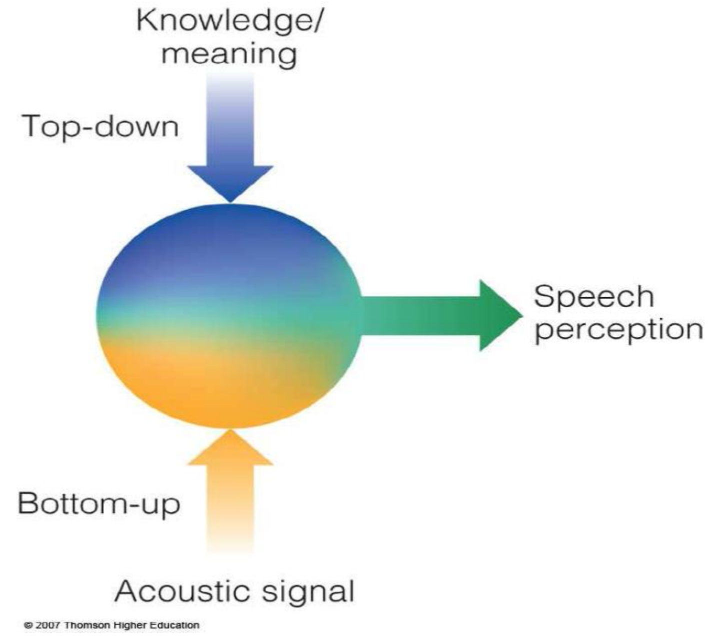

Quiet
Loud
Hearing level dB
-10
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
NORMAL
MILD
MODERATE
SEVERE
PROFOUND
125
250
500
1000
2000
4000
8000
Frequency Hz
Low pitch
High pitch

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

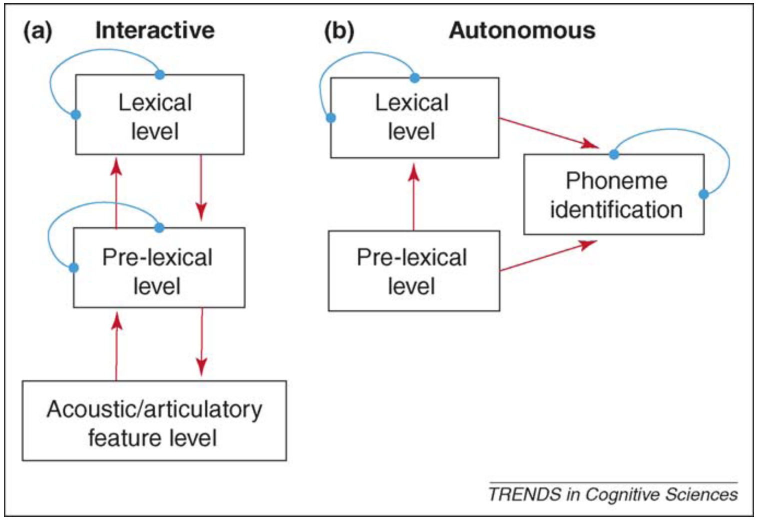


Knowledge driven top-down approach are less resistant to Additive Noise in case of non-sense syllable, nonsense words, incoherent sentence, short utterance, ungrammatical sentence

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 from 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

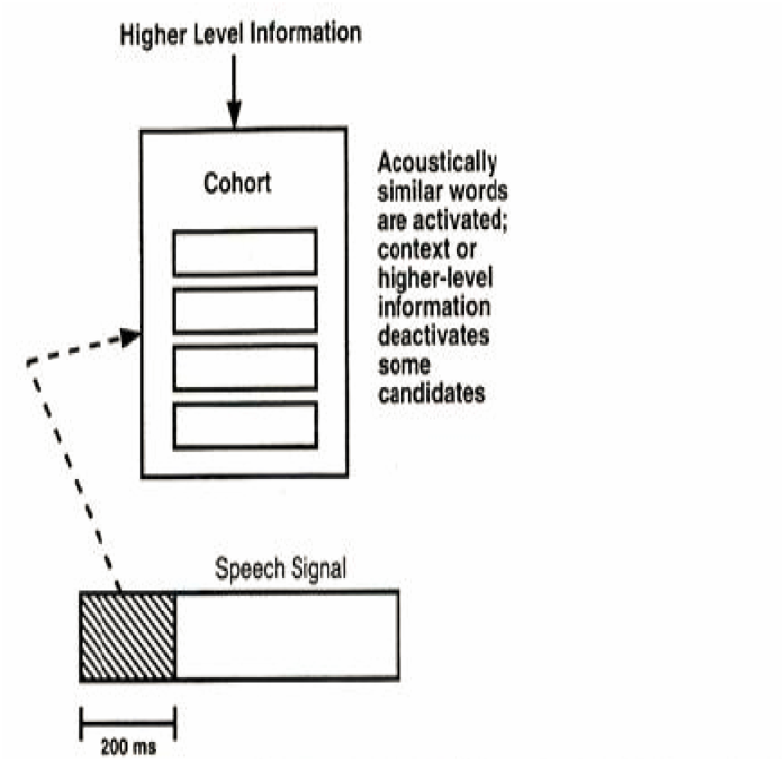


FIGURE 6-11 The Cohort Model of Word Recognition. *(Reprinted with permission by Singular Publishing Group, Inc. Kent, R. [1997]. The Speech Sciences. San Diego: Singular, 388.)*

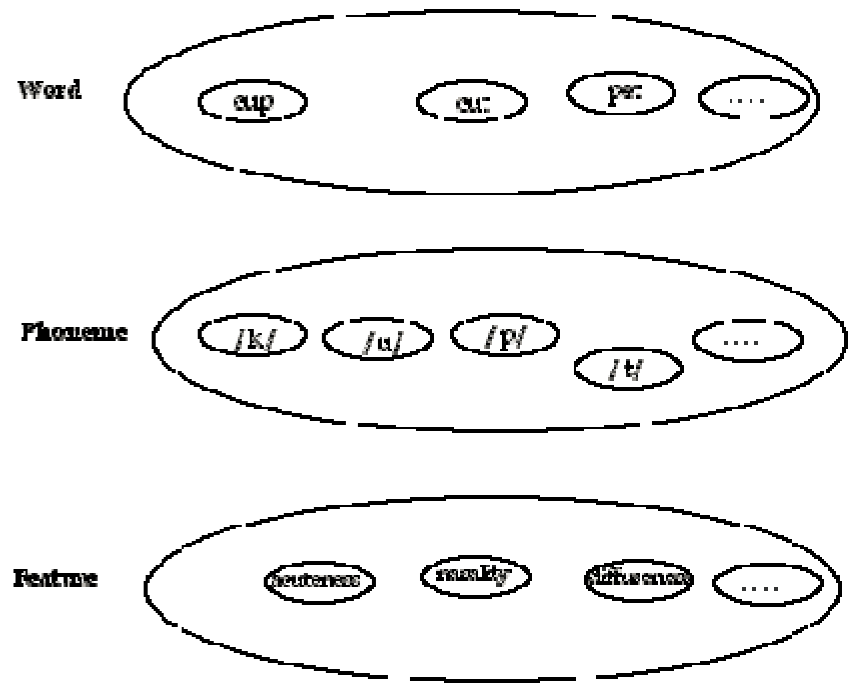
- Mental representations of words activated (in parallel) on the basis of bottom-up input (sounds)
- Can be de-activated by subsequent input
  - bottom-up (phonological)
  - top-down (contextual)

(Marslen-Wilson, 1980)

Theories of Speech Perception

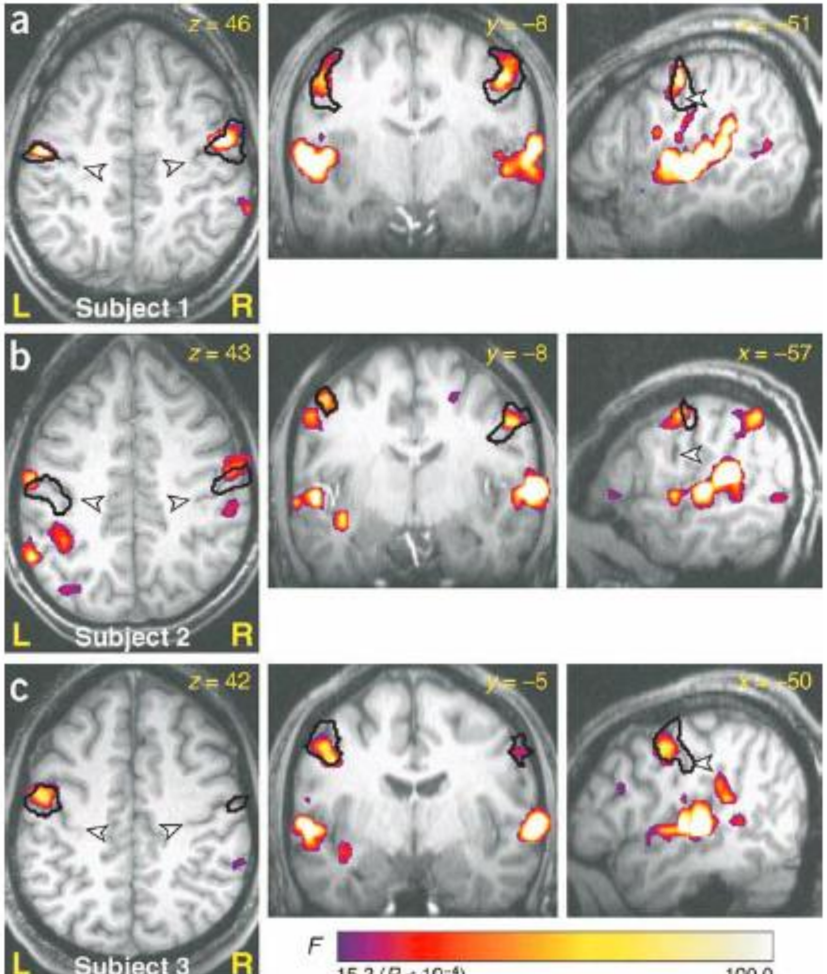
TRACE Model

- 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)

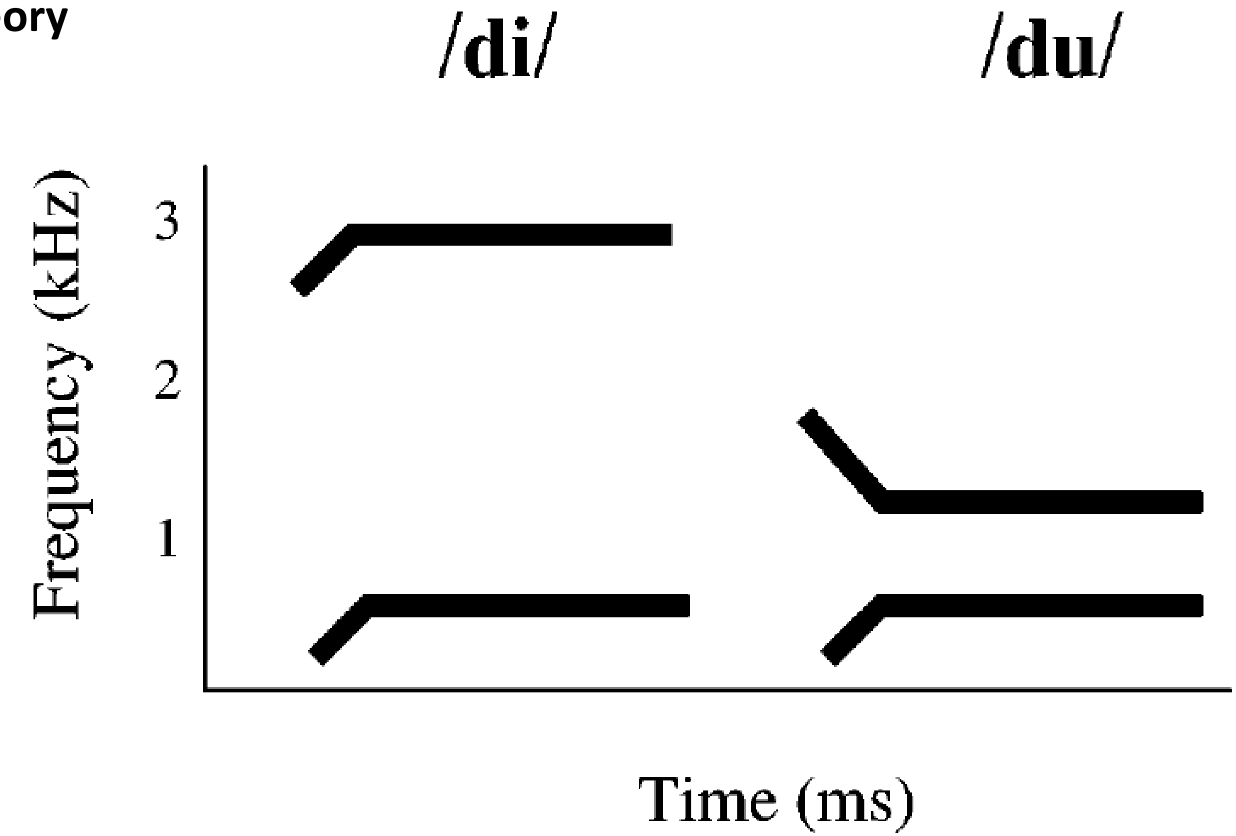
Wilson et al., 2004



- 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).

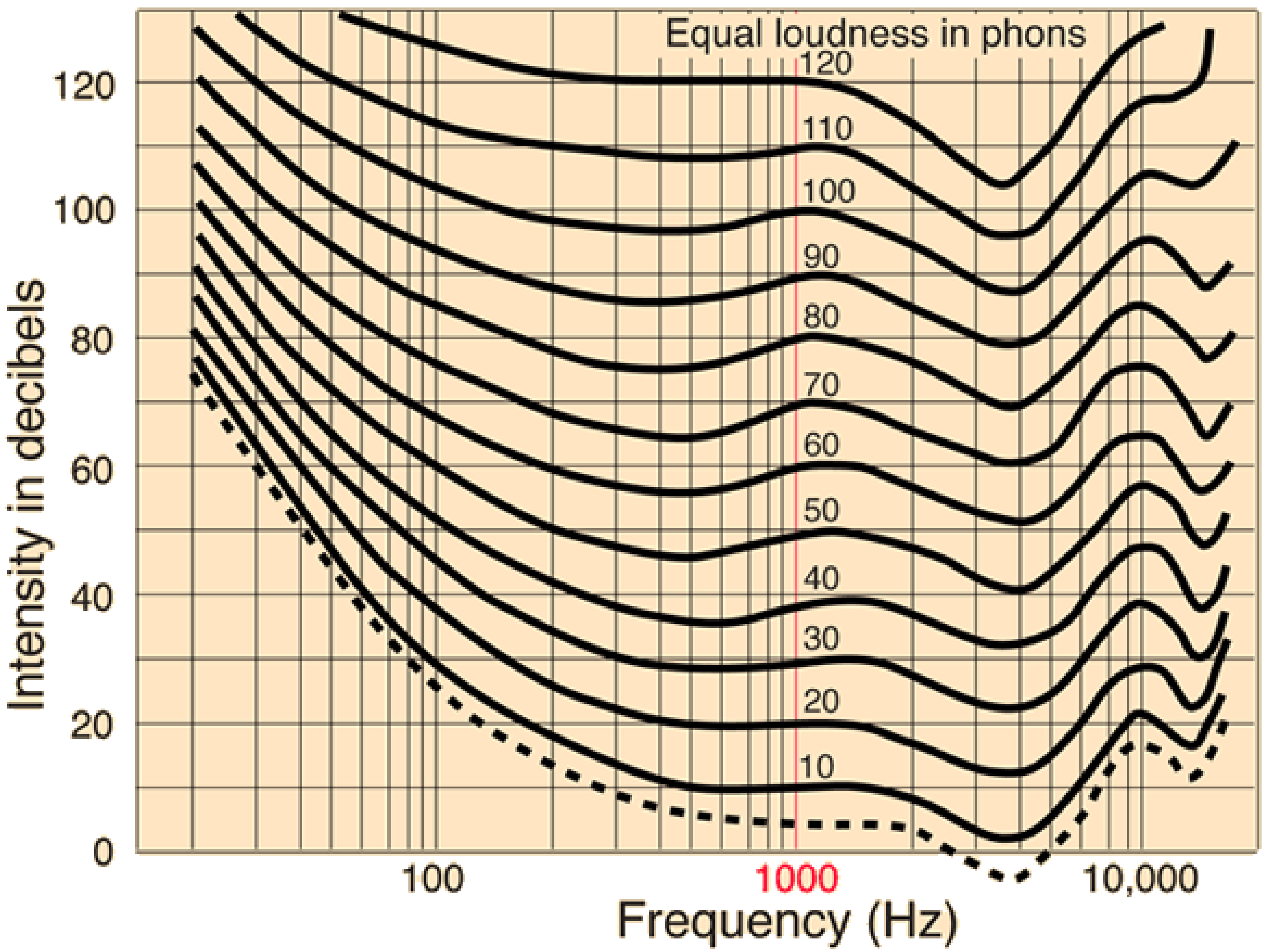
  In vision, you perceive *objects* (e.g., trees, cars, etc.). Likewise with smell you perceive e.g., cookies, roses, etc. Why not in the auditory perception of speech?

  - So, listeners perceive tongues and lips.

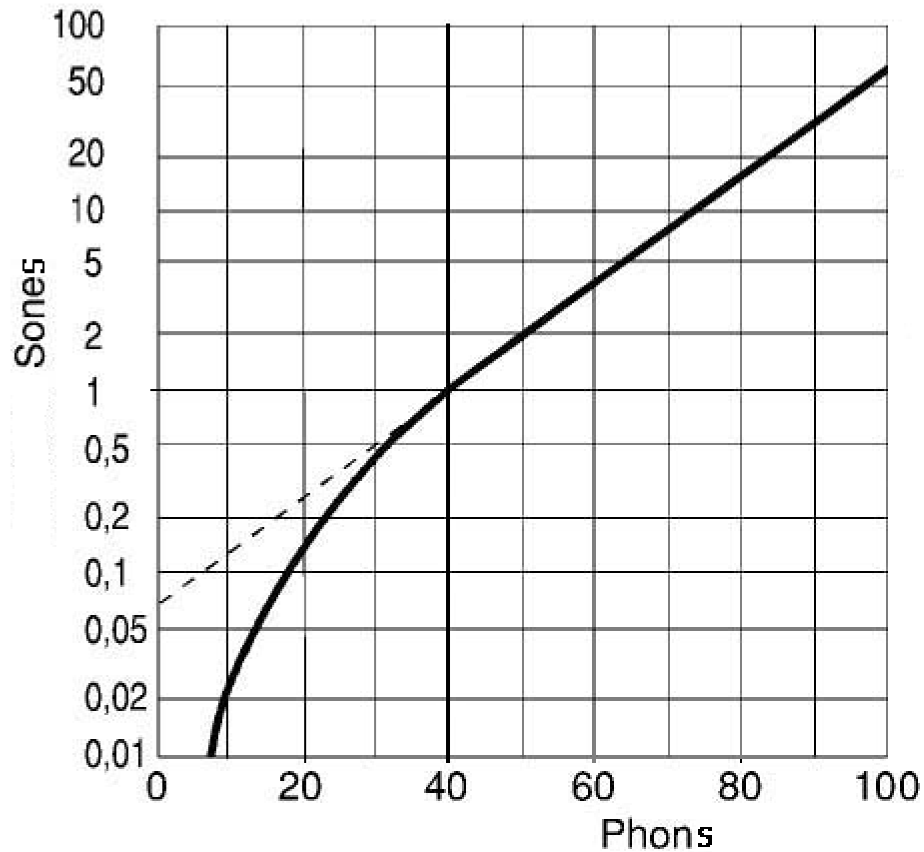
The articulatory gestures that are the objects of speech perception are not *intended* gestures (as in Motor Theory). Rather, they are the *actual* gestures.

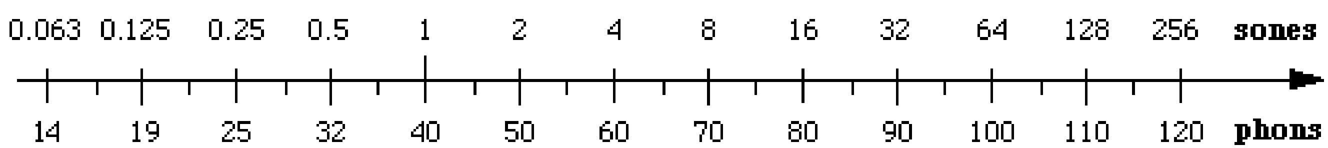
Psycho-acoustic Experiments

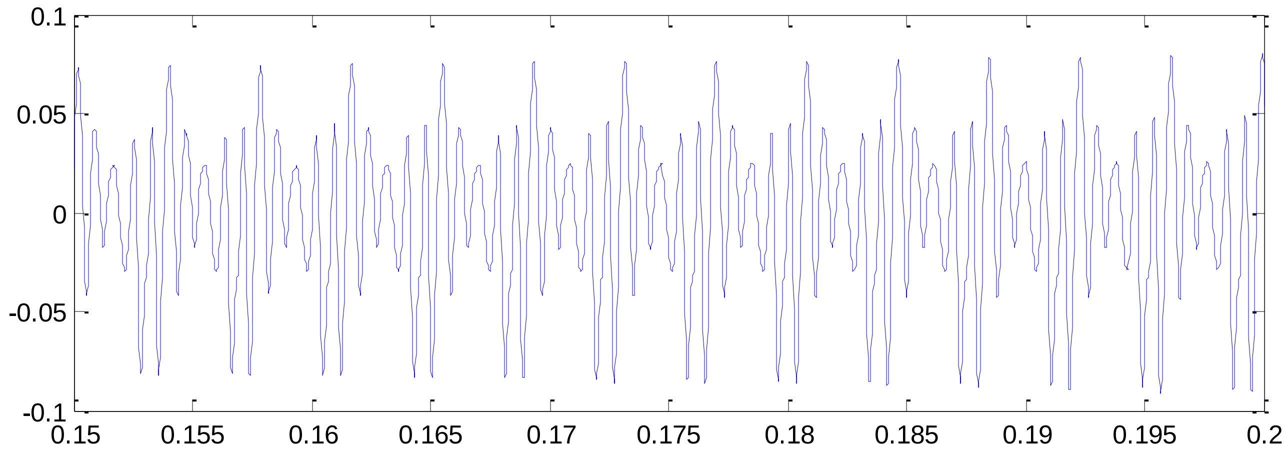
Fletcher Munson Curve – loudness as a function of frequency and intensity

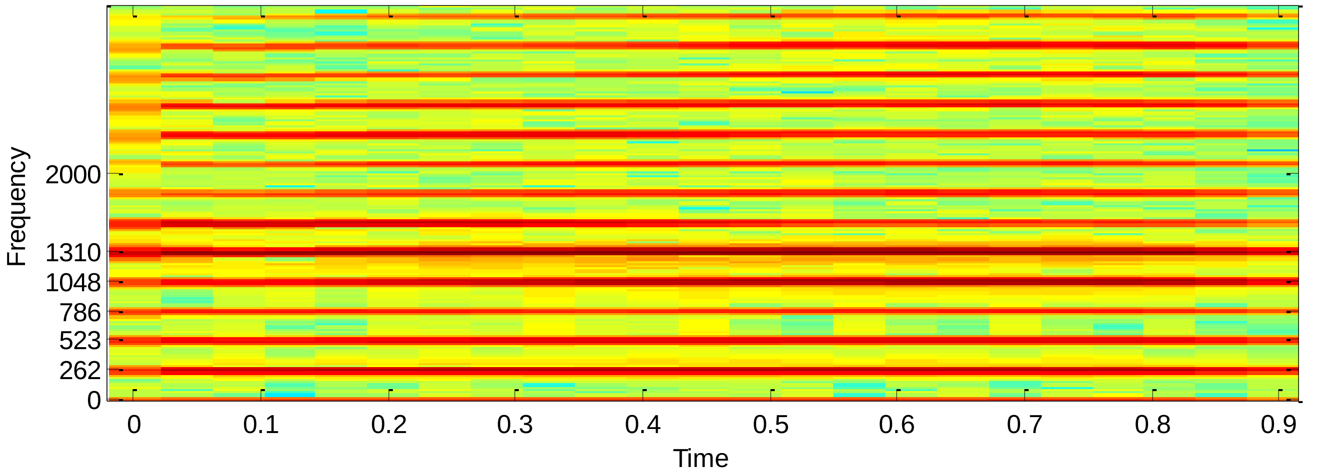


Sone vs Phon


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


0.1
0.05
0
-0.05
-0.1
0.15
0.155
0.16
0.165
0.17
0.175
0.18
0.185
0.19
0.195
0.2


Frequency
2000
1310
1048
786
523
262
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Time

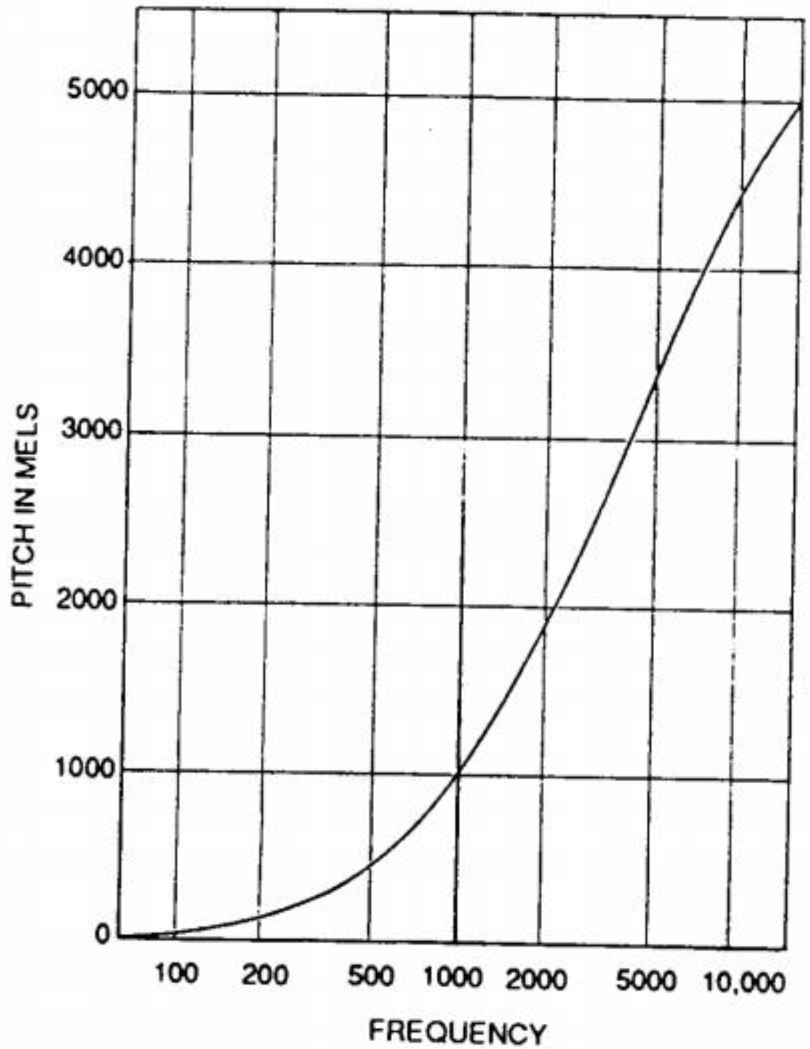
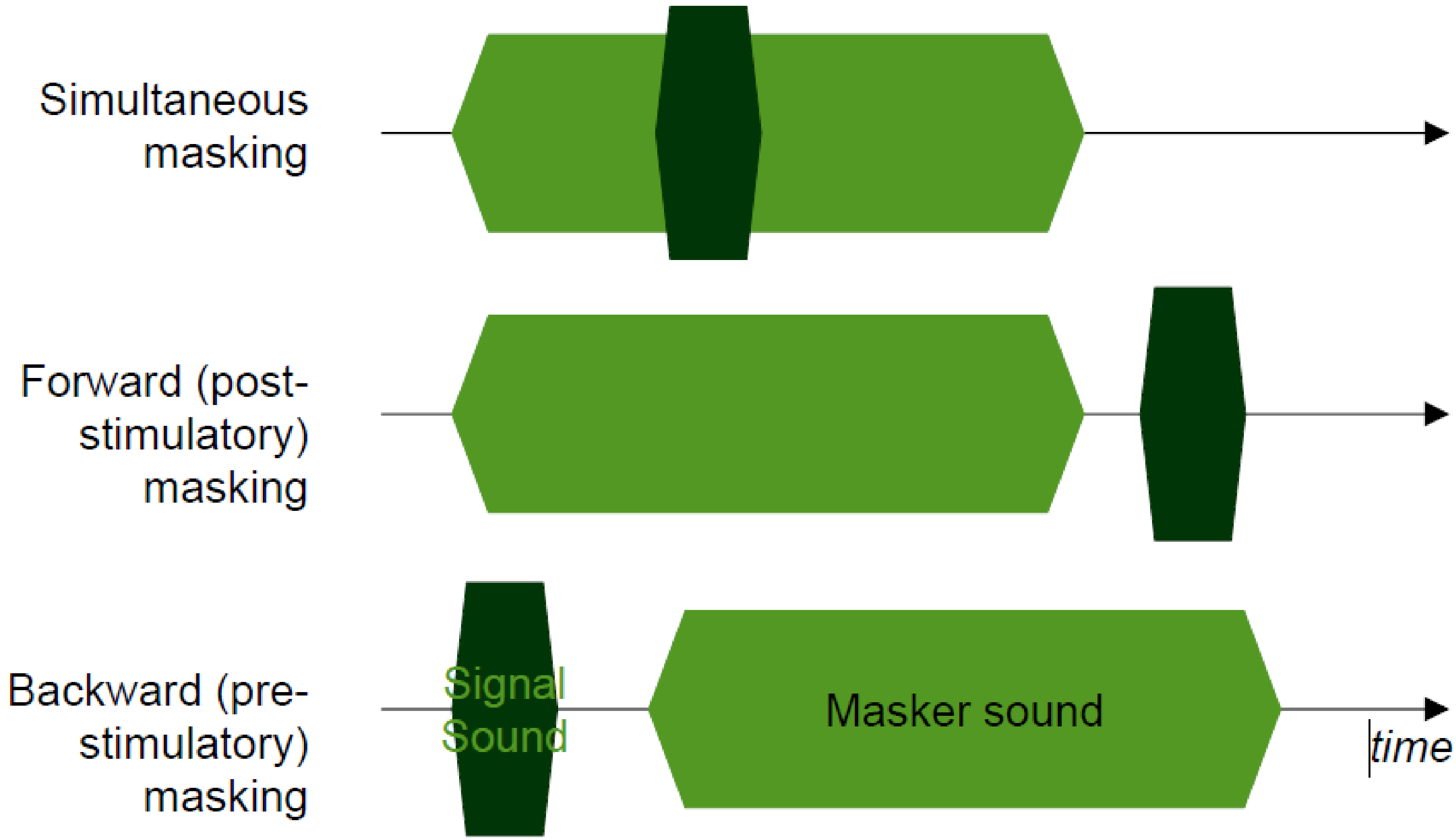
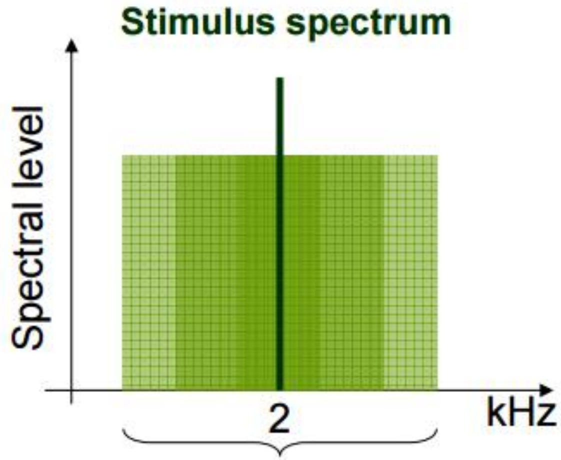


FIGURE 1 The relation of pitch (in mels) to the frequency of a pure tone. A 1000 Hz tone is arbitrarily assigned a value of 1000 mels. (From Stevens et al., 1937, reprinted with permission.)

Types

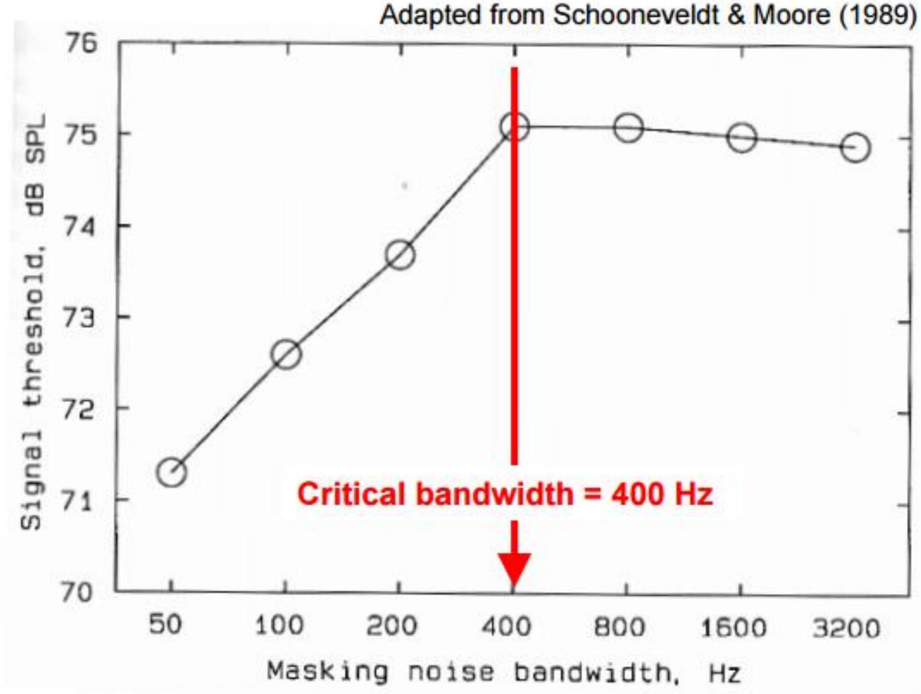

Simultaneous masking
Forward (post-stimulatory) masking
Backward (pre-stimulatory) masking
Signal Sound
Masker sound
time

The "critical band"



Signal detection threshold increases with increasing masking noise bandwidth up to a point after which signal threshold becomes independent of masker bandwidth.

Adapted from Schooneveldt & Moore (1989)



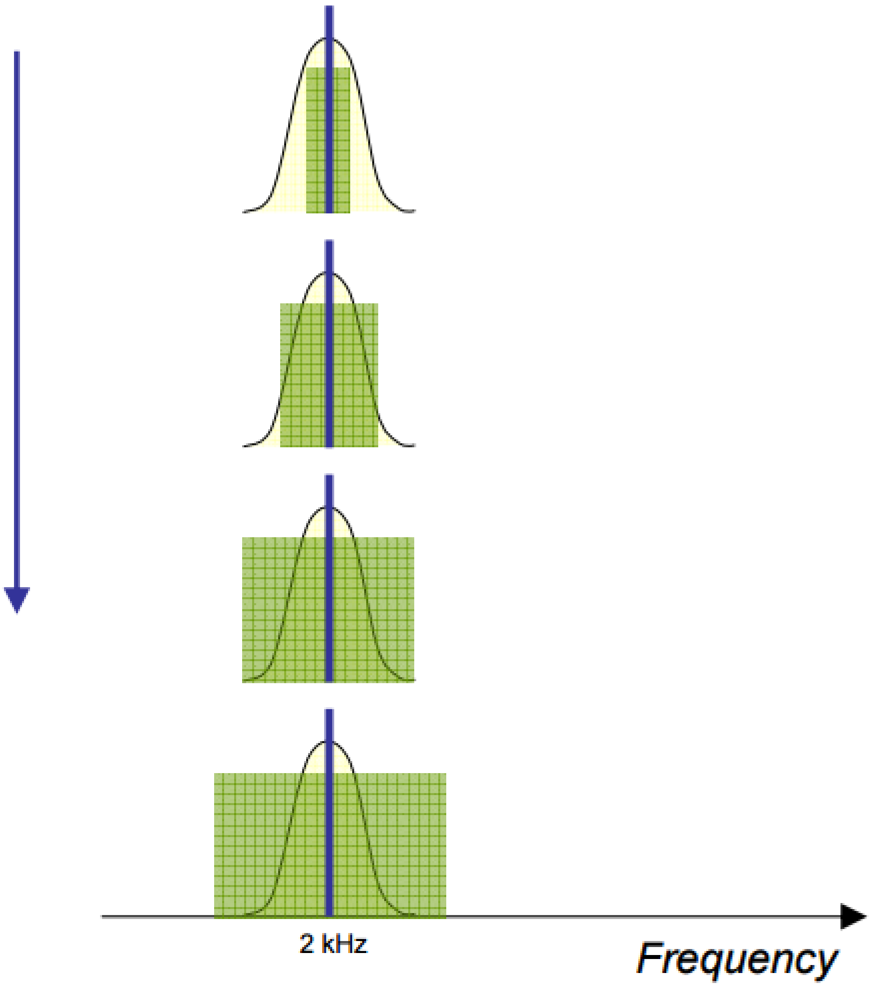
Schooneveldt GP, Moore BCJ. (1989). Comodulation masking release for various monaural and binaural combinations of the signal, on-frequency, and flanking bands. J Acoust Soc Am. 85(1):262-272.

An explanation of the critical band

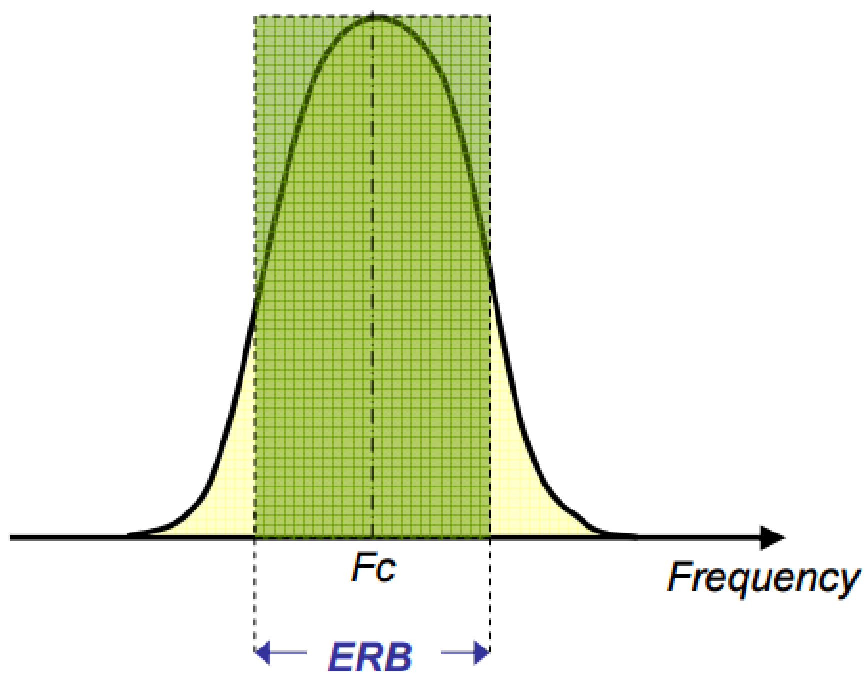
The amount of masking increases with increasing the noise (masker) energy that gets through the filter.

Up to a point…!

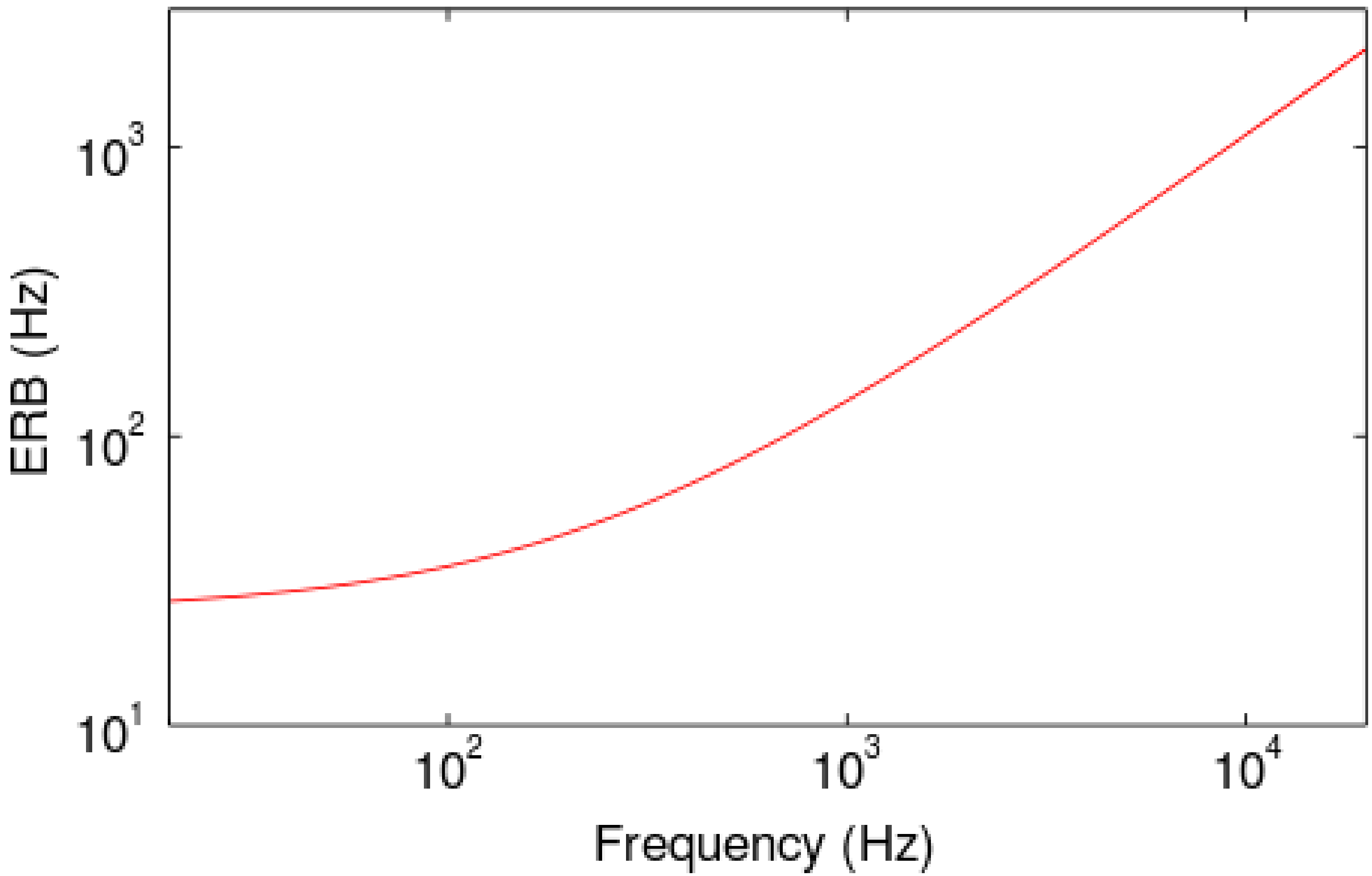
Further increases in noise bandwidth do not increase the masker energy through the filter.



Equivalent rectangular bandwidth (ERB)



An auditory filter (yellow area) and its ERB filter (green area). Both have different shapes but equal height and total area. That is, both let the same energy through.

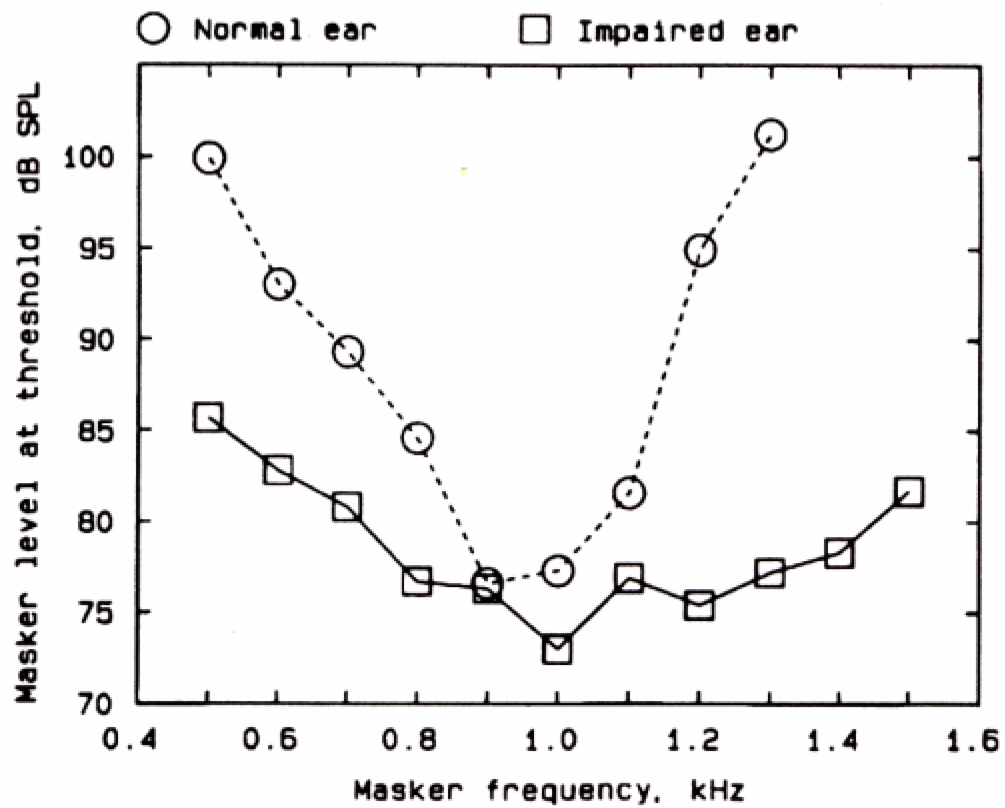


→ BARK scale

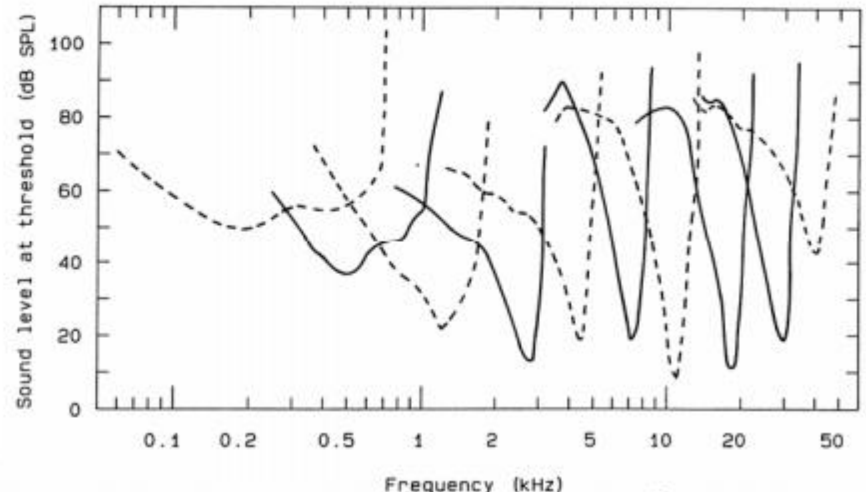
Psychoacoustical tuning curves

Method B produces **PSYCHOACOUSTICAL TUNING CURVES (PTCs)**.

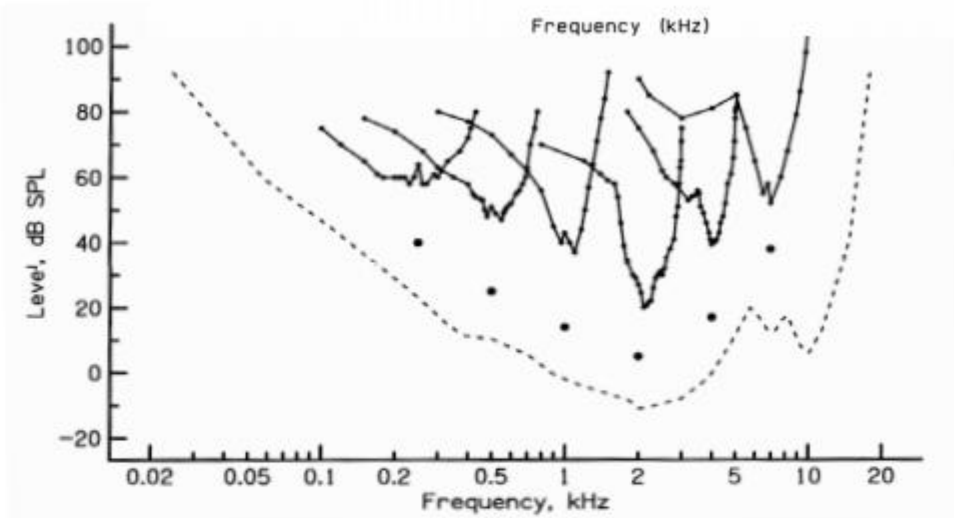
Psychoacoustical tuning curves for normal-hearing and hearing-impaired listeners. Signal was a 1-kHz pure tone at 10 dB SL. Masker was narrowband noise (Moore & Glasberg, 1986).



Moore BCJ, Glasberg, BR (1986). Comparisons of frequency selectivity in simultaneous and forward masking for subjects with unilateral cochlear impairments. J. Acoust. Soc. Am. 80, 93-107.



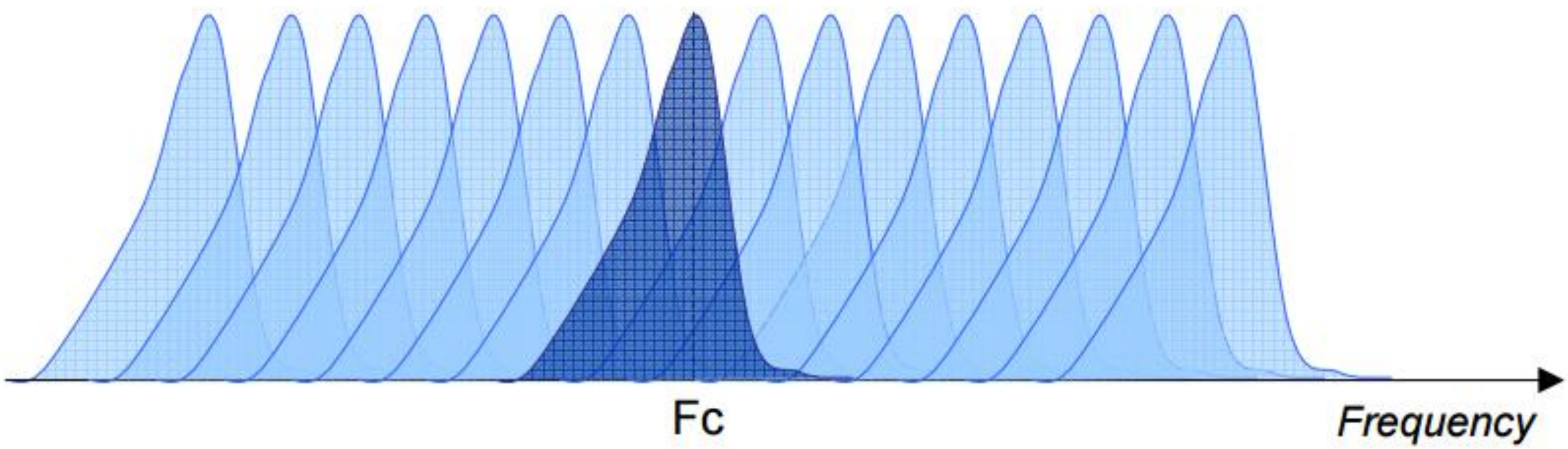
Auditory nerve fiber tuning curves (Palmer, 1987).



Psychoacoustical tuning curves (Vogten, 1974).

Palmer AR (1987). Physiology of the cochlear nerve and cochlear nucleus, in *Hearing*, edited by M.P. Haggard y E.F. Evans (Churchill Livingstone, Edinburgh).
Vogten, L.L.M. (1974). Pure-tone masking: A new result from a new method, in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer-Verlag, Berlin).

Auditory filters



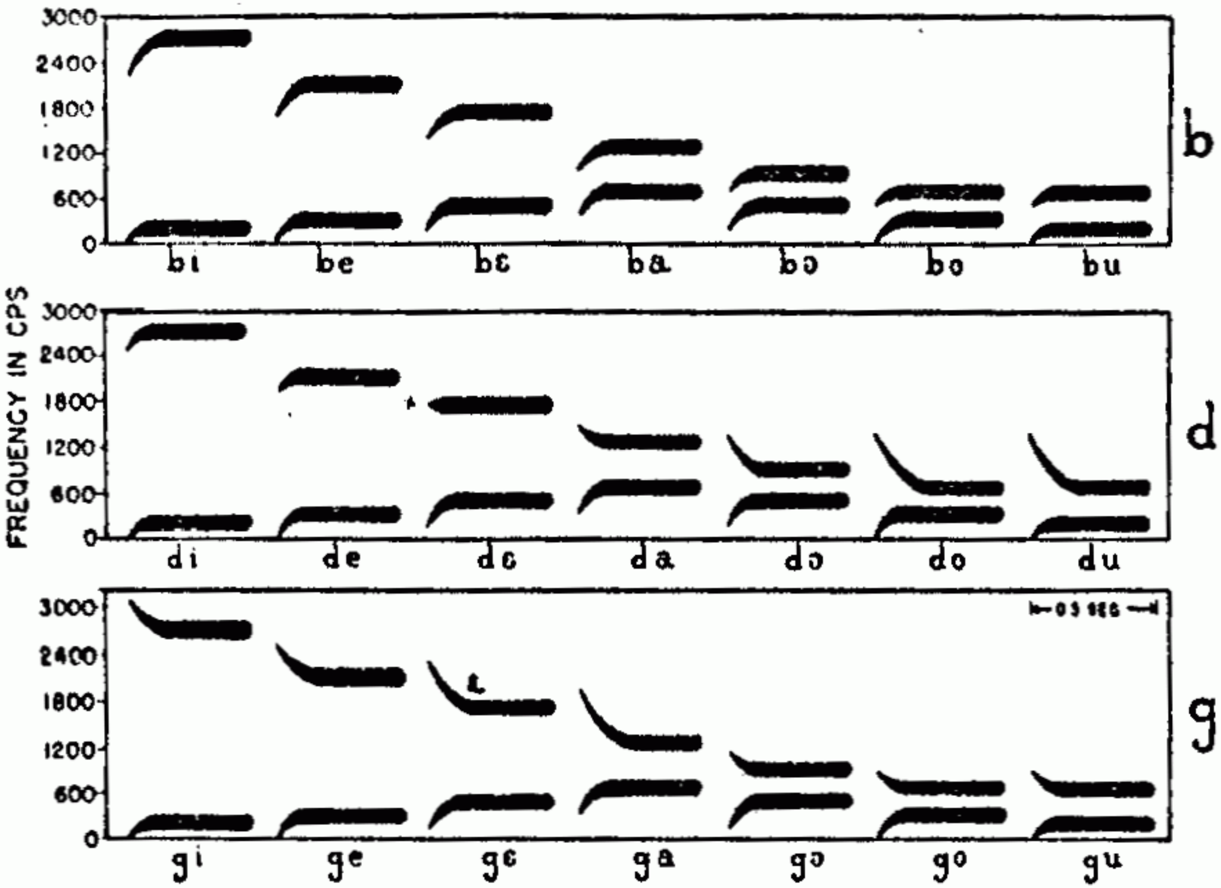
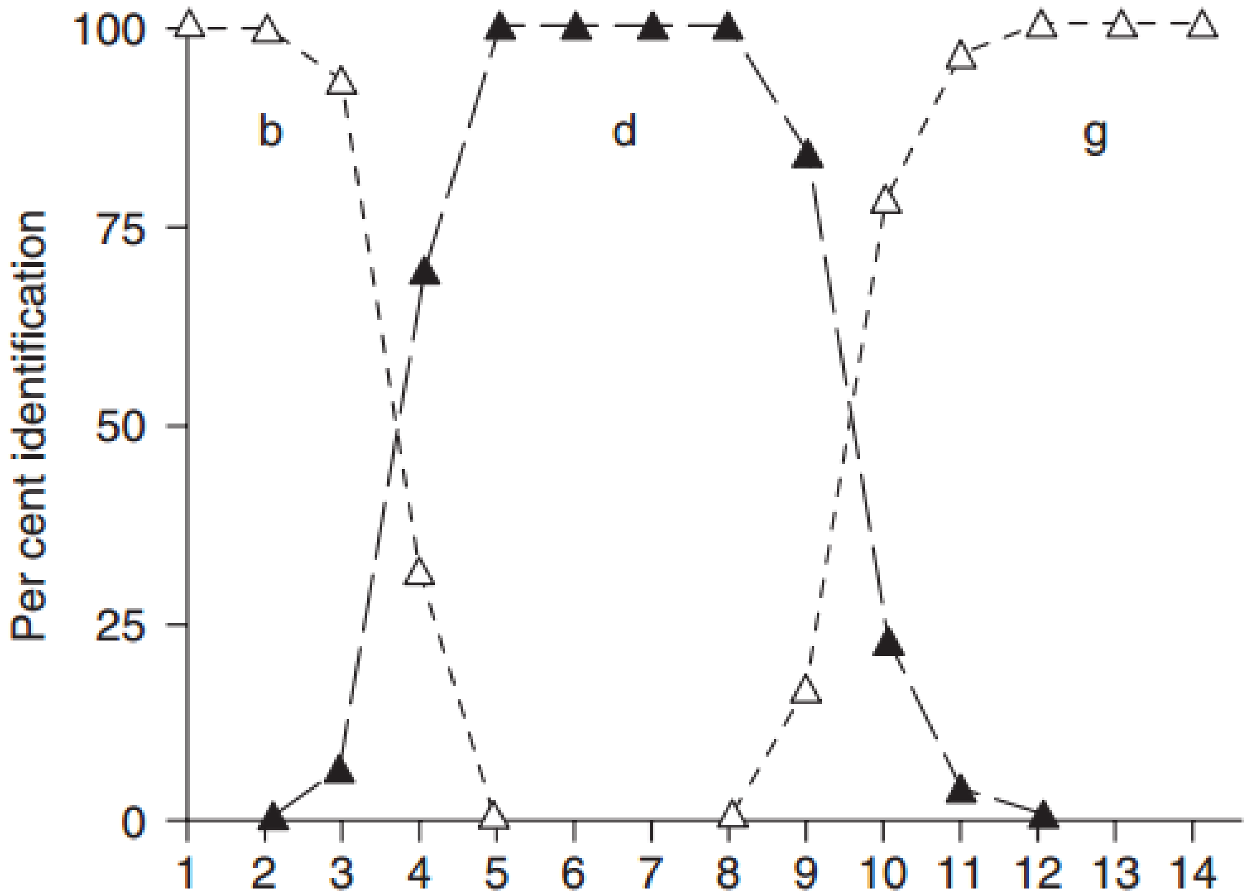


FIG. 1. Synthetic spectrograms showing second-formant transitions that produce the voiced stops before various vowels.



Per cent identification
100
75
50
25
0
b
d
g
1
2
3
4
5
6
7
8
9
10
11
12
13
14