Theories of Speech Perception

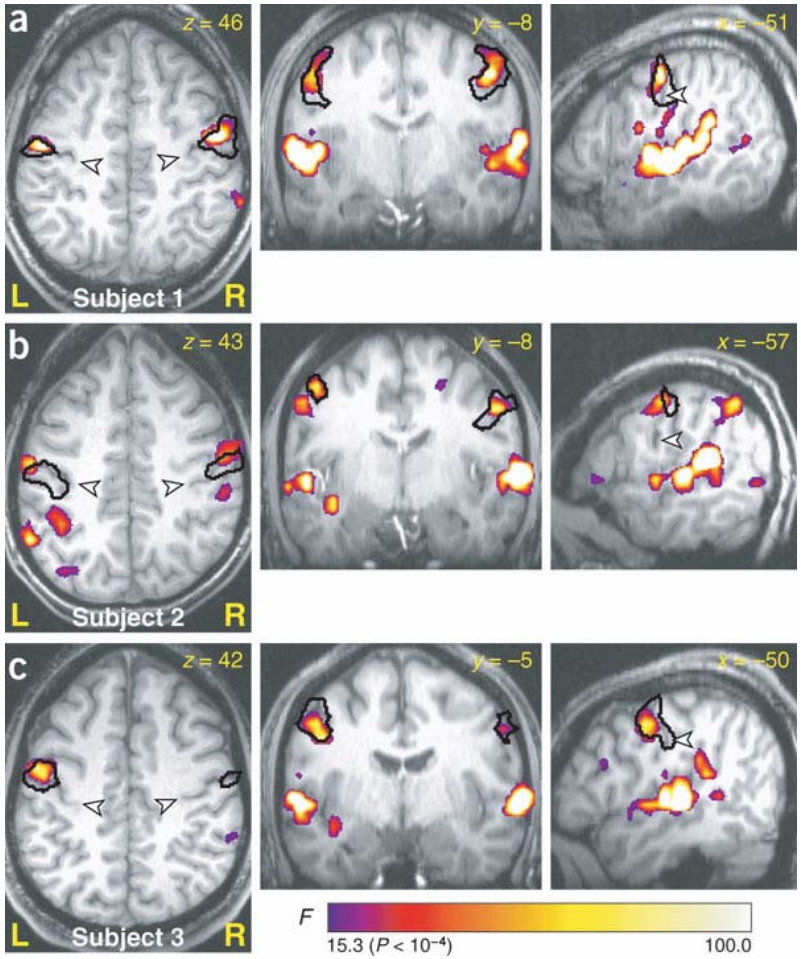
- Motor Theory (Liberman)
  - Close link between perception and production of speech
    - Use motor information to compensate for lack of invariants in speech signal
    - Determine which articulatory gesture was made, infer phoneme
  - Human speech perception is an innate, species-specific skill
    - Because only humans can produce speech, only humans can perceive it as a sequence of phonemes
    - Speech is special
- Auditory Theory
  - Derives from general properties of the auditory system
  - Speech perception is not species-specific

Wilson & friends, 2004

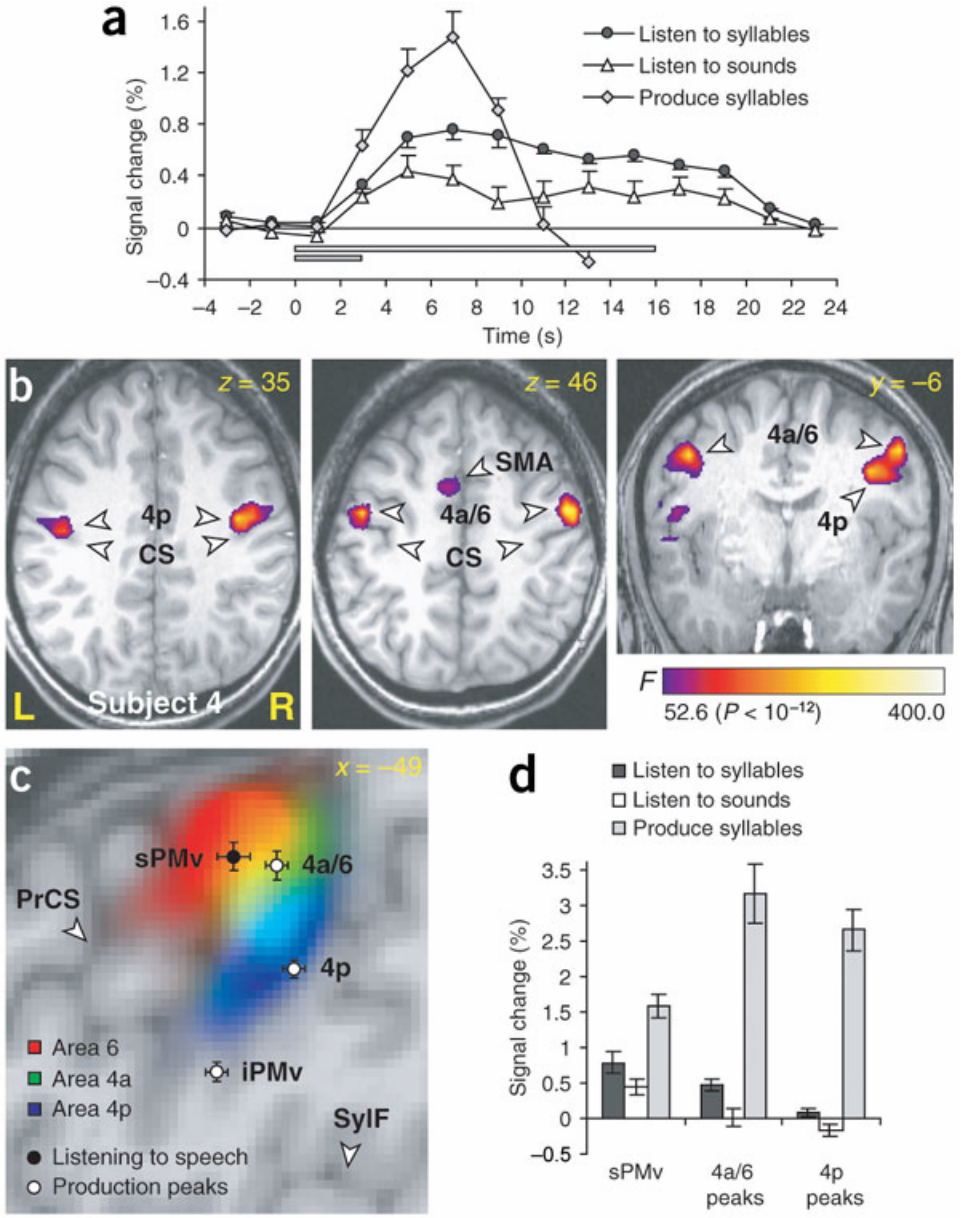
- Perception
- /pa/
- /gi/
- Bell
- Burst of white noise

- Production
- /pa/
- /gi/
- Tap alternate thumbs

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 (!)

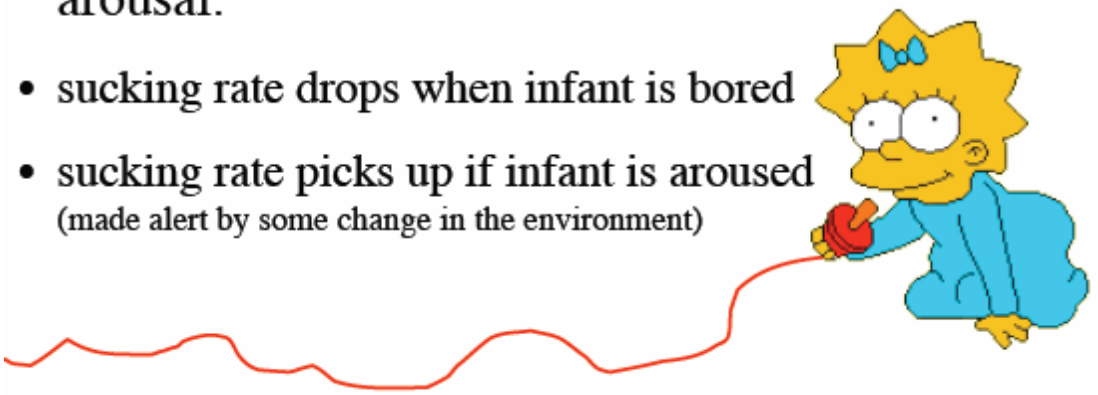


Wilson and colleagues, 2004

Is categorical perception innate?

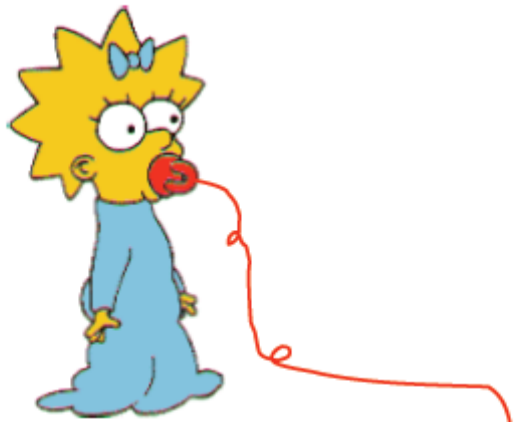
Infant's sucking rate correlates with arousal:

- sucking rate drops when infant is bored
- sucking rate picks up if infant is aroused
  (made alert by some change in the environment)



Manipulate VOT, Monitor Sucking

… ba ba ba ba ba ba ba ba ba ba ba ba ba
ba ba ba ba ba ba ba ba ba ba ba **pa**



4-month-old infants: Eimas et al. (1971)

20 ms
(Different Sides)

20 ms
(Same Side)

0 ms
(Control)

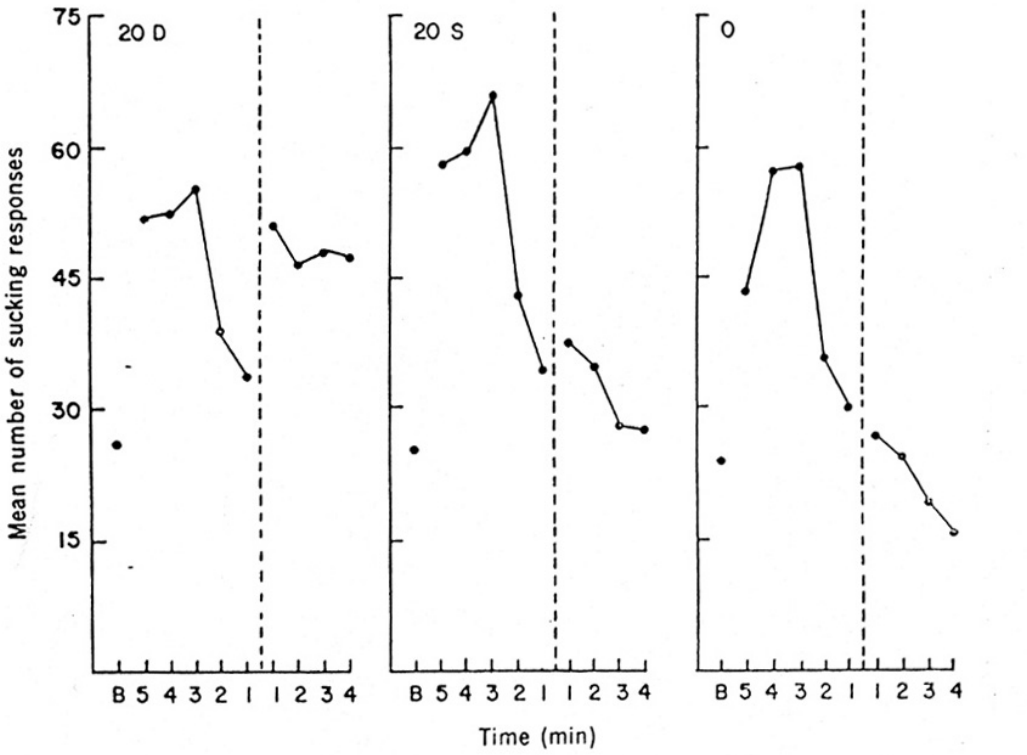


Fig. 2. Mean number of sucking responses for the 4-month-old infants, as a function of time and experimental condition. The dashed line indicates the occurrence of the stimulus shift, or in the case of the control group the time at which the shift would have occurred. The letter *B* stands for the baseline rate. Time is measured with reference to the moment of stimulus shift and indicates the 5 minutes prior to and the 4 minutes after shift.

Might infant performance in the Eimas et al. study be attributable to early exposure to English VOT patterns?

In a follow-up study, Lasky, Syrdal-Lasky and Klein (1975) used a heart rate deceleration procedure to study VOT discrimination among infants being raised in a Spanish-speaking environment.

Three discrimination tests involving VOT differences of 40 ms:

1. **-60 ms VOT/ -20 ms VOT (straddles the Spanish /b/-/p/ boundary)**

2. **-20 ms VOT/ +20 ms VOT (doesn't straddle any boundary)**

3. **+20 ms VOT/ +60 ms VOT (straddles the English /b/-/p/ boundary).**

4-6.5 month-old Guatemalan infants discriminated 1 and 3 but not 2.

What explains the infant categorical perception results?

<u>Two possibilities</u>:

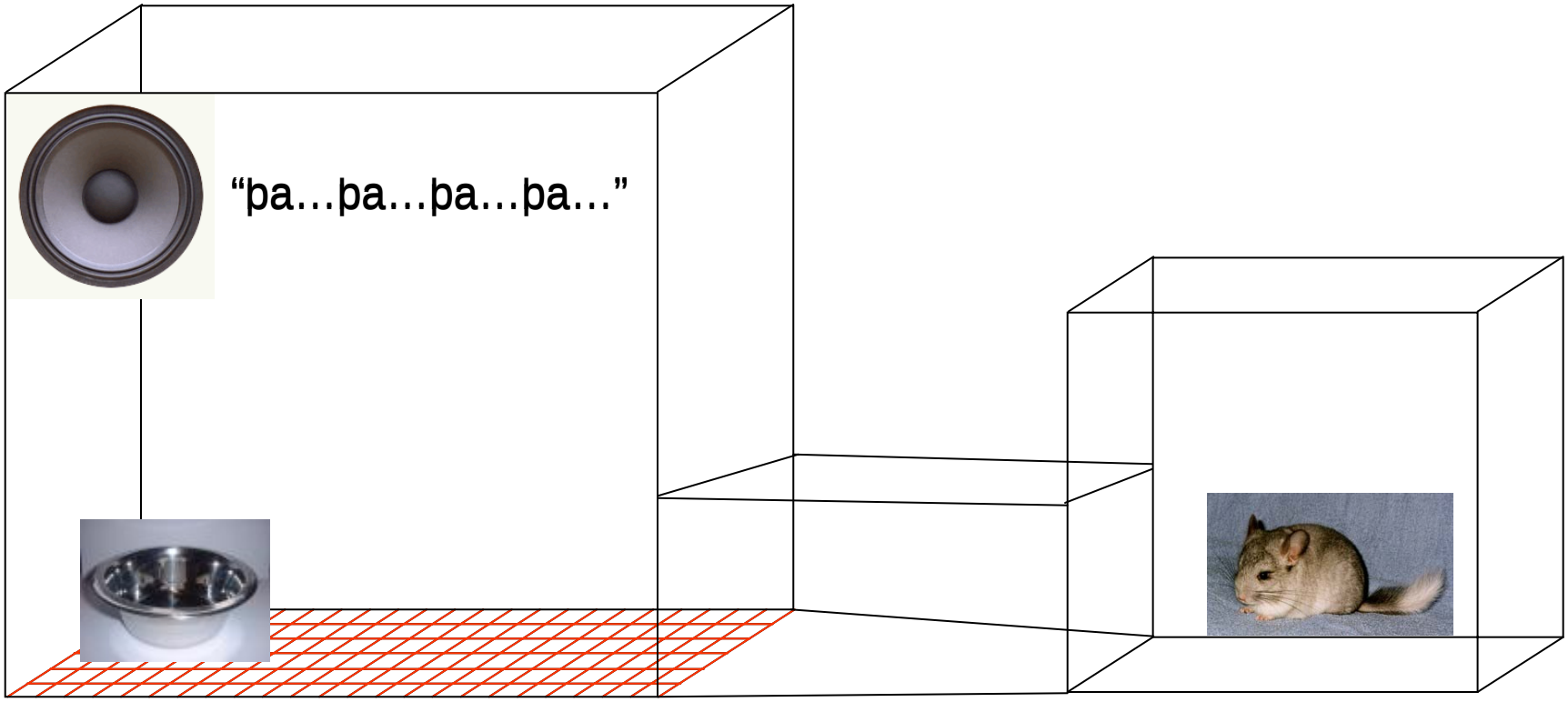
1. Humans have evolved an innate, language-related perceptual mechanism that facilitates the discrimination of speech categories such as [+voice] and [-voice] consonants.

2. Enhanced discrimination observed for human infants at two places along the VOT dimension derives from auditory factors not specific to speech or to humans.

Is categorical perception species specific?

- Chinchillas exhibit categorical perception as well

Chinchilla experiment
(Kuhl & Miller experiment)



- Train on end-point “ba” (good), “pa” (bad)
- Test on intermediate stimuli
- Results:
  - Chinchillas switched over from staying to running at about the same location as the English b/p phoneme boundary

VOT “identification” by chinchillas (Kuhl & Miller, 1981)

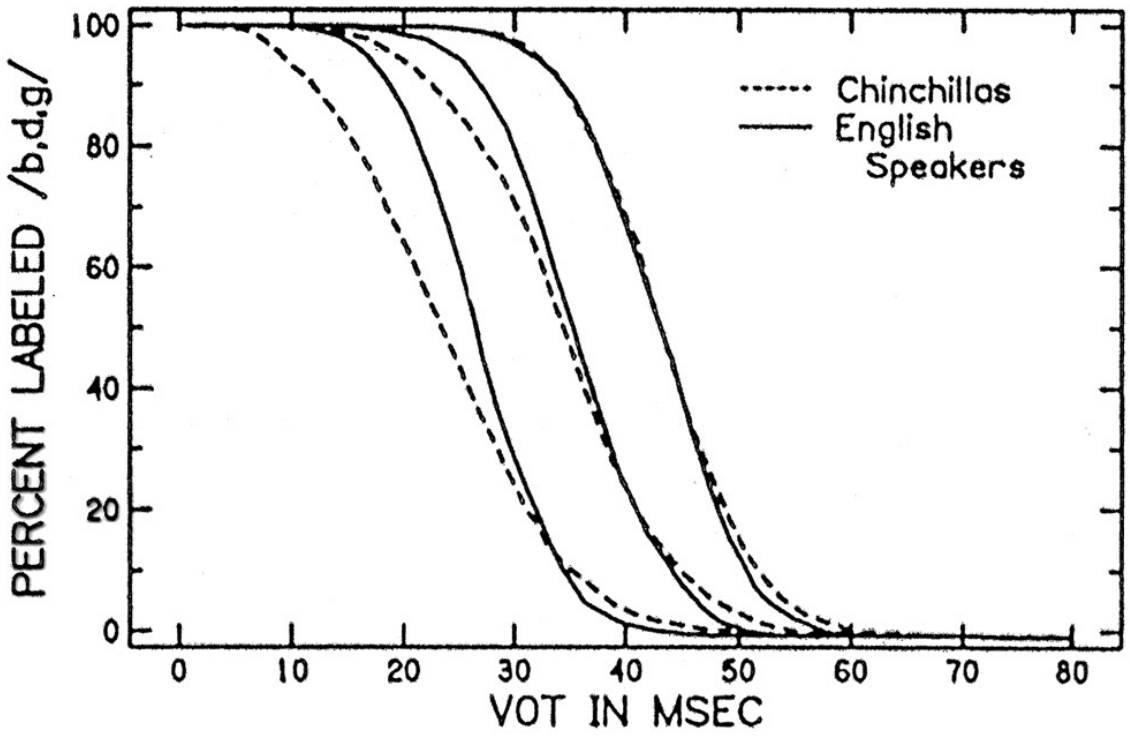


Figure 1

Categorical perception, Take 2

- Natural discontinuities in many sensory systems; many of these are common across mammalian species
- Some stimulus differences are hard; others are easy
- Language takes advantage of “natural boundaries”

Categorical Perception & Auditory Theory

- Categorical perception may arise from rapid decay of auditory memory
  - not unique to speech
- People have some ability to discriminate sounds within a phoneme
  - judgments may reflect decision process rather than perception

Motor Theory *versus* Auditory Theory

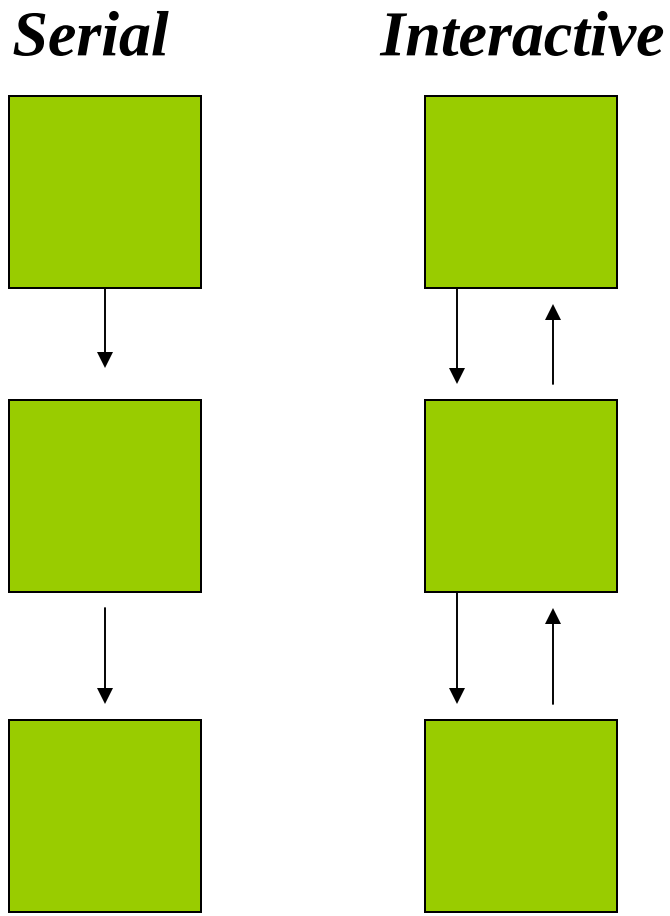
- Close link between speech perception and speech production systems
  - Motor Right!
- Some properties of speech perception (e.g. categorical perception) general auditory properties
  - Auditory Right!
- Speech perception probably not innate species-specific
  - Motor Wrong…

Comprehension

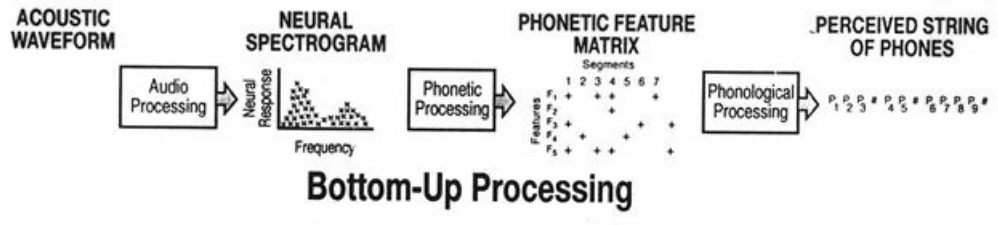
- **Recognize Word**
  - **Phonological Info**
  - Visual Info
- Retrieve Information
  - Syntactic Info
  - Semantic/Pragmatic Info
- Integrate Syntactic & Semantic/Pragmatic Info
- Store Gist Representation

Word Recognition

- Serial
  - Comprehension involves analysis at several different levels in turn
- Interactive
  - Various sources interact and combine to produce efficient analysis

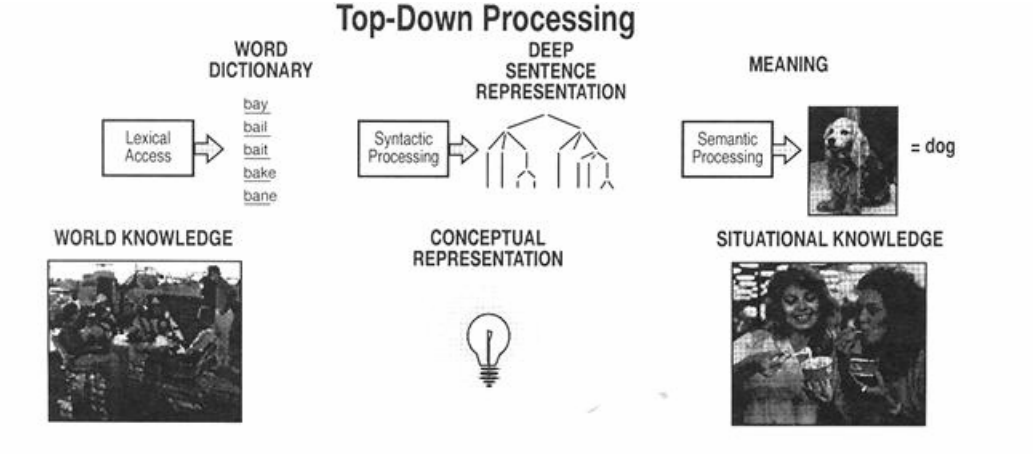


Bottom-up Processes



- Acoustic Info
- Phonetic Info
- Phonemic Info
- Words & Sentences

Top-Down Processes



- To what extent does knowledge of what speaker is saying impact processes necessary for understanding speech?

Phonemic Restoration Effect

- Legislature

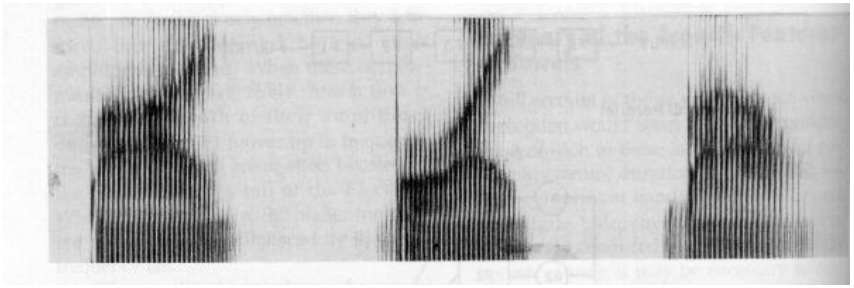

- Sentences



McGurk Effect

McGurk Effect

Lips say “ba”



Sound signal “ga”

- /ba/ bilabial
- /ga/ velar
- /da/ dental

Subjects hear “da”

What’s the relevance?

- What does this stuff have to do with interactive vs. serial models?
- Context Effects
  - Interactive Models use all sources of information for rapid word ID
  - Serial Models inefficient & slow

Marslen-Wilson's Cohort Model

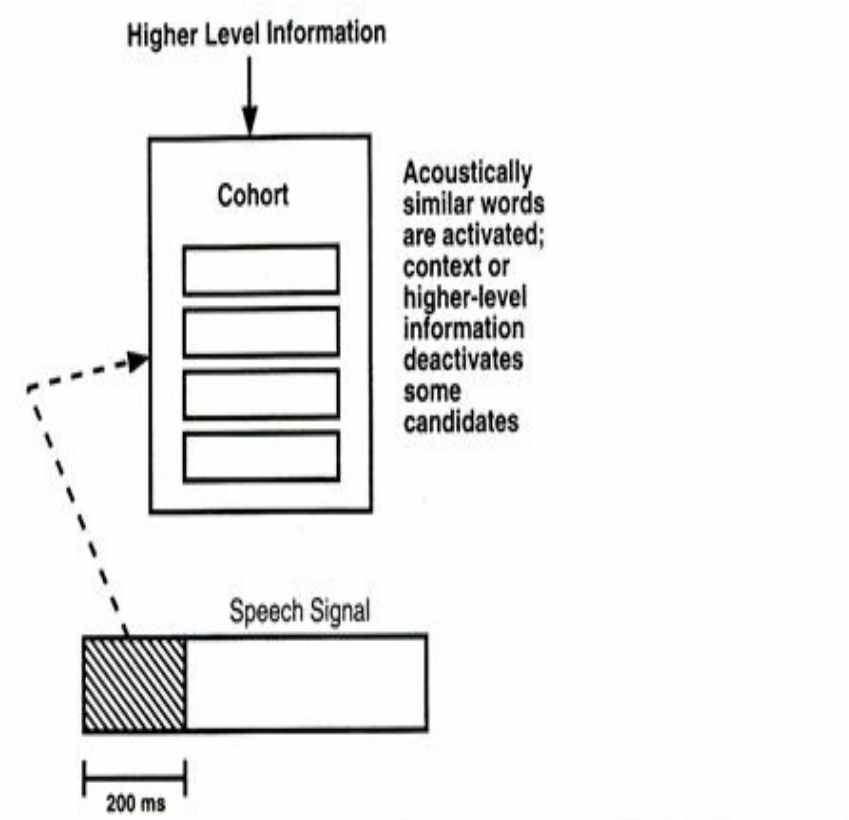


FIGURE 6-11 **The Cohort Model of Word Recognition.** *(Reprinted with permission by Singular Publishing Group, Inc. Kent, R. [1997].* The Speech Sciences. *San Diego: Singular, 388.)*

- 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)

Uniqueness and Recognition

- When we hear the beginning of a word this activates ALL words beginning with the same sound: the “word initial cohort”. Subsequent sounds eliminate candidates from the cohort until only one remains (failure to fit with context can also eliminate candidates)
- t - tea, tree, trick, tread, tressle, trespass, top, tick, etc.
- tr - tree, trick, tread, tressle, trespass, etc.
- tre - tread, tressle, trespass, etc.
- tres - tressle, trespass, etc.
- tresp - trespass.

Uniqueness and Recognition

- The uniqueness point is the point at which a word becomes uniquely identifiable from its initial sound sequence

E.g. “dial” dayl| “crocodile” krokod| ayl
UP UP

- For non-words there is a deviation point: a point at which the cohort is reduced to zero

E.g. “zn | owble” would be rejected with a faster RT than “thousaj | ining”
DP DP

Uniqueness and Recognition

- The recognition point is the point at which, empirically, a word is actually identified
- Empirical studies show that recognition point correlates with (and is closely tied to) the uniqueness point.
  - phoneme monitoring latencies correlate with *a priori* cohort analysis (and one way to recognise word initial phonemes is to recognise the word and to know it begins with e.g. /p/)

Cohort Model (Marslen-Wilson & Tyler)

- Words consistent with input become active
  - Cohort – set of words consistent with first syllable
- Words in the cohort eliminated when they become inconsistent with input
- Words eliminated due to contextual incongruity
- Processing ends when there is one word left in the cohort

/ka/

cat captain catch capitalism

/kap/

captain capitalism

Communism is slightly different from /kap/

capitalism

Marslen-Wilson & Tyler

- Normal

The church was broken into last night.

Some thieves stole most of the lead off the roof.

- Syntactic

The power was located in green water.

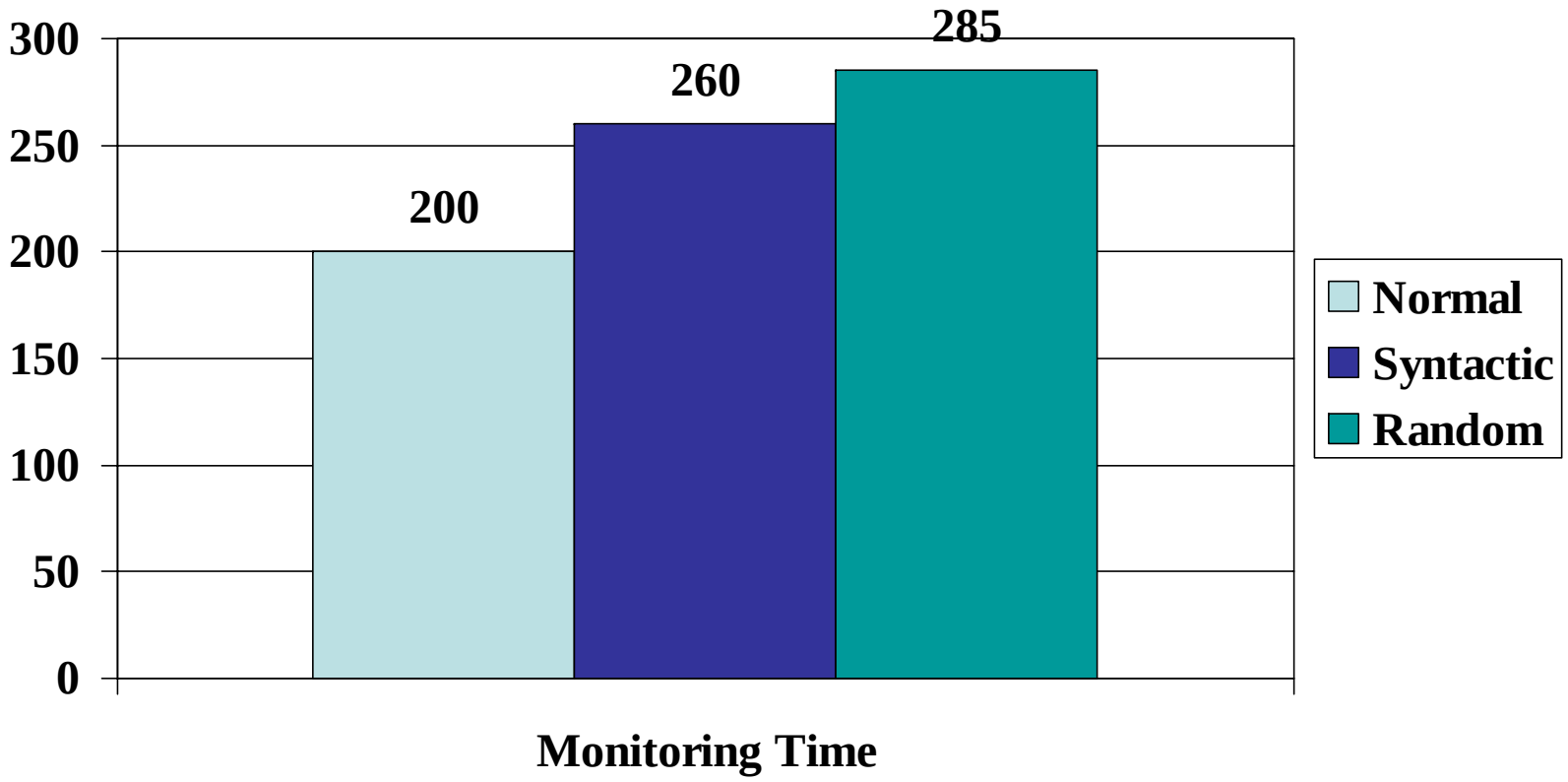
No buns puzzle some in the lead off the text.

- Random

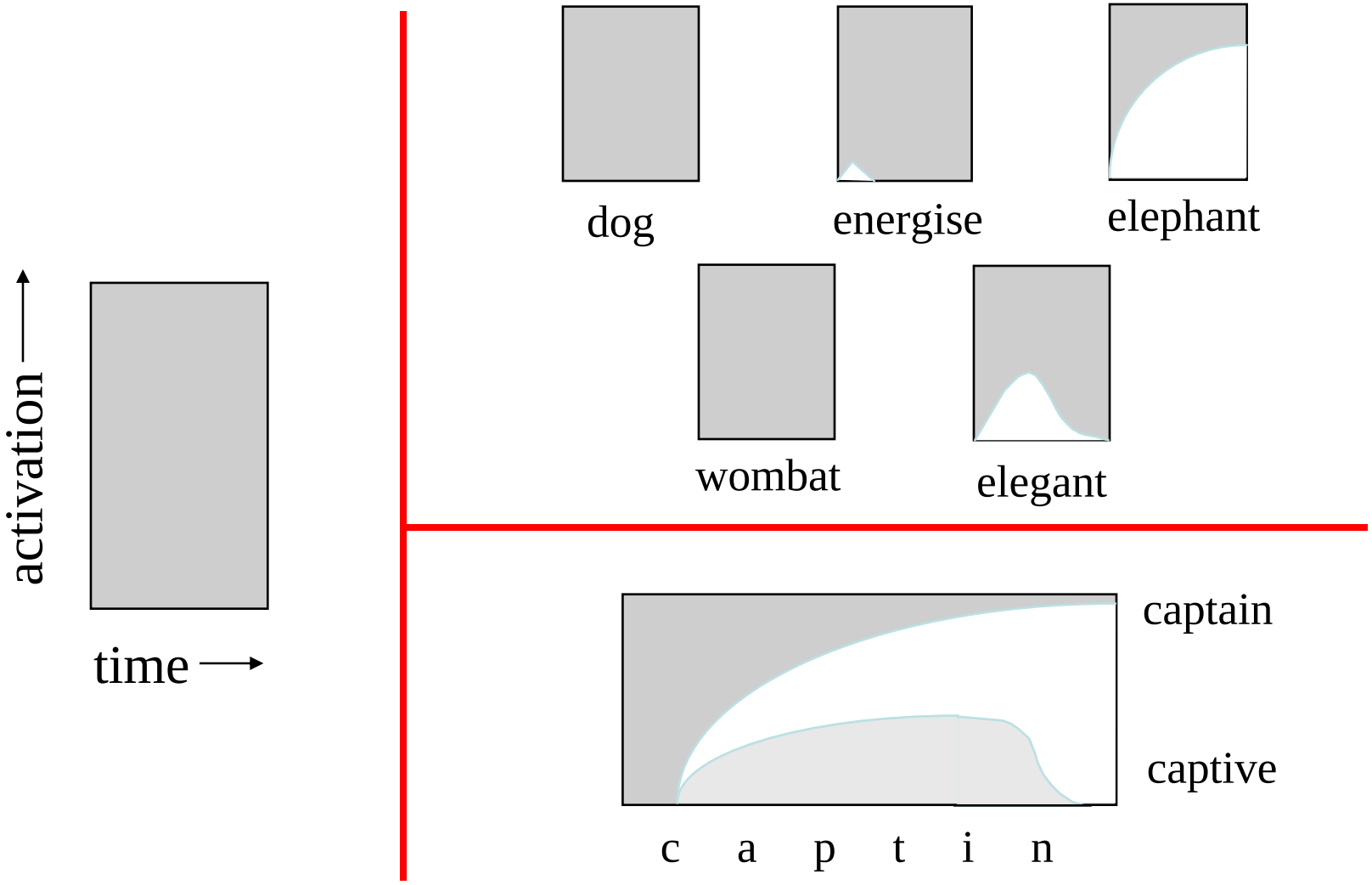
In was great power water the located.

Some the no puzzle buns in lead text the off.

Marslen-Wilson & Tyler

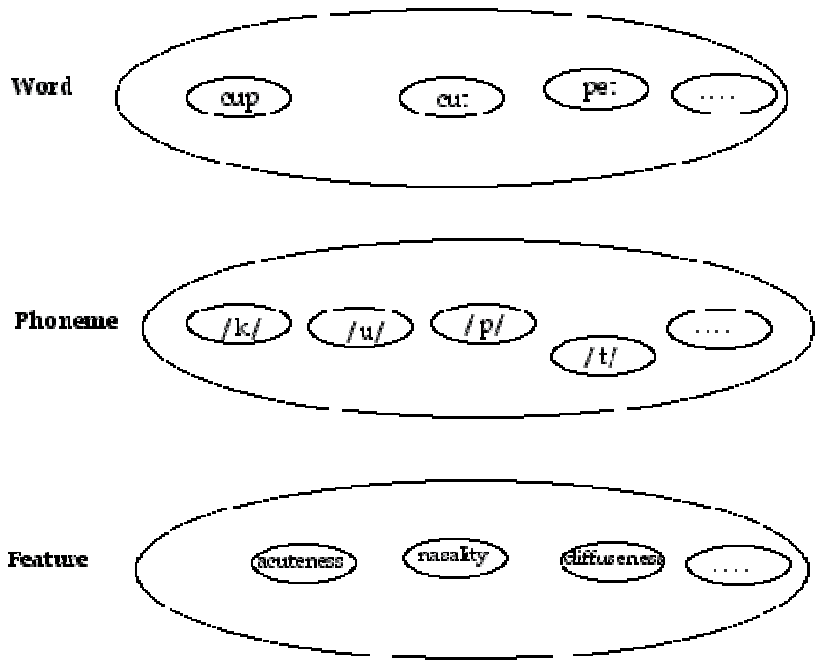


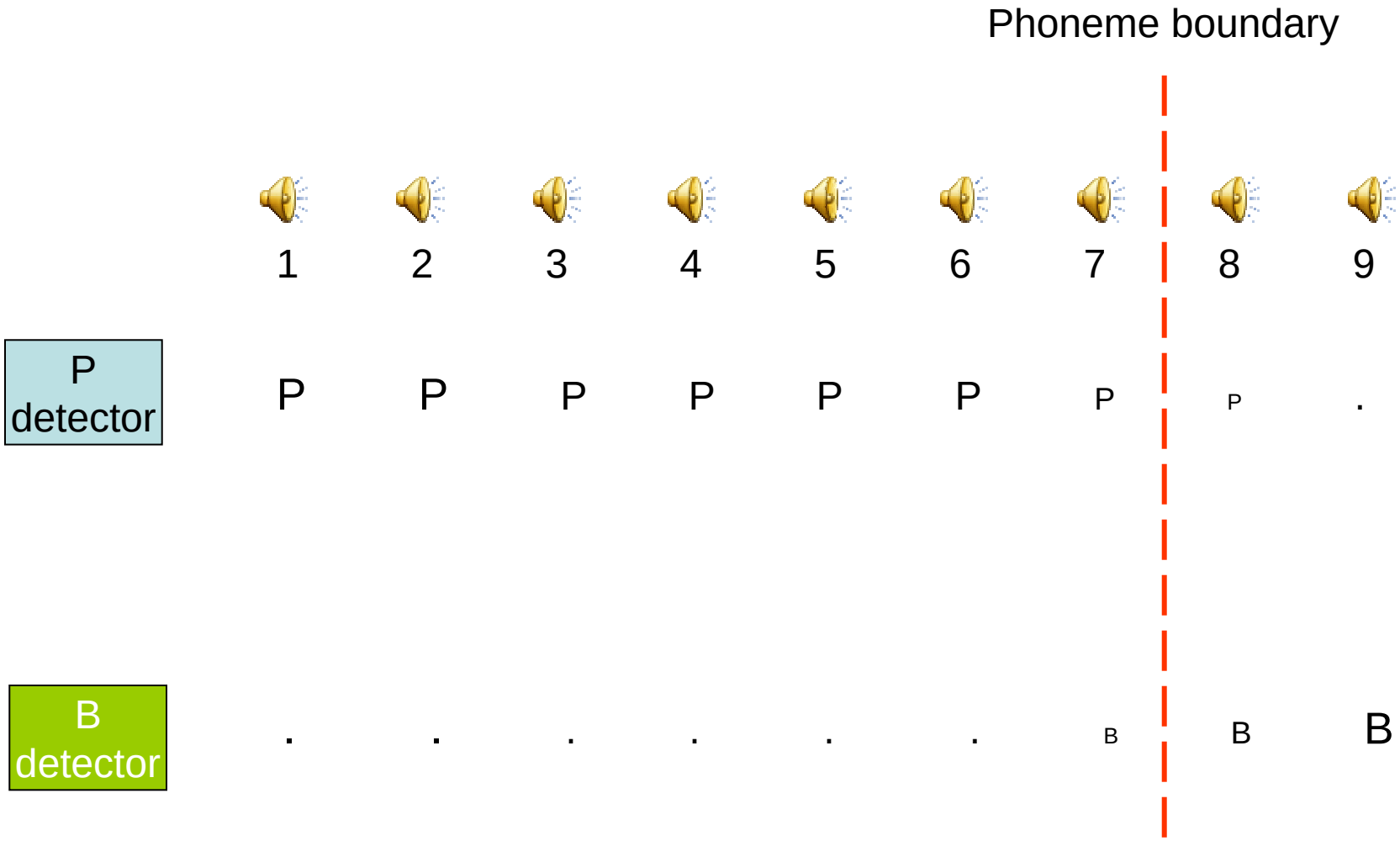
Activation in the Revised Cohort Model



TRACE

- 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.




Phoneme boundary
1
2
3
4
5
6
7
8
9
P detector
P
P
P
P
P
P
P
P
.
B detector
.
.
.
.
.
.
B
B
B

If there are feature detectors, can we tire one of them out?

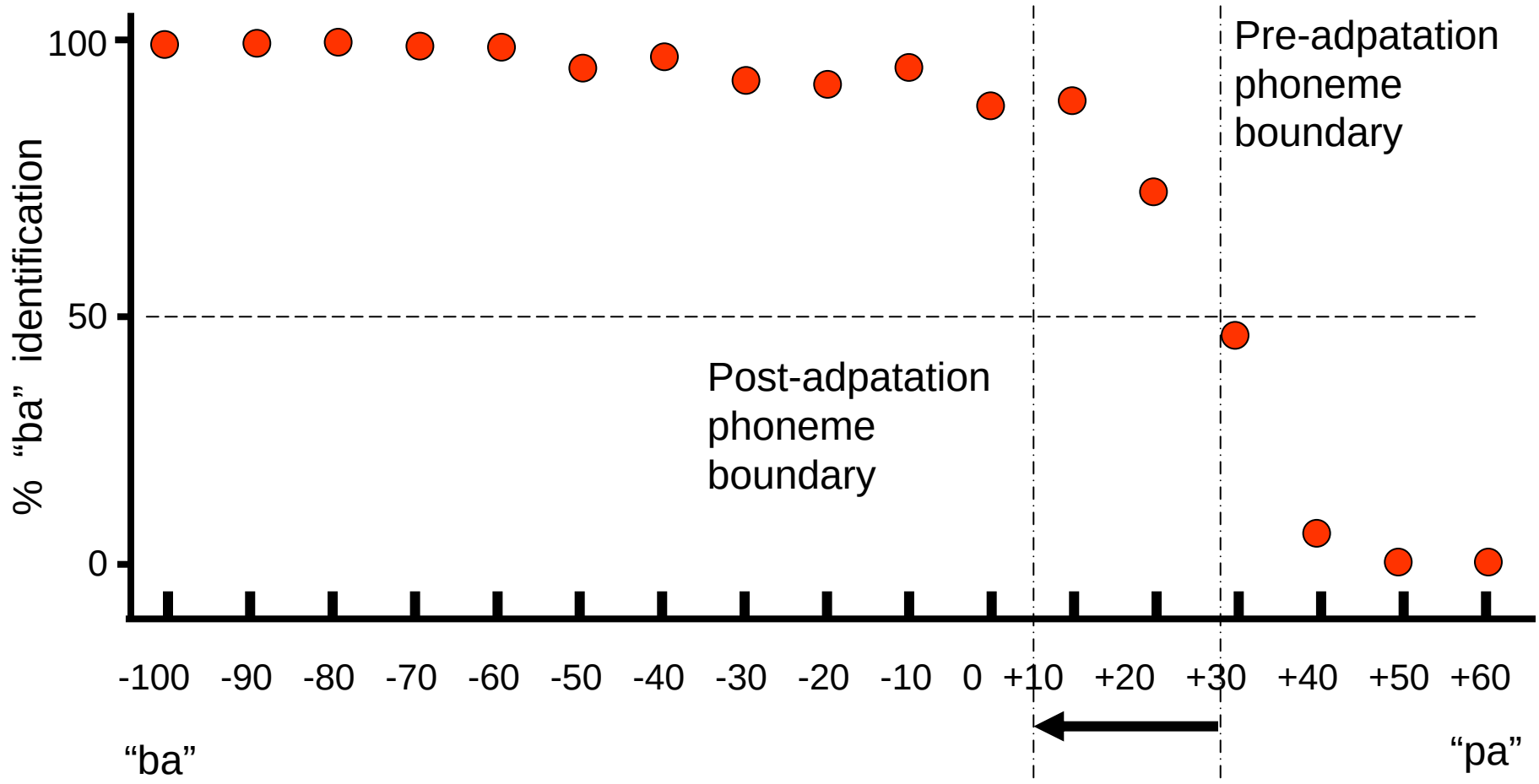
Selective adaptation

1. Do phoneme identification test
   (e.g., “ba-pa” continuum)

2. Play a stimulus from one of the end-points many times (e.g., 100 times)

3. Repeat phoneme identification test

Selective adaptation



REPEAT -100 "ba"
100 times for one minute

Voice Onset Time continuum