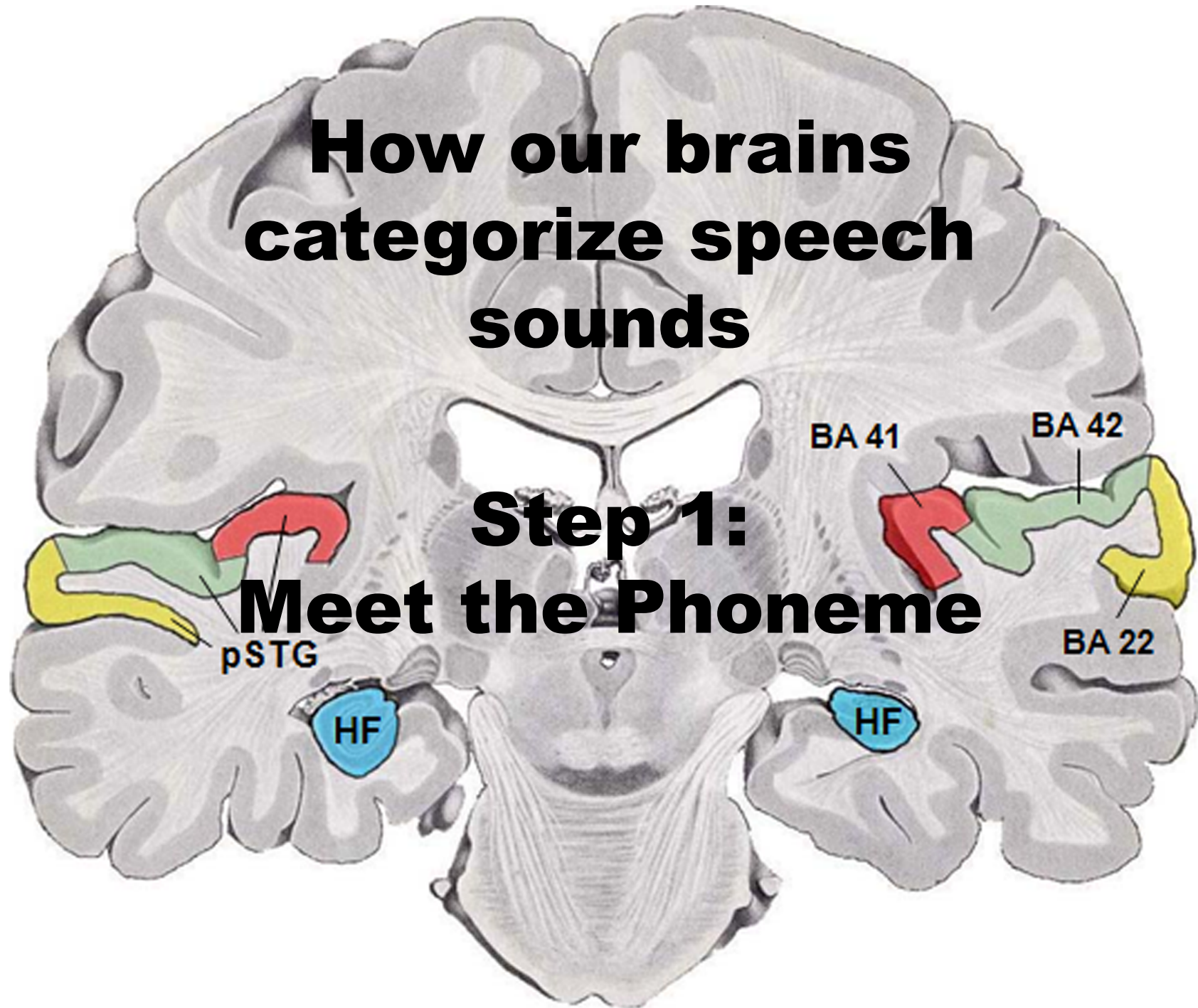
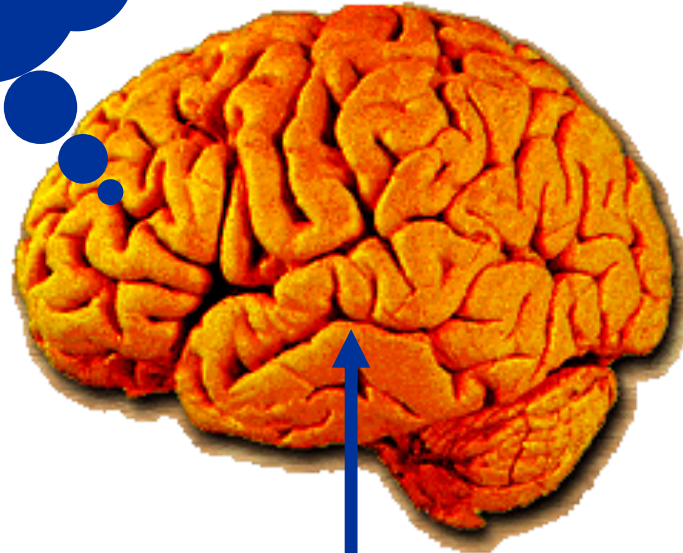


How our brains categorize speech sounds



Step 1: Meet the Phoneme

Task of the listener: Mapping sound to meaning



- Requires converting the raw signal into the kinds of sound representations that connect to meaning.

- **What are those sound representations like?**



Task of the listener: Mapping sound to meaning

- Stored sound representations must abstract away from a huge amount of physical variance in the signal
 - Loudness
 - Pitch
 - Speaker identity
 - Accent and other variance in pronunciation
 - Etc...

Categorization inside and outside language

- What makes a chair a chair?



- What makes a 't' a 't' ?
 - For the listener?
 - For the speaker?
- Narrowing down the question:
 - What makes a 't' a 't' , as opposed to a 'd' ?

Distinctive features

t

d

Major class
features

+ consonantal
- sonorant
- syllabic

+ consonantal
- sonorant
- syllabic

Manner
features

- nasal
- continuant
- lateral

- nasal
- continuant
- lateral

Laryngeal
features

- voiced

+ voiced

Place of
articulation
features

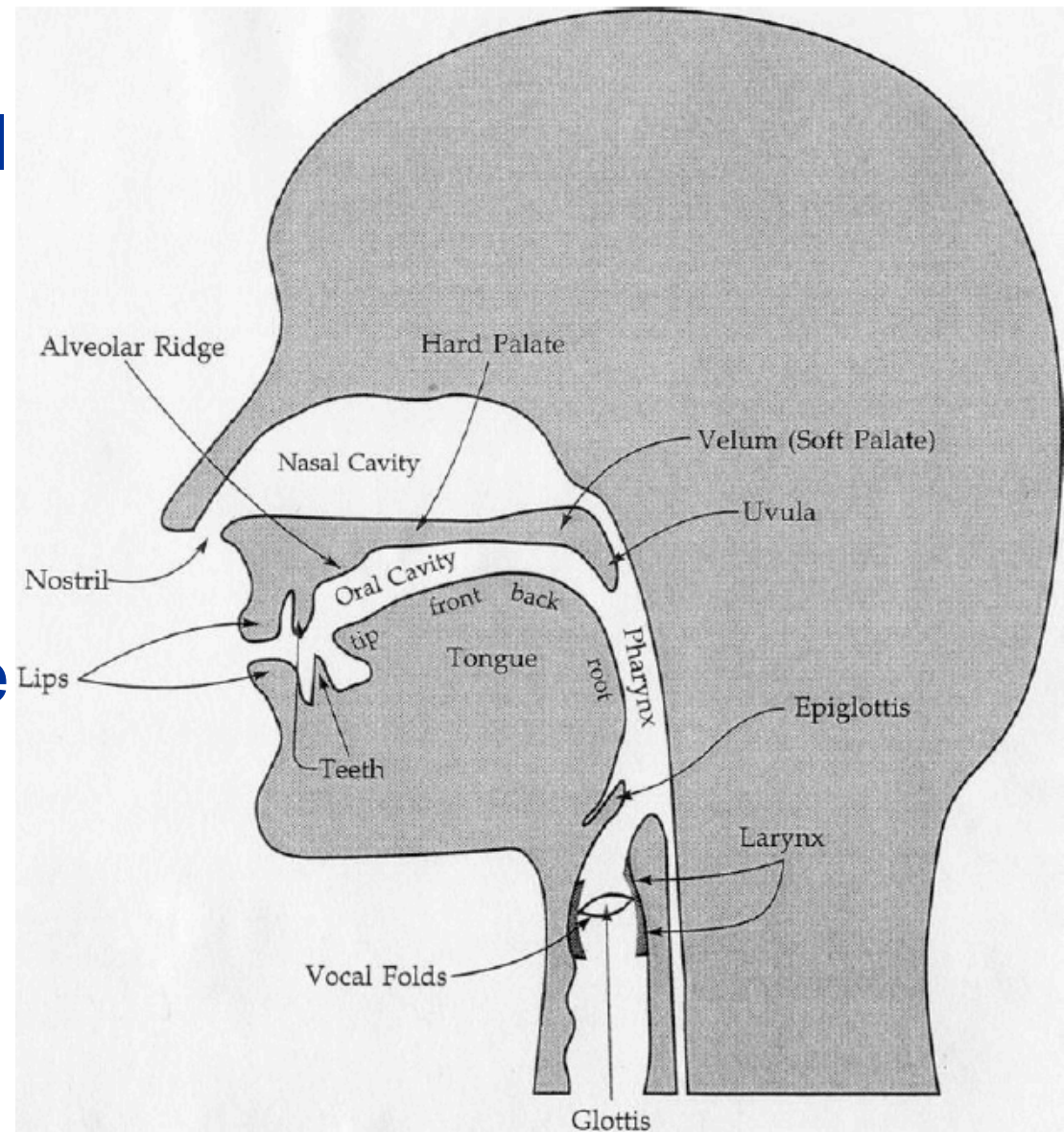
- round
+ anterior

- round
+ anterior

Voicing

Human vocal tract

- [+ voiced] sounds are produced with the vocal folds close together, in such a way that air passing through them causes them to vibrate.
- Touch your larynx while pronouncing [z] vs. [s] to feel that [z] is voiced and [s] voiceless.



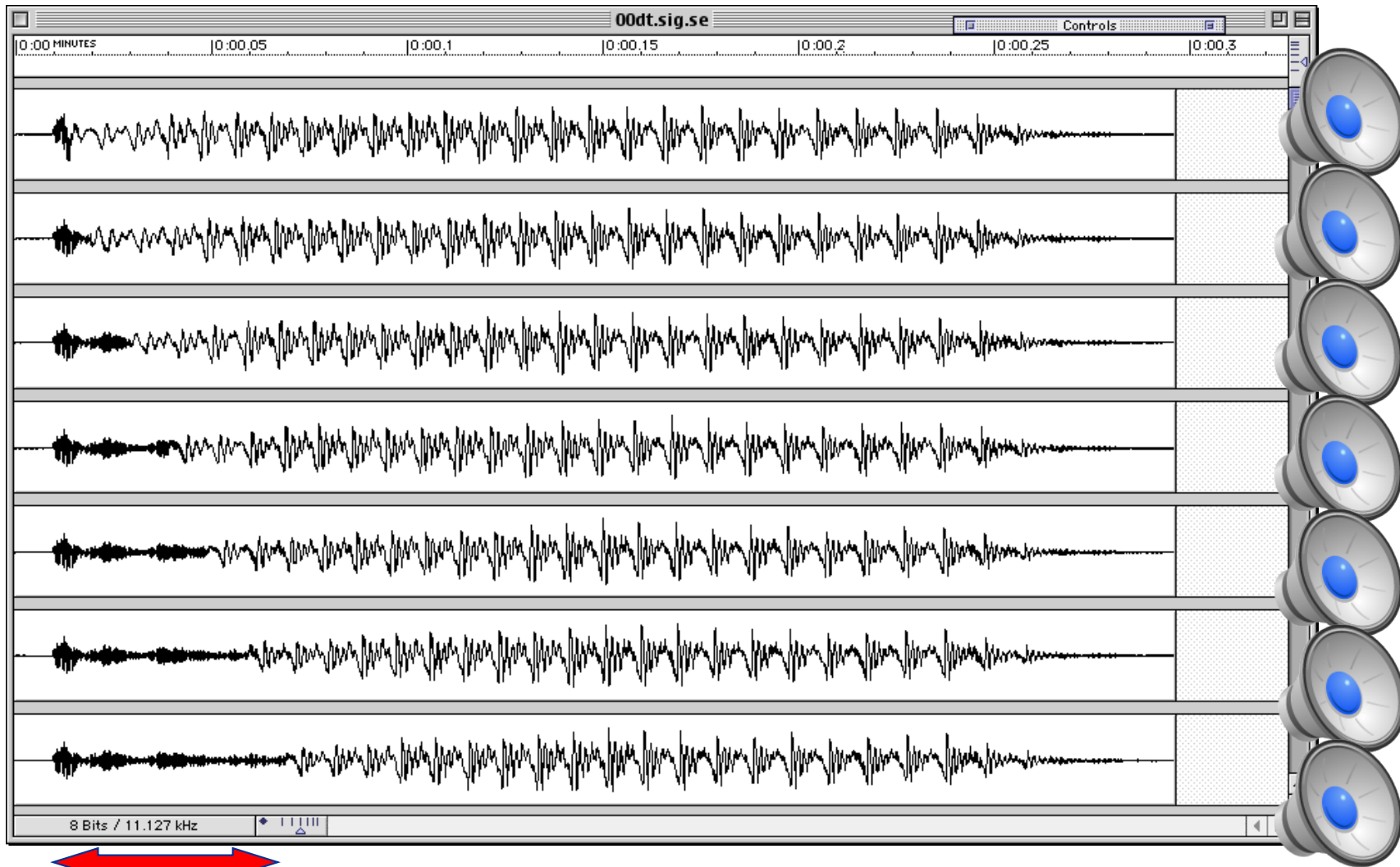
Voicing and stop consonants

- Stop consonants: p, t, k, b, d, g
 - Produced by causing a complete closure of the vocal tract and then releasing it.
- Stops cannot be heard until the closure is released.
- All vowels are voiced.
- When the stop is released quickly, and the voicing of the vowel starts quickly, the stop is considered [+voiced].
- The difference between [d] and [t] is that the onset of voicing of the following vowel takes longer for [t] than for [d].

Acoustically, voiced and voiceless stops differ in their
Voice Onset Time (VOT)

- VOT: The time it takes for the voicing of the vowel to start (voicing = vibration of the vocal folds).

VOT continuum for ta-da



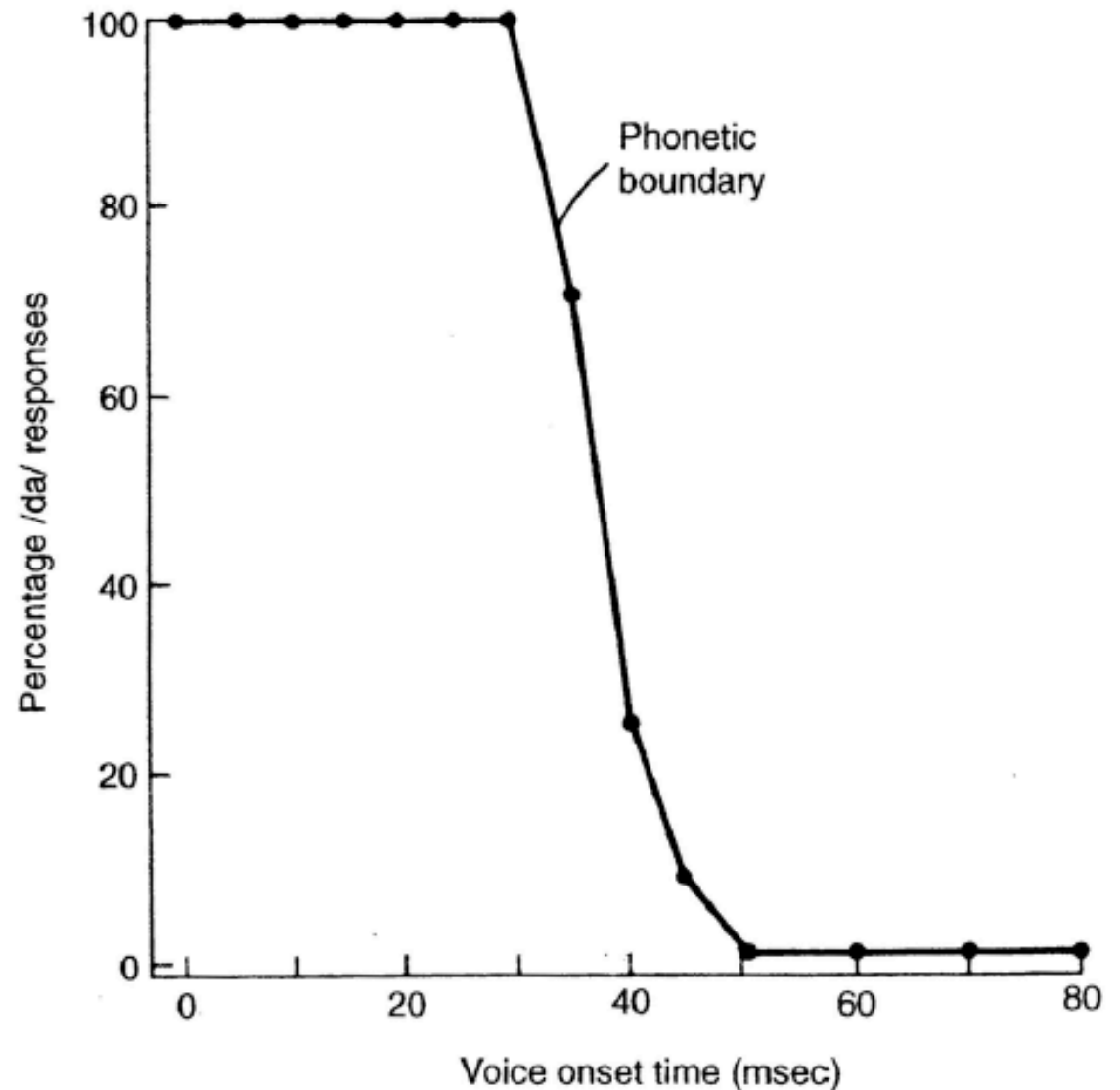
English
boundary

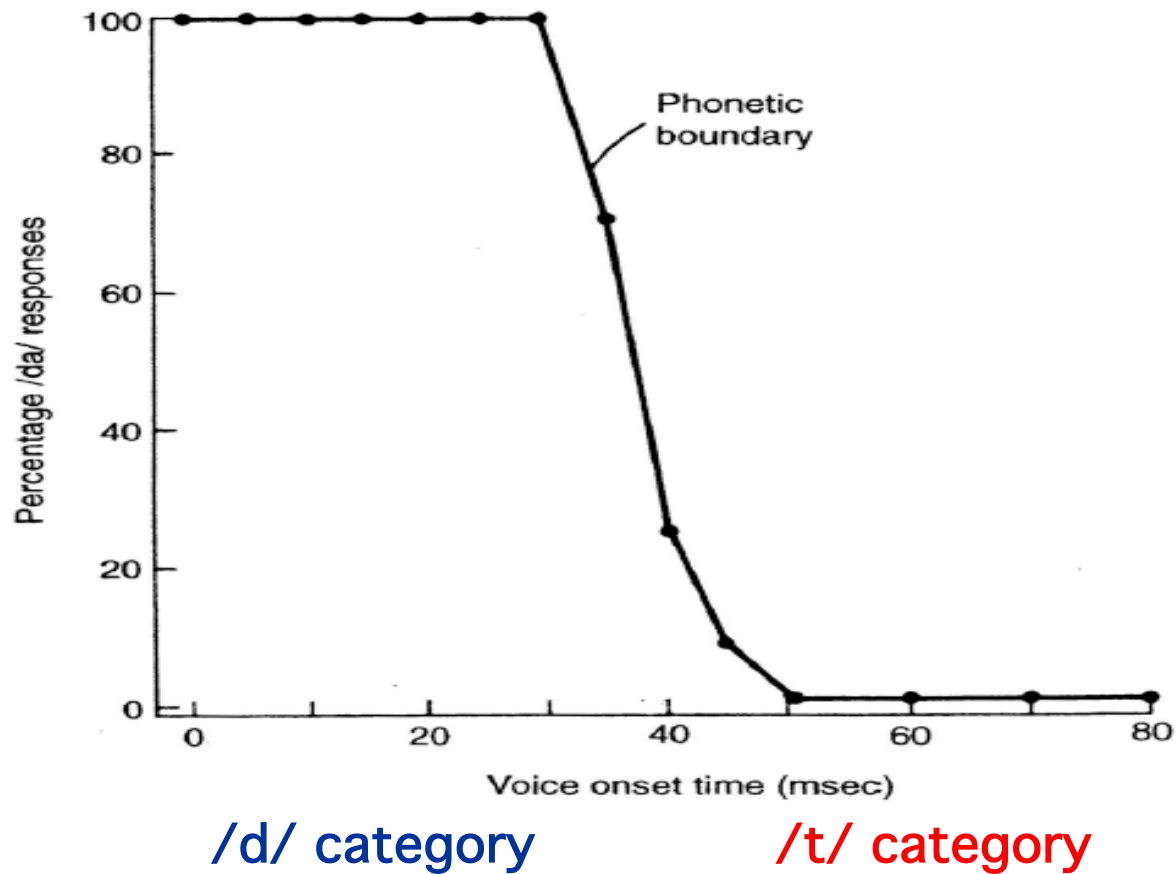


60 msec

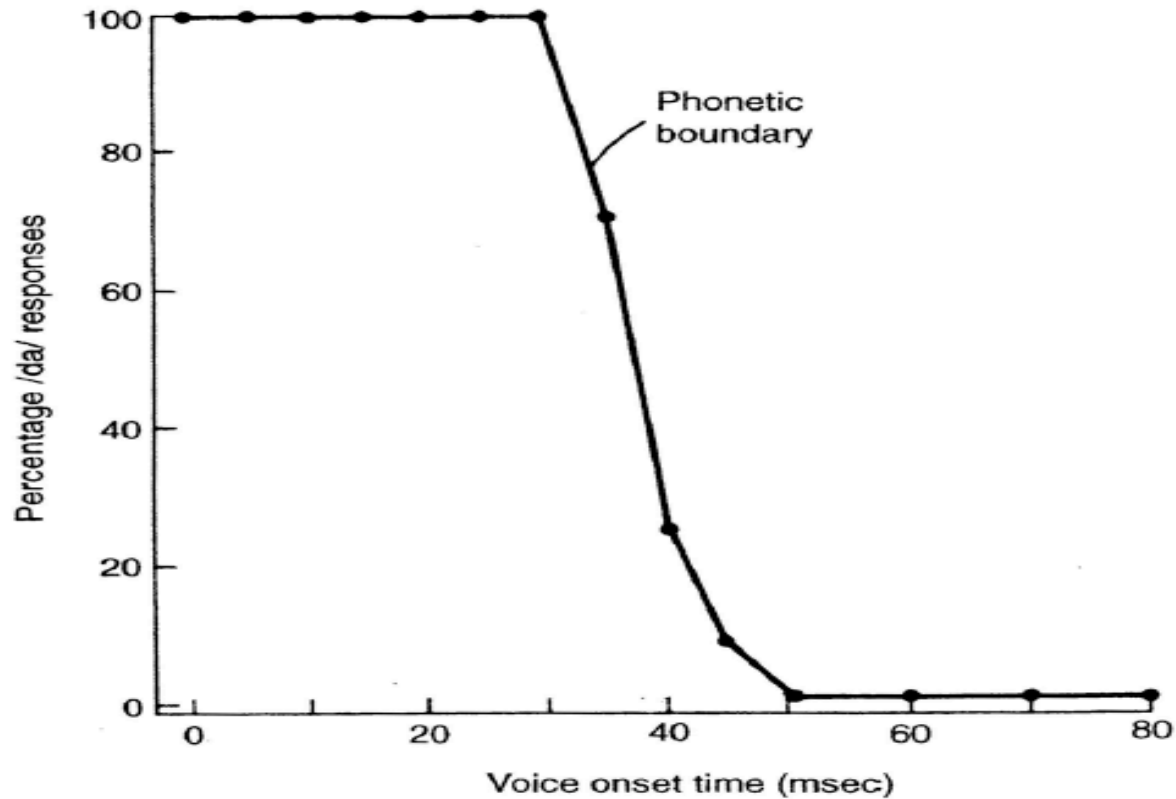
Categorical perception of t-d

- Eimas & Corbitt (1973): Perception of synthesized sounds with VOTs increasing in equal steps from 0 to 80ms.
- Listeners to judge whether they hear a [da] or [ta].





- Within category discrimination is poor.
- How do we learn to group sounds in this way? What is it good for?



/d/ category



A change within
these VOTs ...

/t/ category



... or within
these VOTs

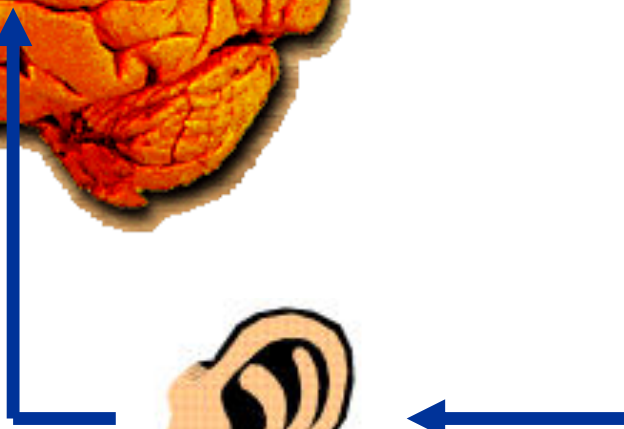
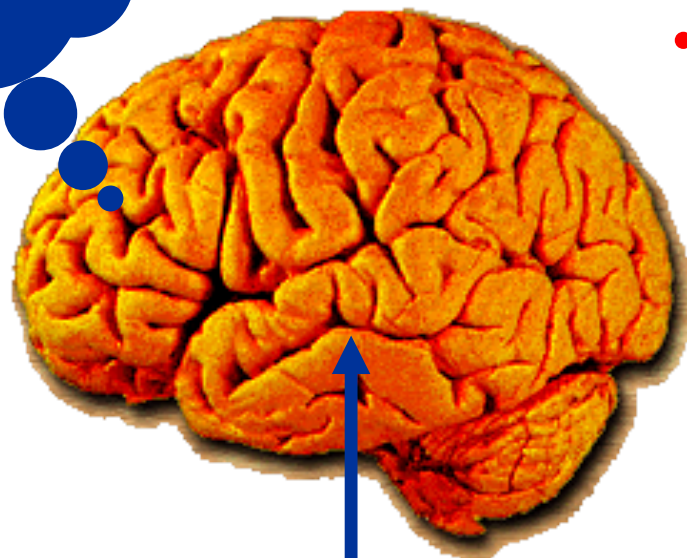
.. is never associated with a change in meaning in English. So being able to discriminate between 10 and 20ms VOTs or 60 and 70ms VOTs won't help you understand English.

Phoneme

- A category of sounds.
- A mental abstraction over a group of sounds within which interchanging one token for another can never induce a change in meaning.
- Since going from /t/ to /d/ in English can induce a change in meaning, /t/ and /d/ are distinct phonemes in English.

A distinction is *phonemic* in a language if it can cause a change in meaning

- Voicing is phonemic feature in English
 - duck vs. tuck
 - zip vs. sip
- Length is not
 - Duck vs. duuck (still means DUCK, just said slowly)
- But it is in Finnish:
 - tuli “fire”
 - tuuli “wind”
 - tulli “customs”



- Requires converting the raw signal into the kinds of sound representations that connect to meaning.

- **What are those sound representations like?**

Sequences of phonemes!