Speech Perception

Phonetic variation:
noise in signal or useful information?

Territory acknowledgment

We should take a moment to acknowledge the land on which we are gathered for the LSA Summer Institute. For thousands of years, this land has been the home of Patwin people, including the Yocha Dehe Wintun Nation today. The Patwin people have remained committed to the stewardship of this land over many centuries. It has been cherished and protected, as elders have instructed the young through generations. We are honored and grateful to be gathered here on their traditional lands.

Pam's office hours (105 Olson)

Thursday, June 27:	3-4 pm
Monday, July 1:	3-4 pm
Wednesday, July 3:	3-4 pm

or send me an email message (beddor@umich.edu) to arrange another time

Traditional main question in speech perception:

How do listeners interpret the input acoustic signal as linguistic forms?

Perception is malleable and dynamic

- is continuously retuned
  - context-dependent
  - cue-dependent
  - talker-specific
  - new experiences
- evolves in real time as input acoustic signal unfolds
- varies across listeners
  - different listeners make different linguistic decisions based on the same acoustic input

This is an an important part of the answer to the ”main question”.

Getting there involves … →

Traditional main question in speech perception:

How do listeners interpret the input acoustic signal as linguistic forms?

… Getting there involves discussion of:

- acoustic variation due to
  - phonetic context (coarticulation)
  - talker differences
- classic perceptual phenomena up through ongoing perceptual studies
  - from categorical perception to …
  - study of moment-by-moment time course of perception
- current theoretical approaches to speech perception

Traditional main question in speech perception:

How do listeners interpret the input acoustic signal as linguistic forms?

Outline for today:

- Nature of acoustic variation
  - What and where are the "linguistic forms"?
  - Variation due to phonetic context (briefly: and due to talker)
- A first stab at: what do listeners do?
  - Early work on perceiving variation: categorical perception
  - Revisiting early findings: not-so-categorical perception?
- Brief foray into theories of speech perception as seen through the lens of two foundational questions:
  - Is coarticulatory variation noise or perceptually useful information?
  - What do listeners recover from the acoustic signal?

What type of "linguistic form" are we talking about?

Feature?

Gesture?

Syllable?

Meaning?

Word?

Phoneme?

Choice of early perceptual research

Goal: acoustics-to-phoneme mapping

Issues that arise in acoustics-to-phoneme mapping

As an initial step (which we'll revise):

Set ourselves the task—like that of the early literature—of determining how listeners extract phonemes from acoustic signal

In tackling acoustics-to-phoneme mapping, researchers encounter

- Difficulty in segmenting acoustic stream (segmentation problem)
- Lack of invariance

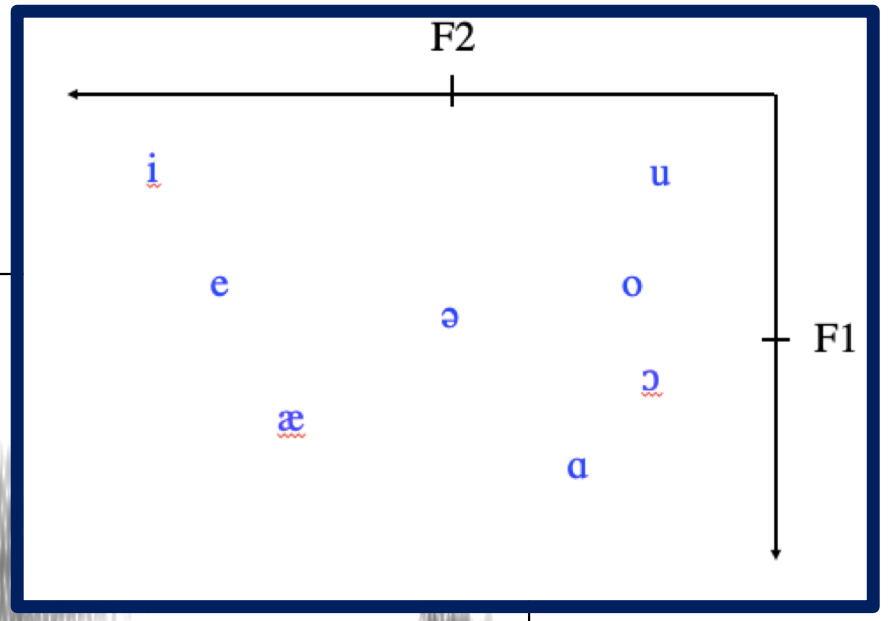
Issues that arise in acoustics-to-phoneme mapping

Segmentation Problem

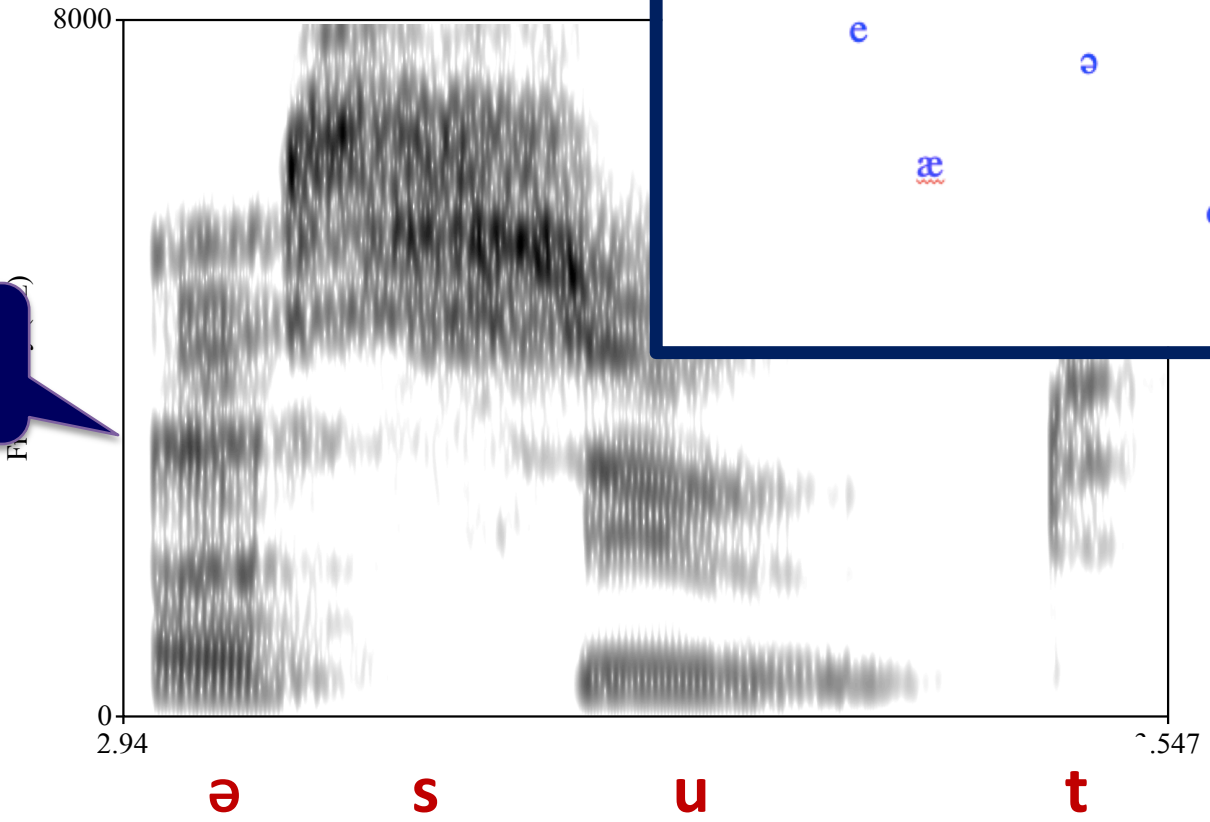
Researchers often cannot isolate segments in the acoustic signal that reliably correspond to units of linguistic analysis and perception.

Issues that arise in acoustics-to-phoneme mapping

Multiple mappings:

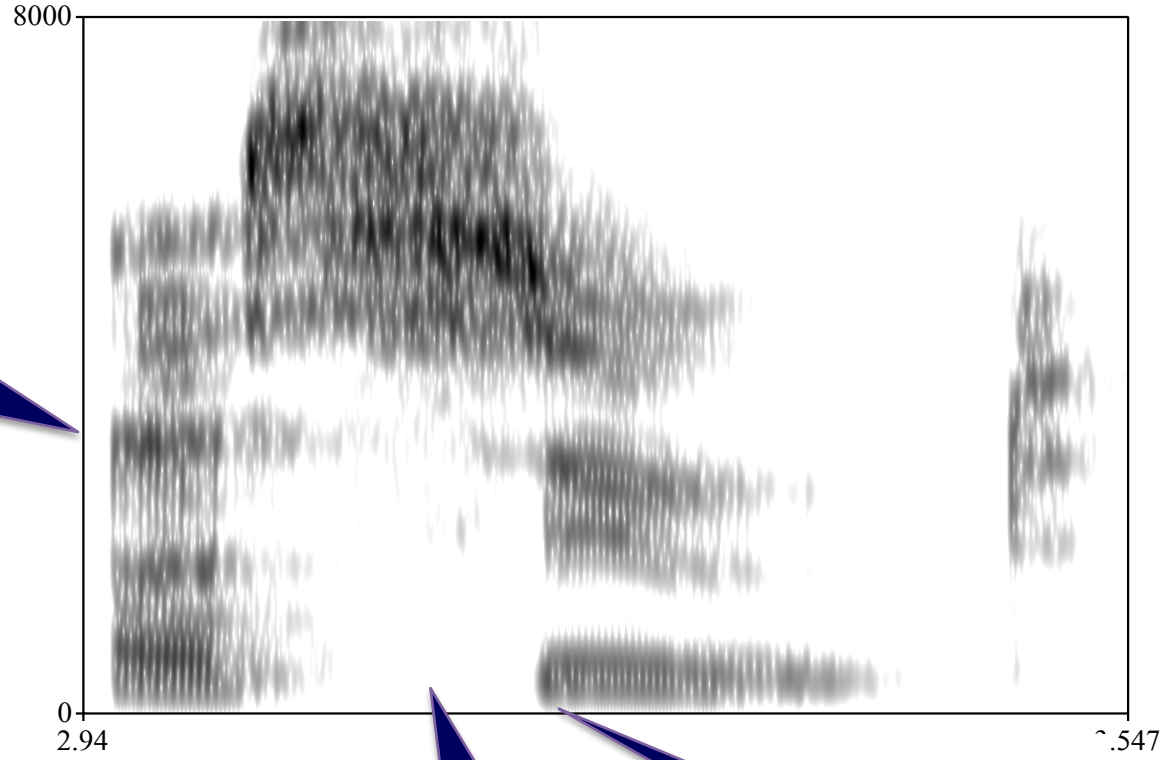


Info about V_2 in V_1



Issues that arise in acoustics-to-phoneme mapping

Multiple mappings:



Info about V_2 in V_1

Info about V_2 in frication

Info about fricative place on transition into V

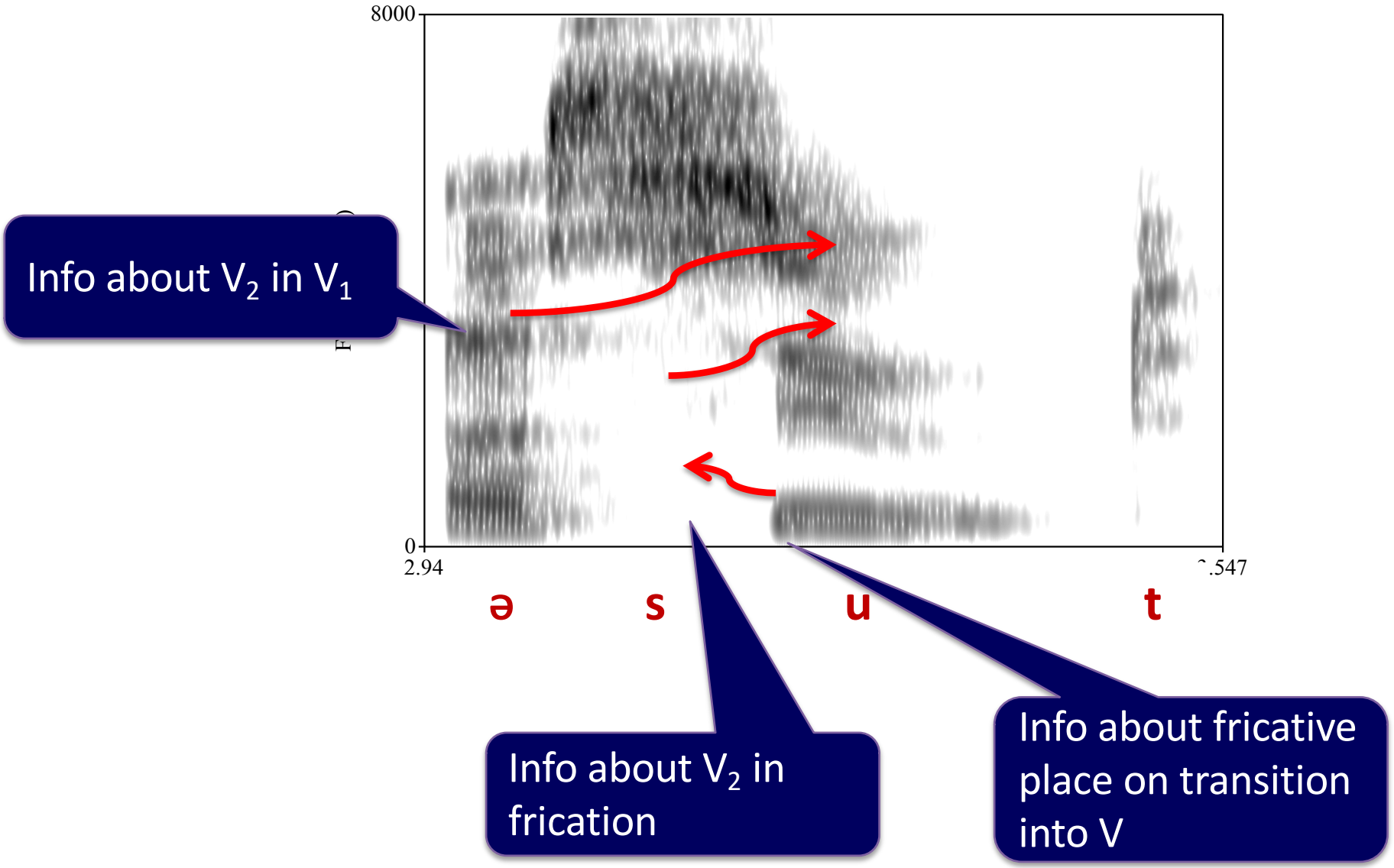
Issues that arise in acoustics-to-phoneme mapping

Segmentation "problem":

NOT a problem for listeners . . .

Issues that arise in acoustics-to-phoneme mapping

Multiple mappings:



Issues that arise in acoustics-to-phoneme mapping

Segmentation "problem":

NOT a problem for listeners . . .

Rather, a challenge for speech researchers trying to understand how listeners integrate information from multiple locations in acoustic signal.

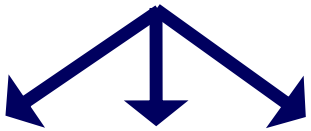
Issues that arise in acoustics-to-phoneme mapping

Lack of Invariance

The problem: In some cases, there appear to be no acoustic properties that reliably correspond to the segments of linguistic analysis and perception.

Contextual influences

Acoustic signal



Multiple percepts

One-to-many mapping

Acoustic signals

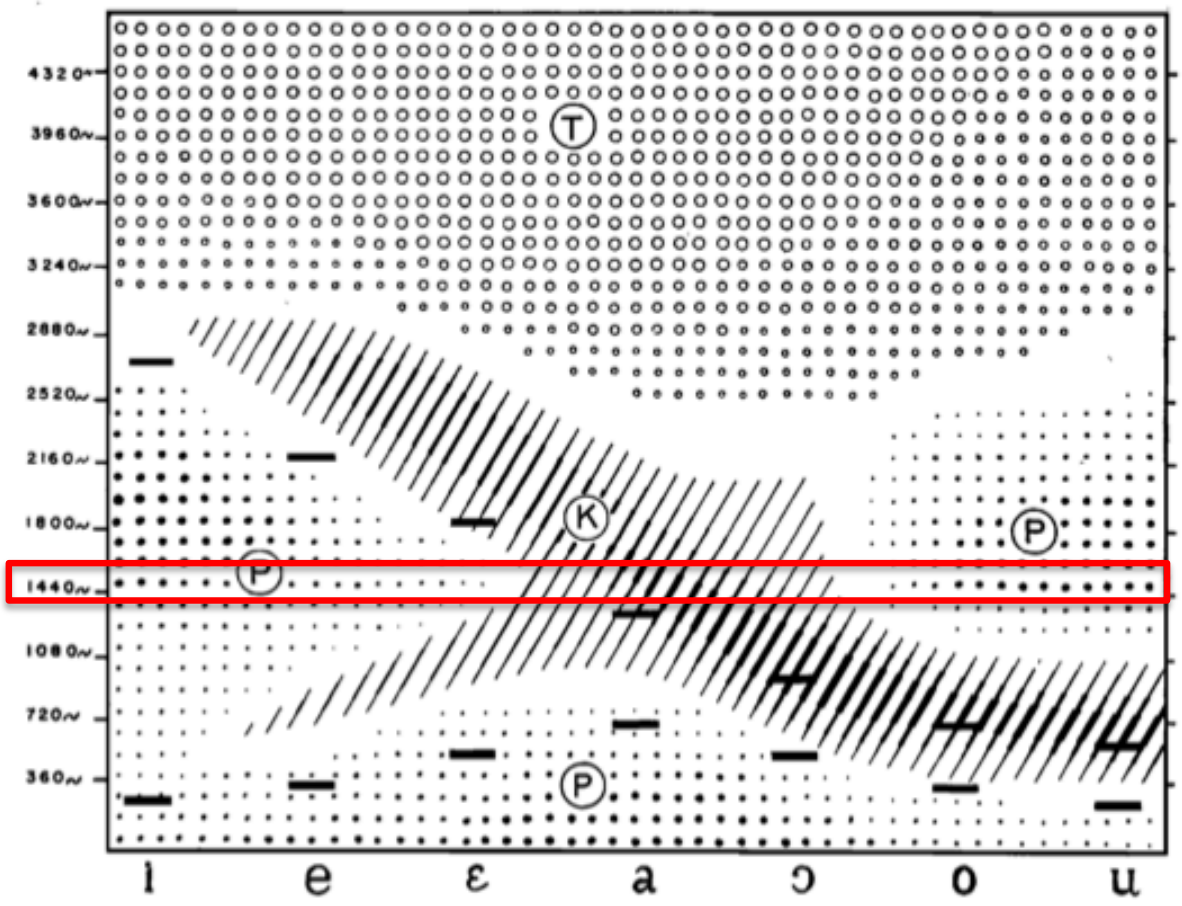
Percept

Many-to-one mapping

Early example of one-to-many mapping: bursts in stops

1440 Hz (synthetic) burst is heard as **[p]** when followed by [i] or [u]
but as **[k]** when followed by [a]

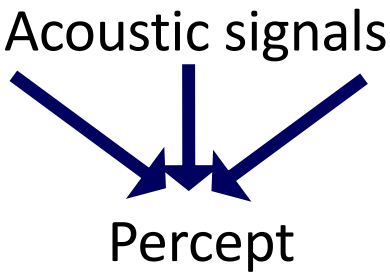
Frequency of stop burst



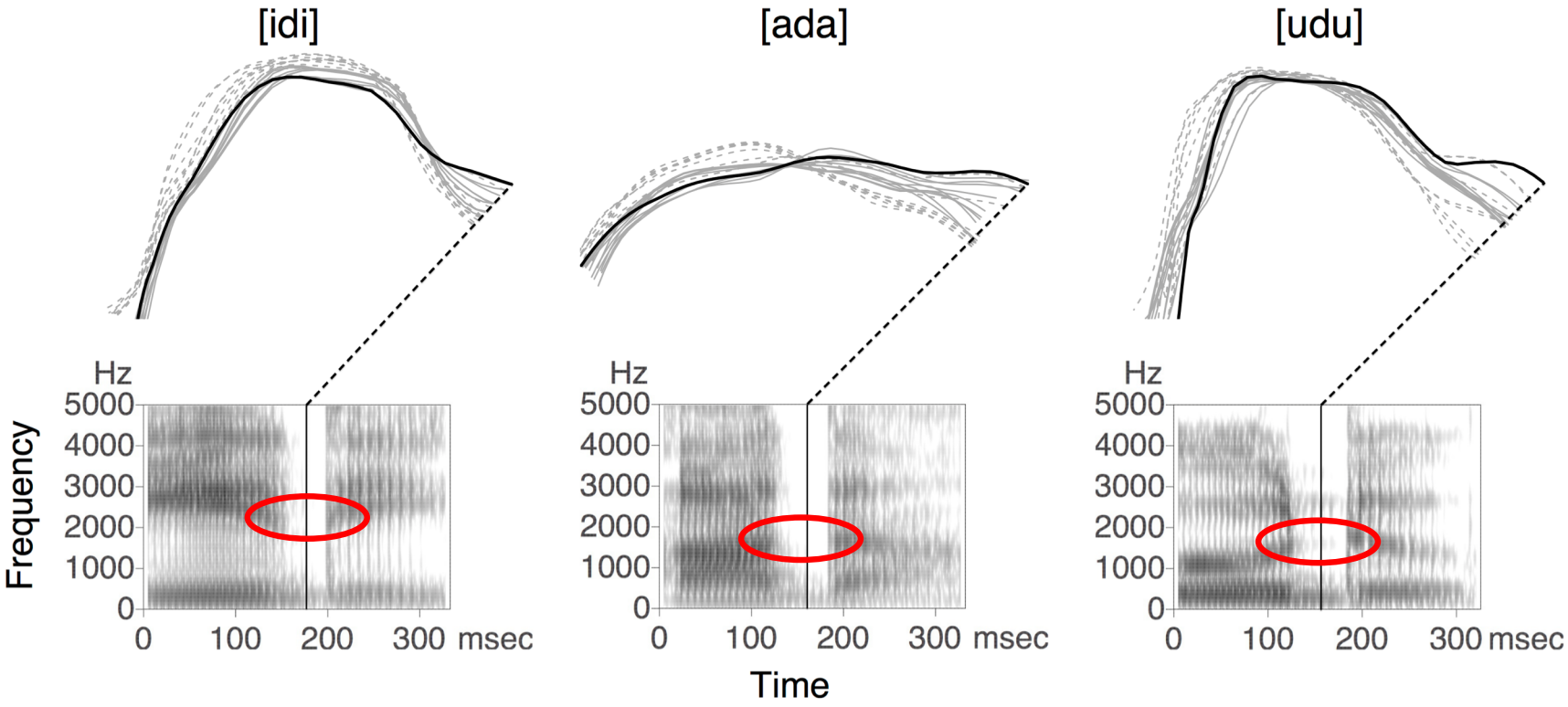
variable percept

From Cooper et al. 1952, *J. Acoustical Soc. of America*

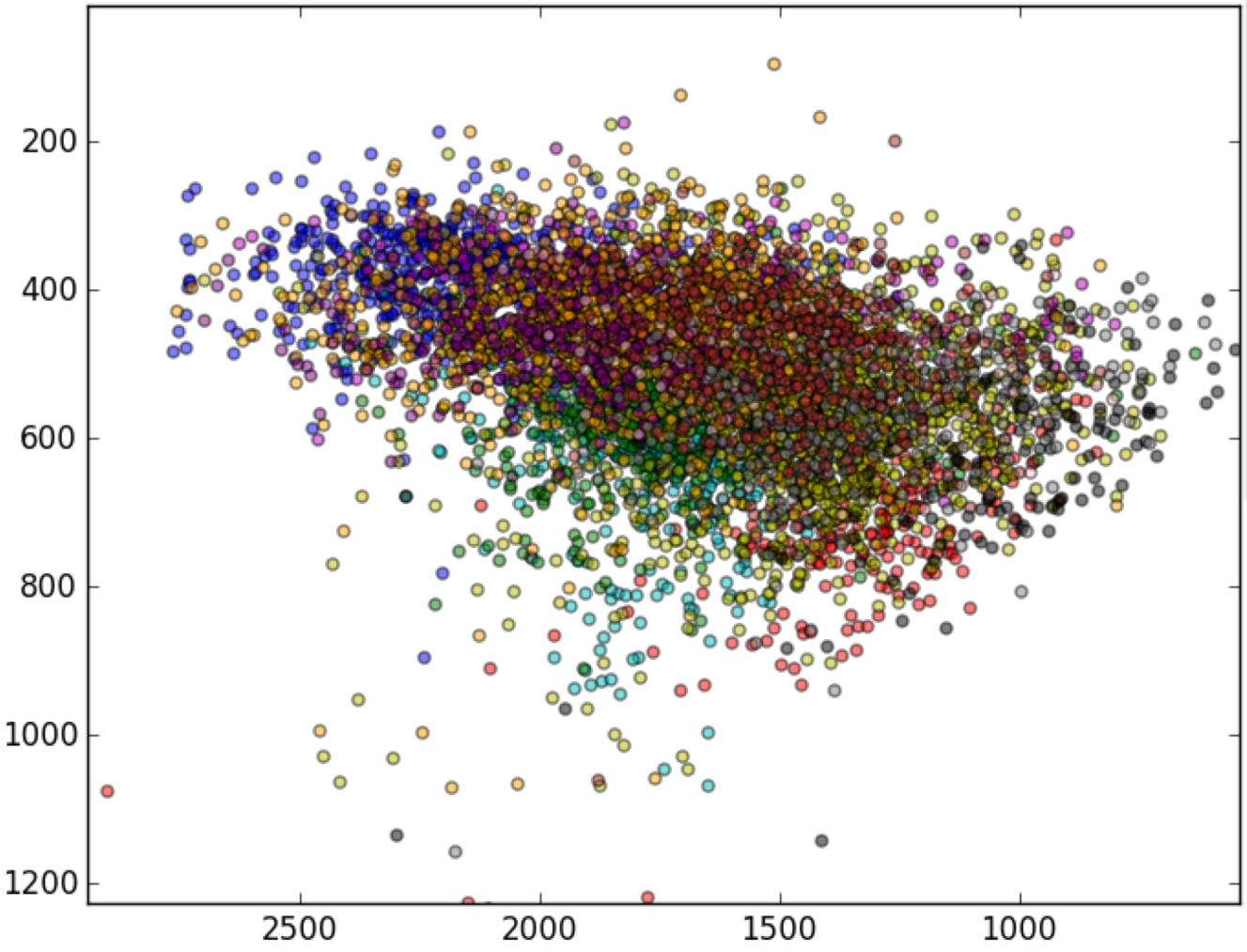
Contextual influences: many-to-one mapping



F2 transitions into / out of flanking vowel(s)



Antetomaso et al. (2017): Vowel overlap (across contexts, speakers) in spontaneous speech (Buckeye corpus)



Across-talker variation

Different sized/shaped vocal tracts for different speakers result in different resonant characteristics and hence different acoustic properties.

Different colors: different vowel phonemes

How *do* listeners reliably interpret the input acoustic signal as linguistic forms?

An early first pass: categorical perception

- Liberman et al. (1957):

  How do listeners "**reduce** the number and variety of the many sounds with which [they are] bombarded"?

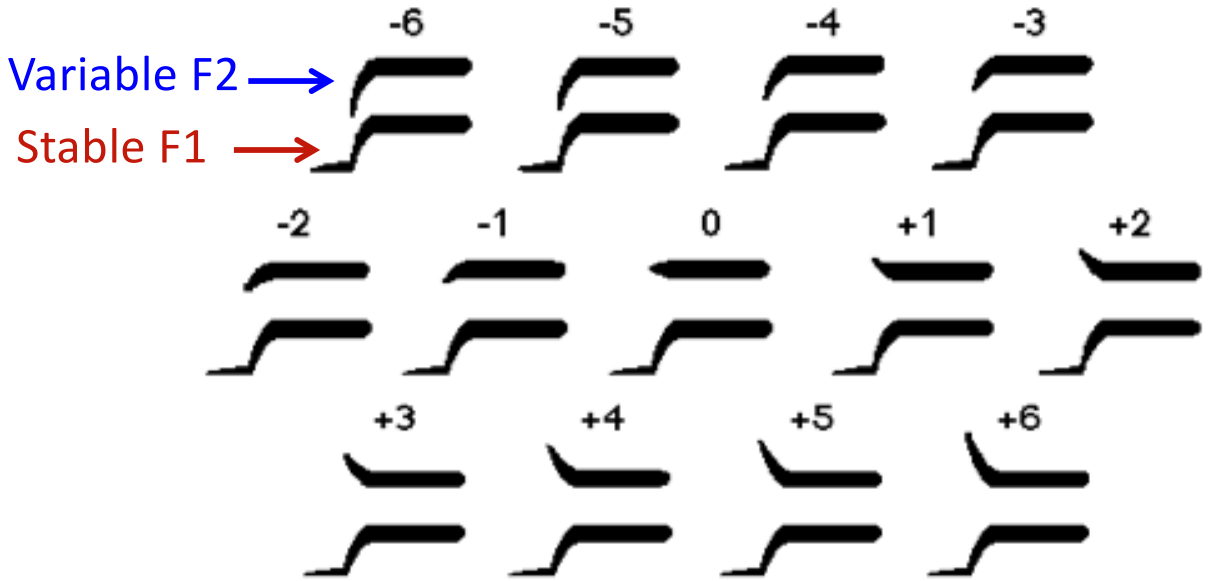
Phenomenon of categorical perception:

- Seemed to offer a tentative answer to Liberman et al.'s question.
- Influenced early theoretical developments
- Served as impetus for a large body of experiments that continue today

Return to historical perspective

Liberman et al. 1957 (J. of Experimental Psychology 54, 358-368)

- Used early Pattern Playback synthesizer to generate physical continuum varying in F2 frequency
- Elicited percepts from /b/ to /d/ to /g/

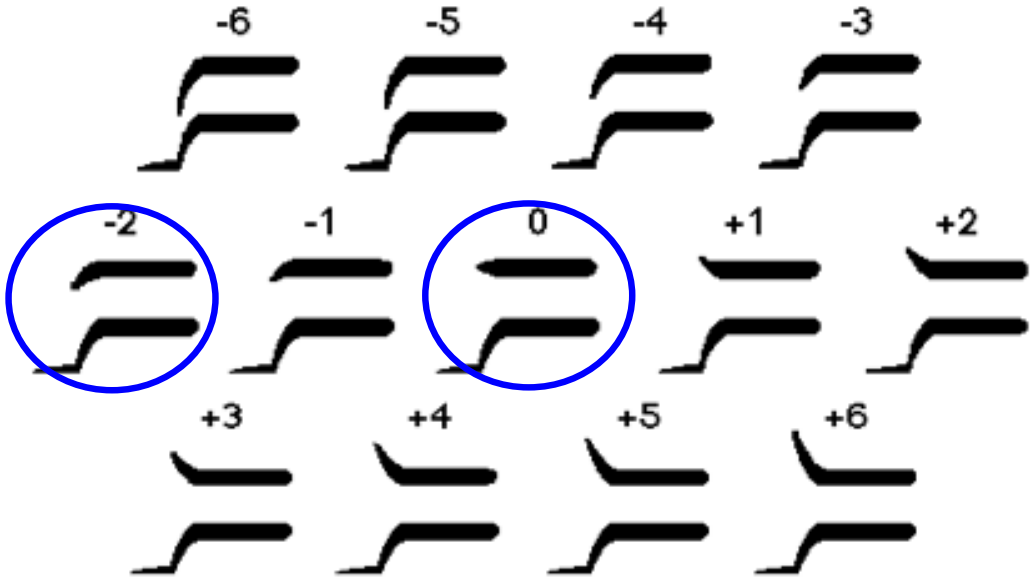


[Stimuli and formant patterns from Haskins Labs website:
http://www.haskins.yale.edu/featured/bdg.php?audio=AIFF#

These are similar to those used by Liberman et al.]

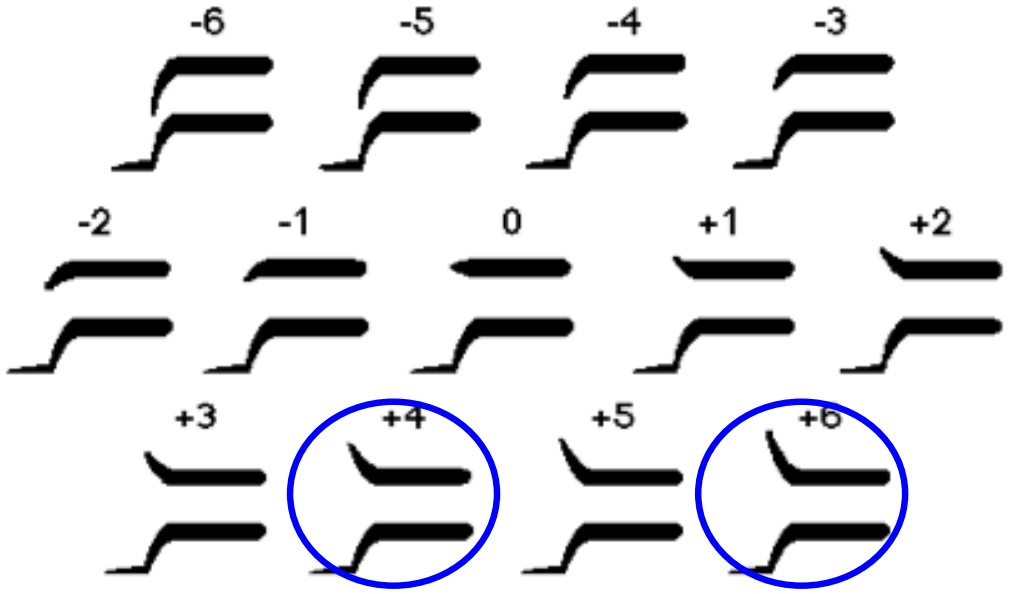
Liberman et al. 1957

- Despite *physical continuum*, listeners tend to abruptly shift from one stop category to another.
- Discrimination (ABX): Peaks of good discrimination occurred for paired stimuli (AB) that were *identified* as different.



Liberman et al. 1957

- Despite *physical continuum*, listeners tend to abruptly shift from one stop category to another.
- Discrimination (ABX): Peaks of good discrimination occurred for paired stimuli (AB) that were *identified* as different



… but not for paired stimuli that were *identified* as the same stop.

Return to historical perspective

Liberman et al. 1957

- Despite *physical continuum*, listeners tend to abruptly shift from one stop category to another.
- Discrimination (ABX): Peaks of good discrimination occurred for paired stimuli (AB) that were *identified* as different.

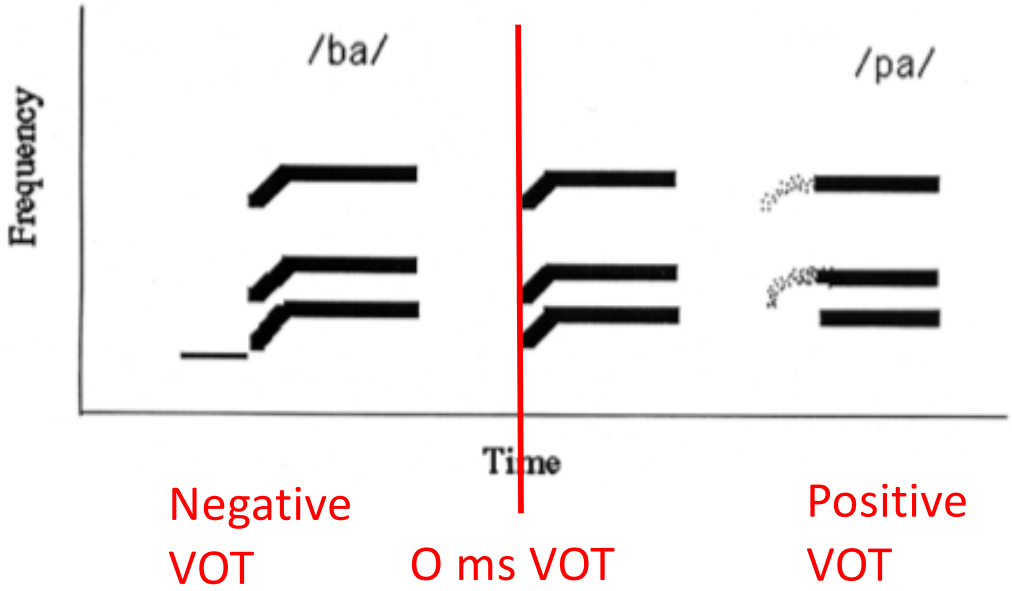
Compare: pure tone perception

- Estimated that (young) human listeners can discriminate about 350,000 different tones
- Only a small fraction of this number can be assigned different labels

VOICE ONSET TIME

Lisker & Abramson (1964, *Word*). "A cross-language study of voicing in initial stops: acoustical measurements."

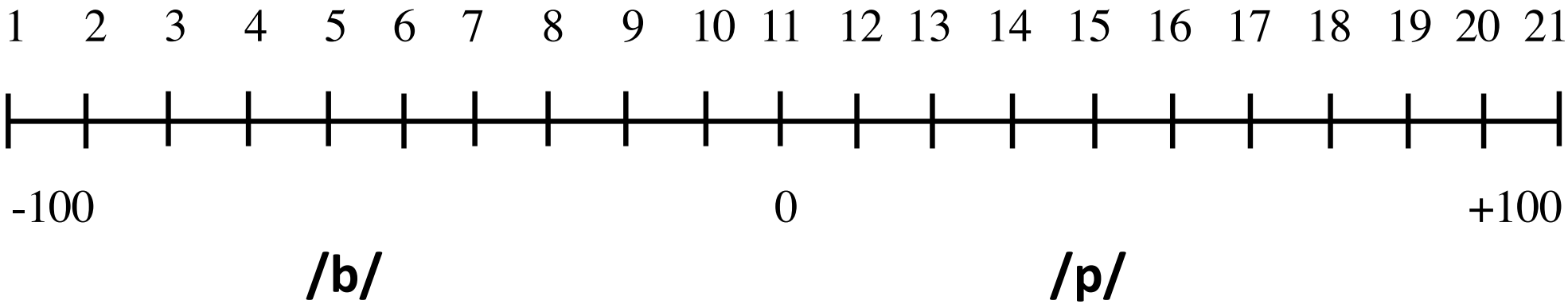
Abramson & Lisker (1967/1970, *Proceedings of 6th ICPhS*). "Discriminability along the voicing continuum: cross-language tests."



- Voice onset time (VOT) = interval between release of articulatory stricture and onset of voicing
- Measure used to describe voicing differences within and across languages

Abramson & Lisker VOT test

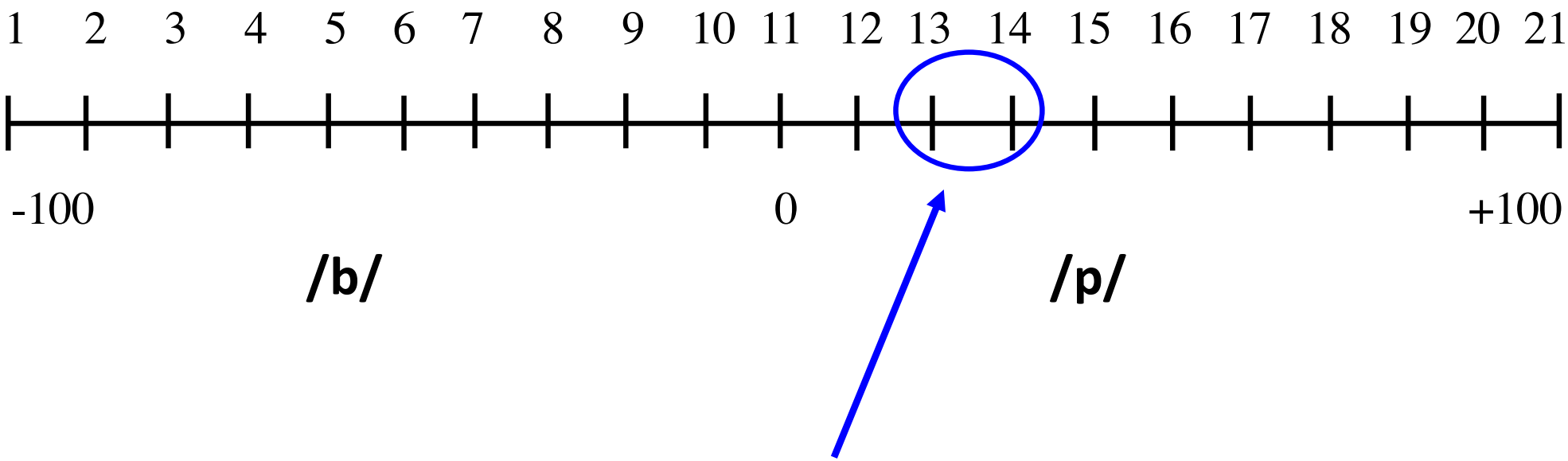
- 21 stimuli varying in voice onset time
- Range from -100 VOT (prevoicing) to +100 ms VOT (voicing lag or aspiration) in 10 ms increments:



Abramson & Lisker VOT test

Identification test:

- abrupt change from /b/ to /p/ percept

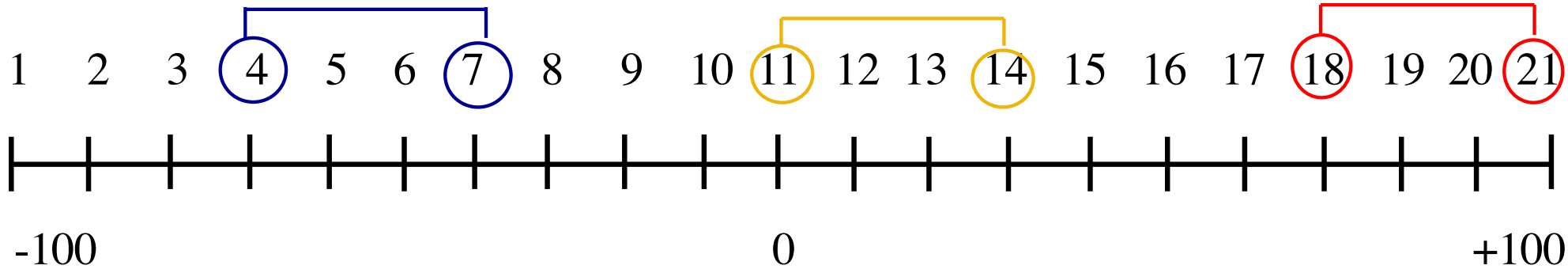


Native speakers of English tend to crossover from /b/ to /p/ between stimuli #13 and 14 (20-30 ms VOT).

Abramson & Lisker VOT test

Discrimination test:

- Paired stimuli: 3 steps apart (i.e., physically, equally distinct) along continuum: Stimulus 1 paired with Stimulus 4, 2 with 5 ….
- Result:
  - Good discrimination of pairs whose members are identified as different
  - Poor discrimination of pairs whose members are identified as same



Categorical Perception

Members of physical continuum perceived as belonging to discrete phoneme categories.

- IDENTIFICATION: Abrupt crossover from 1 category to the other.
- DISCRIMINATION is accurate for pairs that cross the identification boundary; near chance on within-category pairs.

This pattern holds especially for consonant perception (and most especially for perception of stop consonants).

Categorical perception

General picture that emerged:

Many speech contrasts perceived categorically, with differences between *more discrete articulations eliciting the most categorical percepts*.

Important picture in two respects:

- Categorical perception initially interpreted as evidence that listeners ignored the “variety of the many sounds with which [they are] bombarded".
- Differences between *more discrete articulations elicit the most categorical percepts*.

… contributed to two major theoretical questions:

- Is phonetic variation noise to be ignored or is it perceptually useful?
- What do listeners recover from the acoustic signal?

General picture that emerged:

Many speech contrasts perceived categorically, with differences between *more discrete articulations eliciting the most categorical percepts*.

Important picture in two respects:

- Categorical perception initially interpreted as evidence that listeners ignored the “variety of the many sounds with which [they are] bombarded".
- Differences between ***more discrete articulations*** *elicit the most categorical percepts*.

… contributed to two major theoretical questions:

- Is phonetic variation noise to be ignored or is it perceptually useful?
- What do listeners recover from the acoustic signal?

Are listeners recovering articulatory (gestural) information?

Ignoring variation?

Is phonetic variation noise to be ignored or is it perceptually useful?

Fast-forward to the 21st century:

- Some early findings seem to suggest that within-category variation is ignored
- Not so!
- Subphonemic variation (setting aside for the moment whether it is “noise”) is ***not*** ignored.

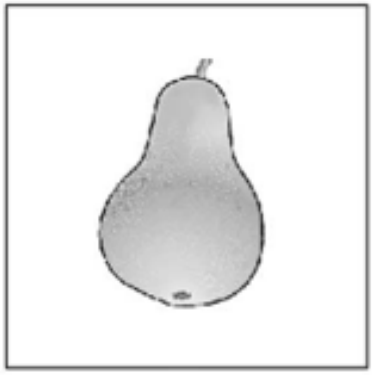
McMurray et al. 2002

McMurray, Tanenhaus & Aslin, 2002, *Cognition*
"Gradient effects of within-category phonetic variation on lexical access"

- Eye movements monitored as participants looked at which 1 of 4 pictures corresponded to auditory stimulus
- Auditory stimuli: varied in VOT in 5 ms steps (0 to 40 ms VOT)
- As VOT approached category boundary, fixations to competitor image increased

McMurray et al. 2002



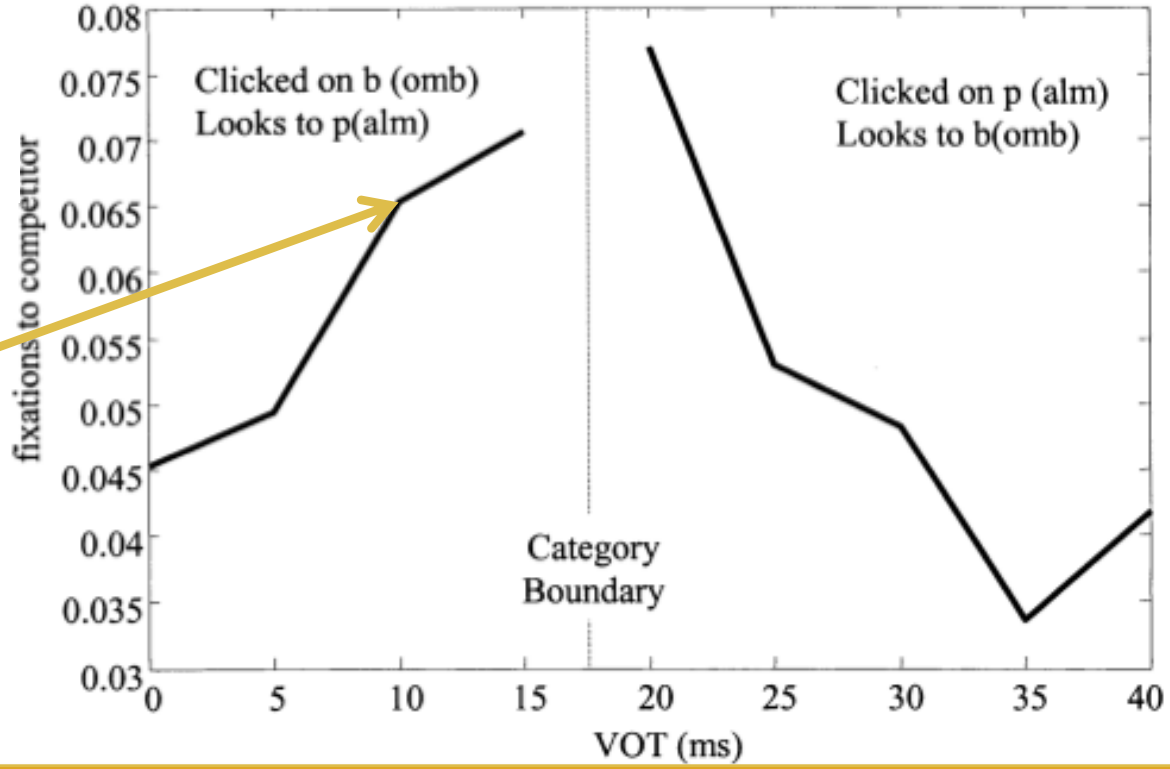



Task: Click on the image corresponding to the auditory stimulus.

Data: mouse clicks
eye movements

Listeners chose [b]-initial word for VOTs of less than 20 ms.

BUT: increasingly likely to look at [p]-initial word as VOT approached 20 ms



Not so “categorical” perception

Thus, listeners *are* sensitive to sub-phonemic differences:

- Although lexical ***decisions*** (as measured by, e.g., identification tasks) can be categorical, lexical ***activation*** appears more gradient.
- Reaction time studies also show outcomes consistent with these findings– e.g., listeners' responses are slowed by inappropriate contextual information (Whalen 1991, Perception & Psychophysics).

Intersection of two major theoretical issues

Is phonetic variation noise or is it perceptually useful?

Contextual (= coarticulatory) variation:

- is *noise* that may interfere with processing
- *facilitates* perception*;* lawful variation that helps listeners arrive at speaker's intended utterance

What do listeners recover from the acoustic signal?

Listeners' processing of contextual variation suggests they recover:

- gestural information
- auditory information

Selected theoretical approaches

General Auditory Theory (Lotto & Kluender 1998; Lotto & Holt 2006, 2016)

- Listeners recover acoustic/auditory information (just as is the case for all other sound perception).
- Auditory processing can accommodate the acoustic effects of – and apparent challenges due to – coarticulation.

= coarticulation as noisiness in signal (that auditory system can handle)

Selected theoretical approaches

General Auditory Theory (Lotto & Kluender 1998; Lotto & Holt 2006, 2016)

- Listeners recover acoustic/auditory information (just as is the case for all other sound perception).
- Auditory processing can accommodate the acoustic effects of – and apparent challenges due to – coarticulation.

Direct Realism (Fowler 1986, 1996, 2006)

- Listeners perceive gestures. They identify the relation between the acoustic signal and the articulatory source of the speech event.
- Due to coarticulation, portions of acoustic signal are shaped by more than one gesture; listeners "parse" acoustic signal along gestural lines.

= coarticulation facilitates perceptual tracking of articulation

Selected theoretical approaches

General Auditory Theory (Lotto & Kluender 1998; Lotto & Holt 2006, 2016)

- Listeners recover acoustic/auditory information (just as is the case for all other sound perception).
- Auditory processing can accommodate the acoustic effects of – and apparent challenges due to – coarticulation.

Direct Realism (Fowler 1986, 1996, 2006)

- Listeners perceive gestures. They identify the relation between the acoustic signal and the articulatory source of the speech event.
- Due to coarticulation, portions of acoustic signal are shaped by more than one gesture; listeners "parse" acoustic signal along gestural lines.

Exemplar Theory (Goldinger 1998, Pierrehumbert 2001)

Listeners store speech events in memory with intact acoustic, contextual, and social information.

= noisy or facilitative nature of coarticulation is largely a non-issue

The nature of phonetic variation:

Massive acoustic variation — but much of this variation is lawful / structured

- Contextual (coarticulation: temporally overlapping gestures)
- Talker-specific (different vocal tracts and socio-indexical characteristics)
- Speaking rate (e.g., selectively shorter durations and more reduced gestures at faster rates)
- Word frequency (again, more reduced gestures for more frequent words)

and other factors …

Perceiving phonetic variation:

- Listeners are good categorizers of variable input (e.g., “I heard [b]” or “I heard 'bear'”)
- Categorization ≠ ignore phonetic (subphonemic) detail
- How do listeners accomplish both?

Answer we’ll consider next time: Listeners *both*

- Use coarticulatory details to anticipate upcoming sound
- Attribute those details to their coarticulatory source

Looking ahead

Next time: perceiving variation due to coarticulation

- Phenomenon: compensation for coarticulation

  Focus: perception of /da-ga/ continuum in /aɹ__/ and /al__/ contexts

  Background acoustics:

  - Main acoustic difference between /ɹ/ and /l/: F3
    /ɹ/: low F3 /l/: high F3

  - /g/ has lower F3 frequency than /d/ has (F2/F3 "velar pinch")

- Thus, due to coarticulation: F3 onset of /g/ is higher after /l/ than after /ɹ/
  That is, */g/ is acoustically more /d/-like after /l/.*

1. What is perceptual compensation for coarticulation?
2. How do different researchers experimentally investigate and theoretically interpret these findings?