Locality & Predictivity

Rebecca L. Morley
Workshop on Emergence of Universals
OSU
February 18, 2018

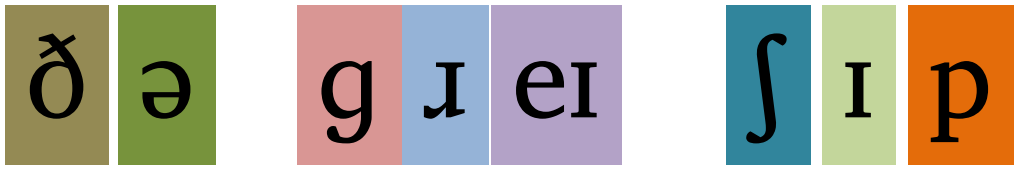
Mental Representations

Levels: degree of generalization, hierarchical structure

- Word
- Foot
- Syllable
- Phoneme
- Features
- Rhymes
- CVC
- Phrases
- …

I.
To Normalize or Not to Normalize

(In)variance Problem



Hockett (1955)

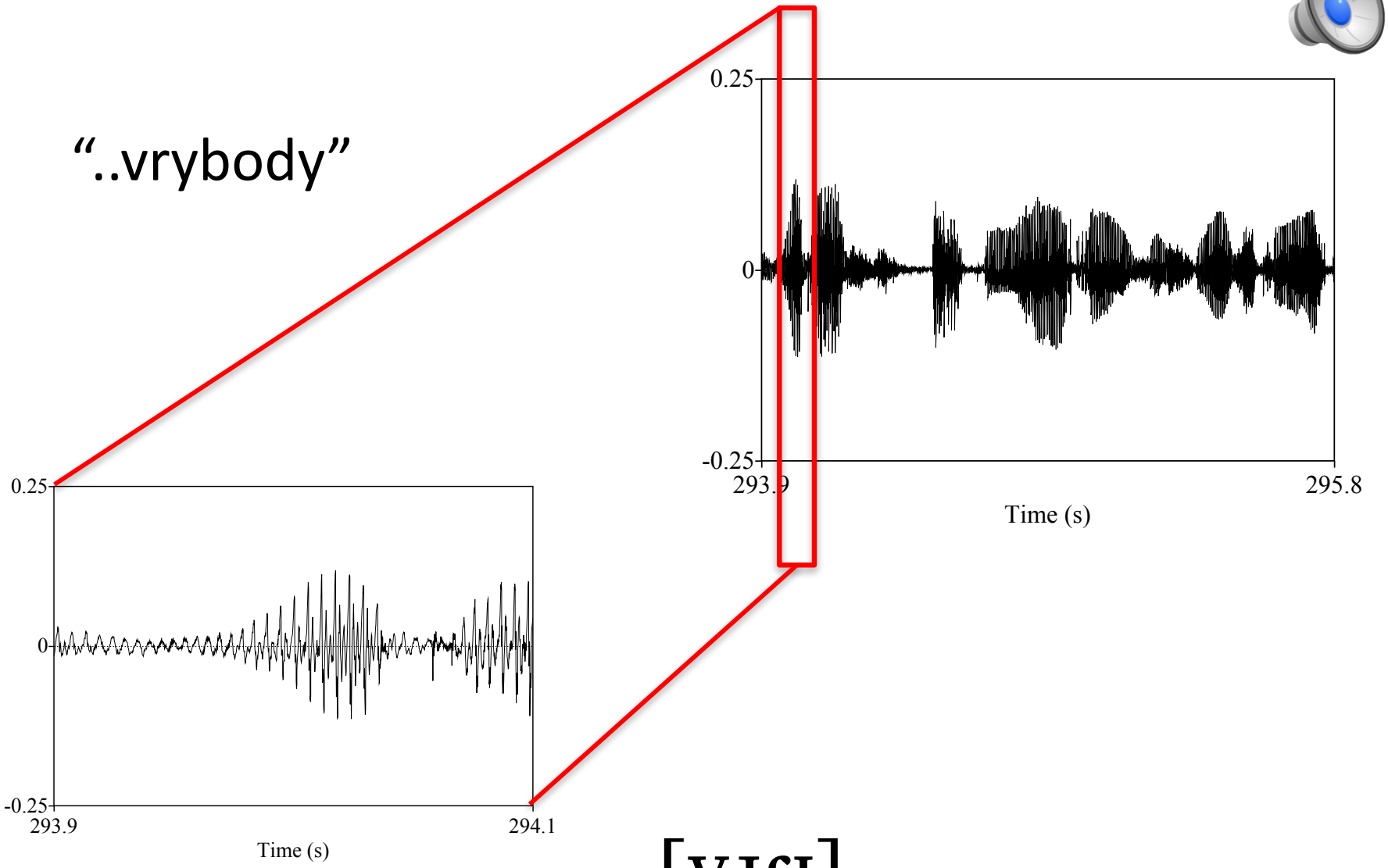
(In)variance Problem

Hockett (1955)

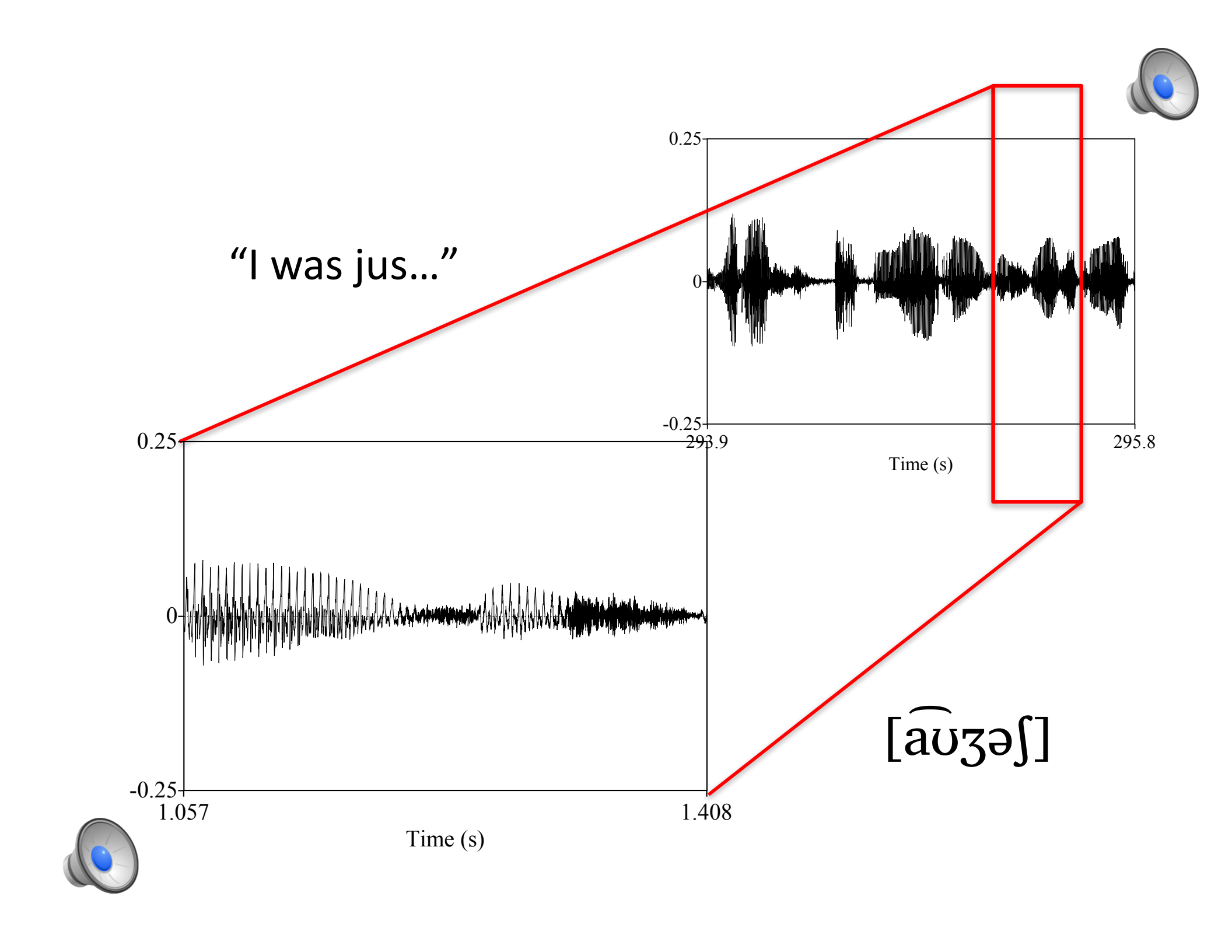
(In)variance Problem

Hockett (1955)

“..vrybody”



[ɹɪɹʌ]


“I was jus…”
0.25
0
-0.25
293.9
295.8
Time (s)
0.25
0
-0.25
1.057
1.408
Time (s)
[aʊzəʃ]

Factors that affect the realization of sound units

- Speaker gender, age, social orientation
- Speaker vocal tract physiology
- Speaking rate
- Ambient noise
- Word frequency
- Part of speech
- Stress
- Prosodic position
- Discourse position
- …

Factors that affect the analysis of data

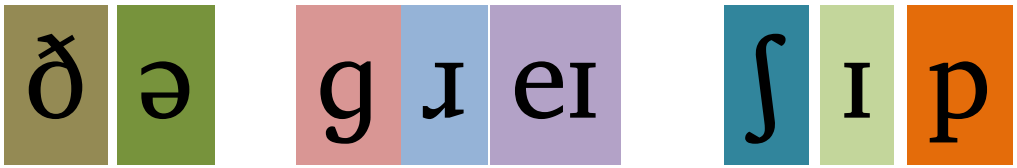
- Inherent language-specific confounds (like the fact that voiceless stops tend to occur in higher-frequency words, for example)
- Confounds from random sampling
- Floor/ceiling effects
- Linear in-separability
- subset/superset relationships

To Normalize

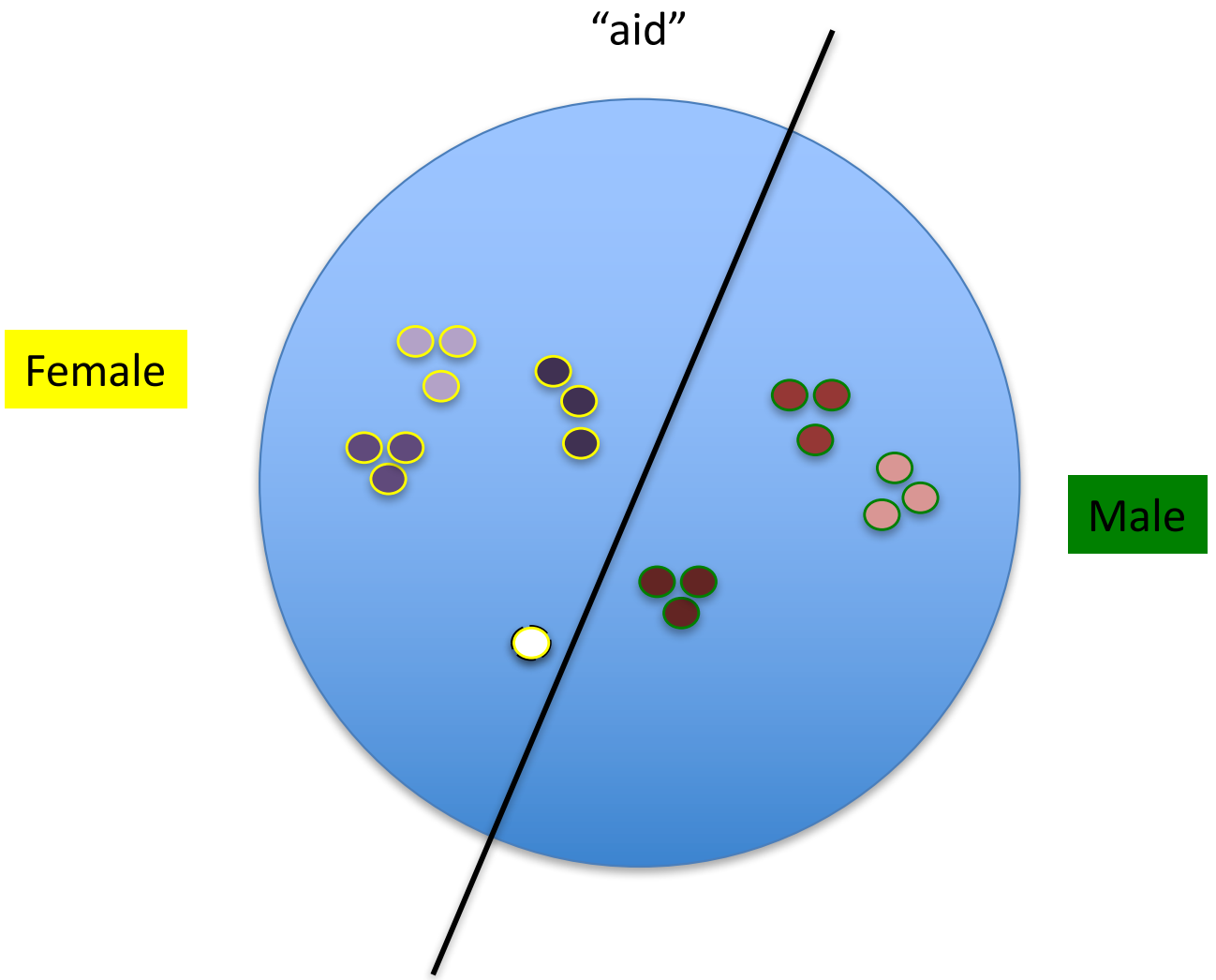
Bulky load:
tap cold:
45 min:
high spin

Nips hgih
:nim 54
:dloc pat
:daol ykluB

1. Deduce the washing machine algorithm from the observed output
2. Run the algorithm in reverse
3. Match to stored abstract representations

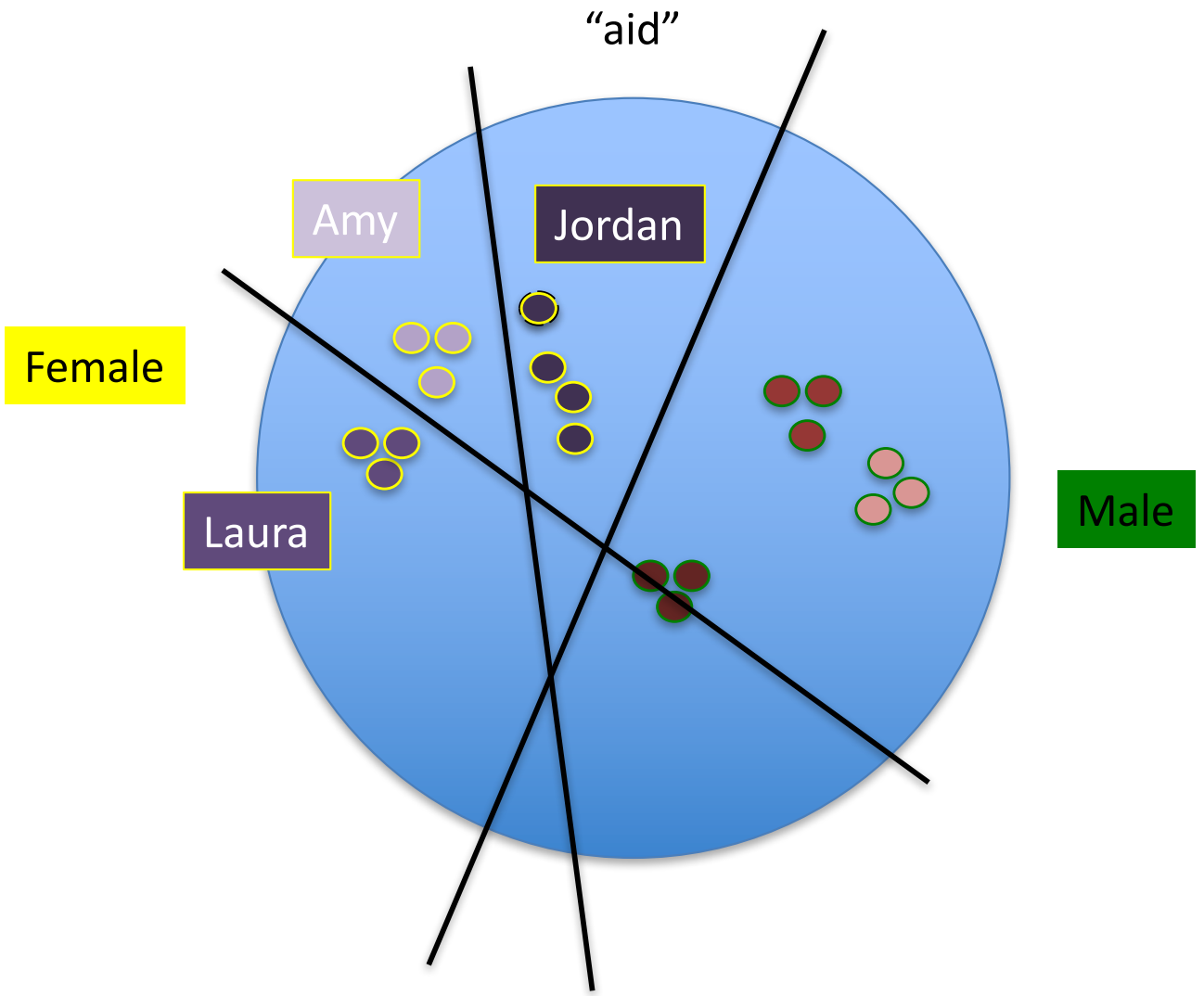


To Not



Goldinger 1996; Johnson 1997;Pierrehumbert 2001,2002;

To Not



Goldinger 1996; Johnson 1997;Pierrehumbert 2001,2002;

Representations

- Normalization:
  - There seem to be an almost unlimited number of factors that must be normalized for
  - Many of which are non-linear
  - One-to-many problem
- Storage:
  - We'd have to store a lot of stuff!
  - We'd have to get lucky with relatively separable dimensions
  - The similarity function probably has to be pretty complicated to make this work

But we already know a lot!

- About visual perception
- About object recognition
- About auditory perception
- About spoken word recognition
- About speech processing
- About phonetic variables
- About stops in American English!

Outline

- **An Introduction to the Data: duration-based effects**
- **The Chicken & The Egg: Duration & Speaking Rate**
- **Locality & Predictivity:**
  - Similar effects across perceptual systems
  - Similar effects across planning/acting systems
- **The Data (Re)Analyzed: An Odyssey**
  - Trading Relations
  - Cue Parsing
  - Categorical Perception
  - Speaking Rate
- **Some speculations regarding underlying representations**
  - Some thoughts about sound change
  - Some constraints on theories of representations
  - Expectations versus perceptual relativity

Managing Expectations

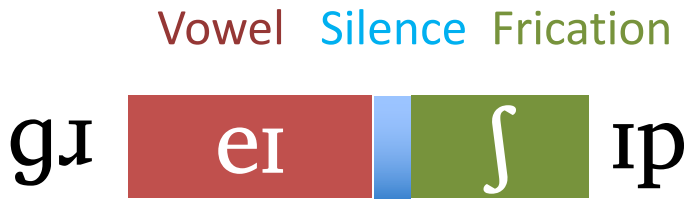
Self-Deprecating

- Almost none of the results or theories I am going to discuss in this talk are original
- I read a lot of stuff
- And tried to make sense of it

- Self-Aggrandizing
  - That wasn't easy!
  - Everyone is working in a different framework
  - You have to make the links between the different results
  - And then you have to translate them into the language of phonological representations
- Self-Deprecating
  - This is very much a work in progress (continuing to progress up to about 12:00 am this morning)
  - So I'm just going to leave you with some hypotheses that I think are testable, and some thoughts on necessary/sufficient conditions for theories of representations

Data

Trading Relations



Start with the phrase “gray ship”

- Manipulate duration of silence between words

If the silence gets long enough, listeners perceive “great ship”

Repp et al. (1978)

Trading Relations

Start with the phrase "gray ship"

- Manipulate duration of silence between words
- Manipulate duration of frication noise

Repp et al. (1978)

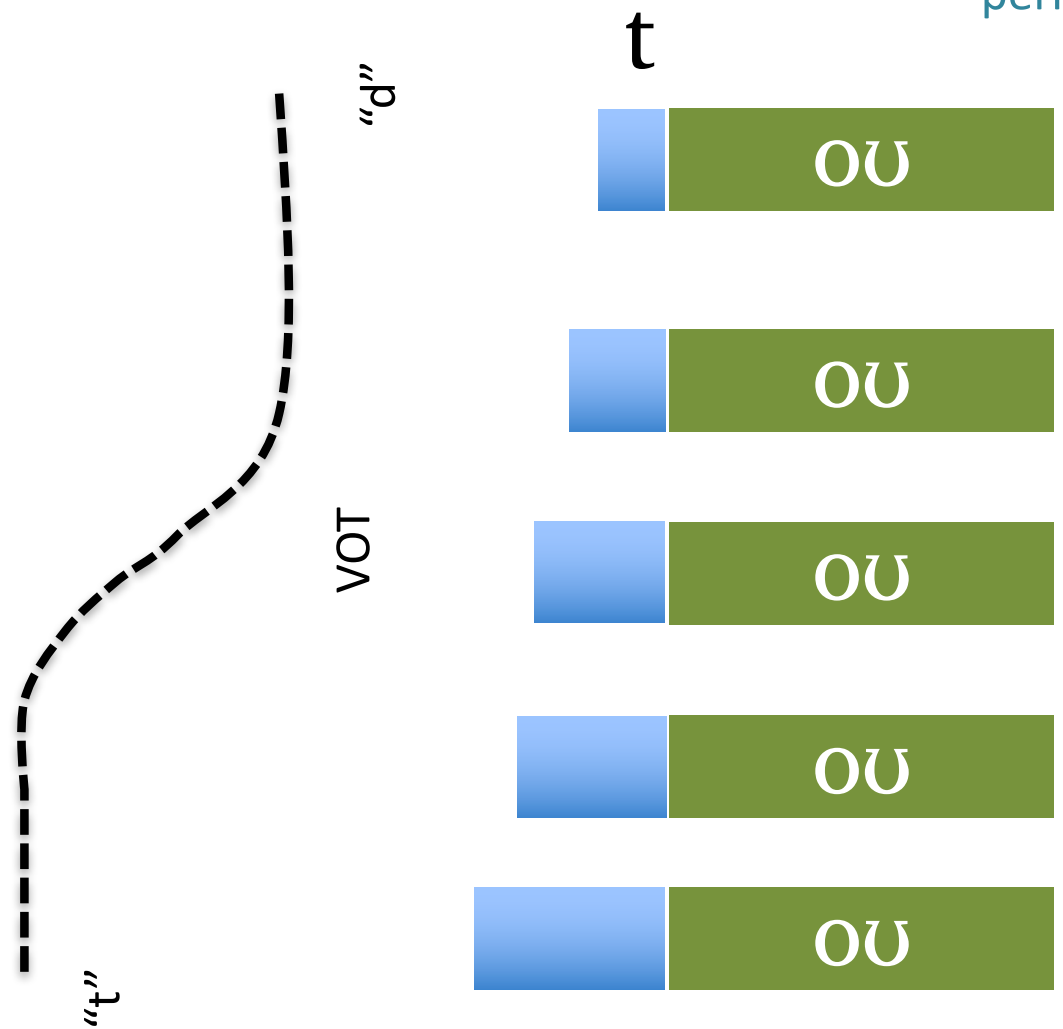
Duration-Based Contrasts

American English plosives

- Phonologically, these are supposed to represent a voicing distinction
  - {t,p,k} [-voice]
  - {d,b,g} [+voice]
- Phonetically, these are described as being realized as the following allophonic variants:
  - Initial Position: long VOT vs. short VOT
  - Medial Position: long closure: vowel duration ratio vs. short closure: vowel duration ratio
  - Final Position: long preceding vowel duration vs. short preceding vowel duration

Initial Stops

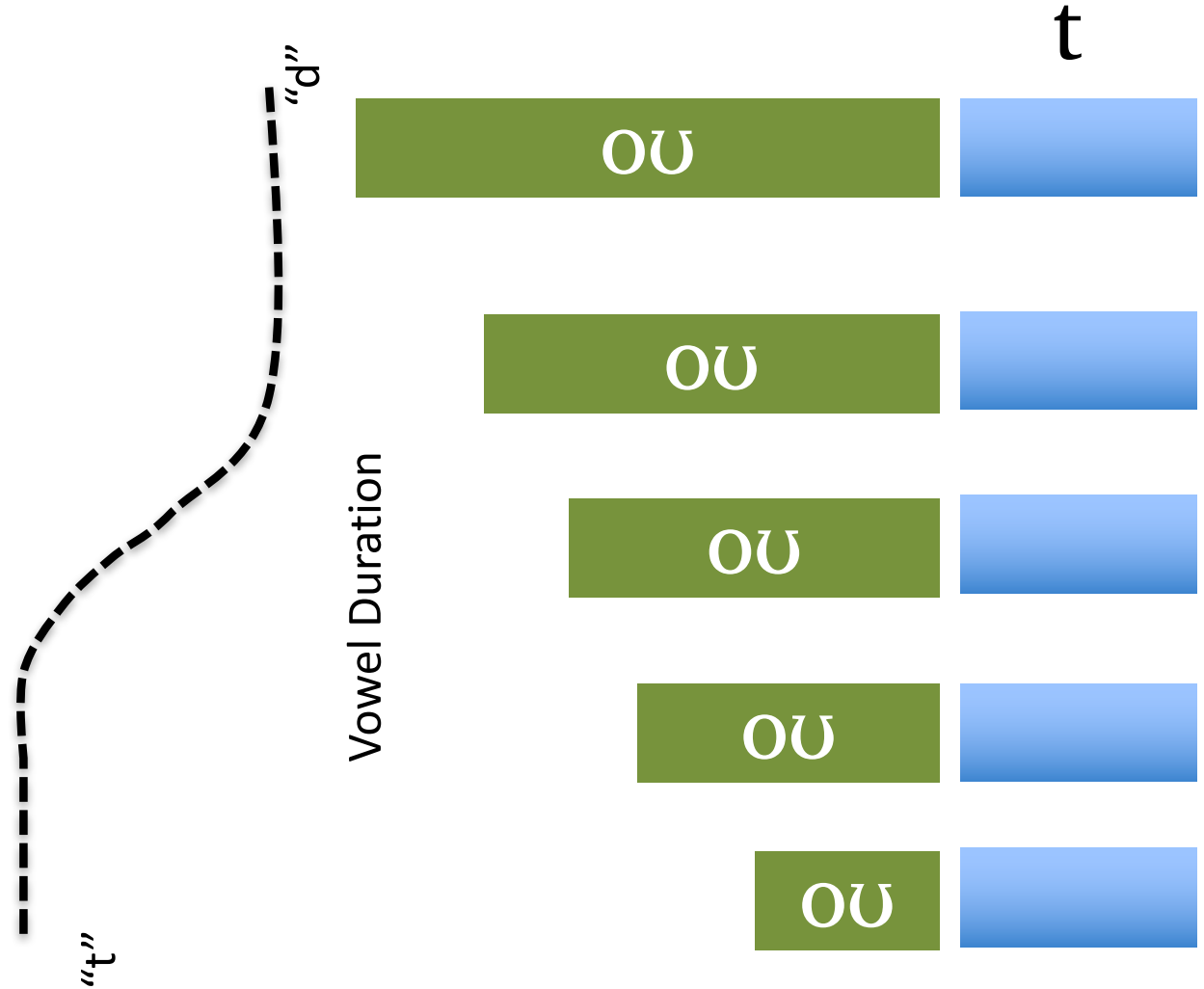
[t] = ambiguous token:
silence, followed by burst, transition period into vowel



Lisker & Abramson (1970); Summerfield (1981); Green & Miller (1985);

Final Stops

[t] = ambiguous token:
silence, followed by burst



House and Fairbanks (1953); Denes (1955,1972); Peterson & Lehiste (1960); Delattre (1964); Chen (1970); Raphael (1972); Umeda (1975); Klatt (1976); Mack (1982); Ohala (1983); Luce & Charles-Luce (1985); Van Summers (1987); Kluender et al. (1988); Fischer & Ohde (1990); de Jong (1991); Crowther and Mann (1992); Laeufer (1992); Smith (2002); Abdelli-Beruh (2004)

The Chicken
&
The Egg

Speaking Rate

- We know that the perceptual boundary in these experiments is affected by changes in speaking rate

  Ainsworth (1971); Umeda (1975); Summerfield (1981); Fitch (1981); Crystal & House (1982); Port & Dalby (1982); Miller et al. (1986); Miller & Volaitis (1992); Volaitis & Miller (1992); Wayland et al. (1993)

- It also seems to be the case that the vowel of the target word itself is the strongest cue to speaking rate

  Crystal & House (1982); Summerfield (1981); Port & Dalby (1982)

- And vowels seems to be affected by speaking rate changes more than consonants

  Gay (1982)

- Duration changes in consonants, with fixed vowel duration can be interpretted as intrinsic duration changes

- But how can changing vowel duration be both an intrinsic duration cue, as well as an extrinsic cue to speaking rate?

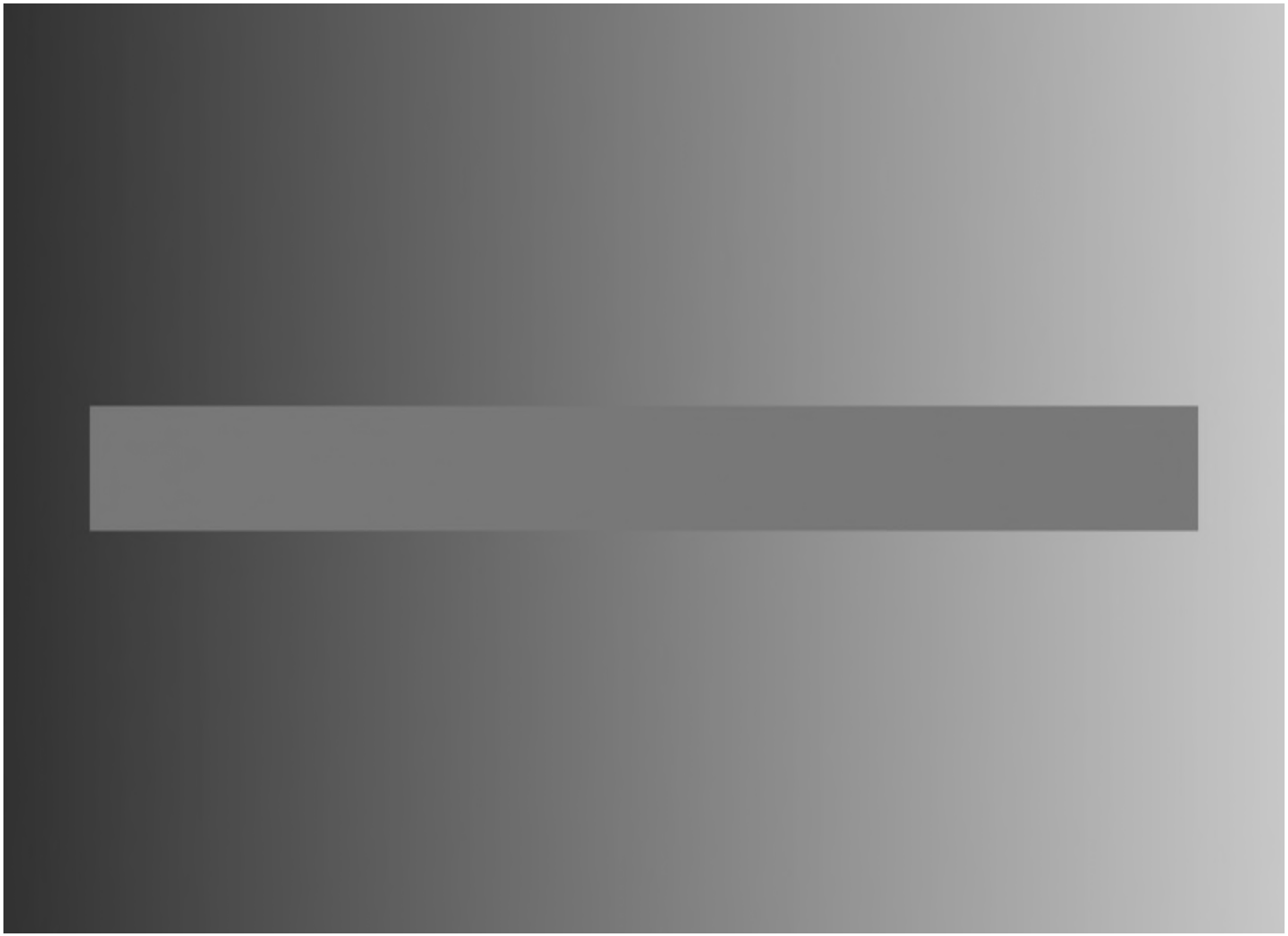
Speaking Rate

- How do we measure speaking rate?
  - Number of syllables per
    - Inter-pause interval containing target
    - 1 sec preceding/following target
  - Difference in word length from the sum of average phoneme durations over corpus
- But how rapidly does speaking rate change?
  - Miller et al (1984): 29 speakers: average syllable duration over a stretch of pause-free speech varied how long? by as much as 100 ms, for 20 speakers, by as much as 300 ms.

Locality & Predictivity

Perceiving/Recognizing

- Aristotle: hot vs. cold
- Color constancy
  - Blue in dim light vs. blue in bright light
  - Gray next to white vs. gray next to black
- Vection: the train next to yours starts moving away from you and you feel like you're moving backwards



Planning

- Motor acts
  - Grasping handshapes at beginning of reaching
  - Fluent reading: saccades forward and back
  - Reading/speaking lookahead: subtitution errors
  - Speech articulation:
    - All kinds of regressive assimilation:
      - Nasalization
      - Palatalization
      - Backing/fronting
      - Rounding
      - ….

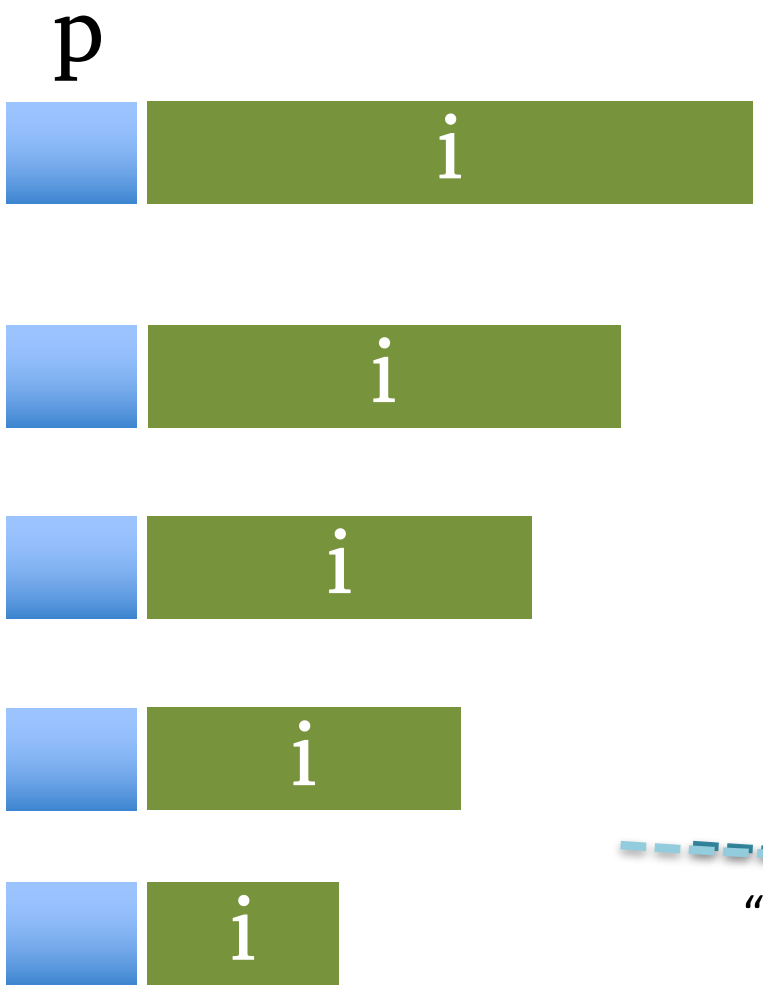
Integration/Analysis

- Speech processing is a lot like:
  - General auditory procesing
  - Even general perceptual processing
- Cues can only be associated with events that are temporally/spatially proximal
- But integration occurs both forwards and backwards in space and time
- General parsing/segmentation problem
  - Which cues/features go with which object/unit
  - Is this object a quiet/bright/short token of X
  - Or a loud/dim/long token of Y

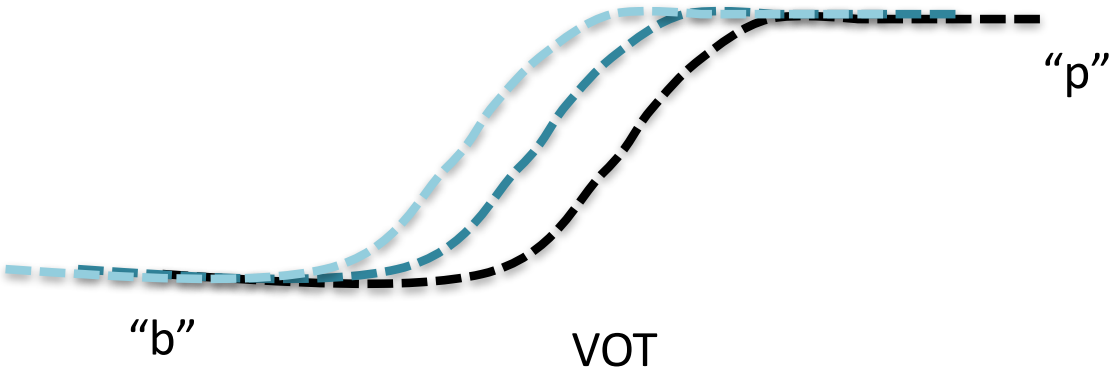
Duration-Based Contrasts Revisited

Initial Stops

For fixed VOT



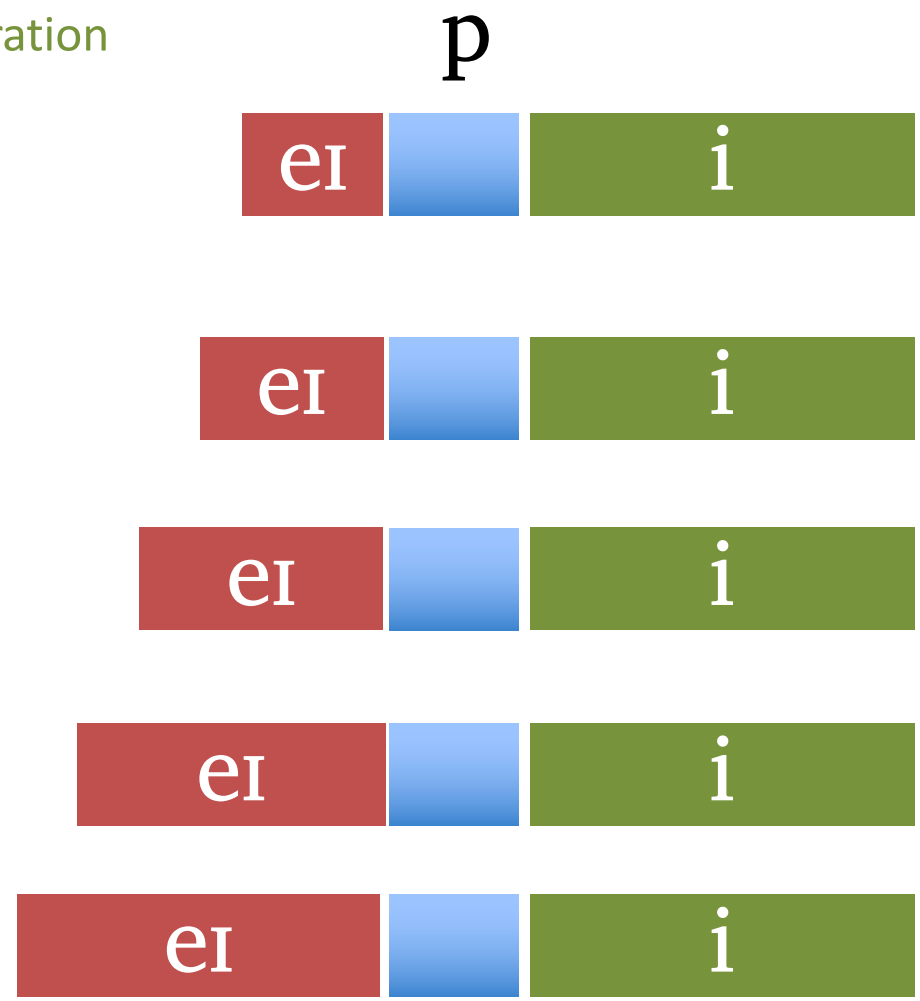
Shortening the following vowel duration shifts the perceptual boundary towards shorter VOT



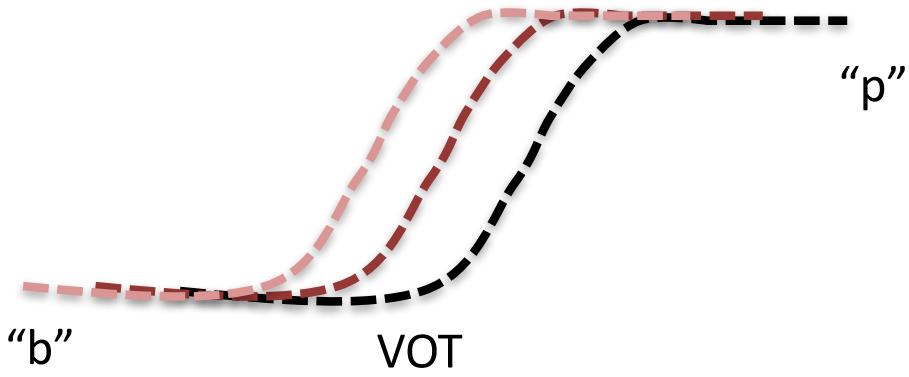
Summerfield (1981); Miller & Volaitis (1982)

Initial Stops

For fixed VOT, and Vowel duration



Shortening the preceding vowel duration shifts the perceptual boundary towards shorter VOT

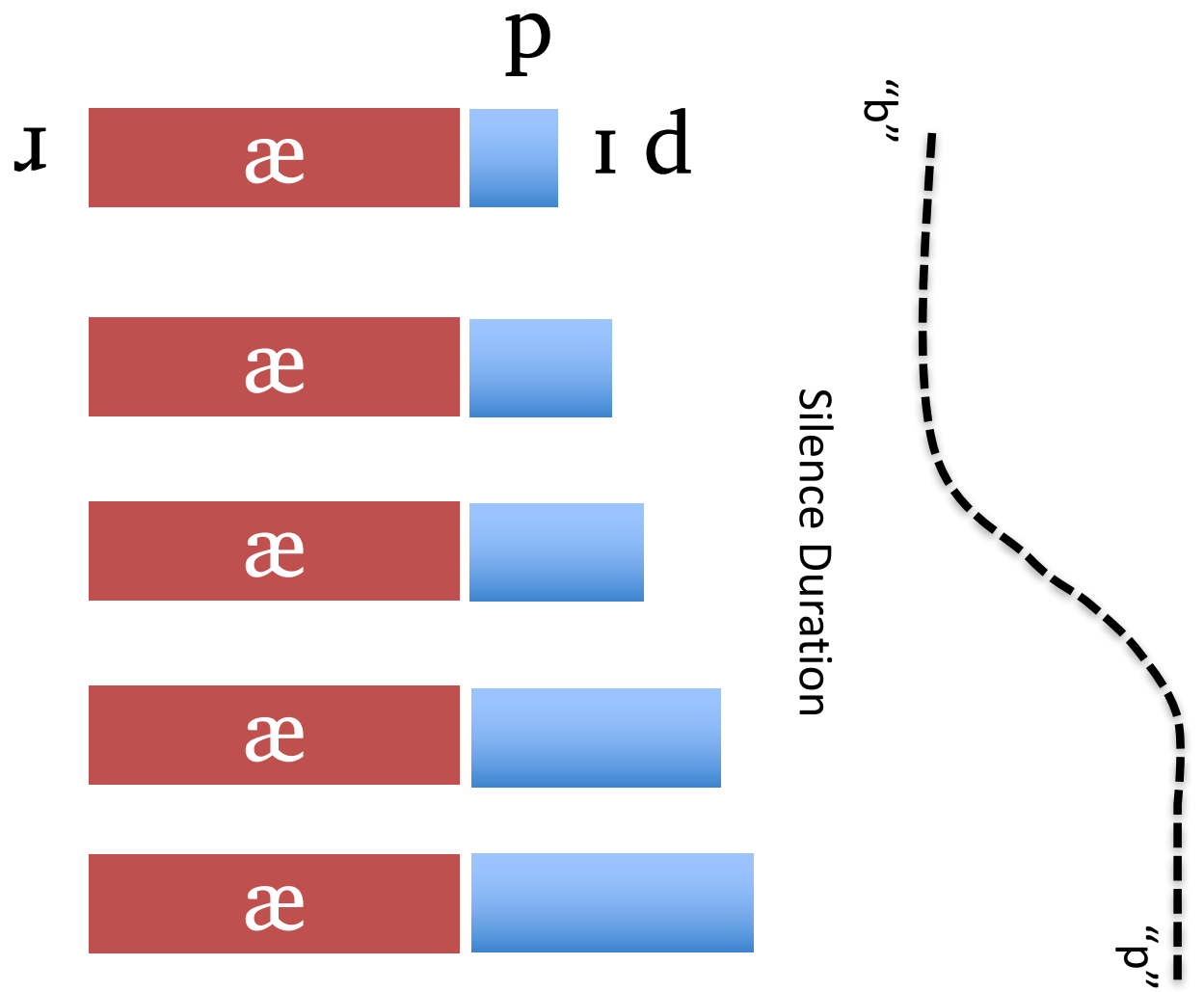


Summerfield (1981); Miller & Volaitis (1982)

Medial Stops

Fixed vowel duration

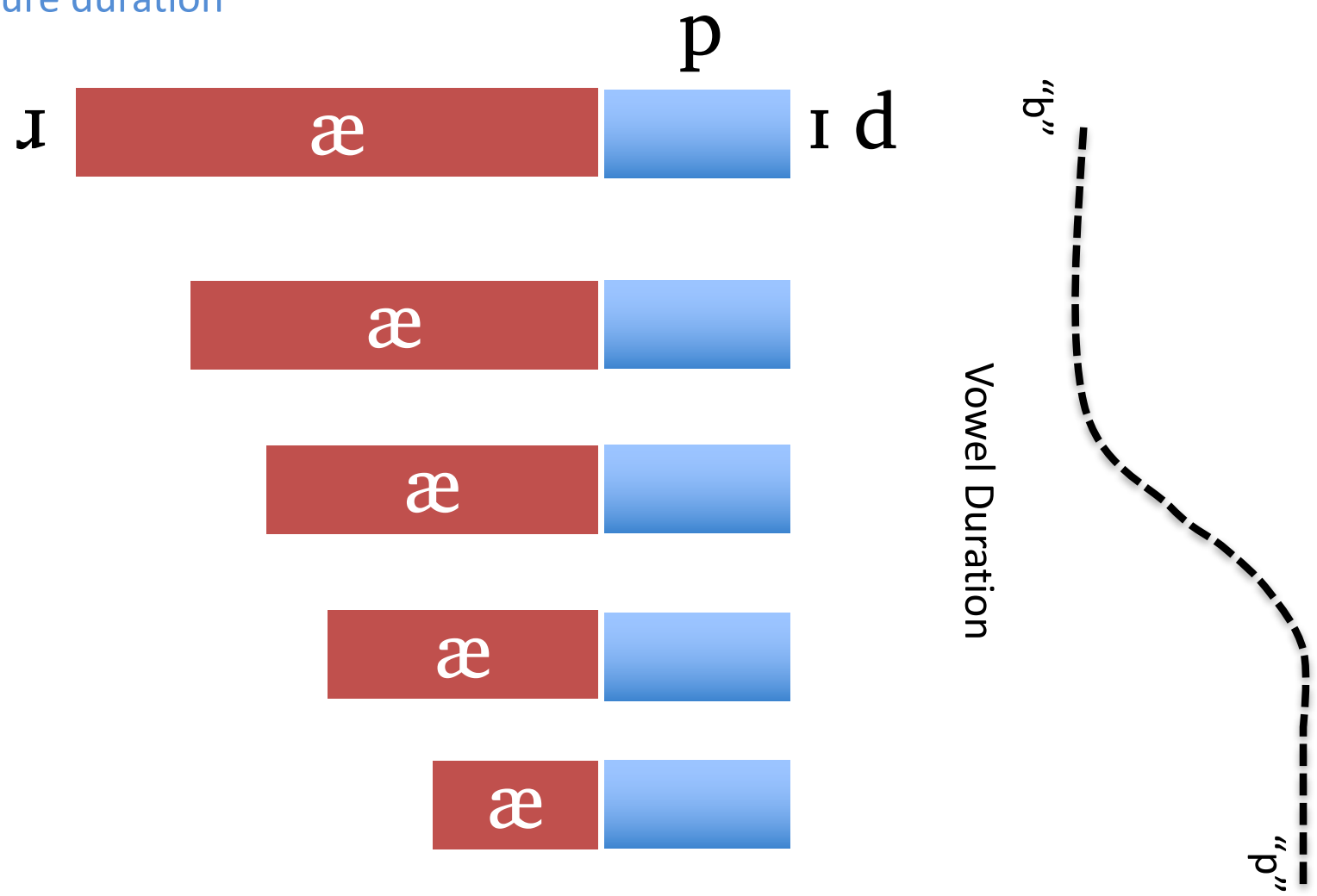
[p] = ambiguous token, silence duration + transitions



Lisker (1957,1978); Port (1979); Fitch (1981); Port and Dalby (1982)

Medial Stops

Fixed closure duration



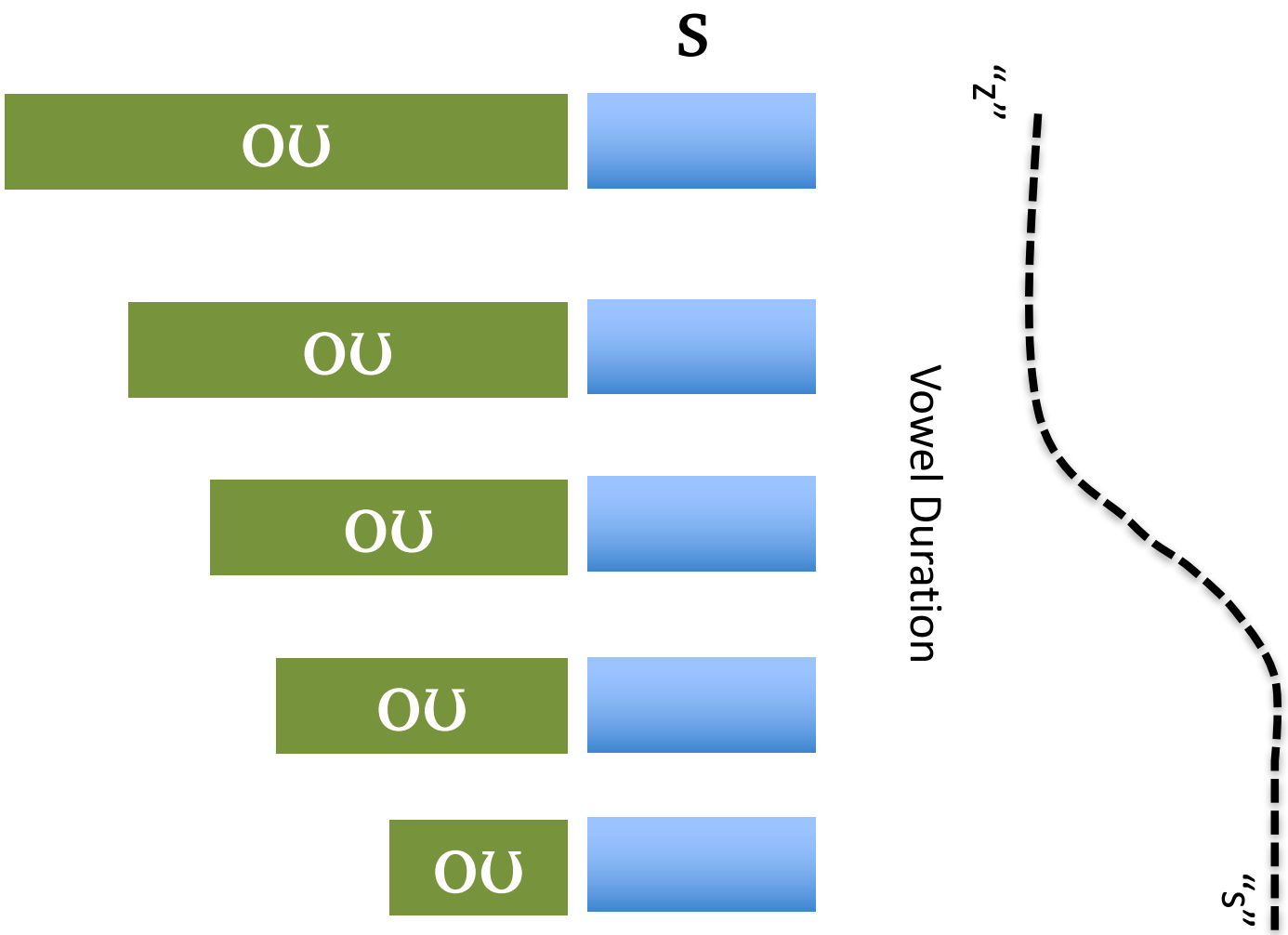
Lisker (1957,1978); Port (1979); Fitch (1981); Port and Dalby (1982)

Final Stops…

Final s

Fixed noise duration

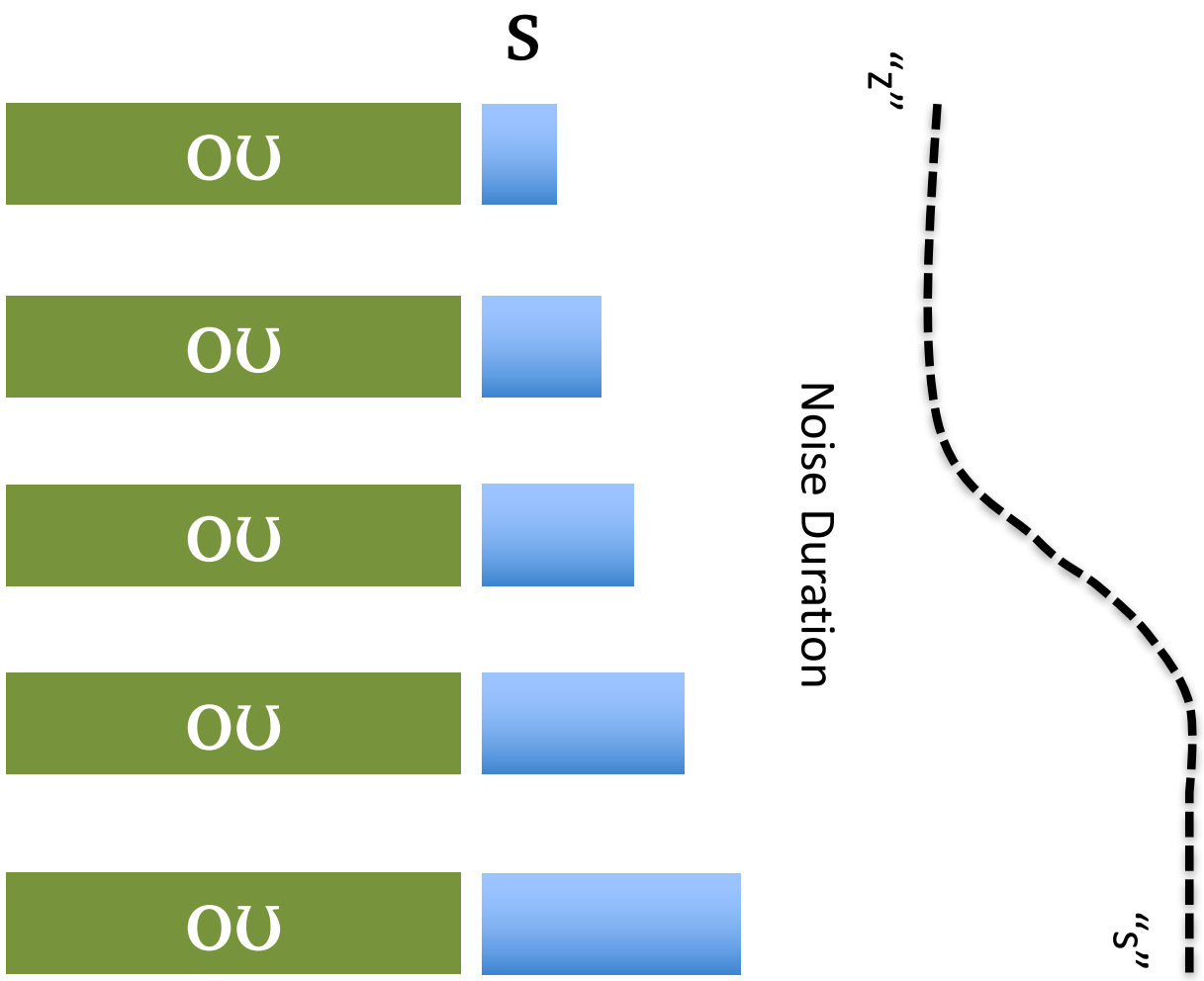
[s]=ambiguous segment, voiceless noise duration



(Denes 1955)

Final s

Fixed vowel duration



(Denes 1955)

The Literature

- Initial stops:
  - VOT is sufficient cue to contrast
  - Shorter preceding vowel moves boundary shorter (more “p”)
  - Shorter following vowel moves boundary shorter (more “p”)
- Medial stops:
  - Closure/silence duration is sufficient cue to contrast
  - Preceding vowel duration is sufficient cue to contrast
- Final stops:
  - Vowels are lengthened before voiced stops (and fricatives)
  - Thus, vowel duration has become a sufficient cue to contrast

Vowel Lengthening

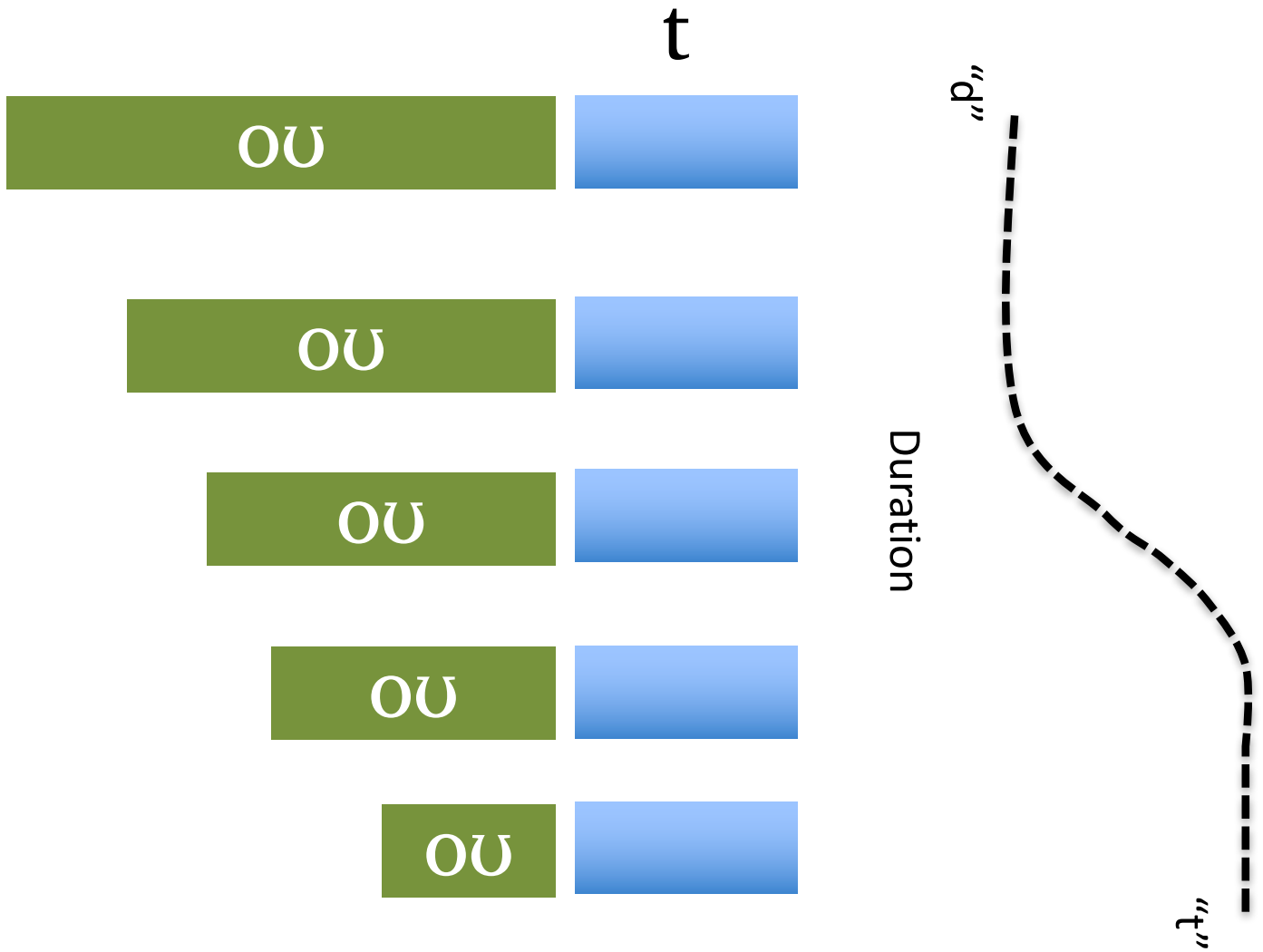
"Vowel Lengthening"

- Word-final stops are primarily cued by preceding vowel duration (e.g., Raphael 1972)
- Phonetic reasons?
  - "probably a result of the natural tendency to make a slightly early glottal opening closure for a postvocalic voiceless consonant in order to insure that no low-frequency voicing cue is generated during the obstruent" Klatt (1976)
  - Conservation of articulatory energy: more energy for voiceless segment = less energy for vowel (Belasco 1953)
  - Laryngeal Adjustment: glottal opening must be widened to maintain non-spontaneous voicing, leading to longer transition time, leading to a longer vowel (Chomsky & Halle 1968)
  - Rate of Closure Duration: lip closure movement is faster in voiceless stops, therefore the preceding vowel has less time (Chen 1970)
  - (Partial) Temporal Compensation: constant syllable duration: longer C closure leads to shorter V duration (Chen 1970; Kozhevnikov & Chistovich 1967)
- Perceptual reasons?
  - Perceptual Enhancement: longer vowel makes short consonants appear even shorter (Kluender 1988)
  - Misparsing? "the continuation of voicing into the consonant causes the perception of greater vowel length (eventually leading to the production of greater vowel length)" (Javkin 1976)

“Vowel Lengthening”

Preceding vowel duration is sufficient cue to voicing contrast

Very consistent effect with boundary around 200ms duration

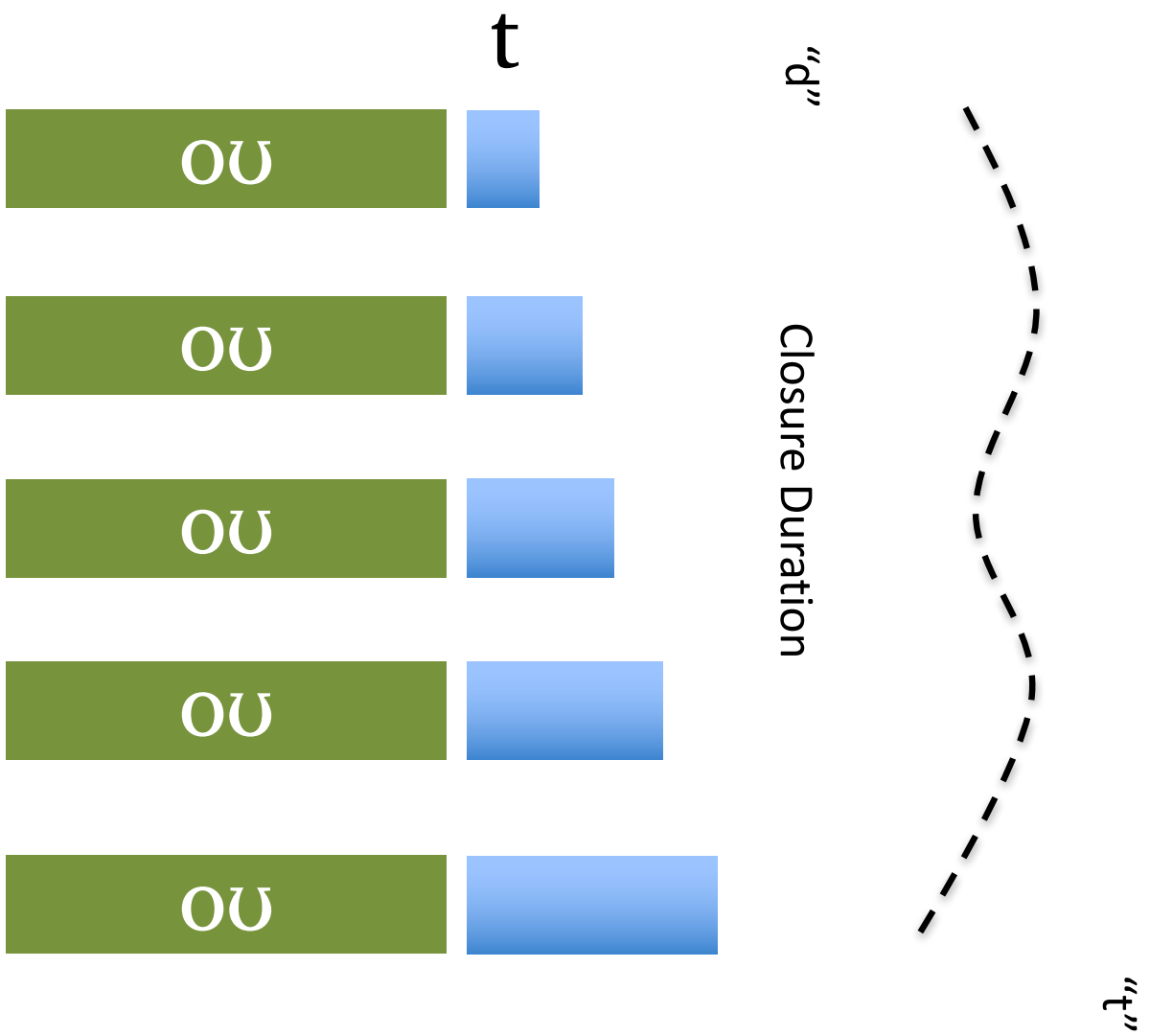


“Vowel Lengthening”

Don’t always see much or any effect of coda duration in obstruents other than s.

Final stops in English are often unreleased, and thus have only weak cues to duration.

Listeners may ignore final release cues if they are not expecting any recoverable information from the coda



Chicken & Egg

- So this is really the beginning of the story
- How can the vowel duration by itself be both the proxy for speaking rate AND the cue to stop voicing??
- Or both the proxy for speaking rate AND the cue to the tense/lax distinction?
- Also, it seems reasonable to combine the shared set of effects under a single explanation

All stops (and fricatives)

1) Voiced obstruents tend to be shorter than voiceless
   Klatt 1976; Lisker 1957; Port 1976

Perceived obstruent duration is affected by:

- Duration of obstruent
- Duration of preceding vowel
- Duration of following vowel

Perceived duration is sufficient cue to voicing in (final) stops

- When there is a following word, effect of stop duration is found
- This is also true to a certain extent in phrase-final (pre-silence?) position when vowel durations are kept constant throughout the experiment

- Changes in the length of the preceding word/vowel affect the location of the category boundary

(Morley, Smith, Kohnlein, Kim. LabPhon 2018)

Local Comparisons

oʊ is long relative to eɪ

t is short relative to oʊ

oʊ is now percieved as shorter

And thus t is percieved as longer

- Same effects for b/w continua with preceding and voweling vowels (Miller & Liberman 1979)
- Short preceding vowels can change the percept of synthetic vowel identity from (ɪ,ɛ,ʊ) to (i,e,u) Ainsworth (1973)

All stops (and fricatives)

1) Voiced obstruents tend to be shorter than voiceless
   Klatt 1976; Lisker 1957; Port 1976

Perceived obstruent duration is affected by:

- Duration of obstruent
- Duration of preceding vowel
- Duration of following vowel

2) Vowels tend to be shorter the longer the coda consonant (and vowels in open syllables are longest)
   Kavitskaya 2002; codas tend to be shorter, the longer the vowel Elert (1964); Arnason (1980); Kristoffersen (2002)

Local Expectations

This seems to me just one type of the phonetic expectations that we know listeners develop about their native language:

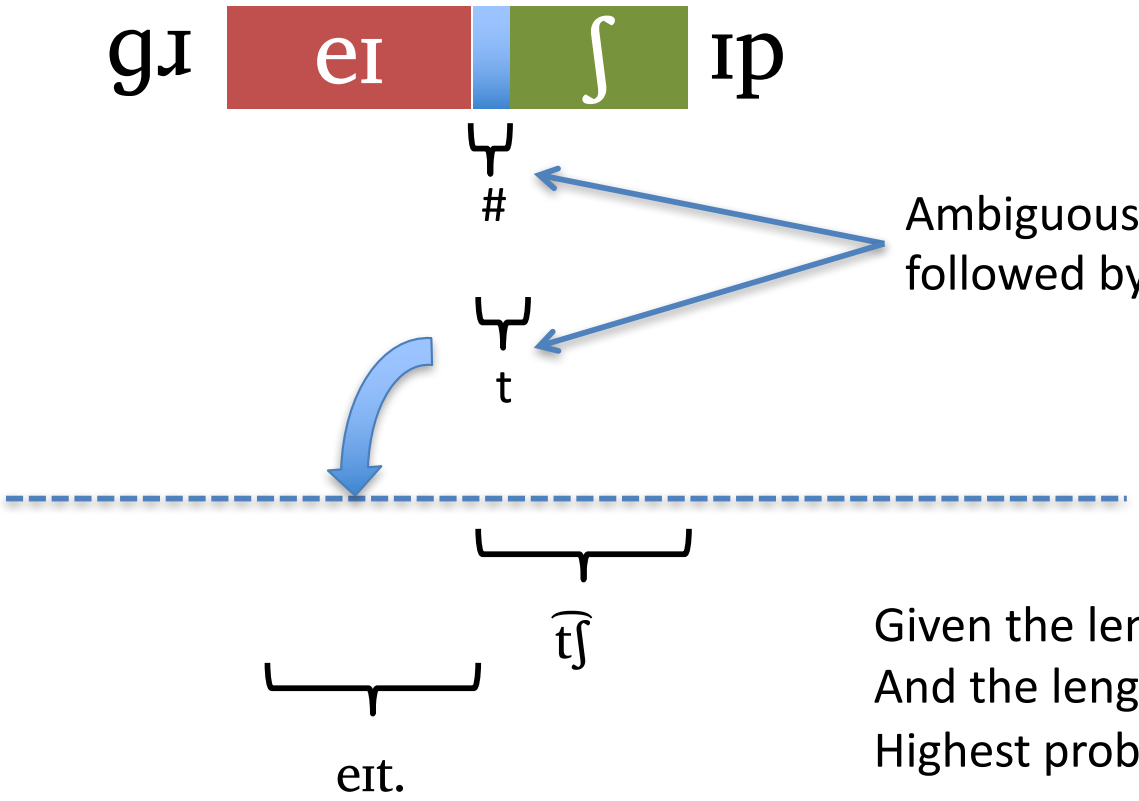
- Vowels are nasalized before nasal consonants
  - Listeners can predict a following nasal; notice when there is a mismatch Lahiri & Marslen-Wilson (1991); Beddor & Krakow (1999)
- n is assimilated to the place of following stops
  - Listeners can predict the place of an upcoming stop Gow (2003)
- As vowel duration increases, the probability of a coda consonant in upper range decreases


lamb/lab


bid/bit

Underlying Representations
&
The Optimal Parse

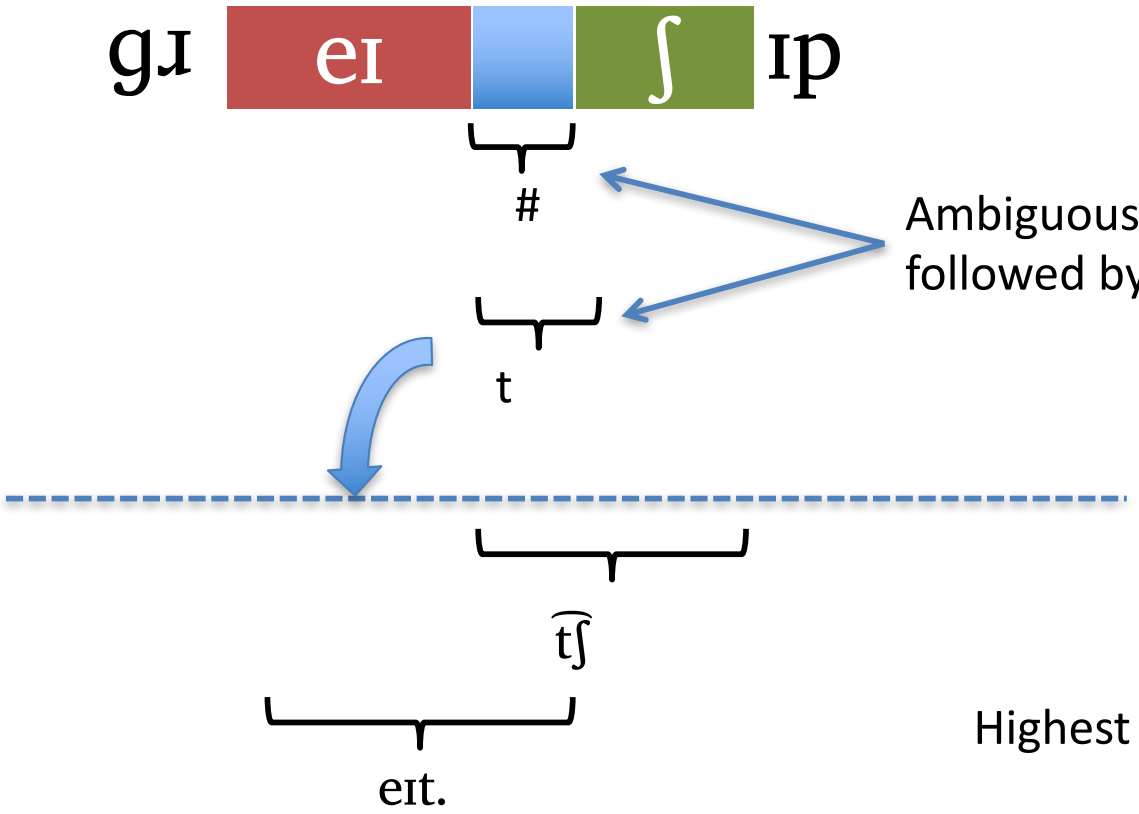
Trading Relations Revisited



Given the length of the silence relative to eɪ
And the length of t relative to ʃ
Highest probability parse is:

"gray ship"

Trading Relations Revisited



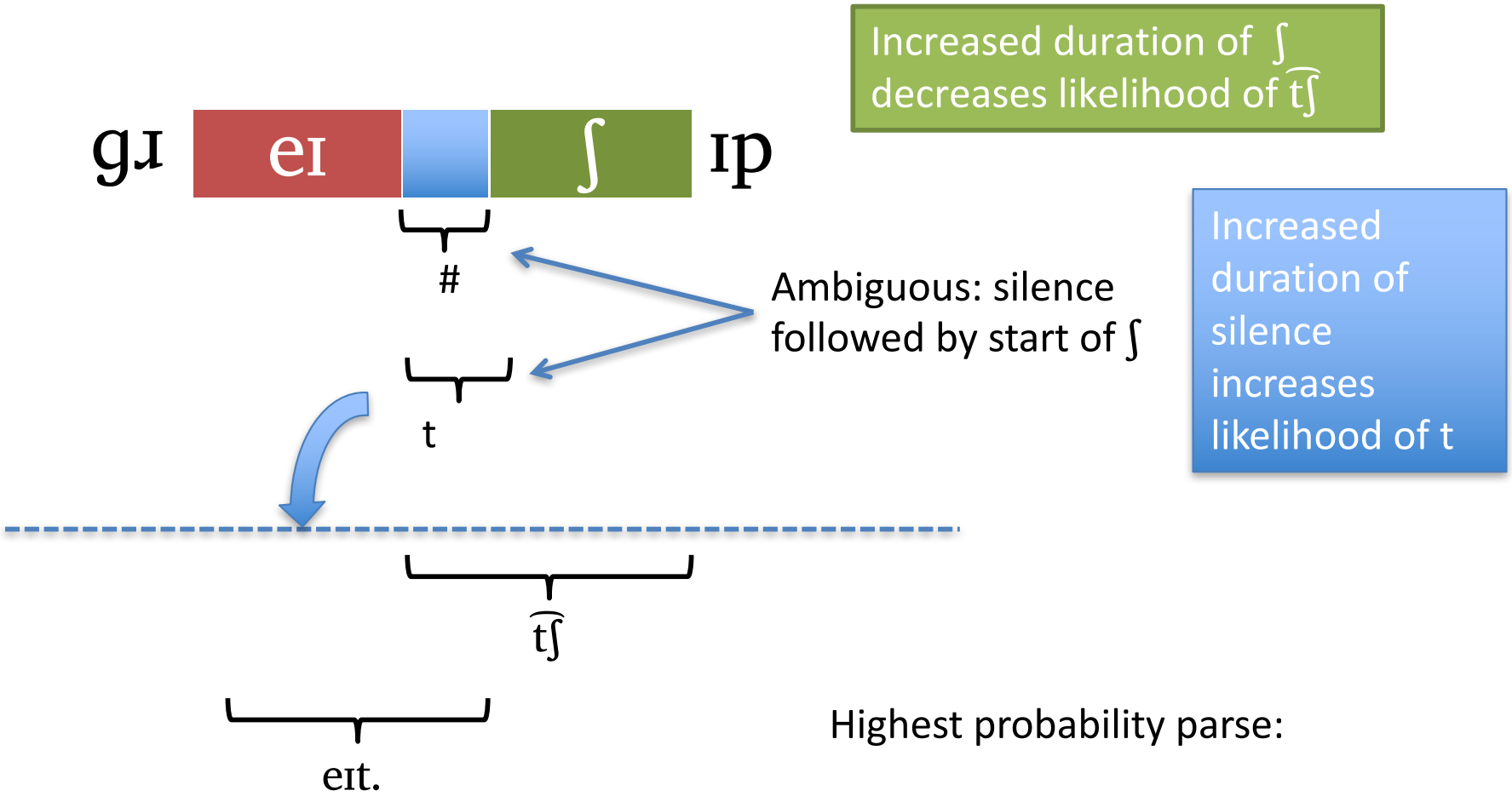
Ambiguous: silence followed by start of ʃ

Increased duration of silence increases likelihood of t

Highest probability parse:

"gray chip"

Trading Relations Revisited



My Attempt at a Coherent Story

Take: 52,567,002

“gray”

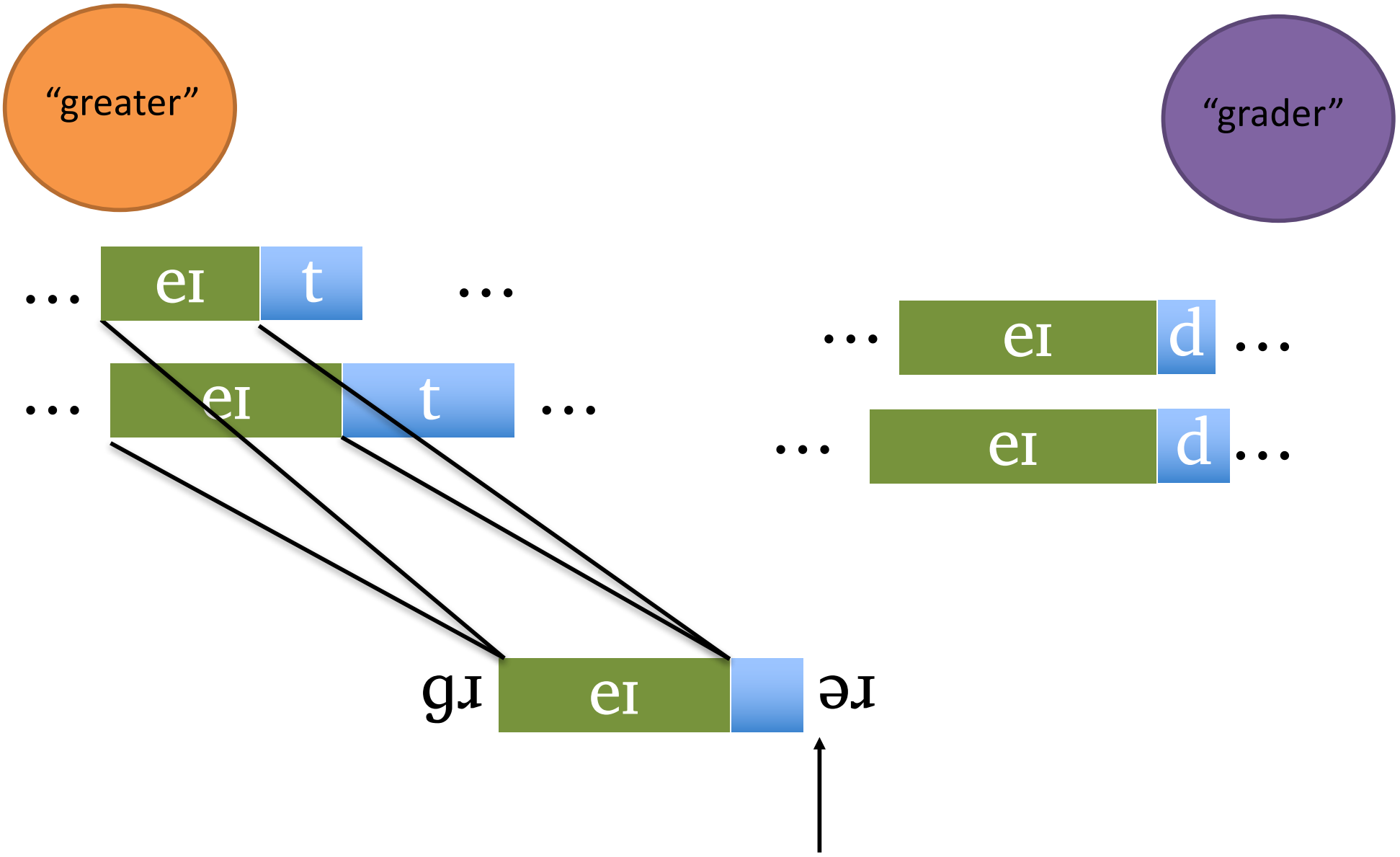
“great”

“greater”

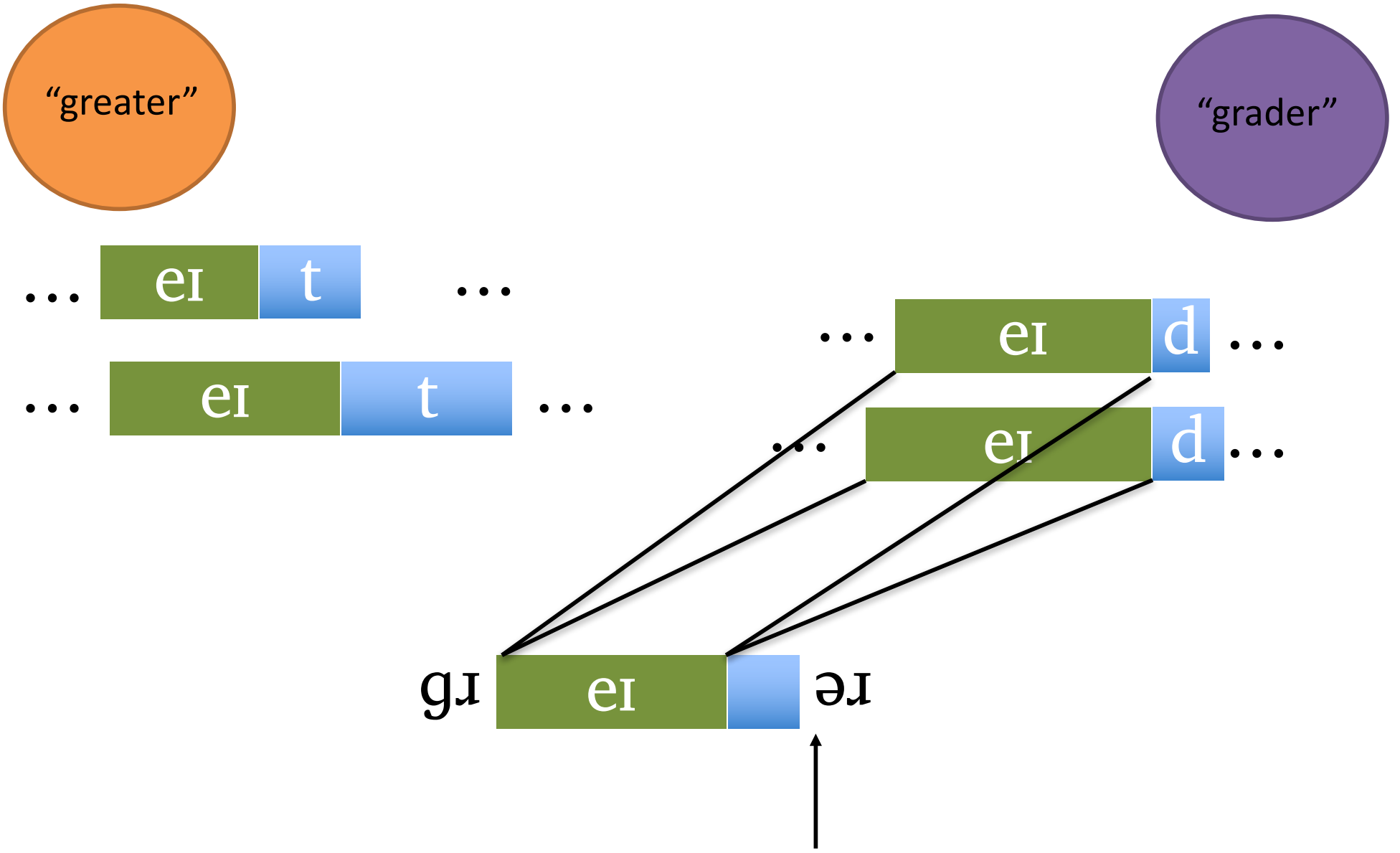
“grader”

“grade”

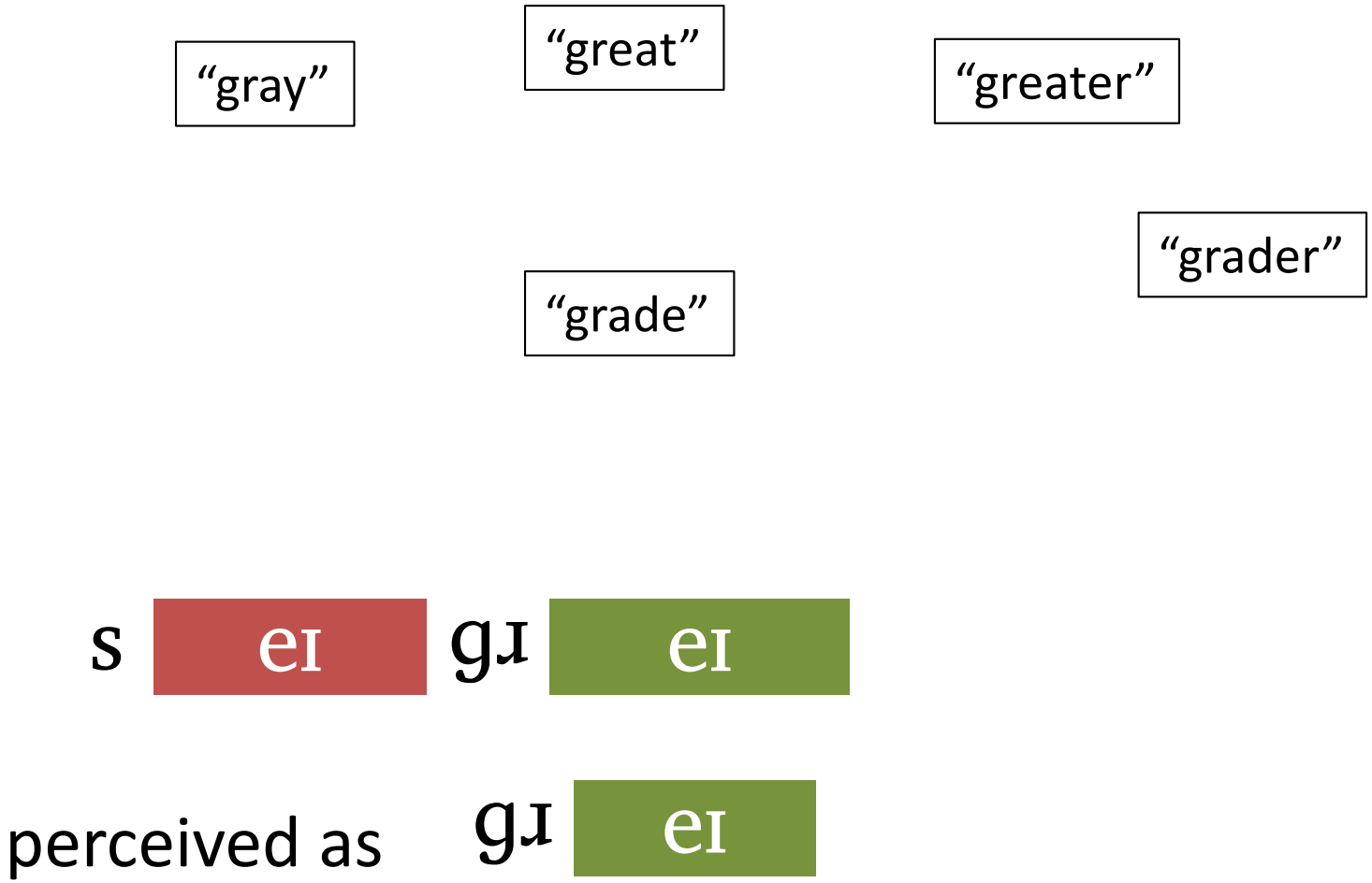
“greater”
“grader”
… eɪ t …
… eɪ t …
… eɪ d …
… eɪ d …
ɡɹ eɪ ɹə



Some kind of minimum edit distance type alignment for similarity computation, or online derivation of boundary ratio



Some kind of minimum edit distance type alignment for similarity computation, or online derivation of boundary ratio


“gray”
“great”
“greater”
“grader”
“grade”
s
eɪ
gɹ
eɪ
perceived as
gɹ
eɪ

“greater”
“grader”
… eɪ t …
… eɪ t …
… eɪ d …
… eɪ d …
gɹ eɪ ɹə

‘Voicing’ Contrast in coda?

So, under what conditions could we say that contrast had shifted to (primarily) the preceding vowel duration?

- obstruent duration shows little to no effect on categorization
- Even phrase-medially

Prediction:
This won’t happen unless the difference in obstruent duration itself is lost (i.e., ceases to have predictive power). Minimally because

- Medial silence/closure must be parsed
- Perceived duration will affect perceived duration of neighboring sounds
- Whether obstruent is parsed into coda or onset will affect perception of vowel

Thank You!

And good night.