

Using Pronunciation Data to Constrain Models of Spoken Word Recognition

Laura C. Dilley,¹ Mark A. Pitt,¹ & Keith
Johnson²

¹Ohio State University

²Univ. of California, Berkeley

Overview

- Spoken word recognition is remarkably successful in spite of variation in pronunciation.
- In *regressive place assimilation*, the variation can be extreme enough to yield a phonologically significant change (*green ball* → *greem ball*).
- **Question:** How do listeners recognize the intended word, *green* given *greem*?

Background

- Experimental work has addressed this issue by testing theoretical proposals that explain how assimilated variants are recognized
- Process-based: Gaskell & Marslen-Wilson 1996, 1998; Gow 2003
- Representation-based: Johnson, 1997; Lahiri, 1999

Present study

Obtaining a better understanding of assimilation and other types of pronunciation variation provides a check on theorizing and aids in evaluating the adequacy of theoretical accounts.

QUESTION: What types of variation do talkers produce in contexts where regressive place assimilation *could* potentially occur?

Approach

Phonological and *acoustic-phonetic* variation were studied in a large corpus of conversational speech.

1) *Phonological* analyses provide information about perceived category membership of a segment (e.g., $n \rightarrow m$).

2) *Acoustic-phonetic* analyses reveal the degree of physical evidence supporting category membership.

Materials and talkers

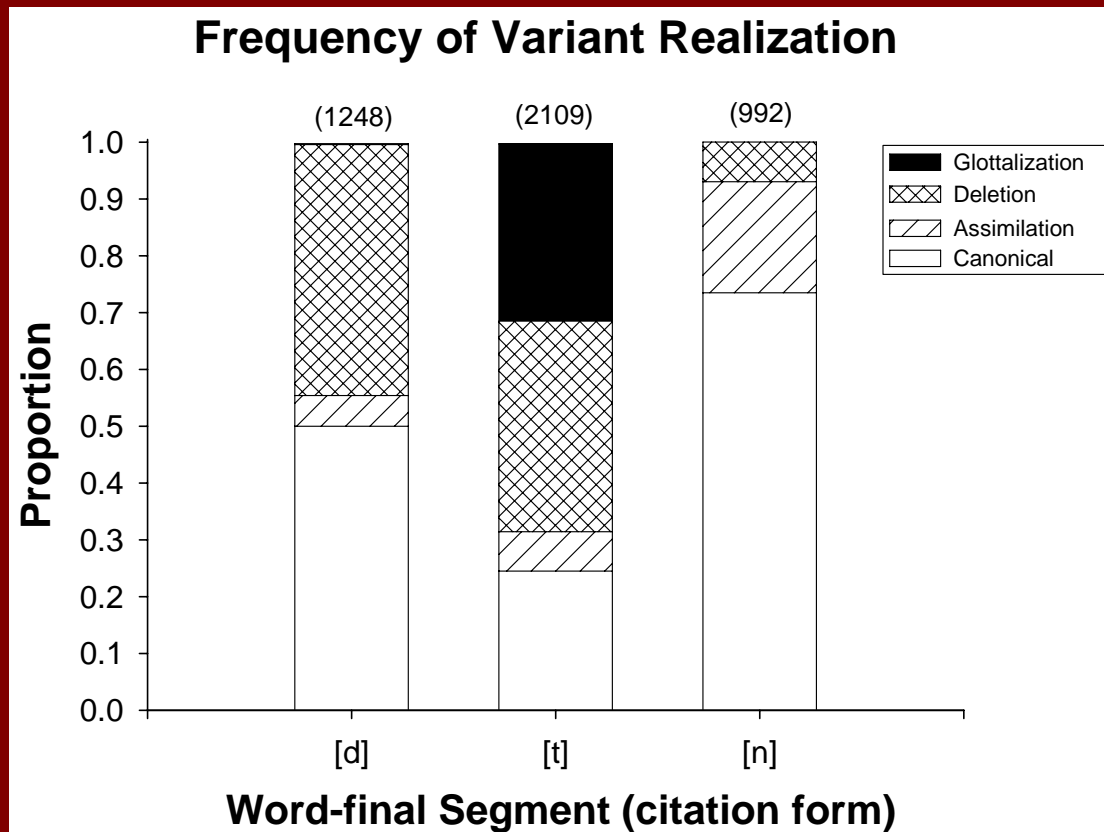
- We used the Buckeye Corpus of conversational American English (www.buckeyecorpus.osu.edu), in which talkers freely converse with an interviewer.
- 19 talkers (~19 hrs. of speech) from Columbus, OH.

Phonological analysis

- Labeling of speech was performed using auditory and spectrographic information.
- We examined *phonological* and *acoustic-phonetic* properties of word-final coronals /t/, /d/, and /n/ in *assimilable environments* (N = 4349):

$$\{t, d, n\}\# \left\{ \begin{array}{l} p, b, m \\ g, k \end{array} \right\}$$

Results: Phonological Analysis



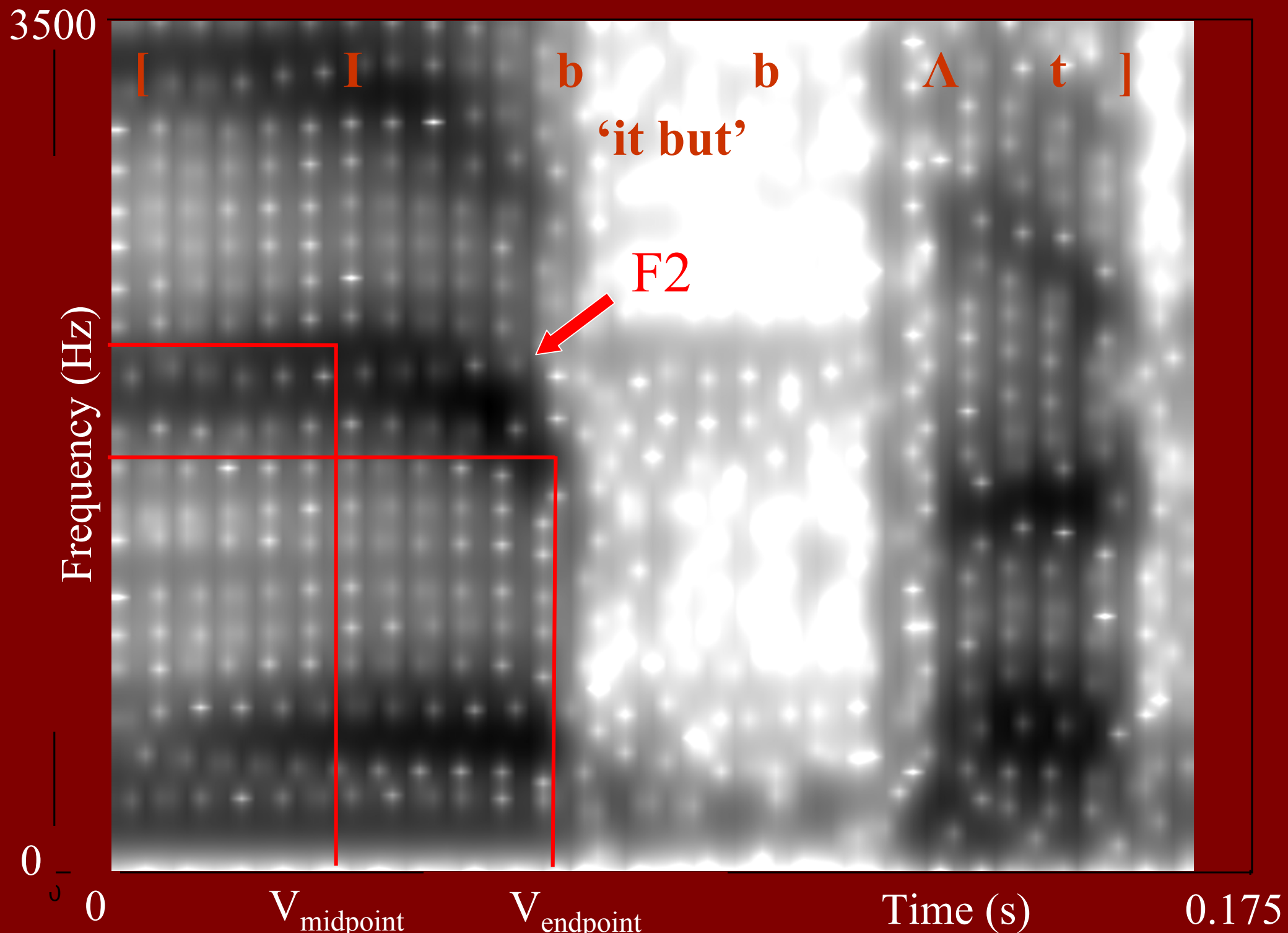
- Assimilation was the least frequent form of variation (9%).
- Deletion (32%) and glottalization (15%) were relatively more frequent.
- Canonical pronunciations predominated (43%).

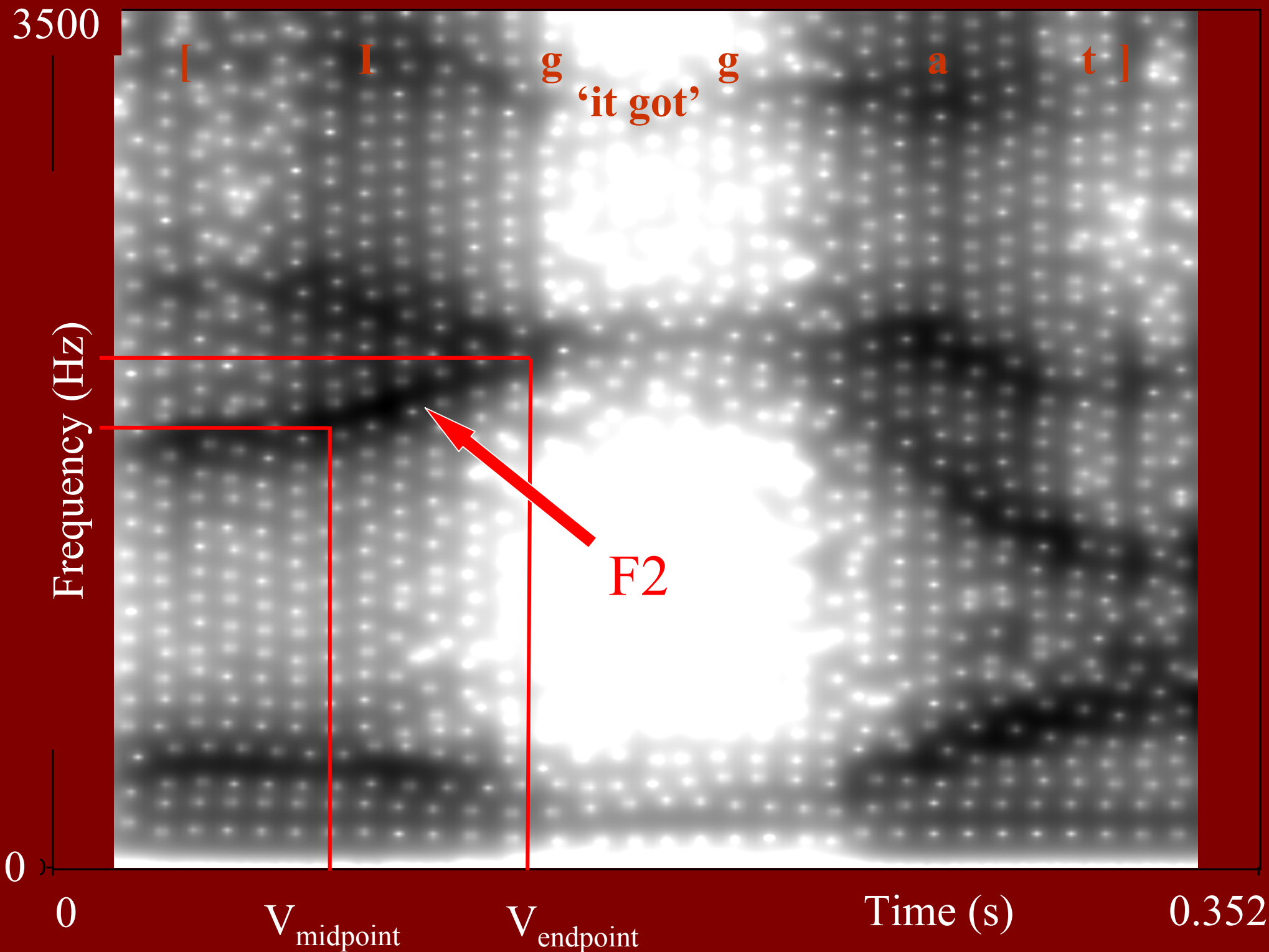
Discussion: Phonological Analysis

- These results suggest that even in a highly constrained environment (word-final coronals in assimilable contexts), variation takes many forms.
- What do acoustic-phonetic analyses reveal about variation in assimilable contexts?

Acoustic-phonetic analysis

- One measure of extent of place assimilation is the direction and extent of an F2 transition to a stop from a preceding segment
- This was determined by calculating the *F2 difference* at the midpoint and end of the vowel preceding the stop.

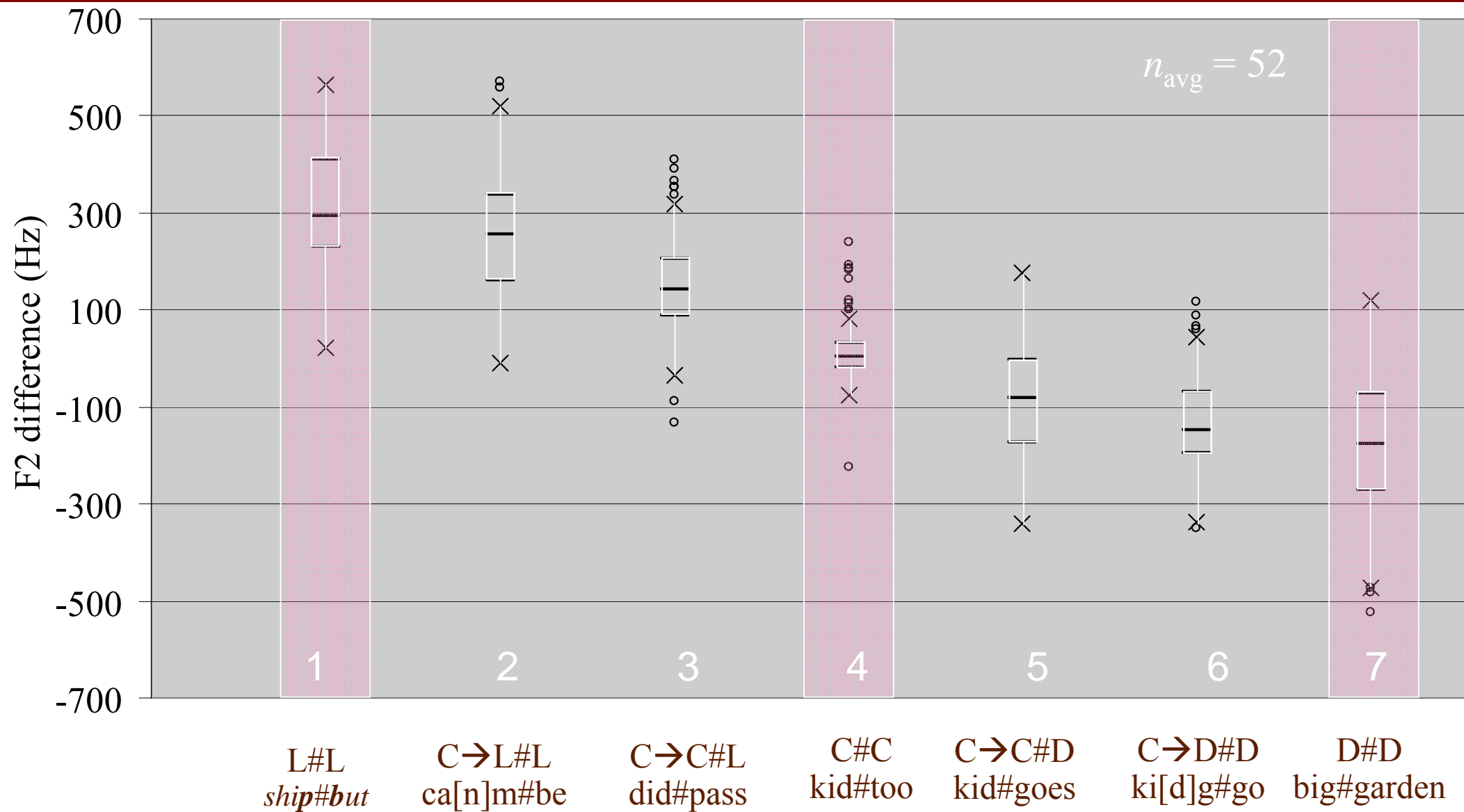




Method: Acoustic-phonetic analysis

- In order to evaluate the relative degree of place assimilation for coronals, the F2 difference for word-final coronals in assimilable environments were compared with word-final consonants in *control environments*:
 - {g, k, ng} # {g, k} (dorsal # dorsal)
 - {t, d, n} # {t, d, n} (coronal # coronal)
 - {p, b, m} # {p, b, m} (labial # labial)
- F2 transition measurements were made for two preceding vowel contexts:
 - high, front vowel ([I])
 - non-high, front vowel ([æ] or [ε])

[I] context

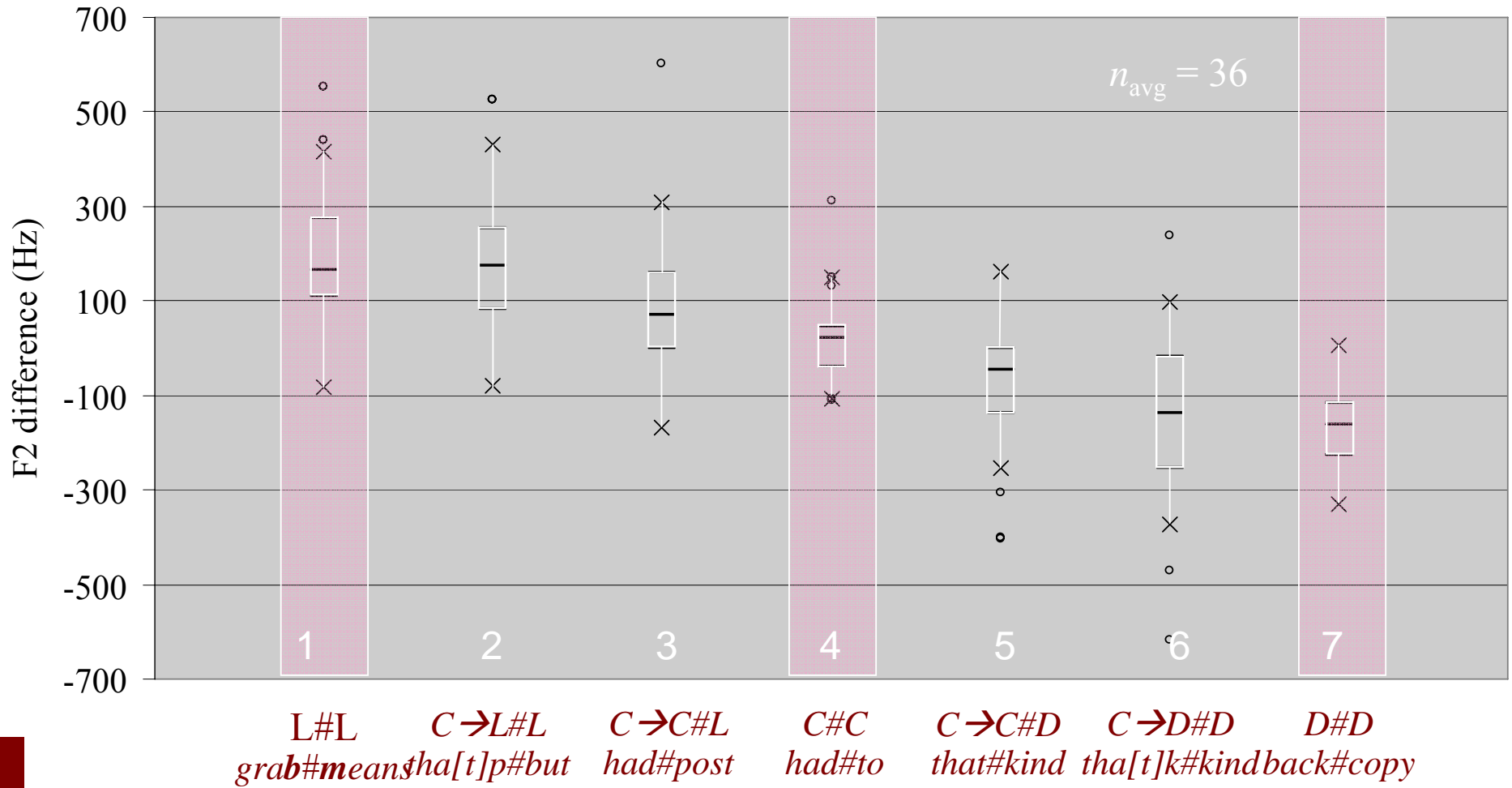


Results: [l] context

- Coronals assimilate almost fully in both following labial (#L) and dorsal (#D) control environments.
- The extent of assimilation for coronals that do not assimilate falls in between that of assimilated coronals and the coronal (#C) control context.

[æ] and [ɛ] context

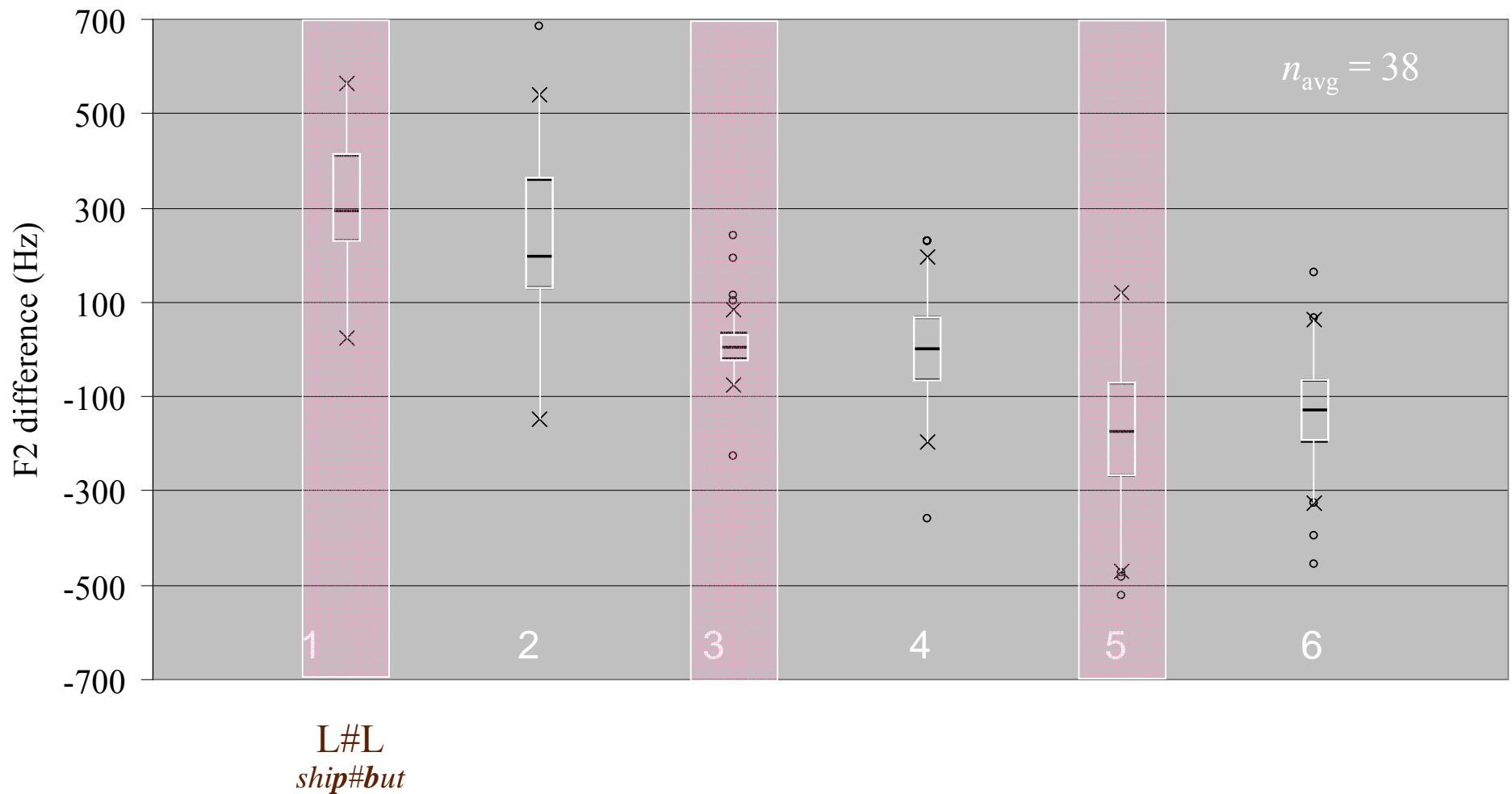
[æ] and [ɛ] context



Results: [æ] and [ɛ] context

- Coronals assimilate to a similar degree in this vowel context.
- Assimilation was similarly graded across contexts, with the median progressively decreasing from labial to dorsal place of articulation.
- **Is there similarly strong acoustic evidence of segment identity when segments were deleted?**

Deletion: [ɪ] context



Results: Deleted cases

- F2 differences for deleted cases in both the labial and dorsal contexts are slightly less extreme than control contexts.
- Even when the segment is imperceptible, remnants of the segment remain.
- The divergence in F2 is comparable whether the coronal is deleted or assimilated.

Summary and conclusions

- Pronunciation variation is complex, even in environments where only assimilation might be assumed to occur.
 - More extreme types of variation, such as deletion and glottalization, are common, while assimilation is relatively rare.
- The extent of place assimilation varies continuously across contexts, for both assimilable and deleted cases.
- Place assimilation is frequently complete.
- Theories of spoken word processing must overcome:
 - The graded nature of place assimilation.
 - The many forms of variation that occur in assimilable contexts.