Title

The Nature of Phonetic Disassociation from Lexical Neighbors

Permalink

https://escholarship.org/uc/item/1f59k0rf

Author

Lefkowitz, Lee Michael

Publication Date

2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Nature of Phonetic Disassociation from Lexical Neighbors

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Arts

in Linguistics

by

Lee Michael Lefkowitz

Abstract

In recent decades, linguists have experimentally demonstrated that the phonetic realization of lexical items and of specific speech sounds within them can be influenced by purely lexical properties such as word frequency (Balota et al 1989; Bybee 1994), contextual predictability (Hawkins and Warren 1994; Lieberman 1963), and, most interestingly, the existence of many phonologically similar words in the lexicon, i.e. lexical neighbors (Wright 1998, 2004; Brown 2001; Scarborough 2004; Baese-Berk and Goldrick 2009).

Several specific phonetic correlates of these lexical factors have been established: the vowel space as a whole is expanded (Wright 1998, 2004); voiceless stops and voiced stops have a larger VOT difference (Goldinger & Summers 1989; Baese-Berk and Goldrick 2009), and the overall amount of coarticulation between local segments is increased (Brown 2001; Scarborough 2004).

The general mechanism that underlies these various effects, however, is not well understood. While they each have the end effect of aiding listener comprehension, and occur under almost precisely the same conditions where word recognition is expected to be more difficult (Luce 1986), there are at least two types of mechanism consistent with this result. The first—the "hyperarticulation" hypothesis—is that speakers diminish processes of reduction, producing

realizations of speech sounds which are *highly faithful*; since the phoneme inventory is generally dispersed, this indirectly facilitates word-recognition. The second—the "dissimilation" hypothesis—is that speakers directly facilitate word recognition by maximizing the perceptual distance between the target word and its lexical competitors, producing realizations of speech sounds which are *phonetically distant* from competing sounds.

An experiment was devised to distinguish between these two possibilities by using a phonetically medial sound: English /ɛ/, which has the potential for competition from, among other vowels, /æ/ and /ɪ/, which are phonetically similar to /ɛ/ and geometrically opposed in formant space. If the realization of words containing /ɛ/ is influenced not only by the existence of minimal pair neighbors, but by the location in phonetic space of the vowels in such neighbors, the second hypothesis will be strongly supported. The results of the experiment were inconclusive; while some data trended in a direction consistent with the dissimilation hypothesis, no lexical neighborhood effects of any kind reached significance, despite a relatively large sample. This fact weakly supports the hyperarticulation hypothesis, at least with respect to vowels. However, the null result is potentially attributable to a number of factors.

The thesis of Lee Michael Lefkowitz is approved.

Patricia Keating

Sun-Ah Jun

Bruce Hayes, Committee Chair

University of California, Los Angeles

2012

iv

Table of Contents

List of Figures

List of Tables

Acknowledgements

Thanks to my committee members Pat Keating and Sun-Ah Jun, and especially to my advisor

Bruce Hayes for their invaluable advice, knowledge, encouragement, patience, and cool-

headedness.

Thanks to Robert Daland and Marc Garellek for explaining how to use R.

Thanks to Henry Tehrani for the MatLab code.

1   Background

The most traditional and widely used model of how speech sounds are realized can be broadly

summarized as follows: a word is constructed from lexemes in the speaker's lexicon, each

consisting of sequence of elements taken from the speaker's phoneme inventory. The word's

segments first undergo phonological processes determined by their morpho-phonological

environment, then undergo phonetic processes which are similarly determined by their

phonetic environment, and are at last articulated. Given perfect phonological and phonetic

models of a particular speaker's idiolect, then, the phonetic realization of a particular speech

sound in a particular word should be predictable, random variation notwithstanding, merely by

knowing the underlying form of that sound, and the relevant details about its morpho-

phonological environment.[1]

In recent decades, however, these basic assumptions about the factors involved in predictable

phonetic variation have been called into question.

---

[1] Prosodic position, which also impacts phonetic realization, is assumed here to be part of the phonological environment broadly defined, despite not being a lexical level phenomenon. However, to the extent that prosodic phrasing can reflect discourse-level information, it could interact with situationally conditioned variation (see next section).

1.1   Other Sources of Within-Speaker Phonetic Variation

Leaving aside variation observed between different idiolects (including those that can be attributed to the existence of multiple dialects within the same speaker), there exist at least two additional sources of predictable phonetic variation.

1.1.1   Situationally Conditioned Variation

Different tokens of a speech sound even within the same lexical item can vary from utterance to utterance based on speech style, and on the speaker's model of the listener. In particular, degree of *phonetic reduction* in speech is correlated with how well the listener is assumed to be able to understand the word or utterance in question. Decreased levels of reduction, and increased use of hyperarticulation, have been observed in the following contexts:

1) When speakers believe their interlocutors to be receiving a relatively weak linguistic signal, due to things like environmental noise (which results in the "Lombard Effect," Lane and Tranel 1971; Lombard 1911), or even perceived lack of visual access to the speaker's face (Anderson et al. 1997).

2) When words are discourse-new (Hawkins & Warren 1994) or otherwise relatively less predictable from syntactic / semantic context (Lieberman 1963; Aylett & Turk 2006).

3) When speakers believe their interlocutors are somehow deficient listeners, e.g. foreigners (Scarborough 2007), children (and, to a lesser extent, household pets) (Kim et al 2006, Burnham et al 2007), and speech-recognition software (Oviatt et al 1998; Soltau et al 2002).

Studies of this sort generally conclude or strongly imply that speakers are employing some sort of listener modeling during speech production, in which difficulty of interpretation is anticipated, and triggers hyperarticulation and/or lack of phonetic reduction.

1.1.2   Lexically Conditioned Variation

This paper, however, is concerned with yet a third source of variation, which has only more recently been investigated. Namely, the purely *lexical* properties of a word[2] being uttered, such as being a low-frequency item (Balota et al 1989; Bybee 1994), and having a large number of relatively high frequency minimal pairs in the lexicon (Wright 1997, 2004) have been demonstrated to influence the phonetic realization of the segments in specific lexical items. Furthermore, there is some evidence that even just having a minimal pair for a particular segment in a word can cause that segment in particular to be hyperarticulated (Baese-Berk and

---

[2] The question of whether it is lexical information about whole words, lexemes, individual morphemes, or some combination that is relevant here is an interesting question which is beyond the scope of this paper, and, incidentally, has never really been investigated. I'll refer to 'word' or 'lexical item' interchangeably, with the assumption that all the examples given, while not necessarily monomorphemic, are at least lexicalized and therefore may have lexical properties.

Goldrick 2009). In addition to reducing reduction and cuing hyperarticulation, competition from lexical neighbors also increases the degree of coarticulation within lexical items (Brown 2002; Scarborough 2004, 2010).

This discovery is striking: the phonetic values of what is assumed to be same underlying speech sound can *differ between lexical items*, even in the same phonological environment! The next section gives an overview of some of the research to date on this topic.

1.2   Lexical Neighborhood Effects: a Brief History

1.2.1   Neighborhood Density and Word Recognition

Luce (1986) was the first to investigate the possibility that words are processed in a context of similar-sounding words in the lexicon, with lexical access potentially being affected by the number of "lexical neighbors"—at this point, it was speech perception rather than production that was under investigation. Luce's definition of "lexical neighbor" included all words which could be derived from the target by adding, subtracting, or replacing one segment with another segment—in other words, all of a target's minimal pairs (p. 6). His calculation of neighborhood density took into account the number of lexical neighbors, and their average lexical frequency; the frequency of the target word was an independent factor.

Figure 1. Lexical neighborhood schematic (Brown 2001, p. 4). Target words are shaded, lexical frequency is represented by bar height.
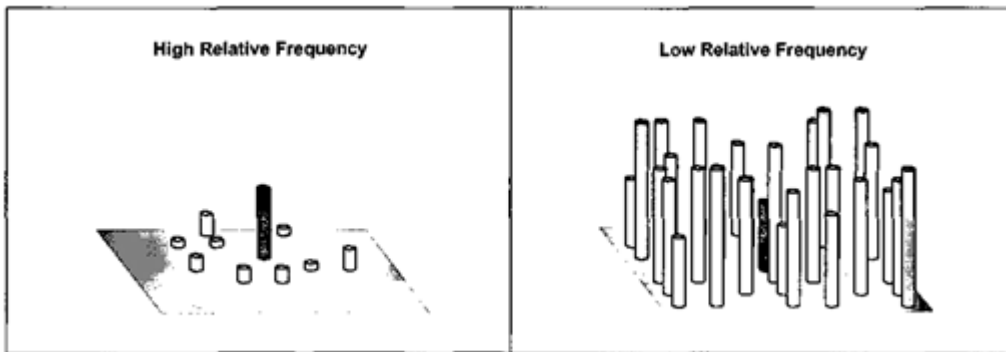
Words with relatively few, relatively infrequent lexical neighbors were termed "easy," while those with relatively many, relatively frequent neighbors were termed "hard" words. In this experiment, all words were monosyllabic, precluding the possibility of competition due to shared syllables, as might be present in longer words.

Luce found that several experimental paradigms saw differing results for hard words and easy words: hard words were less accurately identified in noisy environments (Ch. 2), in lexical decision tasks (Ch. 4), and auditory word-naming tasks (Ch. 5). Based on these results, Luce introduced a model called Neighborhood Activation Model (NAM), in which entire neighborhoods of words are activated during lexical access, with increased competition resulting in inhibited ability to quickly or accurately pick out the target word.

1.2.2   Neighborhood Density and VOT

This observation that dense lexical neighborhoods can inhibit listeners' lexical access, paired

with a body of research demonstrating speakers adjust their speech to accommodate perceived

listener difficulty in a variety of ways, including at the level of individual words with regards to

their predictability (see Section 1.1.1), naturally gave rise to the question of whether lexical

neighborhood density was a factor for speakers as well as listeners.

Goldinger & Summers (1989) investigated this possibility with respect to stop-consonant VOT.

They selected minimal pairs that differed in the voicing of the initial stop consonant, half in

relatively dense neighborhoods and half in relatively sparse neighborhoods, and asked

speakers to read these pairs of words aloud. They used a repetition paradigm where a number

of minimal pairs were displayed (as pairs) in random order, but each pair occurred quite a few

times (16) in each trial. Under the assumption that dense lexical neighborhoods constrain

pronunciation, they predicted that pairs from dense neighborhoods would show a larger VOT

difference, but also that this difference would be more consistent across repetitions and be less

susceptible to reduction over time than words from sparse neighborhoods. It's interesting to

note that in this experiment, altering the VOT of the initial stop (i.e. decreasing it for voiced

and increasing it for voiceless segments) would be equally useful for lexical contrast for both

the "easy" and the "hard" words, since all targets had a minimal pair for voicing, and indeed,

6

participants were made explicitly aware of it. Any effects found on this segment must be therefore be categorized as "whole word" effects, rather than effects tailored to aide specific lexical contrasts.

Goldinger and Summers found that there was a significant interaction of neighborhood density and trials. Neighborhood density affected VOT difference (in the expected direction) only in the second set of eight repetitions for each pair: both categories started out with approximately the same VOT difference, but this difference increased over trials in words in dense neighborhoods, while it decreased in sparse neighborhood words.

1.2.3   Neighborhood Density and the Vowel Space

Wright (1997, 2004) investigated this possibility using the English vowel space. He predicted that dispersed vowel space, characteristic of hyperarticulated speech, could be employed by speakers for "hard" words, facilitating lexical access. Wright's measure of dispersion was the Euclidian distance from the center of the speaker's vowel space (bark); in this study, "dispersed" was simply a measure of peripheralization, not a measure of distance from neighboring vowels in the phoneme inventory.

The recordings were selected from an independently created database of recordings of monosyllabic words spoken in isolation.

Considering the vowel space as a whole, Wright's prediction was born out: hard words, on average, were characterized by a more dispersed vowel space. However, individually, not all vowels showed this effect significantly. In particular, what Wright refers to as "point" vowels (i, æ, ɑ, ɔ, u), which are phonetically peripheral, showed more significant and more reliable effects than relatively phonetically medial vowels (ɪ, ɛ, o, ʌ).

Figure 2a. Amount of dispersion, by vowel (Wright 2004, p. 82)
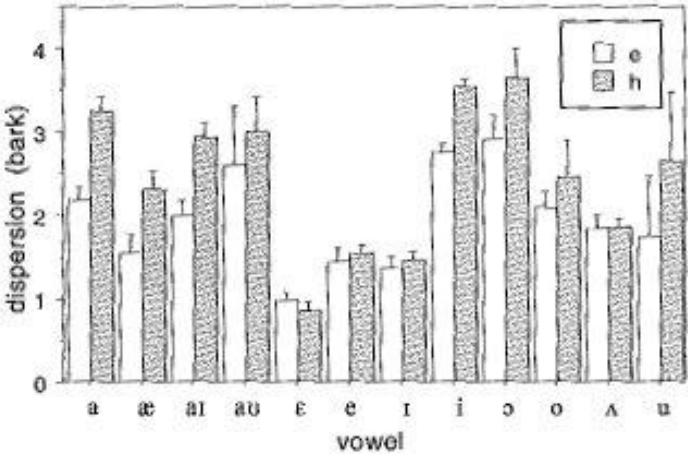
Figure 2b. Formant averages for easy and hard (+) words (Wright 2004, p. 82)
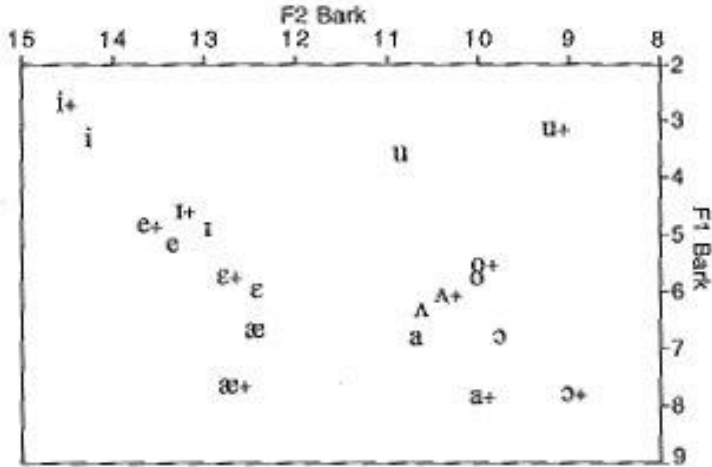
Wright concluded from this that "the expansion of the vowel space occurs in such a way that overall distances between vowels are maximized; only the point vowels, which can move without diminishing the vowel contrasts, become more dispersed while the others remain relatively unchanged across conditions" (p. 84).

Munson and Solomon (2004) confirmed this result, this time controlling for word frequency. They found that the frequency of the target word and its neighborhood density both had effects on vowel articulation and were independent of each other, and of vowel length. The effects of neighborhood density were present even in high-frequency words.

Munson and Solomon also note that vowel duration was not affected by neighborhood density.

1.2.4   Neighborhood Density and Coarticulation

Scarborough (Brown 2001) found that lexical neighborhood density also has an effect on degree of coarticulation. Two kinds of coarticulation were measured. Perseveratory vowel-vowel coarticulation was tested using words of the form CVCi, and measuring the degree of lowering or backing of the onset of the /i/ compared to monosyllabic controls (p. 9). Anticipatory vowel nasalization was tested in CVN words (with a variety of vowels) by

measuring A1-P0 values[3] at the midpoint and at the offset of the vowel to determine its nasality (p. 8-9). In both cases, "hard" words showed higher degrees of coarticulation than "easy" words. Scarborough (2004) later confirmed these effects and elaborated on them, further demonstrating that both vowel-vowel coarticulation and vowel nasalization are bi-directional (pp. 53-68). She also found that the nasal assimilation pattern was similar in French, despite nasalization being a contrastive feature on vowels in that language (pp. 87-95).

In line with Munson and Solomon (2004), Scarborough found no effects of neighborhood density on vowel duration in either the CVCi words or the NVC words. In the CVN words, however, a significant effect was present, with vowels being significantly *shorter* in high-density neighborhoods; this result was surprising, but not discussed.

Since coarticulation is often viewed as a form of reduction, and therefore potentially at odds with comprehension, Scarborough conducted a lexical decision experiment using manipulated tokens of the same words with varying degrees of vowel-vowel and vowel-nasal coarticulation. She found that increased coarticulation resulted in faster decisions times, indicating that it in fact facilitates comprehension. This is in keeping with the general observation that lexical neighborhood effects on production are broadly tailored to aid the listener.

---

[3] A1 - P0 is an acoustic measure of vowel nasality. A1 refers to the amplitude of F1, while P0 refers to the amplitude of an additional low-frequency peak formant found in nasal vowels.

1.2.5   Mismatches Between Perception and Production

Scarborough noted (Brown 2001, pp. 23-24) that not all of the high-density words had

neighbors of the right type to actually be facilitated by the types of coarticulation under study.

33% of the CVN words and 50% of the CVCi words didn't even have neighbors confusable by

the relevant segments.   Nevertheless, speakers increased coarticulation in these words.

Further investigation (2004, pp. 105-116) confirmed that the amount of vowel-nasal

coarticulation in CVN words correlates with lexical neighborhood density overall, and *not*

specifically with number of neighbors that differ by the relevant segments (the final consonant

or the vowel). In other words, phonological contrasts, but, in these cases, no particular lexical

contrasts were being facilitated.

In line with Goldinger and Summers (1989), this again provides some evidence for a "whole

word" effect (at least with respect to coarticulation) whereby lexical neighbors affect the entire

word, not just the segments which are important for the lexical contrast.

Billerey-Mosier (2000) similarly found a lack of an effect of specific lexical contrast. He noted

that each word, treated as a string of segments, has a point in this string at which it becomes

uniquely identifiable in the lexicon (the word's "uniqueness point").   After this point, the

remaining segments are arguably extraneous for listener comprehension.   He therefore

investigated the possibility that the amount of reduction in a particular word might be greater

for segments after this uniqueness point, since listeners would often already be able to identify the word being uttered, and have less need for cues facilitating lexical contrast.

Billerey-Mosier found no such difference; level of reduction (as measured by F1 and F2, and vowel duration) did not differ with respect to position relative to uniqueness point.

Scarborough (2010) measured the interaction of contextual predictability of words in sentential contexts with the purely lexical factors she had already studied on vowel-nasal coarticulation. Predictability of words within specific sentences was determined in advance in a separate, open-choice sentence completion task, and a variety of relatively predictable and relatively unpredictable targets were chosen. Asking speakers to utter these sentences, and measuring the effects on the target words, Scarborough found that both contextual unpredictability and neighborhood density have significant effects on coarticulation. Interestingly, neighborhood density triggered increased coarticulation even for very highly predictable words, where potential listeners demonstrably had no need for additional cues.

Scarborough (2010) ceded that this is not necessarily evidence against listener modeling, but it does demonstrate a mismatch between listeners' needs and speakers' production with regards to neighborhood effects.

1.2.6   VOT Revisited: Specific Lexical Contrast, Lexical Context, and the Underlying Mechanism for LNE

Baese-Berk & Goldrick (2009) reinvestigated stop VOT with the ambitious goal of distinguishing between a number of accounts of the mechanisms underlying lexical neighborhood effects. They sought to determine the effects of both the existence of lexical neighbors, and the inclusion of these lexical neighbors in the production context (i.e. making the speaker explicitly aware of them) on stop VOT.

Baese-Berk and Goldrick measured the VOT of word-initial voiceless stops (p, t, k) in a listener-directed production task described below. Crucially, rather than categorizing targets by overall neighborhood density, they determined *only* whether the target word had a minimal pair with the corresponding voiced-stop. For example, 'pore' has a minimal pair 'bore' for voicing, but 'pork' does not. The two sets of targets, then, differed as to whether the voicelessness of the relevant segment was actually lexically contrastive (p. 6-7).

Participants were paired up as speakers and listeners, and sat opposite each other viewing identical computer displays, each displaying three words. The speaker's display would then indicate one of the three words, and the speaker was asked to instruct the listener to click this word by saying "click on the ___", essentially administrating a closed-set word recognition task. When the target was a "hard" word, half the time, its voiced-stop minimal pair neighbor would be present as one of the two other words on the screen. This effectively created three

conditions: context (min pair shown), no-context (min pair not shown), and no competitor (no

minimal pair in the lexicon).

Figure 3. VOT results (Baese-Berk & Goldrick 2009, p. 24)
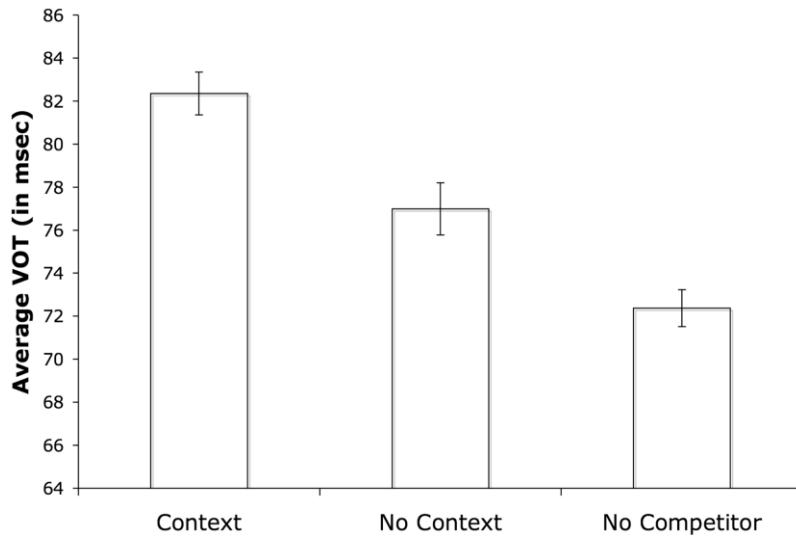
Pooling all three places of articulation, Baese-Berk & Goldrick found a significant three-way

distinction in the VOT of stops in these conditions, showing that both the existance of a single

neighbor for voicing, as well as the inclusion of this neighbor in the context, had significant

effects. Separating the places of articulation, both distinctions were robust for bilabial stops,

but the context / no context distinction was not significant for alveolar stops, and neither

distinction was significant for velar stops; however, all trends were in the right direction.

Baese-Berk & Goldrick consider three accounts of the mechanism underlying lexical

neighborhood effects, and the evidence that their experimental results bring to bear. The three

models they considered were a "speaker-internal" account, perceptual monitoring, and perceptual restructuring.

Their speaker-internal account is an on-line model which doesn't make use of monitoring or listener modeling. Instead, during production, not only the target but its lexical neighbors are activated, causing increased competition in "hard" words—this is analogous to what happens during perception. Increased competition necessitates higher levels of activation of the target word—or some or all of its segments—before it can decisively 'win out' over other activated targets, and finally be selected and produced (p. 3). Baese-Berk and Goldrick further assume higher levels of activation result in decreased levels of reduction.

The perceptual monitoring account, on the other hand, involves on-line listener modeling. Prior to production, speakers "monitor the output of lexical phonological processing" (p. 4). When a speaker anticipates a listener will potentially be confused by an utterance, she will alter her production to make it minimally confusing. Since dense lexical neighborhoods increase the difficulty of perception, speakers will encounter this difficulty during the monitoring process, and hyperarticulate as a result.

The perceptual restructuring account, based on Pierrehumbert (2002), is a purely off-line account that makes use of exemplar theory. Proponents posit that a token of a particular word can only be stored as an exemplar if it is in fact heard correctly and with certainty; words about

which listeners are mistaken or unsure will not be stored as exemplars for the intended word.

Words in dense neighborhoods can more often be misheard, if their pronunciation is not ideal, than words in sparse neighborhoods, for which phonetic deviation is less likely to cause confusion. Therefore, for "hard" words, only more extreme productions will become exemplars, while for "easy" words a larger range of productions can be stored. If exemplar representations also affect production, these differences predict that "hard" words will be pronounced more consistently and with less variation from an unreduced form, and therefore, on average, more extremely.

This last account, being purely off-line, predicts that inclusion of the neighbor in the context should have no effect on articulation; the opposite is true for both the speaker internal account, (since inclusion of the neighbor in the context results in its increased activation) and the perceptual monitoring account (since listeners are at greater risk of confusion when the neighbor is a choice). Conversely, the perceptual monitoring account predicts that, as long as it is not in the context, the presence or absence of a minimal pair in the lexicon should be

irrelevant due to the closed-set nature of the listener's task; the listener is at no risk of

mishearing the word if it has no neighbors on screen.[4] The other accounts predict this effect.

Based on these facts and their results, they throw out both the perceptual restructuring account

and the perceptual monitoring account, based on these models' predictions discussed above, in

favor of the speaker-internal account—the only one consistent with their three-way result.

Since most prior research have hinted at or assumed that speakers in some way have the

listener in mind when they made such phonetic adjustment, this is a relatively novel claim.

Baese-Berk & Goldrick's finding that not just neighborhood density as a whole, but the

presence of a single, lexically contrastive neighbor for a particular segment can in fact robustly

impact that segment's articulation doesn't necessarily contradict the whole-word effect, since

they didn't measure the degree of hyperarticulation of the other segments in the word, but it

does demonstrate that even individual minimal pair neighbors can have significant effects; this

result will be an important factor in the design of the present study.

---

[4] Note that this is a rather tenuous claim: Baese-Berk and Goldrick assume that listener modeling
would be sophisticated enough to take into account the nature of the task.

1.2.7   Neighborhood Density and Lexical Access

While the existence of many high frequency lexical neighbors has an inhibitory effect on perception, Dell and Gordon (2003) summarize a body of research that seems to indicate that the opposite is true during production: lexical items from dense neighborhoods are less susceptible to word-substitution errors (Vitevitch 2002) and ToTs (Harley & Bown 1998) than similar items from sparse neighborhoods, and are also named more quickly in picture-naming tasks (Vitevitch 2002). These results indicating that having many neighbors boosts, rather than inhibits, accurate and efficient retrieval. Dell and Gordon also propose a model that account for this discrepancy between production vs. perception (21-28).

The finding that neighbors facilitate rather than inhibit production is not compatible with Baese-Berk & Goldrick's (2009) account as stated, since the latter specifically emphasizes "competition" from lexical neighbors. However, by both accounts, having neighbors generally speaking causes targets to be more highly activated during production. By whatever mechanism this occurs, if increased activation can be associated both with ease of retrieval and with more extreme segmental realizations, these findings are not irreconcilable.

2   The Research Question

2.1   How Specific is Phonetic Disassociation from Lexical Neighbors?

Most research thus far described has focused on discovering the exact lexical properties that

can have an effect on the articulation of words and the individual segments within them, and

on how they interact with word-internal and contextual factors. However, while frequently

alluded to or assumed, a completely orthogonal question has yet to be formally investigated:

what is the right generalization about the phonetic changes themselves? In other words, once a

speaker has determined that they will adjust the articulation of a particular segment, what

exactly determines the type of adjustment to be made? Most authors point out that the

phonetic effects they are studying will aid listener comprehension, but this end result could be

achieved in a number of ways.

In fact, there are at least two types of potential underlying mechanisms consistent with this

observation, both alluded to in the literature. The first is that lexical neighborhood effects

diminish processes of reduction, resulting in higher degrees of *faithfulness* to a predetermined

underlying target (possibly an exaggerated one), therefore indirectly facilitating word-

recognition.   The second is that speakers are directly facilitating word-recognition by

increasing the perceptual distance between the target word and its lexical competitors in

phonetic space, producing realizations of speech sounds which are *phonetically distant from competing sounds.*

I will call the first the "hyperarticulation" model[5] and the second the "dissimilation" model of lexical neighborhood effects.

Wright (2004), for example, measures dispersion in terms of distance from the center of a speakers vowel space: he is tacitly assuming the hyperarticulation model in the experimental design. However, in the results section, he posits that medial sounds don't move as much from their default positions because of some desire to be distant from other nearby sounds, and *not* from the center, alluding to the dissimilation hypothesis. Baese-Berk & Goldrick (2009) effectively tacitly assume the hyperarticulation model as well, explaining that competition from

---

[5] Johnson et al (1993), in a paper not about lexical neighborhood effects specifically but about hyperarticulation in general, provide some evidence that hyperarticulation can in fact be characterized simply as increased faithfulness; they provide experimental support for the claim that underlying representations are hyperarticulated (the "hyperspace effect"), and all non-hyperarticulated tokens of speech sounds involve some degree of reduction. However, see also Whalen et al (2004), who attempt to refute this claim.

lexical neighbors "boosts" the activation level of the phonological representation, which in turn "leads to more active phonetic representations and consequently more extreme articulatory realizations," although they do not quite explain what 'extreme' is meant to entail (p. 3-4).

It's likely that the difference between these two models hasn't been specifically investigated because their predictions are often highly overlapping. However, they are not indistinguishable, particularly with respect to their predictions about speech sounds that are in some way phonetically *medial* within a phoneme inventory, with competing speech sounds on either side along some phonetic dimension. The hyperarticulation model predicts that there will be no lexical neighborhood effects along this phonetic dimension, or that if there are they will be consistent regardless of the location of lexical neighbors. The dissimilation model however, makes a rather novel prediction: the target speech sound will move in a direction *away* from the corresponding sound in competing words, and the direction and/or degree of movement along the phonetic dimension in question will depend not only on the number of lexical neighbors but on their *location* in phonetic space.

## 2.2   Test Case: English /ɛ/

These two hypotheses differ the most in their predictions about the realization of sounds which are acoustically medial, in some phonetic dimension, between other sounds in the phoneme inventory. As such, languages with a three-way phonemic contrast along some dimension are

needed to clearly test this difference. Several possibilities come to mind: the VOT experiment could be replicated on a language with a three way voicing contrast cued primarily by VOT, e.g. Thai, focusing on the medial, short-lag stop consonants. Languages with at least three sibilants at different places of articulation, like Polish, could also be used.

The present study investigates English /ε/. This choice was made for a number of reasons: not only were English speakers relatively more available at the time of the study than Thai and Polish speakers, but vowels are already known to vary quite a bit in their articulation and to undergo lexical neighborhood effects, while short-lag stop consonants and sibilants are not.

The vowels closest to /ε/ in phonetic space are /ɪ/, /e/, and /æ/. Of these, the monophthongs /ɪ/ and /æ/ lie almost directly opposite each other relative to /ε/ in phonetic space.

Given this fact, the dissimilation hypothesis suggests the possibility that the articulation of /ε/ in various lexical items could be affected by neighbors with /ɪ/ and neighbors with /æ/ in a different or even completely opposite way. The much simpler hyperarticulation account completely precludes this possibility. Therefore, finding an effect that varies not only by the number of lexical neighbors, but by the particular sounds in such neighbors, would provide strong support for the dissimilation hypothesis. Conversely, failing to find such an effect of neighbor location on the realization of any particular segment would be relatively uninformative, but would provide weak inductive support for the hyperarticulation hypothesis.

22

Figure 4a. Predictions of the hyperarticulation hypothesis
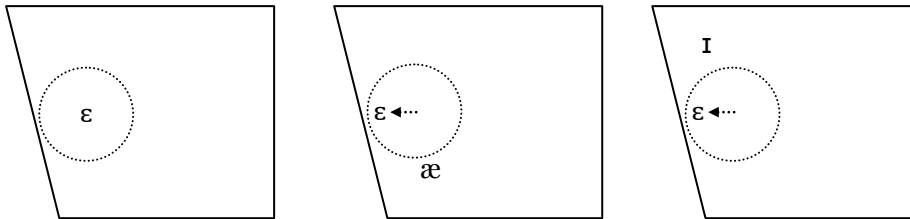
Same effect, if any, regardless of vowel in minimal pair neighbor (in particular, vowel fronting

is predicted by Wright 2004)

Figure 4b. Predictions of the dissimilation hypothesis
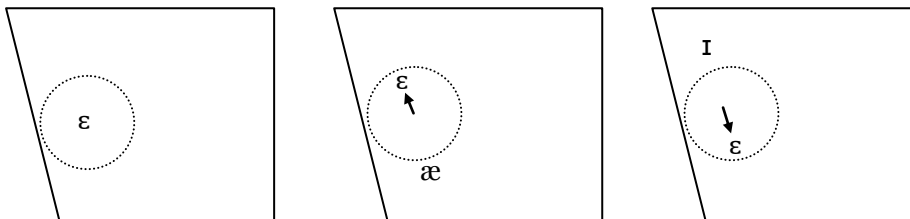
Potentially different effects depending on vowel in minimal pair neighbor.

While Wright (2004) did not find a significant effect of lexical neighborhood density on /ɛ/ in particular (although he did find a trend), this could have been due to the very small number of targets used, but might also have been precisely because his lexical items had lexical competitors with a variety of vowels each affecting /ɛ/ differently, resulting in a null effect on average—for his targets, this is in fact predicted by the dissimilation hypothesis. Its reinvestigation is therefore not unjustified.

3   Methods and Materials

3.1   Participants

49 undergraduate students, 13 male and 36 female, were recruited from a UCLA research

subject pool consisting of students taking introductory psychology and linguistics classes.

All participants were self-reported native speakers of English, and all lived in California at the

time of the study. Many spoke some variety of California English, but a large amount of dialect

variation was present.

3.2   Materials

3.2.1   Target Words

18 target words were selected (listed in Table 1), all of which contained a stressed syllable with

/ɛ/. These fell into three "Neighbor" categories:

1) Ø-targets: no minimal pair with [æ] or [ɪ] - e.g. 'sped' (#spæd, #spd)

2) æ-targets: minimal pairs for [æ], but not [ɪ] - e.g. 'fed' (fæd, #fɪd)

3) ɪ-targets: minimal pairs for [ɪ], but not [æ] - e.g. 'sled' (#slæd, slɪd).

Each category contained six words, and across categories the same immediately post-vocalic

environments, all coronal stops or fricatives, were used to maximally control for the effect of

the following consonant on the target vowel.

Note that preceding environment was not similarly balanced between categories, as fixing both

the following and preceding environments (as well as the vowel itself) would have made it

impossible to find enough targets of the right variety. Instead, preceding environment was

treated as a random effect post-hoc (see section 4.2).

Table 1. Target Words

Post-V Env.	∅-targets	æ-targets	ɪ-targets
_(ɾ/d)ɪ[6]	breading	reddish	bedding
_s#	dress	guess	bless
_st#	nest	vest	rest
_d#	sped	fed	sled
_t#	yet	vet	wet
_ðɹ̩#	feather	leather	weather

These targets were found by searching the CMU Pronouncing Dictionary v0.6 (Carnegie Melon

University, 2010) for ɛ-æ and ɛ-ɪ minimal pairs, using the software MinimalPairs (Hayes,

unpublished), filtering out words for which a minimal trio was found, and selecting minimal

pairs based on post-vocalic environment.

---

[6] These words all contain an underlying /d/ that would normally flap in American English, but was
realized as [d] by some speakers due to the nature of the experiment.

25

3.2.2   Ruling out ultra-low-frequency neighbors

For a few of the target words claimed above to lack a particular minimal pair, a minimal pair

does in fact exist if very low frequency proper nouns and acronyms are admitted (see Table 2).

Table 2. Ultra-low frequency minimal pairs

Target	Potential neighbor	Definition
bedding	Badding	(family name)
bless	Blass	(family name)
sped	1. Spad/SPAD	1. several models of French aircraft
	2. spad	2. neologistic blend of 'spam' and 'ad'
sled	Slad	a village in Gloucestershire

Even if participants had heard these words, their low relative frequency would be predicted to

mitigate any lexical neighborhood effects. Nevertheless, they were asked after the experiment

via a questionnaire whether they knew any of them. The words were accompanied by several

other English words, and several nonce words (see figure 5).

Figure 5. Low frequency word post-experimental questionnaire

Do you know any of the words below?			
If you do, circle them. If not, circle NONE.			
griffy	spad	trough	Slad
covert	frism	Blass	lugs
Badding	colma	Yvoli	etui
NONE			

If participants did claim to know a particular minimal pair, the data from that speaker for the

relevant target word was discarded.

3.2.3   Filler Words

Two kinds of fillers were needed. Due to the nature of the design, not only the target word but

two other words were always displayed on the screen (see Procedure). At least one and

sometimes both were filler words unrelated to the target word. For each target, two fillers were

chosen which had no segments in common with that target, and additionally did not contain a

front vowel of any kind.

The second case involves the large number of filler words actually read aloud by participants

when they weren't reading targets. In these cases, all three words on the screen were fillers,

and were not required to have any particular lexical properties.

See Appendix II for a full list of filler words used.

3.2.4   Paragraph Reading Data

It would be useful in interpreting the results of the experiment to have a general picture of the

normal front vowel space for the subject population.

To this end, participants were asked to read aloud a short story that contained a number of

words with ɛ, æ, and ɪ (listed in Table 3). These words had the same six post-vocalic

environments as the target words, and none of the words were the same as the target words in

the experiment. Furthermore, none of these words had minimal pairs in the front vowel space.

Table 3. Targets for the Paragraph Reading Data

Following Env.	ɛ	æ	ɪ
_(ɾ/d)V	credit	radical	forbidding
_s#	success	class	hiss
_st#	test	outlast	exist
_d#	instead	Chad	skid
_t#	cassette	chat	admit
_ðɹ̩#	together	gather	rhythm

In order to minimize any prosodic differences that might affect formant frequencies, all

reading-style targets were placed in a sentence position predicted to be both pitch accented and

intonation-phrase-medial for most speakers' renditions.

The full text of the short story (which is admittedly rather short on literary merit) can be found

in Appendix I.

3.3   Equipment

Subjects were recorded in one of two UCLA Phonetics Lab sound proof booths. Recordings

were saved directly to disk in .wav format using a head-worn Shure SM10A microphone, and

run through an XAudioBox pre-amplifier and A-D device. The recording was done through PCQuirer, with a sampling rate of 11,000 Hz.

3.4   Procedure

The procedure was almost identical to the experiments in Baese-Berk & Goldrick (2009), with multiple words appearing on a screen, allowing for minimal pair neighbors to be included in the context in some cases. The biggest difference was that no listener was present.

3.4.1   Stimuli and Recording

The stimuli were presented on a computer screen using a MATLAB script.

Participants always saw three words on screen, in a horizontal line, with one word on the left, one in the middle, and one on the right. After three seconds, one of the three words turned red, and participants were asked to read it aloud. After a further three seconds, a new screen with fresh words was presented and the process repeated.

For each set of three words, either all three were filler words, or one was an experimental target.

For each speaker, the order in which they saw various word triplets was randomized, as was the location on the screen of the three words.

3.4.2 Experimental Conditions

When a target word was on the screen, if the target had an æ or ɪ minimal pair neighbor in the lexicon, half of the time, this neighbor was present as one of the other words on the screen (the third word being a filler): I will refer to this as the "Shown" condition. Half of the time, the minimal pair neighbor was not shown (and both of the other words were fillers); I will refer to this as the "Hidden" condition. For ∅-targets, this distinction was obviously not applicable.

3.4.3 Experimental groups

In order to collect both Shown and Hidden condition data for each relevant target, participants were divided into two groups: group A and group B. For participants in group A, the _t#, _s#, and _ɾ/d targets all had their minimal pair neighbors Shown, while the _d#, _st#, and _ðɹ# were placed in the Hidden condition. For participants in group B, the two groups were reversed, balancing the data.

3.4.4 Post-Experimental Questionnaire

Following the recording phase, participants were asked to complete the post-experimental questionnaire described in the Materials section.

3.5   Acoustic Analysis

3.5.1   Vowel Segmentation

Each target was annotated with a Praat textgrid file which indicated the edges of the target vowel. In cases where adjacent sounds were obstruents, the onset or offset of silence (in the case of stops and flaps) or frication (in the case of fricatives) was used to identify the vowel-consonant boundary. When adjacent sounds were approximants, to the extent it was possible, the leveling off of F1 and/or F2 was used.

This was not always a clear cut process, and as a result, the vowel duration values in particular were not assumed to be very reliable.

3.5.2   Vowel Formant Extraction

Using the software VoiceSauce (Shue et al. 2011), the average F1 and F2 from the period starting at one third and ending at two thirds of the way through the vowel were extracted using the Snack Sound Toolkit (Sjölander 2004). Only this middle third was measured so as to maximally avoid influence from the preceding and following consonants, and to mitigate some of the segmentation uncertainty problems mentioned in the previous section. Duration information for the vowel was also extracted.

3.5.3 Outliers and Error Checking

The highest and lowest values of F1 and F2 were hand-checked for correctness. For some of the reportedly most extreme values, the software had failed to identify the formants correctly, and mistook one formant for another. These errors were hand-corrected by taking the formant averages of the middle third of the vowels using Praat (Boersma, P. & Weenink, D. 2011).

4   Results

4.1   The Subject Pool's Front Vowel Space

The formant values for/æ/, /ɛ/, and /ɪ/ collected during the paragraph reading part of the experiment are graphed in Figure 6. They demonstrate a relatively linear short-front vowel space, with a reasonable amount of inter-speaker variation for all three vowels.

Figure 6a. Average F1 and F2 for front short vowels on a Bark scale.
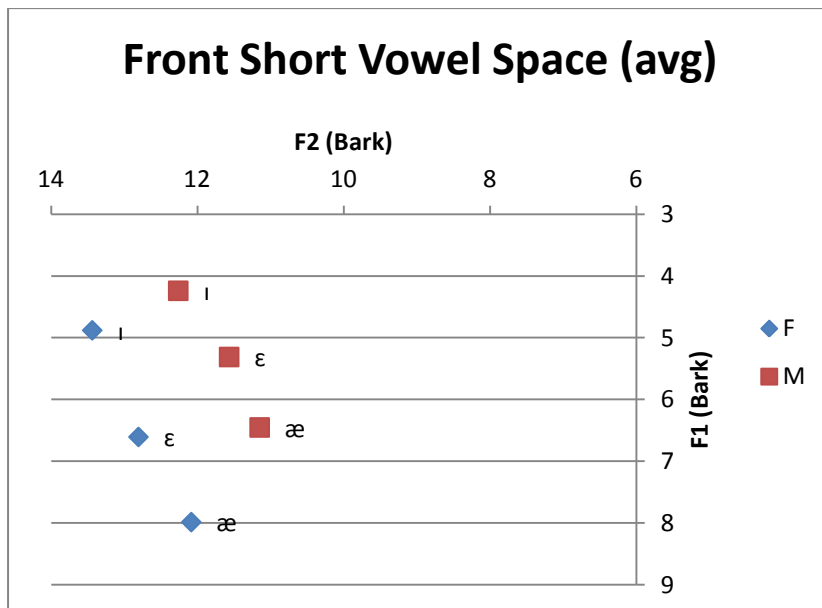
Figure 6b. Individual F1 and F2 Values, by speaker sex, in Hz.
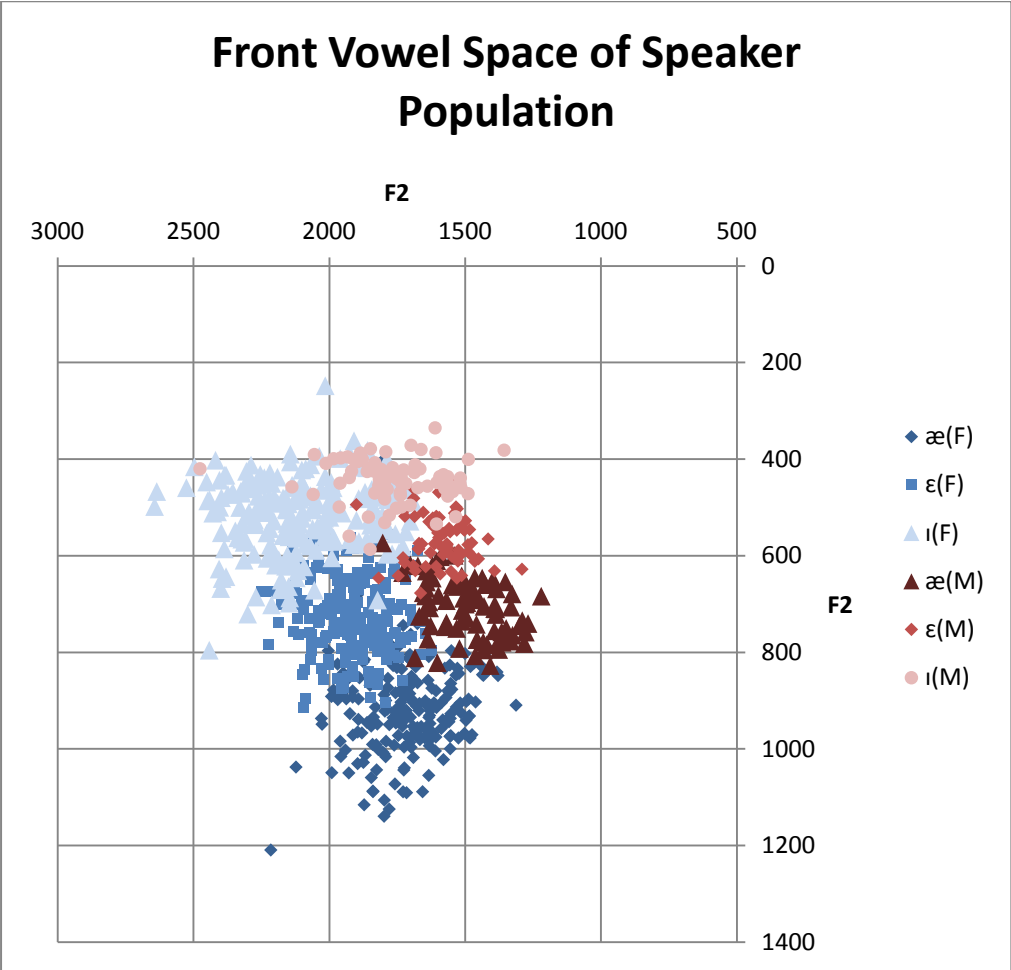
4.2   Statistical Methods

A series of linear mixed effects regressions were performed on the /ɛ/ tokens collected in the

main section of the experiment. The dependent variables were variously F1, F2, and duration.

The fixed effects were neighbor position, existence of an /e/ neighbor, existence of an /i/

neighbor, condition (hidden vs. shown neighbor), speaker sex, and post-vocalic consonant.

The random effects were speaker, and pre-vocalic consonant.

Note that Post-Vocalic Consonant was treated as a fixed effect primarily due to the fact that it was highly controlled in the experimental design, completely intersecting the other fixed effects, whereas Pre-Vocalic consonant was not.

The criterion for significance in all cases was t > 1.96.[7]

4.3   The Experimental Conditions: Hidden vs. Shown

For this comparison, ∅-targets were removed, as these words did not have hidden vs. shown conditions across which to compare.

Three linear mixed effects regressions were run on the data. The dependent variables in these regressions were F1, F2, and duration. The fixed effects were the interaction between condition and neighbor position, existence of an /e/ neighbor, existence of an /i/ neighbor, sex, and post-vocalic consonant. Random effects were speaker, target word, and prevocalic consonant.

Neither condition, nor condition in interaction with neighbor position had a significant effect on F1 (main: t = -0.42, interaction: t = -0.54) or on F2 (main: t = 0.45, interaction: t = -0.18) Condition also did not significantly affect vowel duration (main: t = -1.59, interaction: t = 1.41).

---

[7] 95% of the area under a normal distribution lies within 1.96 standard deviations of the mean. For standard distributions, t > 1.96 therefore corresponds to a p-value of less than 0.05.

In other words, given a particular target, seeing a minimal pair on the screen did not cause speakers to lengthen (in fact, words in the shown condition were on average slightly shorter) or change the quality of the vowel. Based on these results, the condition variable was omitted in the regressions used in the following section, and data from the two conditions was pooled.

4.4   The Influence of Lexical Neighbors

4.4.1   Effects of Neighbors on F1 and F2

The raw averages for F1 and F2 over all tokens of each neighbor type collected in the experiment, normalized by speaker sex and post-vocalic environment, are plotted in Figure 7.

Figure 7. Average F1 and F2 by location of lexical neighbor
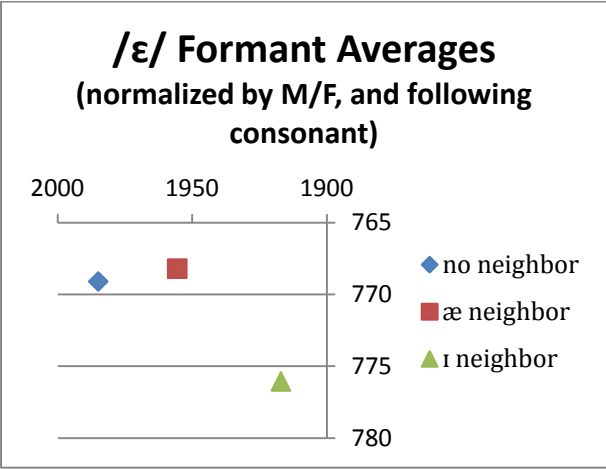
Despite the apparent trend in Figure 7, the actual data are highly overlapping, even when averaged for each speaker and normalized across each speaker's range for /ɛ/ (see Figure 8).

Figure 8. F1 and F2 averages for each speaker, normalized by speaker range
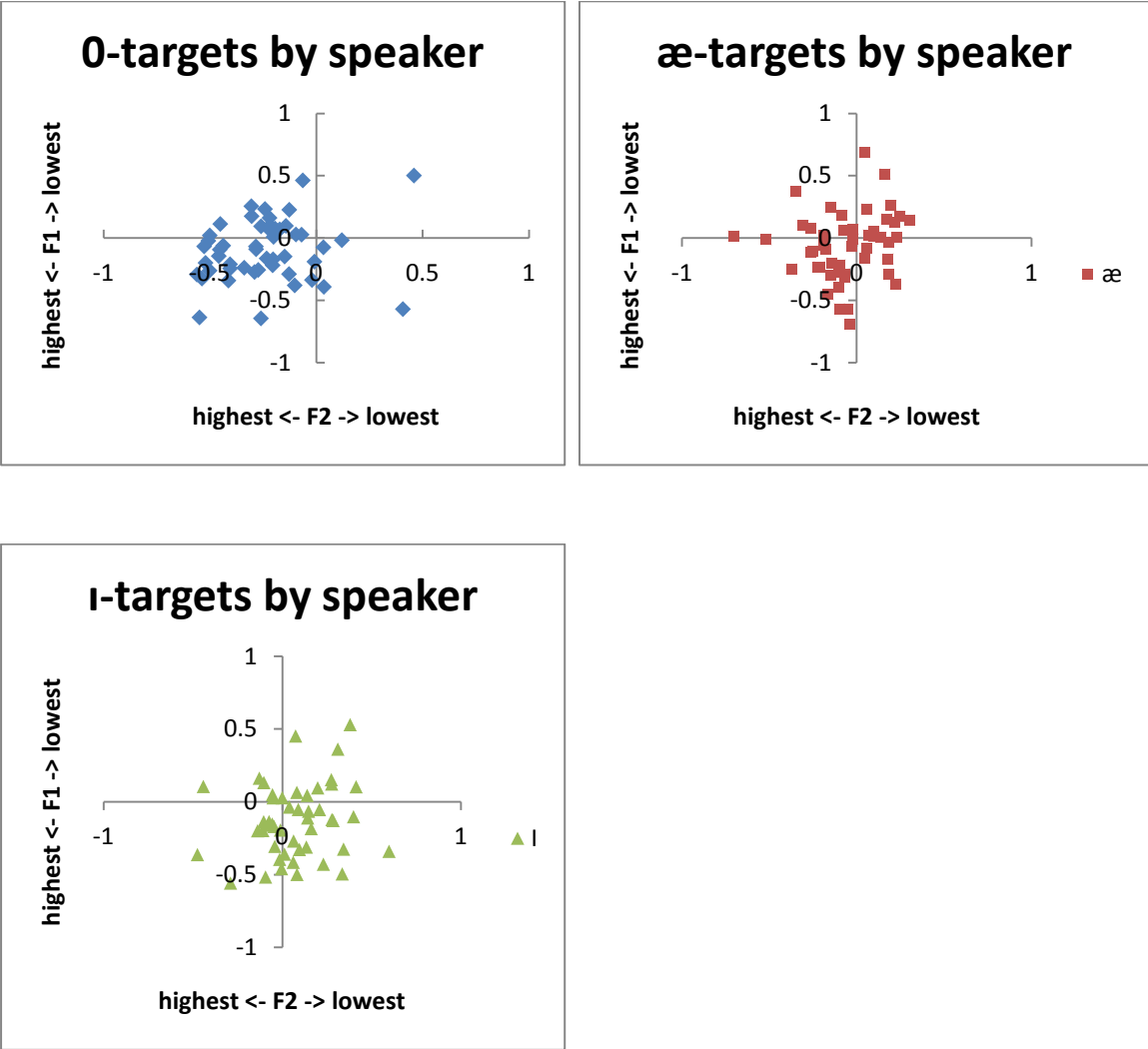
Note, however, that the plots above do not control for factors such as pre- and post-vocalic environment. Since pre-vocalic environment in particular was not balanced, this, as well as individual speaker variation, may be confounding factors.

A series of linear mixed effects regressions were run on the data. The dependent variables in these regressions were F1, F2, and Duration. The fixed effects were neighbor position,

existence of an /e/ neighbor, existence of an /i/ neighbor, speaker sex, and post-vocalic

consonant. The random effects were speaker and prevocalic consonant.

Table 4. The effects of neighbor position on F1 and F2.

	F1 diff (Hz)	t-value	F2 diff (Hz)	t-value
æ - ɔ	-0.9	t = -0.16	0.8	t = 0.05
ɪ - ɔ	7.3	t = 0.99	-32.6	t = -0.84
ɪ - æ	8.3	t = 1.24	-33.4	t = -1.24

Neighborhood position did not have a significant effect on either F1 or F2. However, ɪ-targets

did show a trend of being slightly lower and backer than the other target types (t = 1.24 in each

case), the exact direction predicted by the dissimilation hypothesis.

4.4.2   Effects of Neighbors on a Combined F1/F2 Factor

While the individual F1 and F2 effects were not significant taken separately, the front vowels of

English could arguably be conceptualized as falling along a single axis from front-high to low-

mid because of the trapezoidal shape of the vowel space, with distance along this axis being a

potential primary cue for front vowel height. Using the bark values for the speakers' short

front vowels found in section 4.1, front vowels were found to lie roughly along a line with slope

-2.13 in the bark F1/F2 space.

37

Figure 10. Averages for a combined F1/F2 factor, by neighbor position
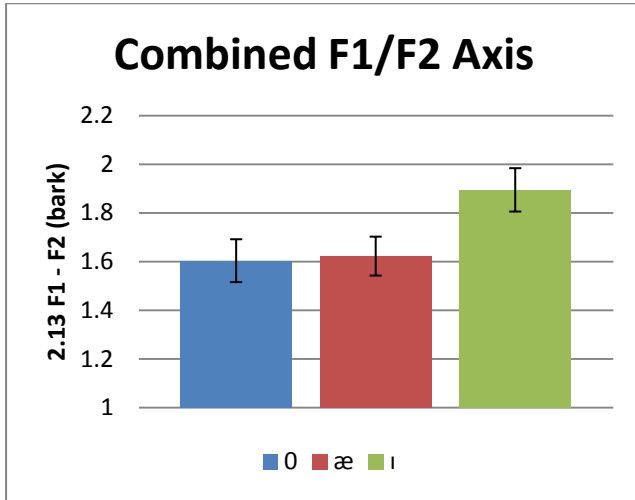
(Error bars indicate one standard deviation.)

Using this slope, F1 and F2 (bark) for the /ɛ/ targets were combined into a single factor ($F1_{bark}$

– 2.13 $F2_{bark}$), which was measured as a dependent factor against the same fixed and random

effects as before.

Table 5. The effects of neighbor position on a combined F1/F2 factor

	F1/F2 (bark) diff	t-value
æ - 0	-0.03	t = -0.25
ɪ - 0	0.16	t = 0.71
ɪ - æ	0.19	t = 1.10

As expected, ∅-targets and æ-targets did not differ along this axis.

ɪ-targets were lower and backer than the Ø-targets and æ-targets, but once again did not reach significance in either case (t = 0.71, t = 1.10). The apparent difference in Figure 10 was therefore partly explainable by other factors, and was not demonstrably due to neighbor location.

4.4.3   Effects of Neighbors on Vowel Duration

Both æ-targets and ɪ-targets had significantly longer duration than Ø-targets, æ by 22 ms (t = 4.0), and ɪ by 17 ms (t = 2.2). There was no duration difference between the æ-targets and ɪ-targets (t = 0.2).

Figure 11. The effects of neighbor position on duration
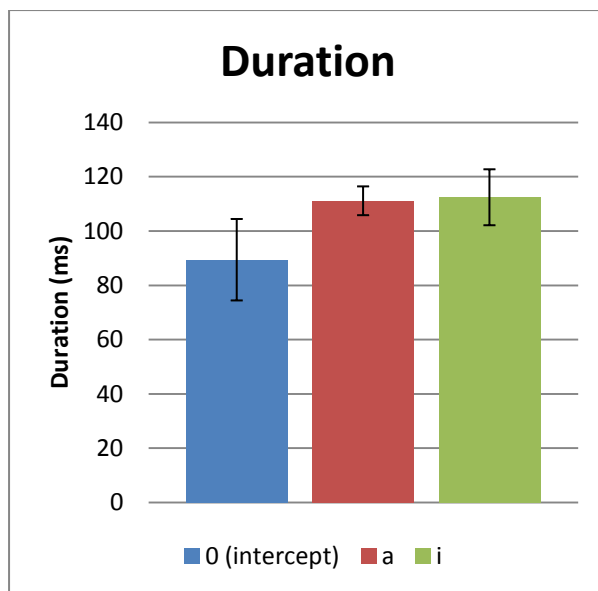
(Error bars indicate one standard deviation.)

Restated, the existence of a lexical neighbor had a significant lengthening effect on the vowel, and type of neighbor did not matter. In the present study, this duration effect was in fact the only result that could potentially be attributed to lexical neighborhood effects at all. This effect, if real, contradicts earlier research indicating that vowel duration is not affected by neighborhoods (Munson and Solomon 2004). However, recall that duration was not precisely measured due to difficulties in segmentation (see section 3.5.1).

5   Discussion

5.1   Discussion of Results

Essentially, the experiment yielded a null result. The fact that neither ɪ-targets nor æ-targets significantly differed from ∅-targets in F1 or F2 effectively provides evidence for a rather weak claim: namely, the type of single-minimal-pair effect found by Baese-Berk and Goldrick (2009) does not detectably apply to the formants of English /ɛ/. This is attributable to several possibilities.

It's possible that this particular vowel, or perhaps medial vowels in general, are not subject to lexical neighborhood effects at all, as suggested by Wright (2004).

The fact that ɪ-targets trended slightly backer and lower than æ-targets, the direction predicted by the dissimilation hypothesis, would suggest that this hypothesis might be correct had this

trend been significant. However, even then, the lack of any formant difference between ∅-

targets and æ-targets would remain to be explained under this account.

The significant correlation between duration and having a minimal pair of either type seems to

suggest vowel lengthening in response to minimal pair neighbors, a correlation that was

specifically found not to exist in earlier work (Scarborough 2004; Munson and Solomon 2004).

While vowel duration measurements were not particularly accurate due to segmentation

difficulties in words with adjacent sonorants, as described in section 3.5.1., this result might

bear further investigation.

In summary, as per Wright (2004), English /ɛ/ either doesn't show significant lexical

neighborhood effects at all, or else the presence of single minimal pair neighbors are not

enough to trigger these effects, or else, if they are, the effects were too weak to reach

significance in the current study. As a result, the central research question about what the

nature of the effect would be for such a segment remains unanswered.

5.2   Future Research

5.2.1   Medial Sounds in Other Languages

Besides English /ɛ/, a number of other phonetically medial sounds come to mind. Thai has a

three way stop systems with respect to VOT, Korean has a stop system for which VOT is a

perceptual cue in at least some environments. French and German have a three-way high-

vowel distinction for which F2 is a perceptual cue, namely /i, y, u/. Russian and Polish each have three-way sibilant distinctions, which are presumably cued at least in part by pitch / center of gravity. The medial sound in each of these cases could be investigated for variability of neighborhood effects due to the location of the competition.

5.2.2 Variations on the Experiment Design

One of the problems with the present study might have been that it relied too heavily on the assumption that investigating individual lexical neighbors rather than the lexical neighborhood as a whole is enough to yield significant results, as it did in Baese-Berk and Goldrick (2009).

An experiment that took into account the entire lexical neighborhood of each target but still made reference to the *location* of lexical neighbors in phonetic space, this time on average (weighted by phonetic similarity), would alleviate this difficulty. If the experiment found that targets move away from the collective center of gravity of their lexical neighbors, the dissimilation hypothesis would be strongly supported.

Additionally, it would be preferable to better balance the pre-target environment in the experiment design. However, this type of control is somewhat restricted by the already small number of appropriate targets.

6 Conclusion

The mechanism whereby words' lexical neighborhoods can influence the realization of their speech sounds during production is not yet well understood. While the end effect of pressure from lexical neighbors is to produce tokens that are more easily distinguished from their lexical competition, whether due to listener modeling or some other mechanism, it is unclear how specifically speakers manipulate segments in order to maximize this perceptual distance. Further research might determine two open questions: whether lexical neighborhood effects are sensitive to the location of lexical competitors in addition to the overall amount of lexical competition, and, if they are, whether the effects themselves vary in a way that exploits these differences.

References

Books & Articles

Anderson, A. H., E. G. Bard, C. Sotillo, A. Newlands, & G. Doherty-Sneddon (1997) Limited visual control of the intelligibility of speech in face-to-face dialogue. *Perception & Psychophysics* 39: 580-592.

Aylett, M., Turk, A. (2006) Language Redundancy Predicts Syllabic Duration and the Spectral Characteristics of Vocalic Syllable Nuclei, *Journal of the Acoustical Society of America* 119: 3048-58

Baese-Berk, M., & M. Goldrick (2009) Mechanisms of interaction in speech production. *Language & Cognitive Processes* 24: 527-554.

Billery-Mosier, R. (2000) *Lexical effects on the phonetic realization of English segments.* UCLA MA Thesis.

Boersma, P., & S. Hamann (2008) The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25: 217–270. Cambridge University Press

Brown [Scarborough], R.A. (2001) *Effects of Lexical Confusability on the Production of Coarticulation.* UCLA MA thesis.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002) What's New Pussycat? On talking to babies & animals. *Science,* 296, 1435

Chen, M. & W. S-Y. Wang (1975) Sound change: actuation & implementation. *Language* 51: 255-281.

Goldinger, S. D., & W. V. Summers (1989). Lexical neighborhoods in speech production: A first report, *Research on Speech Perception Progress Report,* No. 15: 331-342. Bloomington, IN: Speech Research Laboratory, Indiana University.

Harley, T. A., & H. E. Bown (1998) What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology* 89: 151-174.

Hawkins, S., & P. Warren (1994) Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics* 22: 493-511.

Johnson, K., E. Flemming, & R. Wright (1993) The Hyperspace Effect: Phonetic Targets are Hyperarticulated. *Language* 69: 505-528

Kim, H., M. M. Diehl, R. Panneton, & C. Moon (2006) Hyperarticulation in Mothers' Speech to Babies & Puppies. *Paper presented at the annual meeting of the XVth Biennial International Conference on Infant Studies, Kyoto, Japan* <Not Available>.

Labov, W., M. Yaeger, & R. Steiner (1972) *A Quantitative Study of Sound Change in Progress.* Philadelphia: University of Philadelphia Press.

Lane, H., & B. Tranel (1971) The Lombard sign & the role of hearing in speech. *Journal of Speech & Hearing Research* 14: 677-709.

Lieberman, P. (1963) Some effects of semantic & grammatical context on the production & perception of speech. *Language & Speech* 6: 172-187.

Lindblom, B. (1990) Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (eds.) *Speech Production & Speech Modelling. Kluwer Academic Publishers.*

Lombard, E. (1911). Le signe de l'élévation de la voix. *Annales des maladies de l'oreille, du larynx, du nez et du pharynx* 37: 101-119.

Luce, P. A. (1986). Neighborhoods of words in the mental lexicon. Ph.D. dissertation, Indiana University, Bloomington, IN.

Oviatt, S., G. Levow, E. Moreton, & M. MacEachern (1998) Modeling global & focal hyperarticulation during human-computer error resolution. *Journal of the Acoustical Society of America* 104: 1-19.

Peramunage, D., S. Blumstein, E. Myers, M. Goldrick, & M. Baese-Berk (2011) Phonological neighborhood effects in spoken word production: An fMRI study. *Journal of Cognitive Neuroscience* 23: 593-603.

Pierrehumbert, JB. (2002) Word-specific phonetics. In C. Gussenhoven & N. Warner (eds.) *Papers in Laboratory Phonology* 7: 101-139, Berlin: Mouton de Gruyter.

Scarborough, R.A. (2004) *Coarticulation & the structure of the lexicon.* Ph.D. dissertation, UCLA, Los Angeles, CA.

Scarborough, R.A. (2010) Lexical & Contextual Predictability: Confluent Effects on the Production of Vowels. In C. Fougeron, B. Kuehnert, M. Imperio, & N. Valee (eds.) *Laboratory Phonology* 10, (2010) Mouton de Gruyter

Scarborough, R. A. J. Brenier, Y. Zhao, L. Hall-Lew, O. Dmitrieva (2007) An acoustic study of real and imaginary foreigner-directed speech. Proceedings of 17th International Congress of Phonetic Sciences, Germany.

Shiller, N. O., & A. S. Meyer, eds (2003) *Phonetics and Phonology in Language Comprehension and Production. Differences and similarities.* Berlin: Mouton de Gruyter.

Soltau, H., H. Metze, & A. Waibel (2002) Compensating for hyperarticulation by modeling articulatory properties. In *Proceedings of International Conference on Spoken Language Processing* 2: 83–86, Denver.

Vitevitch, M. S. (2002) The influence of phonological similarity in neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28: 735-747.

Wang, W. S-Y (1969) Competing changes as a cause of residue. *Language* 45: 9-25.

Whalen, D. H., H. S. Magen, M. Pouplier, A. M. Kang, & K. Iskarous (2004) Vowel Production & Perception: Hyperarticulation without a Hyperspace Effect. *Language & Speech,* 47: 155-174

Wright (2004) R. Wright, Factors of lexical competition in vowel articulation. *Papers in Laboratory Phonology* 6, Cambridge University Press

Corpora & Software

Audacity Team (2008) **Audacity** [Computer program]. Version 1.2.6, retrievable from http://audacity.sourceforge.net/

Bruce Hayes (unpublished) **MinimalPairs** [Computer program].

Boersma, P. & Weenink, D. (2011) **Praat**: doing phonetics by computer [Computer program]. Version 5.1.18, retrievable from http://www.praat.org/

Carnegie Melon University (2010) *CMU Pronouncing Dictionary*. Version 0.6.

R Development Core Team (2011) **R**: A Language & Environment for Statistical Computing [Computer program]. Version 2.15.0, retrievable from http://wwww.R-project.org/. Vienna, Austria.

Shue, Y. L. (2011) **VoiceSauce**: A program for voice analysis [Computer program]. Version 1.12, retrievable from http://www.ee.ucla.edu/~spapl/voicesauce/.

Sjölander, K. (2004) **Snack** sound toolkit [Computer program]. Version 2.2.10, retrievable from http://www.speech.kth.se/snack. Kungliga Teknisa högskolan, Stockholm, Sweden.

Appendix I: Paragraph Reading Stimulus

The following was the text read aloud by participants in order to diagnose the general shape of

their short front vowel space. Bold added here; it was not present in the stimulus.

*Even though **Chad** didn't feel well today, he still had to go to school. There was a big **test** to take in his **class** on music, and academic **success** was important to him. His teacher had some **radical** ideas about what made for good music, but he had to **admit** that the she gave extra **credit** to Chad a lot, even though he had no sense of **rhythm** or pitch.*

*Chad's parents were **forbidding** him from driving, so he jumped on his bike **instead** to avoid getting lectured at. He didn't mind the workout, but then again, he couldn't play his **cassette** tapes on the bike...*

*Ten minutes later, a bump made Chad **skid** to a stop in front of the school, and he heard a soft **hiss** coming from his tire. "Damn," he thought. Waiting for the students to **gather** in the classroom, he began to **chat** with some friends as they complained **together** about how mean their teachers were. "They only **exist** to make us suffer," one said. "Only two more years...we'll **outlast** them all," Chad replied with a smile.*

Appendix II: Targets and Fillers

Target	Minimal Pair	Filler 1	Filler 2
æ-targets			
vet	vat	moose	gunk
guess	gas	mob	tone
reddish	raddish	punting	soggy
vest	vast	gosh	flood
leather	lather	suffix	coffee
fed	fad	plop	rush
ɪ-targets			
wet	wit	soak	box
bless	bliss	cot	thud
bedding	bidding	roofer	clunky
rest	wrist	guns	palm
weather	wither	poodle	couches
sled	slid	posh	pun
Ø-targets			
yet		probe	slouch
dress		glob	womb
breading		husky	toner
nest		crack	small
feather		glum	loft
sped		clot	guns
Fillers			
crab	crib		toss
brick	brim		nip
chat	chap		spy
slick	flick		bin
smoke	smock		heart
fox	fax		smoke
crow	crew		knack
spat	spot		life
drift	drip		fix
slink	brink		map

pick	sick		pat
prank	drank		soup
drip	drop		seive
spam	spat		glue
back	Bach		red
slip	slop		black
taxes	Texas		daisy
timber	tamber		shilling
soupy	soapy		poppy
piddle	paddle		childish
laughter	lifter		blacken
flip		flown	prime
pod		fly	run
trap		slice	dime
strife		green	stop
true		nice	more
lick		gray	dump
scum		small	grab
smock		God	scoop
slob		arms	lime
scoop		splice	ouch
cab		dog	blood
tarp		cat	man
blah		cruel	Bill
flop		hope	small
glam		Tom	dead
spruce		pug	sty
hay		hog	tick
loved		jacket	club
soda		dip	zipper
clammy		guitar	cooler
llama		clumsy	poncho