Centralization and undershoot in /aɪ/ production: Analyzing durational effects for an American English diphthong

Michael Scanlon
February 5, 2011

Submitted in partial fulfillment of generals requirements
Department of Linguistics
University of Washington

1. Introduction

Duration-dependent vowel reduction (hereafter, 'reduction') is well-documented in a few languages (Joos 1948, Tiffany 1959, Lindblom 1963, Stevens & House 1963, Stalhammar et al. 1973, Koopmans-van Beinam 1980, Beckman et al. 1992, Karlsson 1992, Van Bergem 1993, Moon & Lindblom 1994, Gendrot & Adda-Decker 2007). A reduction effect can be observed as the displacement of formant patterns in vowels of short duration[1][2]; for example, the F2 of the vowel in [dɪd] *'did'* is lower in vowels of short duration than in vowels of long duration. However, studies of reduction have primarily examined duration-dependent effects on monophthongs. It is assumed that monophthongs have a single acoustic vowel target, and durational effects on monophthongs can be modeled as formant displacements at a single point in time. Diphthongal vowels, on the other hand, are believed to have at least two important acoustic characteristics: one associated with the vowel's first target (hereafter, 'onset'), and another corresponding to some feature of the second part of the vowel (hereafter, 'offset'). Thus, modeling reduction effects in diphthongs involves taking into account, potentially, duration-based displacement in formant values at two points of the vowel. This study will test for two kinds of reduction, centralization and undershoot, for the stressed diphthong /aɪ/ in one variety of American English, and will test predictions about the prevalence of each reduction effect on /aɪ/.

/aɪ/ is of particular interest for theories of reduction because several studies (Thomas 1995, Thomas 2004, Moreton 2004, Moreton & Thomas 2007) have proposed that context-dependent phonological patterns observed for /aɪ/ are formed by phonetically-driven reduction. For instance, Moreton & Thomas (2007) propose that Canadian Raising, where the nucleus of pre-voiceless /aɪ/ is raised, is due to increased coarticulation with the high voiceless consonant in vowels of short duration, and that the context-dependent pattern was phonologized and extended to apply to /aɪ/ vowels of any duration. While the current study has little to say about the phonologization of phonetic patterns, it examines the extent to which durationally-based phonetic effects are present in /aɪ/ production, and discusses the implications of the study's results for work by Moreton and Thomas.

Reduction effects are often theorized as the result of natural phonetic processes. However, unconsidered social differences, such as regional accents or dialect differences, may introduce confounding effects. For example, the vowel class represented by /u/ in American English may vary from a high back to a high central variant, depending on regional dialect (e.g., Labov et al. 2006), ethnicity (e.g.,

[1] With this definition of reduction, this study distinguishes between phonetic and phonological reduction phenomena. The latter (where unstressed English vowels change to schwa, as in the stressed second syllable in 'phot<u>o</u>graphy' versus unstressed second syllable in 'phot<u>o</u>graphic'; cf. Chomsky & Halle 1968) will not be discussed here.

[2] While "short" and "long" may be used as descriptors for phonemic length in American English (i.e., as synonymous with "lax" and "tense"), throughout this paper the terms are used to refer to a vowel segment's overall duration.

Eberhardt 2009) gang affiliation (e.g., Fought 1999) or self-identification as a "Valley Girl" (Eckert 2008). A phonetic study investigating formant displacement in American English /u/ production, then, should consider the extent to which F2 displacement reflects natural phonetic processes versus social characteristics of pooled speakers; even when restricting subjects to a relatively small regional area, social characteristics may confound the natural phonetic processes in investigation. It is possible to control for a number of social characteristics while investigating reduction, and to consider reduction simply as a natural phonetic process independent of any social concern. However, a reduction study that accounts for sociophonetic variation not only disentangles natural phonetic and sociophonetic effects, but potentially expands the purview of phonetic theory and contributes to a theoretic model of phonetic processes that integrates sociophonetic variation, of the kind recommended by Foulkes & Docherty (2006). This study will investigate reduction for a vowel that shows variable, context-dependent social variation, and will attempt to isolate social and natural phonetic effects.

Lastly, this study will attempt to expand existing knowledge of centralization and undershoot by examining the extent to which reduction is present in conversational speech. For the most part, hypotheses about reduction have been tested in experimental conditions. The use of carefully controlled test items "may focus the speaker's attention on contrasts, thereby exaggerating them" (Byrd 1994; 40). Phonetic variation in conversational speech, in contrast, is more extensive and less transparent in its systematicity (Engstrand 1992); looking for reduction in conversational speech tests how robust these phenomena are, in speech not specifically meant to elicit particular phonetic properties.

2. Background

2.1. Duration-dependent vowel reduction

There are two kinds of duration-dependent vowel reduction described in the acoustic phonetics literature. In the first, centralization, vowels become more "schwa-like", or "central", or "neutral", at short durations (Joos 1948, Tiffany 1959, Stalhammar et al. 1973, Koopmans-van Beinam 1980, Karlsson 1992, Gendrot & Adda-Decker 2007). Centralization is sometimes explicated with articulatory theories (e.g., Hyper- and Hypoarticulation Theory; Lindblom 1990), but in its most simple form it is described in acoustic terms, as a relation between duration and vowel quality. To give a recent example from the literature, Gendrot & Adda-Decker (2007) examined duration-related vowel reduction in 8 languages (mostly Indo-European, but including Arabic and Mandarin) and found a tendency for reduction for vowels of short duration in all languages, looking at vocalic space contour based on peripheral vowels and dispersion from the center of vocalic space. Centralization effects appear to be closely linked to syllable stress (e.g., Nord 1986), and to languages with lexical stress, but can be found in function vs. content words (Shirai 2005) and in syllable-timed languages (Gendrot & Adda-Decker 2007). Centralization, as context-independent reduction, is expected to affect the entire vowel.

The second kind of reduction, undershoot, is observed as more coarticulatory assimilation with neighboring consonantal sounds at short durations, in the form of vowel formant displacement towards consonantal loci (Stevens & House 1963, Lindblom 1963, Beckman et al. 1992, Van Bergem 1993, Moon & Lindblom 1994). The greater assimilation is a result of articulator constraints:

> "…Undershoot is the automatic response of the motor system to an increase in rate of motor commands. The commands are invariable but when they are issued at short temporal intervals, the articulators do not have sufficient time to complete the response before the next signal arrives and thus have to respond to different commands simultaneously. This induces both vowel shortening and reduced displacement of formants." (Farnetani & Recasens 1999, p. 34-5)

As an example, Lindblom (1963) showed that, for tokens with long vowel durations, /ɵ/ showed similar formant values regardless of phonetic context. But in tokens with shorter vowel durations, /ɵ/ showed different formant values depending on consonantal context, as shown in Figure 1 below. Lindblom described the durational effects as systematic shifts away from the vowel target, in the direction of the consonantal locus.

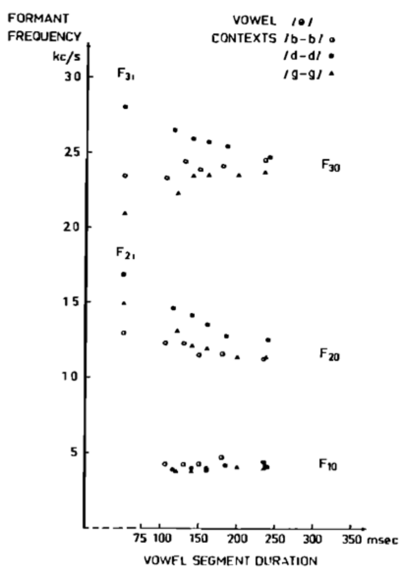

Figure 1. Plot of formant frequency by duration for the vowel /ɵ/, in 3 consonantal contexts (from Lindblom 1963).

Lindblom (1963) argued that the size of the displacement from the vowel target depends on 2 things: the duration of the observed vowel segment and the 'locus-target' distance (i.e., the extent of the CV formant transition). While the theory doesn't say anything about the absolute size of an undershoot effect, predictions can be made about the relative size of undershoot effects in different contexts.

As undershoot is described as an adjacency phenomenon, the most evident undershoot effects on the vowel are expected to be temporally close to the consonant. Moon & Lindblom (1994) expanded the undershoot theory to account for how speakers preserve the vowel target in vowels of short duration when there is a functional reason to do so, by increasing the formant rate of change (by varying the stiffness of or force applied by the articulators). While it seems a truism to say that, relative to longer vowels, shorter vowels display either undershoot or a higher formant rate of change, this formulation can account for speakers' functional adaptation to communicative and sociolinguistic demands. Moon & Lindblom (1994) posit that the undershoot they observed acoustically is due to "an approximate, but straightforward, relationship between F_2 variations and tongue/lip movements" (53) in their study, and argue that articulatory trajectories can be modeled as damped mechanical systems. Moon & Lindblom also observed that undershoot effects appear to be most evident in lax vowels, while tense vowels vary in the presence of undershoot. While they attribute this disparity to durational differences between the two classes, the data actually suggest the influence of some other factor. In their study, while F2 values for the tense vowel /i/ varied little according to duration, evidence of undershoot in the diphthong /eɪ/, in the form of a correlation between F2 values and vowel duration, varied across speakers, even though /i/ and /eɪ/ had similar durational ranges for each speaker. Observed differences across speakers in the application of undershoot suggest that undershoot may be a phonological effect, rather than driven by articulatory constraints, as Moon & Lindblom argue.

Vowel centralization and undershoot sometimes predict similar reduction outcomes (e.g., centralization and undershoot of [pap] both predict a lower F1 at short durations), but in some cases it is possible to distinguish them methodologically (e.g., vowel centralization of [pɪp] at short durations would result in a higher F1, where undershoot would result in a lower F1). In the literature, the two models of vowel reduction are sometimes characterized as opposed to one another. For instance, Flemming and Johnson (2007) found that the quality of schwa varies considerably, from mid central [ə] to high front [i] or even high back [u], and that vowel reduction does not always involve approximation to the center of the vowel space. Flemming (to appear) takes this finding as evidence for an undershoot model of reduction rather than a centralization model. As the current study considers reduction in stressed vowels, it does not seem wise to assume that reduction operates as found in Flemming's work with schwa; but whether centralization and undershoot are two opposing models for describing reduction is a pertinent question. There appears to be robust evidence in the literature for both kinds of reduction. As such, this study takes as a premise that the two models are not of necessity at odds with one another, and that both kinds of reduction can potentially be observed simultaneously.

2.2. Articulatory vs. social variation

When accounting for vowel reduction, it's important to distinguish 2 kinds of vowel displacement: displacement due to natural phonetic constraints, and displacement due to social variation. In sociolinguistic studies where vowel displacement may be indicative of reduction phenomena or of socially meaningful variation, a correlation between duration and phonetic position could signify 2 separate things, potentially: a causal relationship between the two, or simply a co-varying relationship. In a causal relationship, duration (based on speaking style, stress, and phoneme) controls the range of possible acoustic realizations; this would qualify as reduction. In a co-varying relationship, the speaker has control of both vowel quality and vowel duration – both variables can potentially index social meaning. Lisker (1986) found that several phonetic cues may mark a phonemic contrast (in this case, voicing). Analogously, several phonetic cues, including duration, can simultaneously index social meaning; Fox and Jacewicz (2008) found cross-dialectal (i.e., regional) differences in both spectral quality and duration for English vowels.

Dodsworth et al. (2007) provides a method for distinguishing vowel reduction from social variation. The study investigated the role of duration in two variables in Central Ohio: /o/-fronting, which is a change in progress and shows social stratification by age, gender, and task; and /e/-retraction, which was not thought to be socially constrained in Central Ohio. They assumed that undershoot effects could be observed as the interaction between consonantal environment and duration in a regression model, when including social variable terms in the model. The study found evidence of co-occurring social effects and undershoot effects for both /o/ and /e/. Dodsworth et al. argued for the need to control for undershoot effects in sociolinguistic research, when there are durational differences across speech settings or speakers, in order to avoid erroneously interpreting articulatorily-constrained variation as social variation.

It's very difficult to find conclusive evidence of formant displacements for sociolinguistic variables as reduction, because a relationship between duration and vowel displacement could always reflect social variation. For a variable like /aɪ/, which shows great variation across English-speaking communities, we can't overlook the possibility of socially-oriented durational effects rather than reduction. However, evidence for centralization, in the form of a correlation between duration and degree of gliding, or undershoot, in the form of an interaction between duration and context, is made stronger by the extent to which social factors are controlled for.

There are two main social phonological sources of variation in /aɪ/ that must be considered in the present study: /aɪ/ monophthongization and Canadian Raising (CR). /aɪ/ monophthongization is variable context-dependent reduced production of the vowel's glide. Despite the name, the resulting vowel can be monophthongal (i.e., [ɑ]) or show a reduced glide in the offset (i.e., [ɑᵋ] or [ɑᵆ]). Monophthongization can be evident in both F1 and F2 (Oxley 2009). Among speakers who show /aɪ/ monophthongization, monophthongization is probabilistic, rather than absolute, such that any individual vowel may be fully diphthongal or reduced, but the main predictor of whether a vowel is monophthongized is the following context. The contexts under which monophthongization occurs vary by speaker's regional dialect (Labov et al. 2006) and by other speaker variables such as ethnicity (Thomas 2007), but monophthongization occurs commonly before voiced codas and in open syllables. For Yesler Terrace[3], the Seattle community from which the current study draws its data, Scanlon and Wassink (2010) found evidence of /aɪ/ monophthongization; they found that /aɪ/ tokens before liquid codas were most likely to monophthongize, /aɪ/ tokens before voiceless codas were least likely to monophthongize, and /aɪ/ before nasals, voiced obstruents, and in open syllables monophthongized at intermediate levels. /aɪ/ monophthongization will need to be controlled for in the analysis below.

CR is an allophonic process whereby the height of the /ai/ onset is raised (i.e., has a lower F1), conditioned by coda voicing. /ai/ is realized with a higher onset before a voiceless coda consonant. Depending on the regional dialect, raising may also result in a displacement along F2, though the direction of the F2 displacement isn't consistent across dialects (e.g., backing in Fruehwald 2008 but fronting in Hall 2005). Unlike /aɪ/ monophthongization, CR is often a regular phonological phenomenon. Despite the name, CR is found in many regions of the United States (cf. Moreton and Thomas 2007). The phonetic and phonological motivation for a raised prevoiceless /aɪ/ onset, and even whether CR is truly a "raising" or a "failure to lower", has been the subject of much debate (cf. Schilling-Estes 1996). Moreton & Thomas (2007) contend that CR arises because

> voiceless codas promote assimilation of the /ai/ nucleus to the offglide, while voiced ones promote assimilation of the offglide to the nucleus. Since the offglide is high and the nucleus low, assimilation creates higher pre-voiceless allophones and lower pre-voiced ones (39).

The proposed assimilation is most evident in vowels of short duration. Ostensibly, the proposed assimilation is a universal phonetic effect, and should be observable in the vowel onset in prevoiced contexts even for speakers who do not display phonologized CR in their speech. The assimilation of the nucleus to the offglide is also proposed in Thomas (1995, 2004). The present study examines whether the proposed assimilation is present in a sample of speakers who do not display CR.

Table 1 outlines studies of the diphthong /aɪ/, giving examples of variation in production of the vowel in American English. Note that the vowel glide ranges from fully diphthongal to fully monophthongal, and that the vowel onset may be raised, fronted, or backed.

Table 1. Attested variants of /aɪ/ in American English (adapted from Schilling-Estes 1996)

Region	*Phonetic variants*	*Study*
New England (@1900)	ɐɪ, aɪ	Kurath & McDavid (1961)
Mid-Atlantic states (@ 1900)	aɪ	Kurath & McDavid (1961)
South Atlantic states (@	aɪ, ɐɪ, ɑɪ, ɒɪ, aɛ, aə	Kurath & McDavid (1961)

[3] More details on the sample are below in Methods.

1900)	(/aɪ/ monophthongization)	
New York, Minnesota	aɪ, əɪ ("Canadian Raising")	Vance (1987)
Northeastern Ohio	aɪ, əi ("Canadian Raising")	Thomas (1995)
Western Pennsylvania	aɪ, əi ("Canadian Raising")	Thomas (1991)
Philadelphia	aɪ, əi ("Canadian Raising")	Payne (1980)
Various Southern states	aː~ɑɪ (/aɪ/ monophthongization)	Labov (1994)
African American Vernacular English	aː~ɑɪ (/aɪ/ monophthongization)	Thomas (1995)
New York City	aɪ~ɔɪ	Labov, Yaeger, & Steiner (1972)
Johnstown, Ohio	aɪ~æɪ	Thomas (1995)

2.3. Phonemic status of English diphthongs

The phonemic status of American English diphthongs has been studied as far back as Pike (1947). Linguists are in general agreement that English diphthongs act as long vowels with two elements. For example, Lehiste & Peterson (1961) observed two identifiable steady-state periods[4] in a corpus of American English diphthongs. Additionally, a number of perceptual studies show higher identification rates for diphthongs in a number of varieties of English when formant slope (Assman, Nearey, & Hogan 1982) or offset position (Fox 1983; Hillenbrand et al. 1995; Andruski & Nearey 1992; Hillenbrand & Nearey 1999; Hillenbrand, Clark, & Nearey 2001) information is given, relative to single-point stimuli. But there is still some dispute as to the relevant specification of the two elements of a diphthong, particularly in the second part of the vowel. For example, the reliable steady-state cues found in Lehiste and Peterson (1961) are often absent from the vowel offset (Holbrook & Fairbanks 1962). Morrison and Nearey (2007) review three competing hypotheses of the perceptual cues relevant to identification of vowels with vowel-inherent spectral change (VISC)[5]; the *onset+offset* hypothesis, where the formant values at the end of the vowel are a relevant perceptual cue; the *onset+slope* hypothesis, where the rate of change of formants over time is perceptually relevant; and the *onset+direction* hypothesis, where only the direction of formant movement

---

[4] In a time-by-frequency representation of the vowel formants, a formant's "steady state" is within the time interval where the formants are parallel to the time axis; in other words, a formant's steady state is the point at which slope of that formant is zero.

[5] While theories of VISC are meant to account for features of both nominally monophthongal and truly diphthongal vowels, the current study considers only those studies that have dealt specifically with truly diphthongal vowels.

in F1-by-F2 space is perceptually relevant. While there is not definitive evidence for one model over the others, the literature appears to support the view that, perceptually at least, the *onset+offset* model adequately differentiates English diphthongs. Bladon (1985) compared identification rates for the RP diphthongs [eɪ əʊ aʊ ɔɪ ɪɔ], and found evidence supporting the *onset+offset* rather than the *onset+slope* hypothesis. Jacewicz, Fujimura, & Fox (2003) provide evidence for the *onset+offset* rather than the *onset+direction* hypothesis in a variety of American English. Gottfried, Miller, & Meyer (1993) compared the three hypotheses for American English monophthongs and diphthongs, and found that the *onset+offset* hypothesis had relatively high correct classification rates in two experimental conditions, while the *onset+slope* hypothesis had relatively high correct classification rates in only one of two experimental conditions; the *onset+direction* hypothesis had relatively low correct classification rates in both experimental conditions.[6][7] The present study, then, assumes two targets for diphthongs: the onset target and the offset target.

2.4. /aɪ/ and reduction

This section reviews previous studies comparing /aɪ/ production and vowel duration. It hypothesizes, based on extrapolation from reduction studies of monophthongs, what constitutes evidence for reduction for /aɪ/. Lastly, it discusses possible phonetic contexts that may result in undershoot for /aɪ/ vowels.

Gay (1968) examined the effect of speaking rate on the production of American English /aɪ/.[8] He found that at shorter distances, /aɪ/ offsets did not show prominent steady states. He found that onset formant frequencies did not vary greatly between the fast and slow conditions (i.e., for durationally short and long vowel tokens, respectively). He did find differences in offset formant frequencies between the two conditions, though. While both durationally short and long vowel tokens followed the same trajectory, F1 offsets were higher and F2 offsets were lower in the durationally short tokens, as shown in Figure 2. The short /aɪ/ trajectory resembles a “truncated” long /aɪ/ trajectory, such that short tokens follow the same trajectory as long tokens, but do not reach the extreme values that the long tokens do. Gay’s results suggest an effect of centralization, since the /aɪ/ offsets in durationally short vowels showed more central F1 and F2 values. Gay examined duration differences according to coda consonant voicing, but did not discuss coda voicing differences in formant values.

---

[6] A number of studies also support the *onset+offset* hypothesis for English nominal monophthongs that show vowel-inherent spectral change (cf. Morrison 2008).
[7] Gay (1970) promotes the *onset+slope* hypothesis rather than the *onset+offset* hypothesis, but his methodology and interpretation of the study’s results have been questioned by Bladon (1985) and Morrison (2008).
[8] The study examined diphthongs /ɔɪ aɪ aʊ eɪ oʊ/, but only /aɪ/ is discussed here.

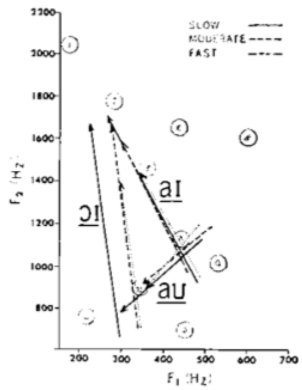

FIG. 2. F1–F2 grid for /ɔɪ, aɪ, aʊ/.

Figure 2. Formant trajectories for /ɔɪ,aɪ,aʊ/ at 3 conditions (from Gay 1968).

Thomas (1995, 2004) looked at the role of duration in production of /aɪ/ in Johnstown, Ohio. He found that the onset F2 of /aɪ/ is fronted at shorter durations for many speakers, ostensibly due to a truncation of the onset. Furthermore, this fronting at shorter durations was stratified by age; younger speakers have fronter /aɪ/ onsets than older speakers. Thomas attributed this difference to a change in progress for a few reasons. The change strongly correlates with fronting of the onset of (ar), and /a/ in Johnstown was undergoing shift as part of the Northern Cities Shift. Also, females lead males in fronting; this is a familiar (though not universal) pattern in linguistic change. Thomas argued that the change in progress is due to phonetically motivated coarticulation, such that "the onset assimilates to the offset" (59), and suggests that the phonetically motivated fronting will become phonologized at a later stage.

However, a correlation between duration and vowel quality in a sociophonetic study is not in itself evidence of reduction in sociophonetic studies, and one could just as easily interpret Thomas' (1995) findings as purely social variation. Johnstown speakers may be fronting the /aɪ/ onset and shortening the vowel simultaneously to signal the same (age-related) social meaning. Individual speakers in the Thomas study showed systematic variation in /aɪ/ onset production, and within-speaker variation might be taken as evidence of articulatory variation; however, such within-speaker variation is also common for changes in progress. Evidence of phonologized fronting in a later generation, where the vowel onset is fronted for tokens of both short and long duration, would be stronger evidence for the phonetically-motivated social variation Thomas proposes; but absent this data, it is not possible to distinguish between articulatorily-constrained vowel shifts and vowel shifts due to social variation in the Thomas study.[9]

[9] Additionally, Thomas' study used tokens from a range of linguistic tasks (reading passage, minimal pairs, conversational data, storytelling), which varied from speaker to speaker and were not balanced for consonantal context; this may be a confounding factor in the analysis.

In order to study reduction in /aɪ/ production, the data need to be constrained, on the one hand, such that there is a suitable range of variation in the data set to observe reduction, and, on the other hand, to control for a number of factors which might otherwise confound the data. To the first point, /aɪ/ tokens should have a wide range of durations; they should show variation in degree of gliding, potentially reflecting both reduction effects and socially meaningful variation; and tokens should have both large and small locus-target distances. To the second point, tokens should be relatively controlled for stress (Nord 1986); tokens should be normalized across speakers, in order to neutralize any physiological differences between speakers (Adank et al. 2004); and any speaker variables should be tested for correlations with duration and environment and controlled for, if necessary.

What constitutes evidence of reduction for /aɪ/, then? Although the above-cited studies of vowel reduction used different experimental conditions to test for reduction – e.g., by varying word stress, varying tasks, and comparing content vs. function words – each study observed a correlation between duration and reduction. Furthermore, Moon and Lindblom (1994) maintains the primacy of duration in undershoot effects. For this reason, the current study examines reduction exclusively as a relation between duration and formant values, and controls for word stress, task, and word class, as much as is possible using conversational data. If reduction effects are robust, I expect to see them independent of any stress, task, or word class effects. Evidence of centralization can be observed as a context-independent effect of duration on vowel position, in the form of a shift towards the center of acoustic vowel space for shorter vowels relative to longer vowels. Evidence of undershoot can be observed as an interaction between duration and environmental context, when regressing vowel position on duration and context. Methodologically, this kind of evidence for undershoot is qualitatively different from that utilized in other phonetic studies, and is necessitated by the fact that the current study investigates centralization and undershoot effects concurrently. Ideally, if both centralization and undershoot effects are present in the data, they should be observable and testable as separate terms in a single regression analysis, as a "main effect" of duration and as an interaction between duration and phonetic context, respectively. It remains, however, that the most definitive evidence for vowel undershoot, rather than centralization, is formant displacement towards a consonantal target and away from the center of the acoustic vowel space.

In the case of undershoot, there are potentially a number of environmental contexts that may contribute to undershoot. The degree of undershoot for any /aɪ/ token should be related to the locus-target distance (Lindblom 1963), and is dependent here on the identity of the coda consonant. This study will consider differences in locus-target distance based on coda place of articulation and coda voicing, and each of these are discussed in turn below.

This study will consider /aɪ/ in two different places of articulation – pre-labial and pre-alveolar – and there is strong evidence that, all else being equal, labial and alveolar consonants have different F2 loci, particularly for vowels in the range of /aɪ/ offsets (i.e., ranging from [ɪ] to [æ]). Delattre et al. (1955) found, using synthetic CV sequences, that consonant place of articulation can be identified from

F2 transitions between the consonant and the ("steady-state") vowel, and identified largely vowel-independent F2 'loci' for [b],[d],[g]. Kewley-Port (1982) confirms, using naturalistic data, that formant transition cues are different for [b] and [d], for front vowels [ɪ e ɛ æ] (see Table II, p. 383). By way of illustration, Delattre et al.'s proposed locus for [b] F2 was 720 Hz, and the proposed [d] F2 locus was 1800 Hz; the mean /aɪ/ F2 at 80% for the data used for this study (before normalization) is 1767 Hz. [show histogram of F2_80 values]. It appears safe to assume a larger F2 locus-target distance for /aɪ/ before labials than for /aɪ/ before alveolars; [10][11] thus, undershoot effects on F2 for /aɪ/ should be observable, dependent on coda consonant place of articulation. F1, on the other hand, should not show any undershoot effects due to place of articulation, because the F1 target values associated with [b] and [d] are similar (Delattre et al. 1955, Kewley-Port 1982).

There is also evidence that, all else being equal, voicing differences in the following consonant contribute to different phonetic realizations of English diphthongs. Thomas (2000) observed that /aɪ/ is raised in pre-voiceless contexts, relative to pre-voiced contexts. Moreton (2004) hypothesizes a general phenomenon, Pre-Voiceless Hyperarticulation, that accounts for patterns associated with coda voicing in both monophthongs and diphthongs. Moreton, looking at English diphthongs (/aɪ ɔɪ eɪ aʊ/), found that prevoiceless vowels are more peripheral than prevoiced ones along both the F1 and F2 dimensions, and that voicing effects increase with proximity to the consonant – that is, that effects at vowel offset are larger than effects at vowel onset, and effects are less consistent at vowel onset. He confirmed that both F1 and F2 at offset serve as perceptual cues to consonant voicing in diphthong-consonant sequences. Moreton attributes these differences to hyperarticulation of the vowel in prevoiceless contexts, and posits that coda voiceless consonants are also hyperarticulated. In terms of locus-target distance, it appears safe to assume that voiceless consonants have more peripheral loci than their voiced counterparts (e.g., lower F1 and higher F2 for alveolars); thus, undershoot effects on both F1 and F2 for /aɪ/ should occur, based on coda consonant voicing. Moreton found the most robust coda voicing effect in the vowel offset; while he did observe coda effects on /aɪ/ onsets, he did not control for

[10] Note that Delattre et al. (1955) and Kewley-Port (1982) dealt with CV transitions, but this /aɪ/ data considers consonant-to-vowel influence in VC# sequences. While CV and VC# sequences do have different consonant place cues, I'd expect the transition cue to be, if anything, more acoustically robust in VC# than CV sequences. The dominant perceptual cue for consonant place of articulation in VC# sequences is the formant transition, while CV cues are more varied (e.g., usually include a release burst) and transitions may be less perceptually salient (due to frication before vowel periodicity) (Ohala 1990).

[11] Note that all the above experiments dealt with voiced stops, and present study looks at other labial-alveolar pairs as well. However, Liberman et al. (1954) found that F2 transitions can serve as consonantal cues for [ptk] and [mnŋ] as well, so there's some evidence that place of articulation is the factor that influences locus-target distance in the studies considered.

Canadian Raising; rather, he asserted that the Canadian Raising clearly found in his sample is a phonologized undershoot effect. I contend that the onset effects Moreton (2004) observed were phonologized CR, rather than the result of a natural phonetic process, and that CR confounds his onset results. I propose that using a sample of speakers that do not exhibit CR would be clearer evidence for the natural phonetic effect in /aɪ/ onsets for which Moreton argues. If Moreton is correct in expecting an effect of coda voicing on onset position, we should see a similar effect of coda voicing on the onsets and offsets of durationally short vowels for speakers whose dialects do not have CR.

2.5. Overview of the present study

This study will examine acoustic evidence of reduction in /aɪ/ production, and considers three hypotheses: (1) that /aɪ/ is subject to centralization independent of context, (2) that /aɪ/ is subject to undershoot based on coda place of articulation, and (3) that /aɪ/ is subject to undershoot based on coda voicing. Each of these is discussed in turn below.

Hypothesizing about how reduction affects a diphthongal vowel requires extrapolating from results in studies on nominally monophthongal vowels. One major difference between reduction effects on monophthongs and diphthongs is the temporal domain of each effect. Centralization and undershoot are expected to operate in different temporal domains. Undershoot effects on vowels are expected to be temporally close to the consonant, as undershoot is explained as gestural overlap between a consonant and an adjacent vowel. In the present study, I take durational effects localized in the vowel offset as evidence of undershoot. Contra Moreton (2004), I do not expect to see undershoot effects in the vowel onset. Centralization effects, on the other hand, are expected to occur throughout the vowel, and should be present in both the vowel onset and offset.

This study will only consider /aɪ/ production in closed syllables. In some dialects, /aɪ/ offset values for tokens in open syllables would be the clearest indication of the unobstructed vowel offset target, because there is no coarticulation with a coda consonant. In these dialects, it would be informative to compare /aɪ/ in other contexts to the “default” /aɪ/ in open syllables, as any excursions from the default production could be interpreted as a consonantal effect. However, /aɪ/ in open syllables is subject to /aɪ/ monophthongization in the community under study, and the difference between /aɪ/ values in open syllables and /aɪ/ values in closed syllables is not as straightforwardly interpretable. As /aɪ/ monophthongization is found to some extent in all phonetic contexts in the data set, I base my predictions on the relative effect of duration on groups of /aɪ/ tokens, rather than assume that vowel targets are directly observable from measured formant values.

I predict centralization throughout the /aɪ/ vowel, along both the F1 and F2 dimensions; all else being equal, vowels with a short duration should show centralization for both the onset and the offset. For the vowel onset, I expect centralization to be most evident along the F1 dimension, as F2 values for /aɪ/ onset are already relatively central in the F1-by-F2 vowel space. For the vowel offset, I expect centralization to be most evident along the F2 dimension, as F1 values for /aɪ/ offsets are already relatively central. The predicted centralization effects are demonstrated in schematic form in Figure 3 below.

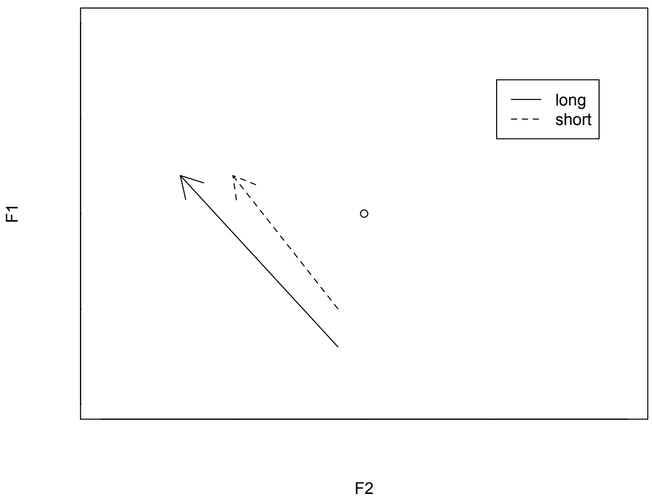

Figure 3. Schematic – predicted effect of centralization

I predict undershoot in the /aɪ/ offset based on coda place of articulation. For long vowels, I expect pre-labial and pre-alveolar /aɪ/ offsets to have similar values. Based on the large locus-target distance between /aɪ/ F2 offset and [b] F2 locus, I predict that the high F2 offset target will not be reached in vowels of short duration, due to increased coarticulation between the offset and the coda consonant. As a result, the vowel offset will show a relatively large durational effect for labials, in the form of a lowered offset. I predict, based on the small locus-target distance between the /aɪ/ F2 offset and [d] F2 locus, a relatively small durational effect for /aɪ/ before alveolars. F2 shifts due to coda place of articulation should give the least ambiguous evidence of undershoot, for two reasons. First, all else being equal, I expect similar offset F2 targets for pre-labial and pre-alveolar tokens at long durations. Second, it is relatively straightforward to control for social factors when looking at coda place of articulation. Speaker variables such as ethnicity and gender can be included as terms in a regression model, and social-phonological phenomena such as /aɪ/ monophthongization and CR can be controlled for by comparing groups with comparable levels of /aɪ/ monophthongization and CR.

This study hypothesizes, then, that centralization and undershoot based on coda place of articulation will be observed simultaneously. While previous studies (e.g., Moon and Lindblom 1994) compared centralization and undershoot as competing hypotheses – i.e., by looking at cases where the theories predicted displacements in opposite directions – the current study contends that we can observe undershoot by comparing across the pre-labial and pre-alveolar contexts, and observe centralization by looking for a consistent main effect of duration.

Durational effects based on coda voicing will also be investigated, though this part of the study will be necessarily more exploratory than the observation of durational effects based on coda place of articulation. It is difficult to make predictions about undershoot based on coda place of articulation. In samples of speakers that show variable /aɪ/ monophthongization, pre-voiceless and pre-voiced /aɪ/ often display different rates of /aɪ/ monophthongization. So, any differences across coda contexts in offset position, for instance, could be due to either reduction or /aɪ/ monophthongization.

With this caveat, I predict undershoot in the /aɪ/ offset based on the voicing of the coda consonant. It is not possible to make predictions comparing the locus-target distances for /aɪ/ in prevoiced and prevoiceless contexts, but predictions about the offset vowel position can still be made, based on the degree to which the coda consonant promotes or inhibits the diphthong's glide. For vowels of short duration in glide-inhibiting contexts, the offset target will not be reached, due to increased coarticulation between the offset and the coda consonant, and the vowel will show a smaller degree of gliding. For vowels of short duration in glide-promoting contexts, I may expect a smaller degree of gliding, but the effect of duration should be smaller than in glide-inhibiting contexts.

From the Pre-Voiceless Hyperarticulation hypothesis and /aɪ/ monophthongization, I assert that the prevoiceless context is a glide-promoting context, and that the prevoiced context is a glide-inhibiting context, relatively speaking; I expect a disproportional effect of duration on the two groups. I expect prevoiceless /aɪ/ to have a higher and fronter offset (i.e., lower F1 and higher F2) than prevoiced /aɪ/ in vowels of long duration. I predict a small durational effect on prevoiceless /aɪ/, and a larger durational effect before voiced codas.

The predictions for the first experiment are summarized in schematic form in Figure 4 below. For each graph, the x-axis represents vowel duration, and the y-axis represents predicted formant values (though the numbers on the y-axis are arbitrary). Predicted centralization is displayed as a non-zero slope; overall, I expect centralization for onset F1 (Figure 4a) and offset F2 (Figure 4d). The figure shows a difference in slope between two conditions when more undershoot is expected in one condition than the other. Figure 4c shows that an undershoot effect based on coda voicing is expected for offset F1, and Figure 4d shows that undershoot effects are expected for both coda voicing and coda place of articulation for offset F2.

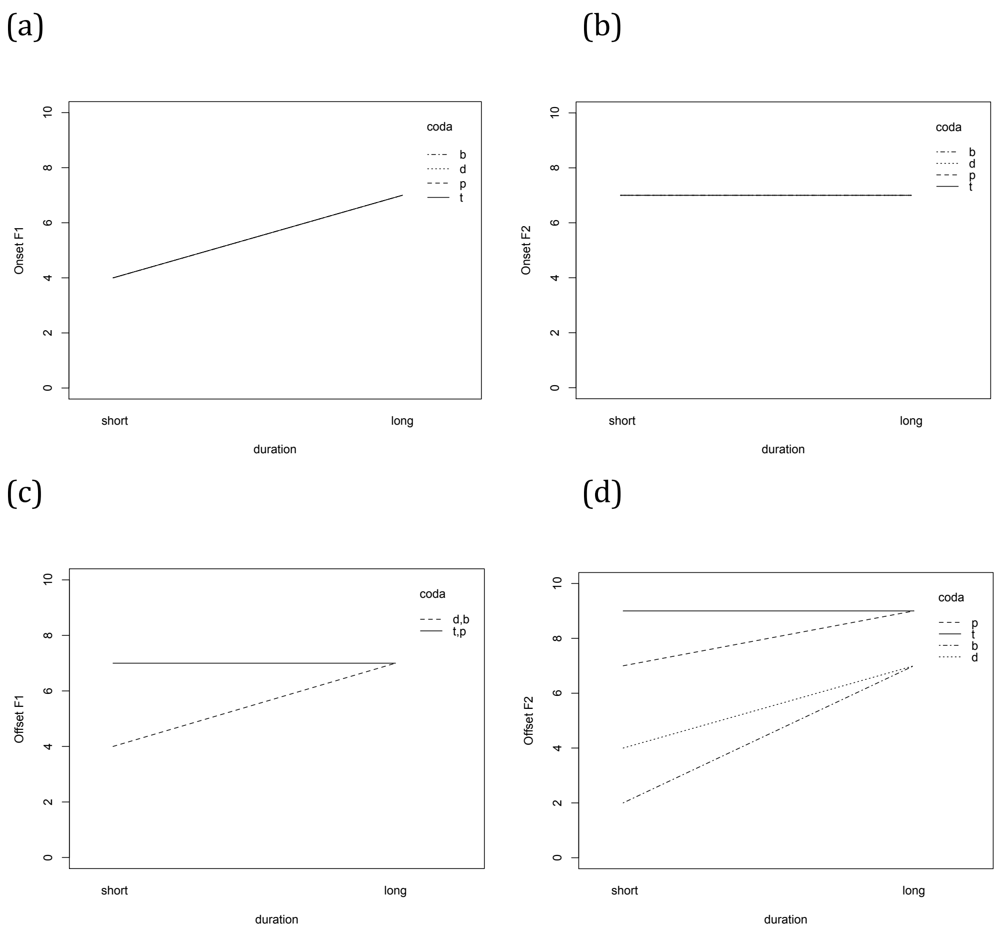

Figure 4. Schematics – predicted durational effects for (a) onset F1, (b) onset F2, (c), offset F1, and (d) offset F2, by coda environment

Any results from the first experiment may be interpreted as either natural phonetic effects, as unexamined social effects, or some combination of the two. The methodology described below accounts for some social variation, such as effects of subject gender and subject ethnicity, and to some extent, /aɪ/ monophthongization. However, Canadian Raising is not accounted for in the methodology of Experiment I, and may confound any /aɪ/ onset results from the first experiment. A second experiment tests whether Canadian Raising is present in the sample. If the sample displays Canadian Raising, we could expect to see evidence of a phonological effect for F1 or F2 at the 20% point, in the form of a raised and/or fronted onset in the pre-voiceless context. It may be, also, that there is a lexical basis for CR, such that particular lexical items show consistently raised and/or fronted onsets. Lexicalized CR based on phonetic environment been found by Vance (1987), Dailey-O'Cain (1997), and Fruehwald (2008). Fruehwald (2008) also notes that CR underapplies at morpheme boundaries. Hall (2005) suggests lexical neighborhood effects have a role in variable raising.[12] If this is the case, we would expect to see one of two kinds of evidence for lexically-based CR: 1) a bimodal distribution for onset F1 values for pre-voiceless tokens, or 2) tokens of a particular lexical item that show consistent extreme formant values along F1, F2, or both.

One implication of previous research on vowel reduction is that the phonetic realization of vowels is context-dependent. /aɪ/ in particular may be subject to relatively high levels of phonetic variation because confusions between diphthongal and nominally monophthongal vowels are rare (Assmann et al. 1982). As a result, it is not clear which, if any, acoustic parameters remain consistent across realizations of /aɪ/, independent of duration. The first experiment will attempt to describe /aɪ/ variation at onset and offset for the sample. However, onset and offset values represent only a small portion of the acoustic parameters associated with the vowel. Onset and offset values say nothing about, for instance, if /aɪ/ onsets and offsets are pronounced steady states, and if overall duration has any effect on the presence of any steady states. Also, onset and offset are approximated in the first experiment by regular intervals in the vowel (i.e., 20% and 80% points of the vowel, respectively); while any observed differences at these intervals are ultimately interpretable, these intervals may not reflect the actual onset and offset targets.

Gay (1968) addressed the question of how duration affects the contour of American English diphthongs, including /aɪ/. Gay concluded that onset target position and F2 rate of change were relatively fixed features of diphthongs, and were independent of duration; offset formant positions, on the other hand, were duration-dependent. Furthermore, he found that while overall vowel duration appeared to correlate with the duration of the onset steady state, the glide portion of the vowel, and the offset steady state for /aɪ/, offset steady states were negligible or not present for short /aɪ/ vowels.

While it is beyond the scope of this study to address these questions exhaustively, a third experiment takes an exploratory look at /aɪ/ formant contours for the sample using Smoothing Spline ANOVAs (SSANOVAs). In its most general use, the SSANOVA is a statistical method that allows for the holistic comparison of curves. Gu (2002) is an in-depth explanation of how SSANOVAs work; Davidson (2006) provides a more intuitive description of the statistical process. This method has been used for comparing formant contours over time in linguistics by Nycz & De Decker (2006) and Koops (2010). Formant data are fitted to a curve in a duration (x-axis) by frequency (y-axis) plane, and 95% Bayesian confidence intervals (i.e., the range within which the mean curve lies, stated with 95% confidence) are constructed around each curve. A significant difference is found between two curves when the confidence intervals of the two curves do not overlap. As an example, Figure 5 shows F1 and F2 for a set of /aɪ/ tokens over time. The x-axis reflects time, and ranges from the 0% point to the 100% point of the vowels. The y-axis displays formant values. So, the figure resembles a spectrogram, except that the two groups are scaled so that one point on the x-axis reflects the same proportional point for each contour. The tokens are divided into 2 groups, long (red) and short (blue). 95%

---

[12] Each of these studies, though, found a lexicalized effect in communities for which raising in the pre-voiceless context was common.

confidence intervals are displayed as dotted lines above and below the main contour. At the 20% point of the vowels, the F1 values for the 2 groups are not found to be significantly different; but at about three-quarters of the way through the vowel, the two contours diverge. Also, in this example there are two distinct steady-state portions in the long F2 contour (i.e. points where the slope of the contour is 0), but the short F2 contour shows a more consistent slope, with no clearly discernable steady states.

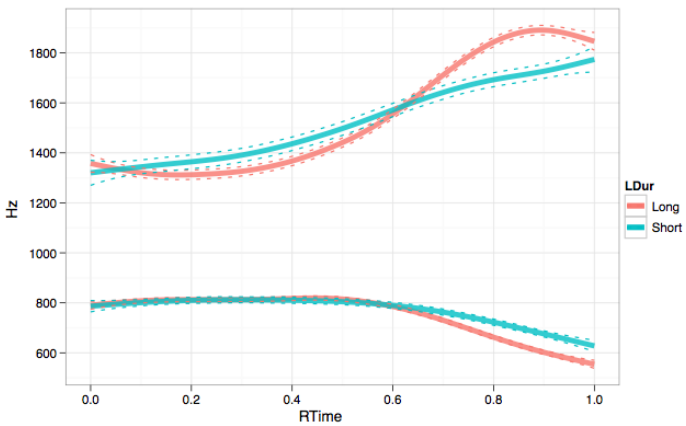

Figure 5. Example – SSANOVA for /aɪ/

The present study uses SSANOVAs as a graphical tool to ask preliminary questions about the effect of duration on /aɪ/ formant contours in the sample. A few caveats pertaining to interpretation are necessary. First, the sample data in the present study reflects /aɪ/ production for a very small community of American English speakers; as such, it would be presumptuous to assert that the formant contour patterns here reflect American English usage as a whole. Also, differences between long and short contours may not be statistically significant, but overall may reflect trends of interest. There may be variation between contours at the 0% or 100% point of the vowel; however, these points of the vowel most likely say more about the flanking consonants than the vowels, so the contour analysis was restricted to contours between the 20% and 80% points of the vowel.

3. Experiment 1

3.1. Methods

838 total stressed /aɪ/ tokens were extracted from unscripted, conversational interviews with 17 Seattle-born speakers. Speakers were controlled for age and social class; speakers were born between 1940 and 1960, and all speakers grew up in a low-income public housing community. No more than 3 tokens of a word were extracted from a single interview. After the restrictions outlined below, 280 tokens were used for the analysis. Tokens were restricted to /aɪ/ in closed syllables. Tokens

with flanking liquids were excluded from the analysis, as were tokens in coda environments with low token counts. The tokens were relatively controlled for stress, as they all came from stressed syllables; however, there is likely some variation in stress, due to factors such as focus.

For the analysis, tokens with following labials ([b p m v]) and alveolars ([d t n z]) were examined. Each token was judged as either fully diphthongal or reduced[13]. Vowels with labial codas were compared to vowels with the corresponding alveolar coda (i.e., pairs with corresponding manners of articulation and voicing were compared – e.g., /aɪb/ compared with /aɪd/, /aɪp/ compared with /aɪt/, etc.) using chi-square tests of independence, to ensure that the study compared groups with comparable rates of /aɪ/ monophthongization. These tests ensure that /aɪ/ monophthongization is a controlled factor when analyzing Where at least one cell in the chi-square table had a count smaller than 5, Yates' continuity correction was applied. (Tallies and results are found in Appendices I and II.) Nasals had to be excluded from the analysis, because /aɪm/ was found to be reduced at a different rate than /aɪn/ (χ^2=10.994, p=0.001). /aɪ/ preceding voiced stops [b d] did not show differences in reduction from /aɪ/ preceding voiced fricatives [v z], so voiced stop and fricative environments were pooled for the analysis.

F1 and F2 values were measured for each token at the beginning of the vowel segment, the 20%, 50%, and 80% points of the vowel segment, and end of the vowel segment. Duration for each token was measured. Tokens were divided along the mean duration into durationally "short" and "long" categories, for graphing purposes and for the SS-ANOVA analyses.

The data were normalized using Nearey's (1978) formant-intrinsic method. All formant data were log-transformed. Then, the log-mean for each formant (F1, F2), for each speaker, was calculated. The log-transformed formant values were subtracted by the relevant log-mean (e.g., F2 values for a particular speaker were subtracted by the F2 log-mean for that speaker – what Wassink (2006) refers to as a speaker-intrinsic grand mean). This normalization method scales the data to account for human physiology. The logarithmic scale reflects more accurately than a linear (Hz) scale the sensitivity of the human ear to changes in frequency (e.g., humans perceive a change of 100 Hz in a low-frequency range as a larger change than 100 Hz change in a high-frequency range). The range of formant values for any two speakers may be different, due to differences in vocal tract physiology; the log-mean formant values are used as a correction factor, making the center of acoustic space (i.e., mean F1 by mean F2) for all speakers equivalent, and defining the formant data in terms of distance from the speakers' respective centers of acoustic space (i.e., the center of each speaker's formant space is represented by the normalized coordinates [F2=0, F1=0]).

[13] Here I assume that perception compensates for coarticulatory effects, after Liberman et al. (1954), Beddor et al. (2002); judgments, then, are evaluated according to their phonetic context, and should reflect the degree of gliding for the vowel itself, rather than consonantal effects on the vowel. Note, though, that Liberman et al. (1954) tested for consonantal identification rather than vowel identification, and Beddor et al. (2002) tested for vowel-vowel coarticulation.

Duration was found to correlate with speaker ethnicity (t=-2.125, p=0.034) and speaker gender (t=2.315, p=0.023) in the data set; to control for confounding speaker effects, ethnicity, gender, and their interactions with duration and following phone were controlled factors in the statistical analysis. For the regression analyses described below, these factors were included in the regressions (however, results for these factors are not discussed below). For the Smoothing Spline ANOVA analyses, these factors were also controlled, using a method described below.

A number of descriptive statistics and graphics were examined. The data set was tested for the presence of CR. Plots of duration by F1 and F2 at 20% and 80%, for the whole sample and for individual speakers, were examined. Plots of duration by F1 and F2 at 20% and 80% were also plotted for each coda context, to compare the relation between duration and formant values across coda contexts.

Four stepwise multiple regressions were performed[14] in order to assess reduction effects on the dependent variables. The dependent variables were F2 at 80%, F2 at 20%, F1 at 80%, and F1 at 20%. For each regression, the following independent variables were included (but due to the stepwise process, the final regression models did not include all of these, and may have differed from one another): duration (as a continuous variable), following environment, speaker gender, speaker ethnicity, and interactions. There was some question as to the proper way to model coda voicing and coda place of articulation in the regressions. The first option is to include 'voicing' and 'place of articulation' dummy variable terms, each with two levels (voiced/voiceless and labial/alveolar); the second is to include a variable for coda phone containing 4 levels (b/v, d/z, p, t). It is unclear which model is more parsimonious, and thus which model offers the best generality of conclusions. The best-fitting model was selected using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) statistics. For both statistics, a smaller value indicates a better model. A difference of 10 indicates very strong evidence for one model over another (Raftery 1995). For all regressions, the models using dummy variables for coda voicing and coda place of articulation were as good or better a fit than those using dummy variables to code each coda phone.[15] Of particular interest for hypothesis testing are main durational effects, main voicing effects, main place of articulation effects, interactions between duration and coda place of articulation, and interactions between duration and coda voicing. There were a total of 5 factors x 4 regression models = 20 statistical tests performed. Due to the large number of statistical tests performed, the likelihood of obtaining a Type I error is much larger than if only one hypothesis were tested, when using the standard $p \leq 0.05$ to infer a significant effect. To ensure that the odds of making a Type I error remains at .05 for all tests combined, a Bonferroni

[14] Using the `step` and `lm` functions in the R statistical environment (R Development Core Team 2010).

[15] Models using the log of duration were also considered, because duration showed a skewed distribution; the models using duration as an independent variable were as good or better a fit than those using the log of duration.

correction was applied, such that a result is inferred as significant when $p \leq 0.0025$. (Near-significant values are reported where $p \leq 0.005$.) Predicted values for each of the dependent variables, based on the regression results, were calculated for the purpose of graphing /aɪ/ trajectories predicted by the statistical tests. Predicted values were controlled for the social factors of subject gender and subject ethnicity, so that the resultant graphs represent the trajectories the models predict for African American males.[16] The predicted values were calculated to highlight duration, coda place of articulation, and coda voicing effects. The graphed trajectories will be of the most use in explaining any interactions in the data.

3.2. Results

3.2.1. Descriptive statistics

The tokens ranged from 73 to 480 ms, with a mean of 192 ms; the durational distribution was skewed right, as shown in the histogram in Figure 6.

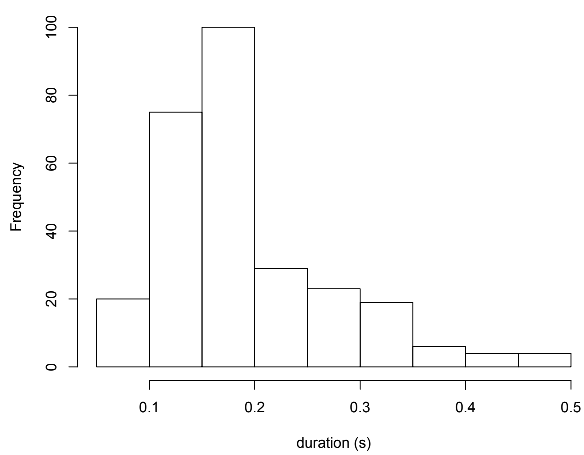

Figure 6. Histogram of sample duration

Tokens were divided along the mean duration into durationally "short" and "long" categories. Table 2 shows the distribution of short and long tokens by following phone.

[16] African American males were used as the controlled social group for these graphs because there was a sufficient number of tokens spoken by the group in each coda context to make the resultant trajectories visually informative; otherwise, the selection of African American males here has no significance.

Table 2. Distribution of short and long tokens, by coda context

following phone	*duration*		*Total number of tokens*
	short (<192 ms)	*Long (>192 ms)*	
b	3	14	17
d	42	40	82
p	1	13	14
t	18	67	85
v	18	19	37
z	24	21	45

Note the paucity of short /aɪ/[p] tokens sampled; the analysis included /aɪ/[p] tokens, but due to low token counts, descriptive analyses including /aɪ/[p] cannot be considered representative of the population. (The inferential statistics are more accurate in this respect, as they deal with duration as a continuous variable.)

Figure 7 shows the position of /aɪ/ onsets and offsets, relative to the reference vowels /i a u/, in a normalized F1-by-F2 plot. Reference vowels are plotted in black, and /aɪ/ values are plotted in blue. The mean for each vowel class is represented by a large letter in a lighter shade (gray for reference vowels, light blue for /aɪ/). The point where both normalized F1 and F2 are zero represents the center of the vowel space. The left side displays the position of /aɪ/ onsets, and the right side displays the position of /aɪ/ offsets. /aɪ/ onsets, as a whole, appear raised and fronted relative to the /a/ vowel. Normalized onset F1 values range from 0.77 (a low vowel close to [a]) to -0.10 (a mid-high vowel between [ə] and [ɨ]), with a mean of 0.41. Normalized onset F2 values range from -0.42 to 0.32, with a relatively mean of -0.02. /aɪ/ offset positions vary greatly, but range between the position of /aɪ/ onsets and /i/. Normalized offset F1 values range from 0.64 to -0.28, with a mean of 0.23. Normalized offset F2 values range from -0.04 to 0.59, with a mean of 0.21. The mean offset approximates the cardinal [æ] or [ɛ] vowel.

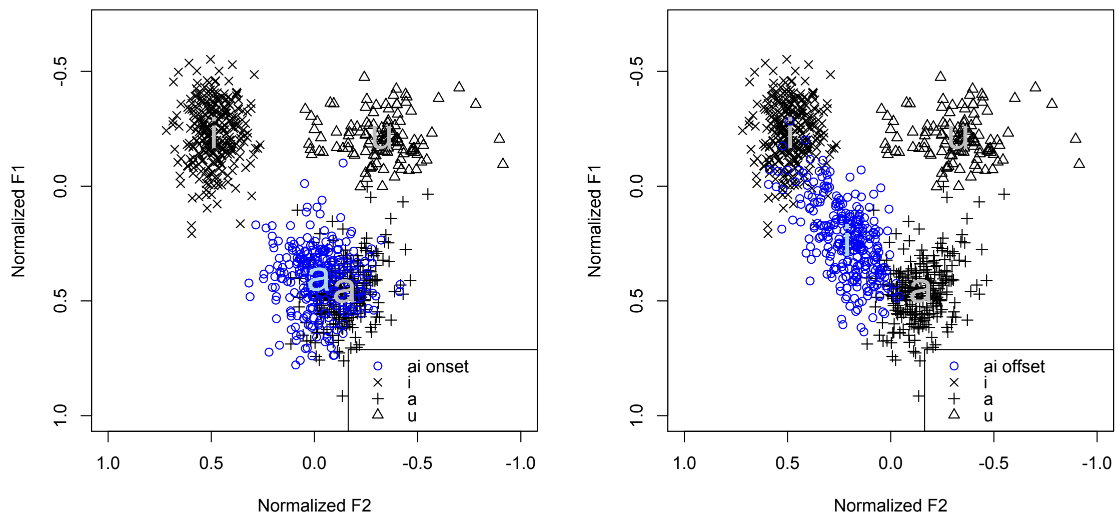

Figure 7. Normalized F1-by-F2 plots of /aɪ/ onsets (left) and offsets (right), with reference vowels

3.2.2. Regression results[17]

A regression of F2 position at the 80% point of the vowel found a significant positive effect of duration (p=0.0000); no significant main effect of voicing; a significant effect of place of articulation (p=0.0018); no significant interaction between duration and coda place of articulation; and a significant interaction between duration and coda voicing (p=0.0000).

A regression of F2 position at the 20% point of the vowel found no significant effect of duration; no significant main effect of voicing or place of articulation; no significant interaction between duration and coda place of articulation; and no significant interaction between duration and coda voicing.

A regression of F1 position at the 80% point of the vowel found no significant effect of duration; no significant main effect of voicing or place of articulation; no significant interaction between duration and coda place of articulation; and a significant interaction between duration and coda voicing (p=0.0000).

A regression of F1 position at the 20% point of the vowel found a significant positive effect of duration (p=0.0000); a near-significant effect of voicing (p=0.0040); no significant effect of place of articulation; no significant interaction between duration and coda place of articulation; and no significant interaction between duration and coda voicing.

A summary of the regression results is displayed in Table 3.

Table 3. Summary of regression results

[17] Full regression results are given in Appendix III.

Test (regression)	*Main duration effect?*	*Main coda voicing effect?*	*Main coda POA effect?*	*Interaction b/t duration and coda POA?*	*Interaction b/t duration and coda voicing?*
F2 at 80%	Yes (p=.0000)	No	Yes (p=.0018)	No	Yes (p=.0000)
F2 at 20%	No	No	No	No	No
F1 at 80%	No	No	No	No	Yes (p=.0000)
F1 at 20%	Yes (p=.0000)	Near-sig. (p=.0040)	No	No	No

Figure 8 demonstrates the significant main effect of duration on F2 at offset. Vowel duration is plotted on the x-axis, and normalized F2 is plotted on the y-axis. The left side of the figure shows all tokens in the sample, as well as a trendline for the data. The right side of the figure shows all tokens, broken down by speaker.

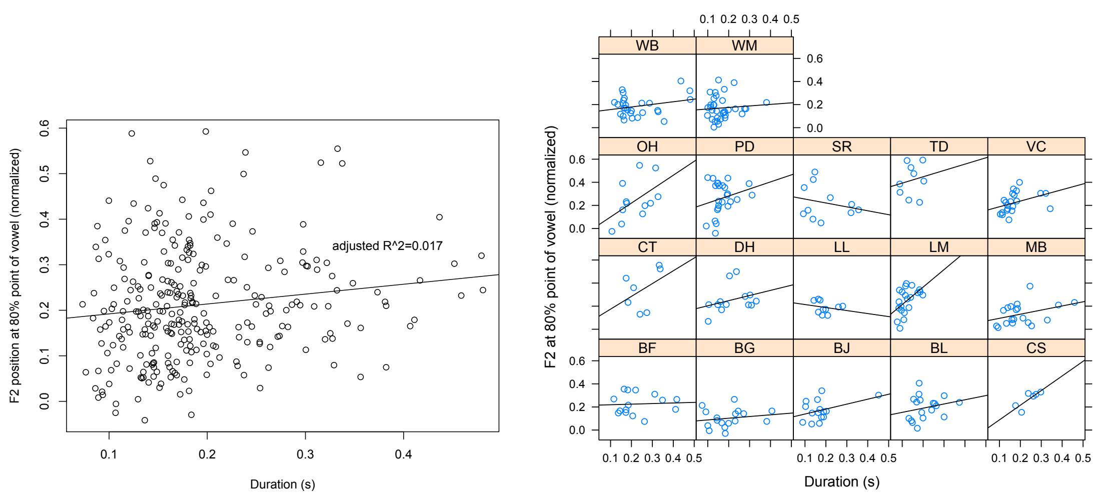

Figure 8. Scatterplots of duration by offset F2, for whole group and individual speakers

Overall, there is a positive correlation between duration and offset F2, though the size of the effect is small. Individual speakers vary greatly in the significance and size of durational effect, but the majority of speakers (15 out of 17) show a positive correlation between duration and F2 position.[18] When these plots are broken down by coda context, the significant main effect of place of articulation and interaction

[18] Trendlines for individual speakers most likely vary because the tokens for each speaker were not balanced for coda context.

between duration and coda voicing are demonstrated, as shown in Figure 9 below.[19] The graph on the left shows the relation between duration and F2 position at 80% for prevoiced tokens, and the graph on the right shows the relation for prevoiceless tokens. Labial and alveolar tokens and trend lines are plotted in black and red, respectively.

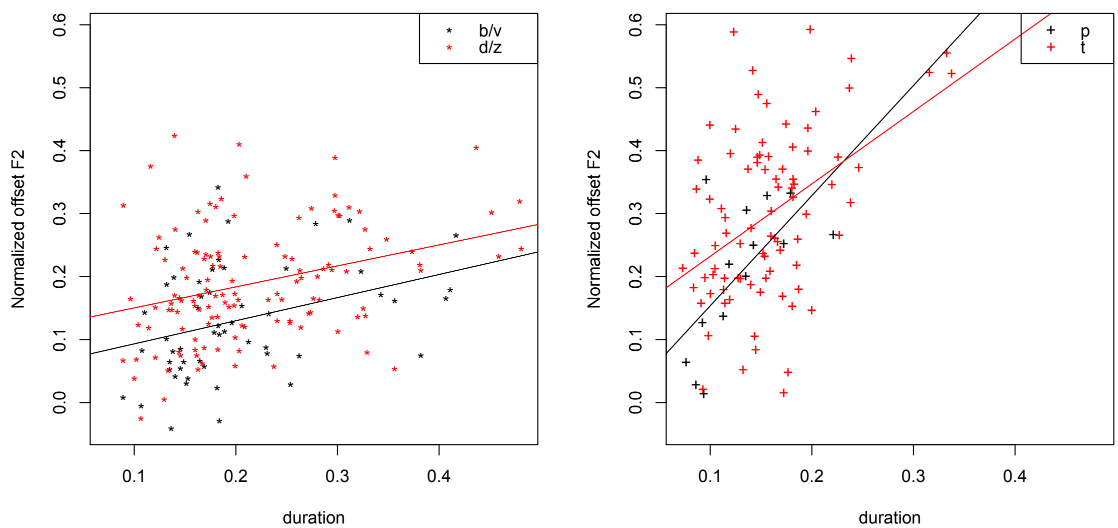

Figure 9. Scatterplot of duration by offset F2, by coda environment

Both groups show significant correlations between duration and gliding (pre-voiced: r=.34; pre-voiceless: r=.49). However, pre-voiced tokens have a more "flat" distribution (i.e., a smaller-size effect of duration on gliding), while the pre-voiceless group shows a steeper slope (i.e., a larger-size effect of duration on gliding). The voiced tokens show the clearest difference between the pre-labial and pre-alveolar conditions: pre-labial and prealveolar tokens are showing similar durational effects, as evidenced by the slope of the trendline for each, but pre-alveolar /aɪ/ has a higher F2 than pre-labial /aɪ/. (The small number of /aɪ/ tokens before /p/ makes comparison of the voiceless pre-labial and voiceless pre-alveolar groups tenuous.)

Figure 10 demonstrates the significant interaction between duration and coda voicing at F1 offset. Again, the graph on the left shows the relation between duration and normalized F1 for prevoiced tokens, and the graph on the right shows the relation for prevoiceless tokens. Labial and alveolar tokens and trend lines are plotted in black and red, respectively.

[19] A caveat: tokens from White and Black speakers have significantly different durations, particularly for pre-voiceless tokens; ethnicity may be a confound for this data. The inferential statistics below will account for subject ethnicity, however.

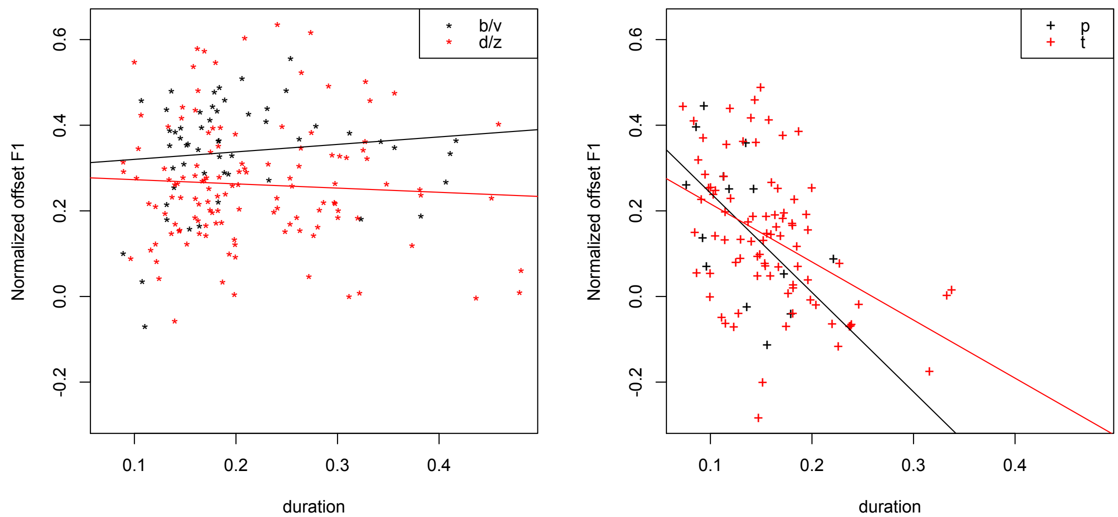

Figure 10. Scatterplot of duration by offset F1, by coda environment

No significant main effect of duration on offset F1 was found. From the left side of the figure, we can see that there is no relation between duration and F1 position for pre-voiced tokens, but prevoiceless tokens show a negative correlation between duration and F1 position (r=-.451). In other words, in the prevoiceless context, shorter vowels have a lower offset F1 than longer vowels. Though there appear to be slight differences between pre-labial and pre-alveolar trend lines, these differences are not found to be significant.

Figure 11 demonstrates the significant effect of duration on onset F1 values. The left side of the figure shows all tokens in the sample, as well as a trendline for the data. The right side of the figure shows all tokens, broken down by speaker.

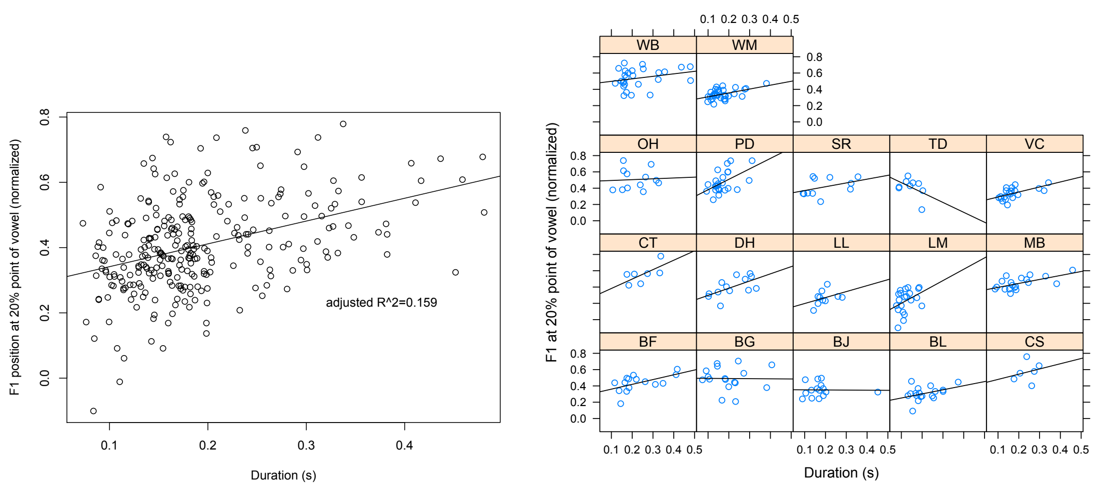

Figure 11. Scatterplots of duration by onset F1, for whole group and individual speakers

Overall, there is a positive correlation between duration and F1 position. Individual speakers vary greatly in the significance and size of durational effect, but the majority of speakers (14/17) show a positive correlation between duration and F1 position.

Figure 12 visually summarizes the "main" effect of duration found in the regression analyses, in a normalized F1 (y-axis) -by-F2 (x-axis) vowel space. It plots the predicted trajectories for long and short tokens, for the entire data set. For each trajectory, the length of the arrow indicates degree of gliding; the symbol reflects the mean 20% point of the vowel, and the arrow head reflects the mean 80% point of the vowel. The long vowel groups are shown in blue, and the short vowel groups in red. The left figure plots predicted trajectories for individual tokens; the right figure plots the mean predicted trajectories for durationally long and durationally short vowels. The right figure demonstrates that shorter vowels have a lower onset F1 and a lower offset F2.

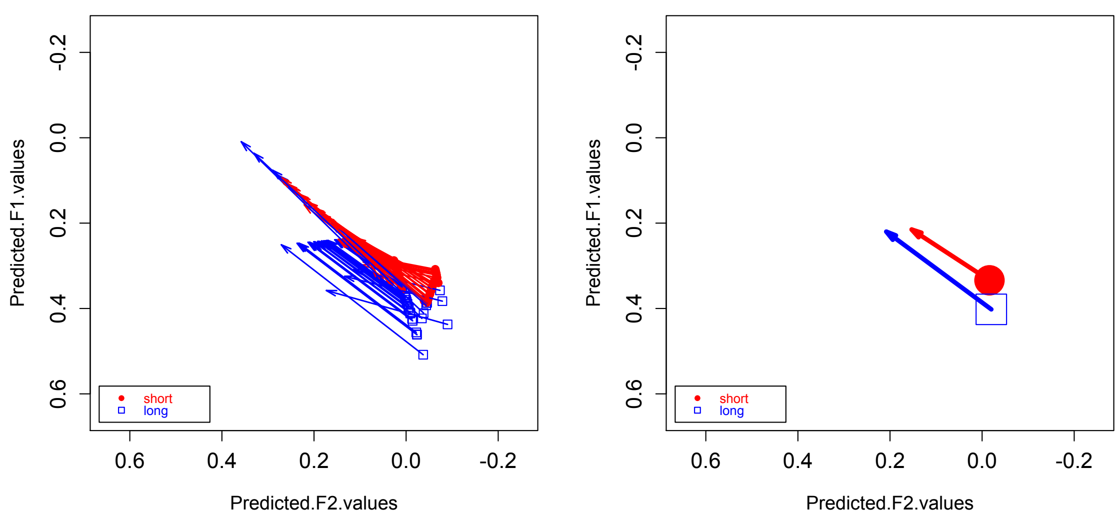

Figure 12. Predicted trajectories by duration, for whole sample

Figure 13 visually summarizes the main durational effects and interactive effects for duration and coda place of articulation, in a normalized F1-by-F2 vowel space. It shows plots of the predicted mean trajectories for /aɪd/ and /aɪb/ tokens, for long and short vowels. The /aɪb/ trajectories are represented by red arrows, and /aɪd/ trajectories are represented by blue arrows. Long trajectories are represented by solid lines, while short trajectories are represented by dashed lines. Again, the plot displays the same main durational effects as above: shorter vowels have a lower offset F2 and a lower onset F1. The plot also displays a significant differences between the /aɪb/ and /aɪd/ trajectories. There are no differences in the durational effects on the two groups, but /aɪb/ offset F2s are lower than their corresponding /aɪd/ F2s. (It appears that the two groups have different onset F2 values, but this difference was not found to be significant.)

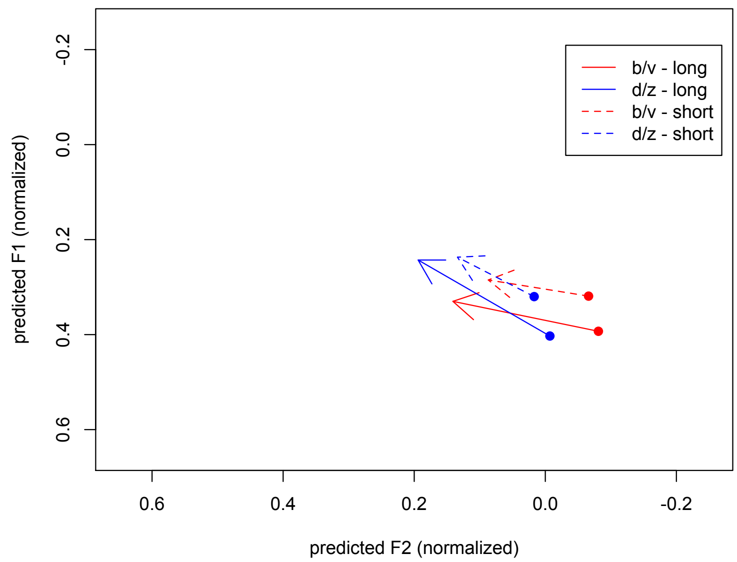

Figure 13. Predicted trajectories by duration, for prelabial and prealveolar voiced tokens

Figure 14 visually summarizes the main durational effects and interactive effects for duration and coda voicing, in a normalized F1-by-F2 vowel space. It displays plots of the predicted mean trajectories for /aɪd/ and /aɪt/ tokens, for long and short vowels. The /aɪd/ trajectories are represented by red arrows, and /aɪt/ trajectories are represented by blue arrows. Long trajectories are represented by solid lines, while short trajectories are represented by dashed lines. There are main durational effects for offset F2 and onset F1, such that shorter vowels have a lower offset F2 and a lower onset F1. The plot also displays significant differences between the /aɪd/ and /aɪt/ trajectories. There is an interaction effect on offset F2, such that durational differences have a larger effect on /aɪt/ than /aɪd/. /aɪt/ shows a durational effect for offset F1, but /aɪd/ does not. No significant durational effects were found for F2 at onset.

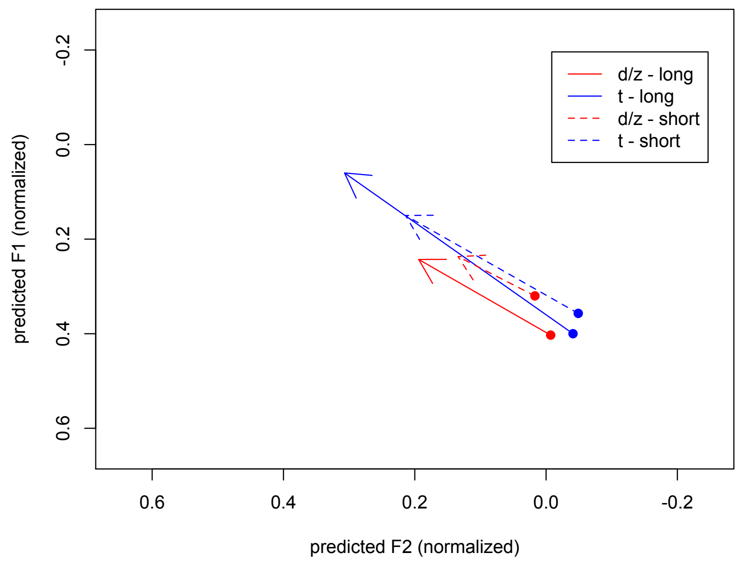

Figure 14. Predicted trajectories by duration, for prealveolar voiced and voiceless tokens

3.3. Discussion

I should be careful to say this study's null results don't necessarily indicate a lack of POA undershoot in the data; the data aren't nearly as controlled as a number of more experimental studies. For instance, the data does not control for preceding phone, and words from which tokens were extracted have varying degrees of focus. However, there is not much evidence for the predicted undershoot effects on the offset in this data set, based on locus differences in place of articulation. Duration appears to affect pre-labial and pre-alveolar /aɪ/ in comparable ways; in both contexts offset F2 is lower and onset F1 is lower. Nor do social factors appear to confound the data. The speaker variables of gender and ethnicity were controlled in the statistical analysis. Only pre-labial and pre-alveolar groups with comparable levels of /aɪ/ monophthongization were compared.

This study does find some evidence of centralization. The significant main effect of duration, in the offset for F2 and in the onset for F1, strongly suggests centralization. These durationally-based displacements are in the direction expected for centralization. The durational effects occur throughout the vowel, i.e., in the temporal domain hypothesized for centralization. The durational effects are found in a sample of vowels in stressed syllables, so it is not likely that word stress plays a role in the patterns observed. Lastly, the durational effects are found for offset F2 and onset F1, which occupy relatively peripheral positions in the vowel space, but not for offset F1 and onset F2, which occupy relatively central positions in the vowel space; these results conform with the hypothesis that relatively central vowel formants will not show durationally-based displacements.

In light of these results, one could make a strong or weak conclusion about centralization versus undershoot theories. The strong conclusion would be that centralization occurs and undershoot does not. While I do think there is strong support for centralization in the current study's results, the current study does not offer a strong rejection of the undershoot theory, especially given the strong past research on undershoot. The current study seems to support a weaker conclusion, in support of centralization and ambiguous about undershoot.

There is strong evidence of an interaction between duration and voicing of the following segment at vowel offset. As predicted, in durationally long vowels, prevoiceless /aɪ/ had a higher and fronter offset than prevoiced /aɪ/. Counter to the prediction, though, duration had a larger effect on prevoiceless /aɪ/ than on prevoiced /aɪ/, along both the F1 and F2 dimensions. It seems unwise to characterize this interaction as undershoot, for two reasons. First, undershoot predicts a larger effect on prevoiced than prevoiceless /aɪ/. Second, the durational effect on offset F1 for prevoiceless /aɪ/ – such that vowels of shorter duration have a higher F1 – is in the opposite direction predicted by undershoot. Nor is the prevoiceless F1 offset displacement in the direction expected for centralization. Two possible explanations for the unexpected results are considered.

First, it's possible the observed displacements are socially-oriented displacements, in particular, due to differing levels of /aɪ/ monophthongization in the short prevoiceless versus long prevoiceless context. The criticisms leveled at Thomas (1995, 2004) are mitigated somewhat; there is no reason to assume a change in progress, and a number of major speaker variables were controlled for in the data collection methodology (age, social class to some extent) or accounted for in the statistical methodology (speaker gender and ethnicity). But the presence of higher levels of monophthongization in short tokens would result in less gliding along both F1 and F2, and would account cleanly for the observed pattern of a higher offset F1 and lower offset F2. As seen in Table 4, short vowels are disproportionately monophthongal. Thus, it seems plausible that /aɪ/ monophthongization acts as a confound for these results.

Table 4. Tally of auditory judgments, by coda context and duration

Coda context	*Duration*	*Coded as diphthongal*	*Coded as monophthongal*
prevoiceless	short	73 (91%)	7 (9%)
	long	19 (100%)	0 (0%)
prevoiced	short	64 (68%)	30 (32%)
	long	80 (92%)	7 (8%)

Second, it's possible the differences between /aɪd/ and /aɪt/ trajectories are due to an effect of F1 and F2 cutback, where the trajectory into the following consonant is shortened by either a preglottalized or glottalized release. We might expect a fair amount of word-final [t] glottalization, but don't know much about word-final [d] glottalization for Pacific Northwest English speakers, especially for African Americans (cf. Koops and Niedzielski 2009 for glottalization in Houston). Vowel tokens used for this study were coded for aperiodicity during the vowel ("creaky voice"), but not for coda glottalization. It's certainly possible that cutback occurs disproportionately in /aɪt/ tokens; it also could be that shorter durations result in more gestural overlap between the vowel offset and the coda glottalization, such that the glottalization masks movement towards the offset target. Crucially, this explanation accounts for the prevoiceless offset F1 pattern. Ultimately, though, the relationship between glottalization and duration needs to be more thoroughly investigated.

In either the first or second case, a social and phonological difference between contexts acts as a confound for the data. Considering either account, there is not much evidence for undershoot based on coda voicing.

Why didn't this study find evidence of undershoot, either due to coda place of articulation or coda voicing? A few possibilities are suggested by Moon & Lindblom (1994). Since /aɪ/ is a tense vowel, it may only appear in vowels of relatively long duration, such that it is not subject to undershoot. It's also possible that there are situations where undershoot would not be expected for /aɪ/, as speakers can control the degree of undershoot somewhat when there is a functional reason to do so. However, Moon & Lindblom are vague about what constitutes a functional reason to avoid undershoot. There are a few studies that have taken up this question. For example, Pearce (2008) shows that duration-dependent centralization does not occur in domains where there is vowel harmony, for a number of Chadic languages. Shirai (2005) found undershoot effects in Japanese for /a/ but not for /o/ or /e/, suggesting vowel height regulates undershoot somewhat for Japanese. Here, however, it is not clear what circumstances motivate an increase in articulatory force or speed of response to motor commands. Another possibility is that the undershoot effects observed in experimental conditions do not occur in conversational speech. Moon & Lindblom (1994) suggest that the articulatory effort of a speaker may be task-dependent. Although there is recent interest in using conversational speech corpora to investigate phonetic phenomena, the relation between experimental findings and conversational speech is not clear for many such phenomena (Byrd 1994). Lastly, it could be that Moon & Lindblom's acoustic formulation of undershoot is an oversimplification of a complex articulatory phenomenon. It may be that the relationship between articulator position and acoustic output is not necessarily as straightforward as Moon and Lindblom's model of the tongue as a damped mechanical system suggests. For example, Johnson (2003) notes that both differences in consonant locus positions and vowel formant patterns are due to properties of the vocal tract cavities, and suggests that a two-tube (Helmholtz resonator) model is more appropriate for describing F2 variation than a model of the tongue as a damped mechanical system. At the very least, Moon & Lindblom's model may not be appropriate for examining the effect of labial obstruents on vowels, because tongue position operates relatively independently of the jaw and lips in labial constrictions.

While the reasons for the observed /aɪ/ offset patterns based on coda voicing are ambiguous, onset patterns for prevoiced and prevoiceless /aɪ/ are more straightforwardly interpretable. Again, one should be careful not to attach too much importance to a null result; but this study does not support Thomas' (1995, 2004;

Thomas and Moreton 2007) assertion that the vowel onset assimilates to the vowel offset for durationally short prevoiceless diphthongs. The study did find lower F1 onset values for vowels of short duration, which could be interpreted as assimilation to the offset; but the more parsimonious explanation would be centralization, as centralization is a well-documented phenomenon, and serves as an explanation for observed shifts in both F1 and F2. If we consider displacements along F2, we see offset displacements for /aɪt/, but observe no onset displacements at all. The assertion that one portion of a diphthong may assimilate to another portion of the same diphthong is a unique claim within the phonetics literature; its lack of utility as an explanation for patterns in the present study sheds some doubt on its usefulness as a general phonetic theory. Along similar lines, the study finds no evidence of either phonological or phonetically-based CR; these results suggest that Moreton's (2004) finding of lowered onset F1s for prevoiceless /aɪ/ may be due to phonological CR, and may not reflect a natural phonetic process of assimilation to offset.

4. Experiment 2

4.1. Methods

Boxplots of normalized F1, F2, and F2-F1 at vowel onset were examined, to visually investigate differences in formant values based on coda voicing. T-tests were conducted to test for statistical significance of any differences between onset formant values in the pre-voiced and pre-voiceless conditions. Histograms of onset formant values were examined, in order to visually investigate the distribution of formant values, and Shapiro-Wilk tests of normality were conducted. Tokens with extreme F1 or F2 onset values were identified, and the distribution of tokens extracted from the same lexical entry were visually examined.

4.2. Results

Figure 15 shows normalized F1 at the 20% point of the vowel for pre-voiced and pre-voiceless tokens; so, an unraised [aɪ] vowel might have an onset F1 around 0.4, while the raised variant (e.g., [ʌɪ]) might have an onset F1 around 0.0. Overall, neither group appears to be raised, and the two groups were not found to be significantly different (t=-.1509, p=0.8802).

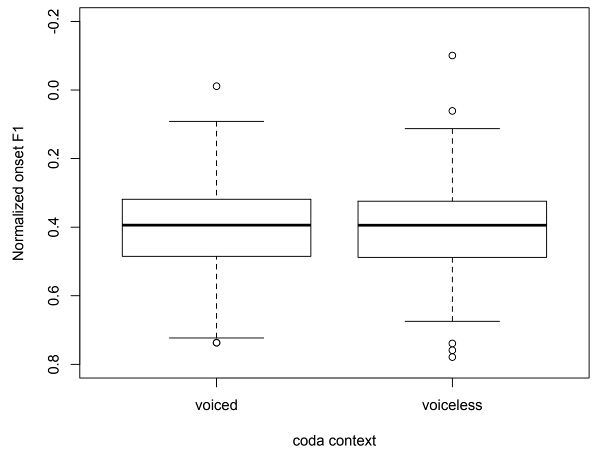

Figure 15. Boxplot of F1 at 20% point of vowel, for prevoiced and prevoiceless contexts

Figure 16 shows histograms for normalized F2 at the 20% point of the vowel for pre-voiced and pre-voiceless tokens. Each group has a central mean onset F2 (i.e., near 0.0). The two groups were not found to be significantly different (t=0.8750, p=0.3826).

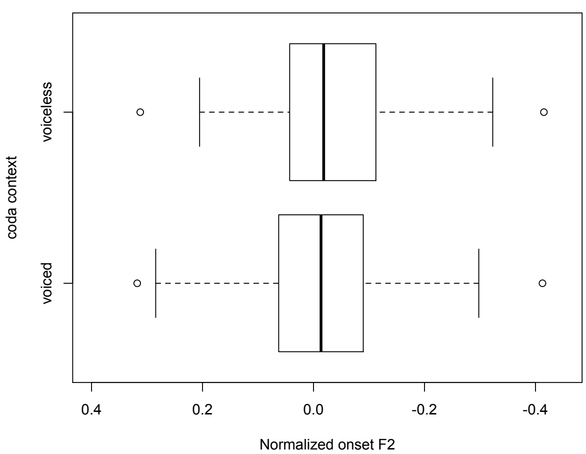

Figure 16. Boxplot of Normalized F2 at 20% point of vowel, for prevoiced and prevoiceless contexts

Figure 17 shows the difference between normalized F2 and normalized F1 at the 20% point of the vowel for pre-voiced and pre-voiceless tokens; the two groups were not found to be significantly different (t=0.6775, p=0.4989).

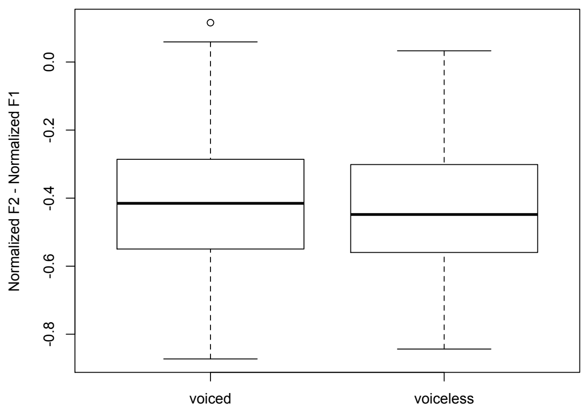

Figure 17. Boxplot of Normalized F2 – Normalized F1 at 20% point of vowel, for prevoiced and prevoiceless contexts

Figure 18 shows histograms for F1 and F2 at onset. While a Shapiro-Wilk normality test found that onset F1 values fall outside a normal distribution (W=0.9879, p=0.0187), the distribution of F1 values is not clearly bimodal either. A Shapiro-Wilk normality test found no evidence that onset F2 values fell outside a normal distribution (W=0.9952, p=0.5430).

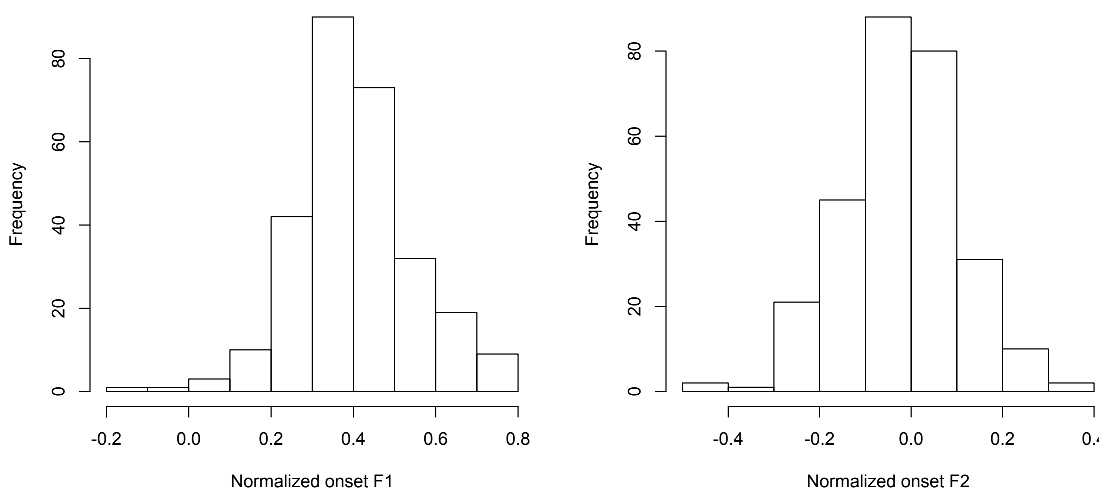

Figure 18. Histograms for normalized onset F1 (left) and normalized onset F2 (right).

There is one extreme onset F1 value for the data set, for the lexical item *invite*, which has a normalized onset F1 of -0.101. Figure 19 displays a normalized F1-by-F2 plot of /aɪ/ onsets; each token is represented in the plot by the word in which the vowel occurred. /aɪ/ in pre-voiced contexts is plotted in black; most /aɪ/

tokens in pre-voiceless contexts are plotted in red. There are three tokens of the word *invite*, plotted in blue. As can be seen in the plot, the three *invite* tokens are not clustered towards the lower end of the range of onset F1 values; instead, they are found among the high, middle, and low values for onset F1. Nor do any of the *invite* tokens show extreme F2 values.

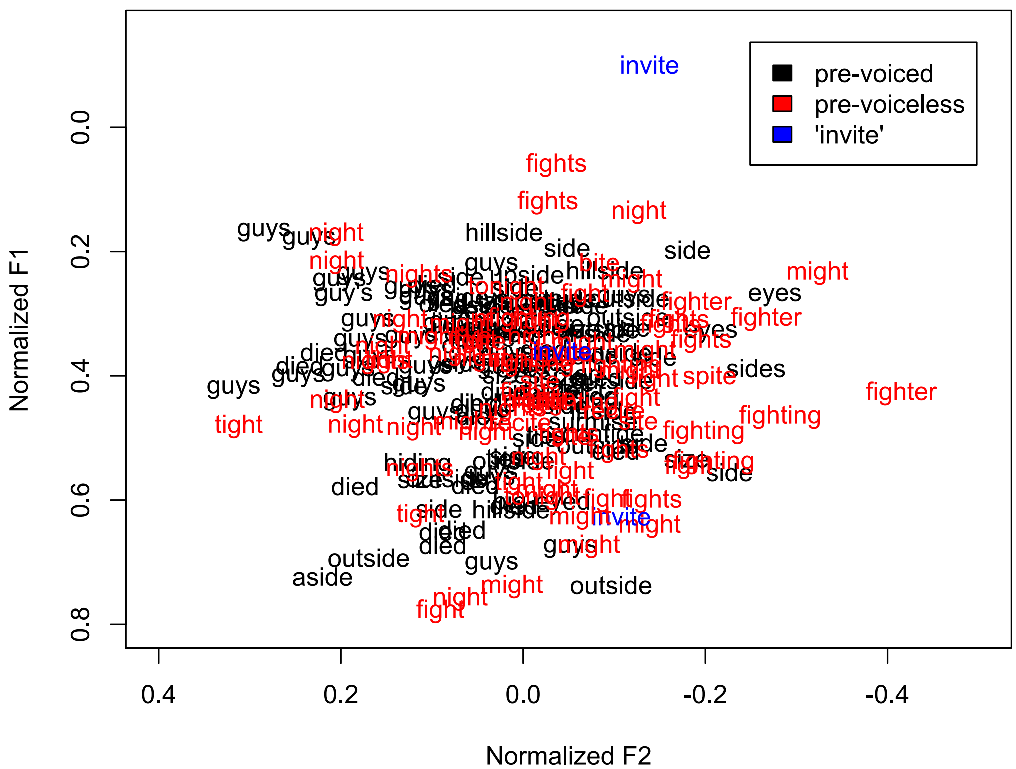

Figure 19. F1-by-F2 plot of /aɪ/ onset values, by voicing

4.3. Discussion

As shown above, pre-voiceless tokens as a whole are not found to have different onset F1, F2, or F2-F1 values than pre-voiced tokens. Additionally, there is no evidence that lexically-influenced CR is in effect for this study's data set. Again, one should be cautious not to infer too much from a null result, but there is not any evidence that the onset results from Experiment 1 are confounded by CR.

5. Experiment 3

5.1. Methods

The /aɪ/ tokens used in Experiment 1 were plotted as SSANOVAs, using the `gss` and `ggplot2` packages in R (R Development Core Team 2010). Smoothed contours were created using F1 and F2 measures at the 0%, 20%, 50%, 80%, and 100% points of each token. F1 and F2 contours were examined separately. To control for effects of coda environment, contours were examined in the pre-voiced labial, pre-voiced alveolar, and pre-voiceless alveolar contexts separately. (Pre-voiceless labial contours were not examined, due to low token counts.) Speaker ethnicity was controlled: only vowel contours from tokens produced by African Americans were examined, because the sample has the highest number of tokens from African American speakers. To control for gender, African American male contours and African American female contours were examined separately.

5.2. Results

Figure 20 shows three F1 contours for African American males, for three coda environments: voiced labial (b/v), voiced alveolar (d/z), and voiceless alveolar (t). Contours and 95% confidence intervals for long tokens are shown in red; contours and confidence intervals for short tokens are shown in blue. The short and long contours in each environment appear similar, and even reach their highest F1 values at similar points of the vowel, proportionally speaking. It appears that /aɪb/v/ tokens reach their highest F1 point relatively late in the vowel, proportionally speaking (i.e., close to the 50% point of the vowel); /aɪt/ tokens appear to reach this point relatively early (i.e., around the 35% point of the vowel). There does not appear to be an F1 offset steady state for any of the groups.

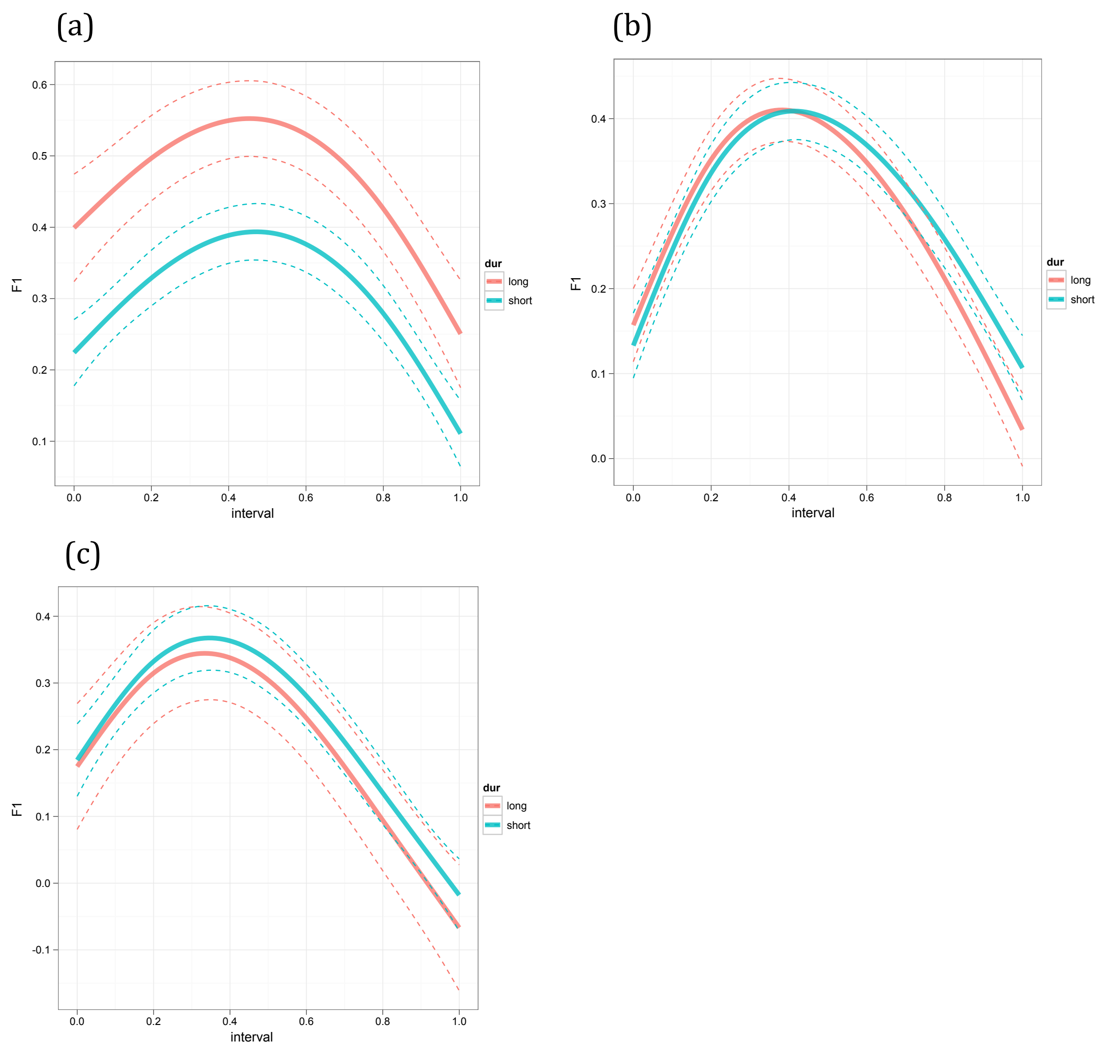

Figure 20. SSANOVAs for F1 in long and short vowels for African American males, in pre-voiced labial (a), pre-voiced alveolar (b), and pre-voiceless alveolar (c) contexts.

As shown in Figure 21, similar F1 contours are found for African American females in the three coda environments: voiced labial (b/v), voiced alveolar (d/z), and voiceless alveolar (t).

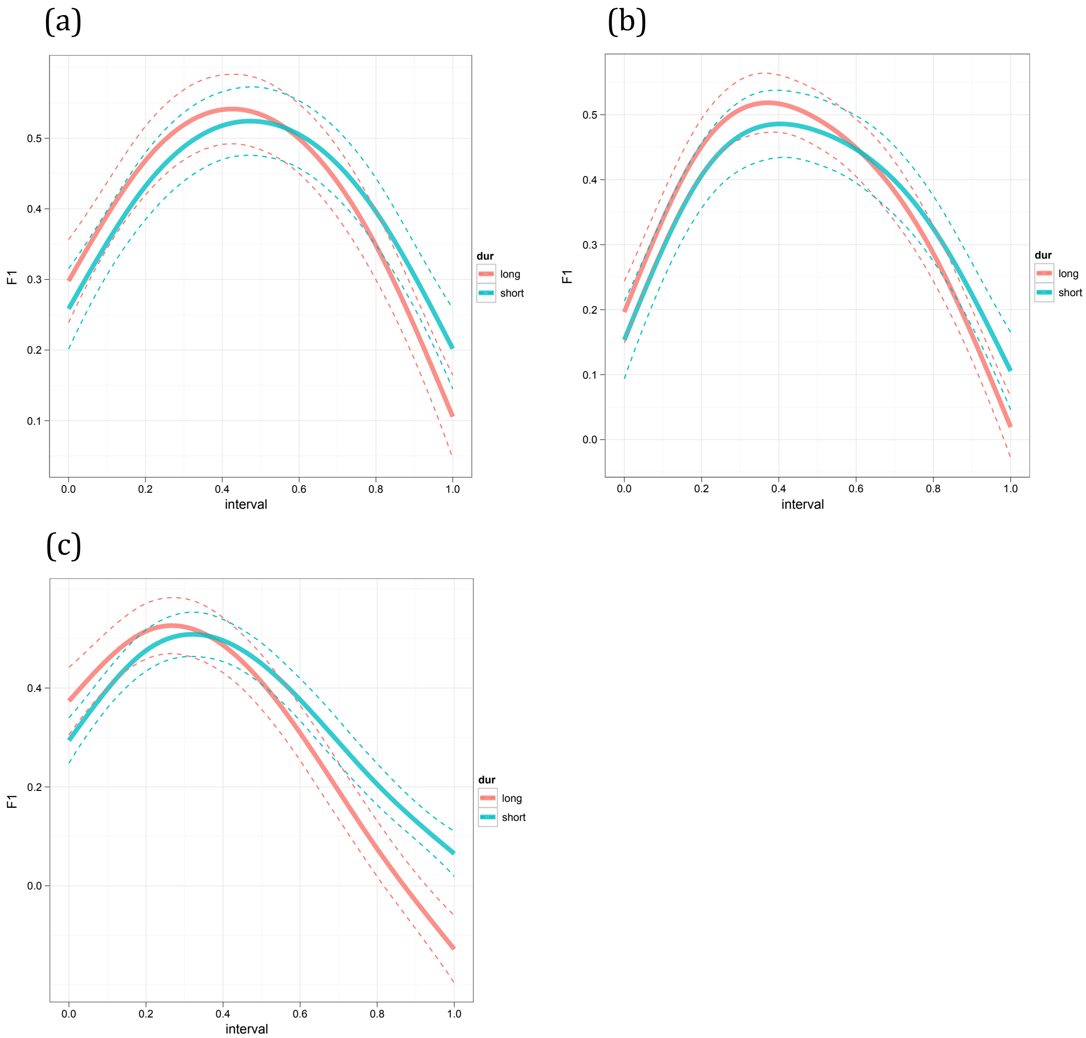

Figure 21. SSANOVAs for F1 in long and short vowels for African American females, in pre-voiced labial (a), pre-voiced alveolar (b), and pre-voiceless alveolar (c) contexts.

Figure 22 shows three F2 contours for African American males, for three coda environments: voiced labial (b/v), voiced alveolar (d/z), and voiceless alveolar (t). Upon first glance, it may appear that F2 contours before voiced alveolars display prominent onset steady states, while F2 contours in the other two contexts do not. However, this may be an effect of preceding environment, as /aɪd/z/ tokens have a disproportionate number of preceding alveolars, relative to other coda contexts (see Appendix IV). In most coda contexts, an offset steady state is not detectable; however, towards the end of the vowel, each context shows a lessening of slope for the short tokens, while the slope for the long contour remains relatively consistent. Surprisingly, the short pre-voiced labial contour does appear to have a steady state. Note that the F2 values at the 80% point are similar for short and long /aɪb/v/. Given the results of Gay (1968), it seems odd to find a steady state in a short vowel but not in a comparable long vowel. Rather than an offset steady state in the short vowels, the difference in slope between the short and long /aɪb/v/ groups towards the end of the vowel may reflect a greater effect of the labial coda for short tokens.

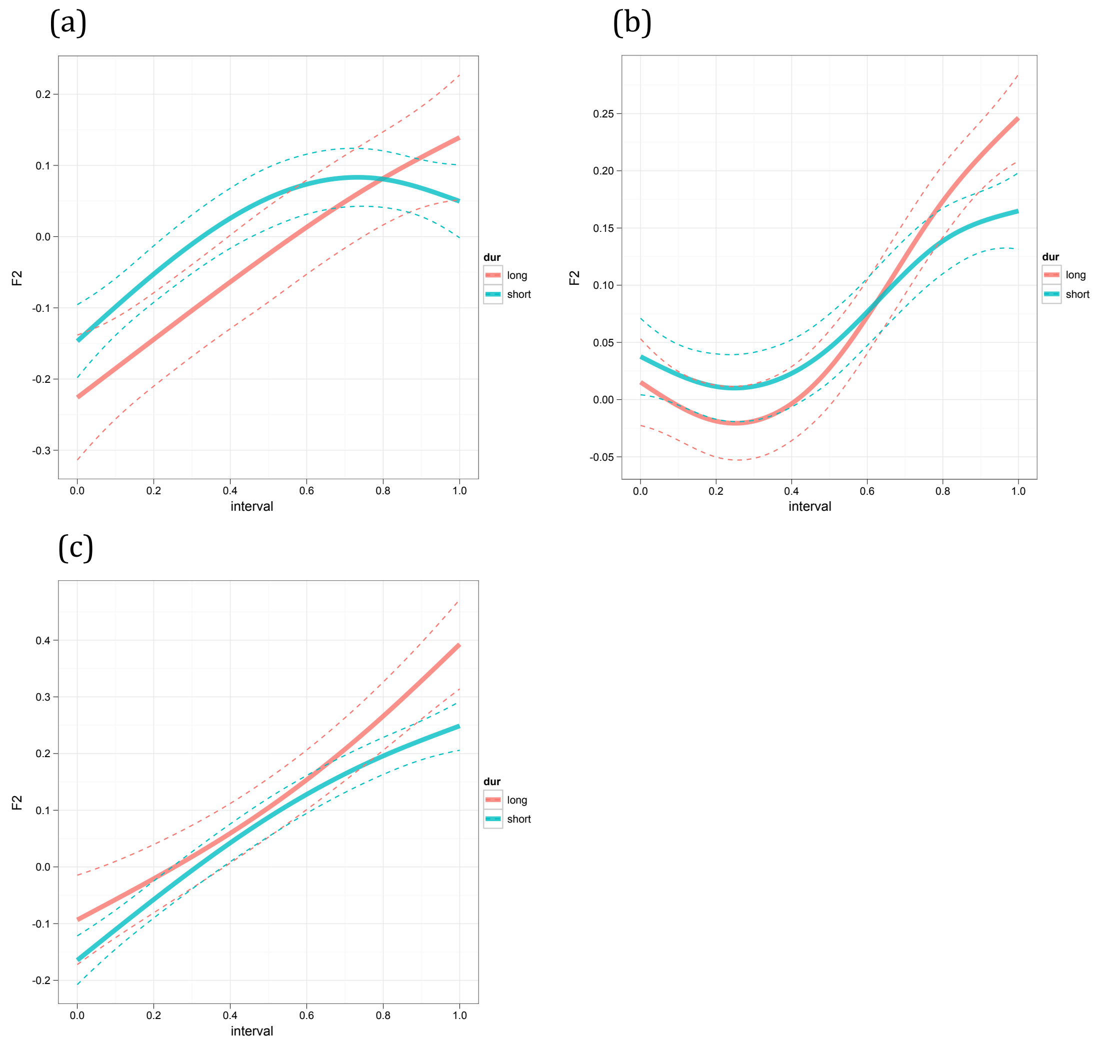

Figure 22. SSANOVAs for F2 in long and short vowels for African American males, in pre-voiced labial (a), pre-voiced alveolar (b), and pre-voiceless alveolar (c) contexts.

Figure 23 shows three F2 contours for African American females, for three coda environments: voiced labial (b/v), voiced alveolar (d/z), and voiceless alveolar (t). Again, we see what looks like a clear onset steady state for pre-voiced alveolar tokens; again, while this may be an actual effect of coda context, it may also be an artifact of unbalanced sampling. Interestingly, while there isn't a visible offset in any of the F2 contours in Figure 23, the slope of the contours appears to lessen towards the end of the vowel; note the contrast with the contours for males, as shown in Figure 22, particularly in the long vowels, where the slope of the contour is relatively consistent towards the end of the vowel. This difference in slope near the vowel offset may reflect a subtle yet consistent difference in F2 production between males and females in this sample.

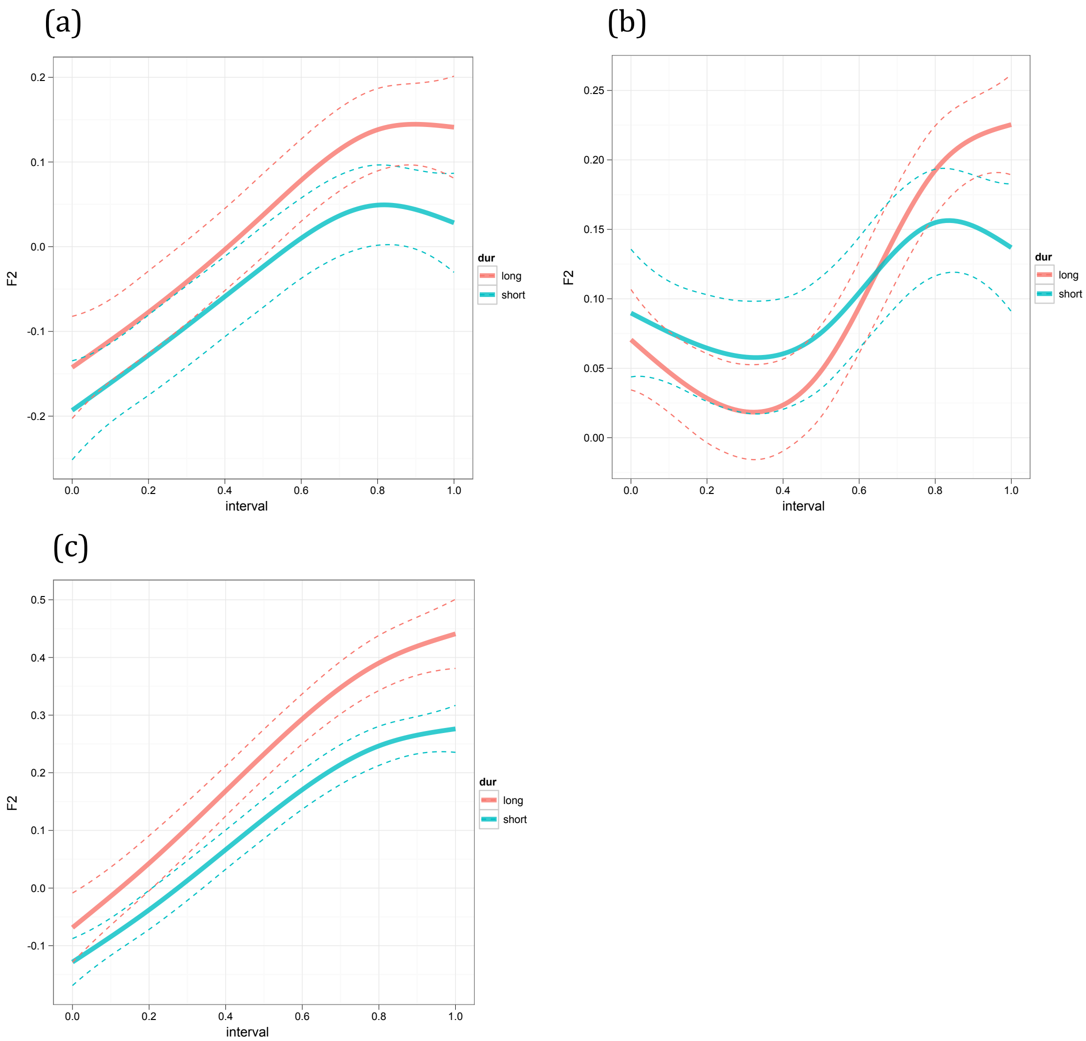

Figure 23. SSANOVAs for F2 in long and short vowels for African American females, in pre-voiced labial (a), pre-voiced alveolar (b), and pre-voiceless alveolar (c) contexts.

5.3. Discussion

Although the results from the SSANOVA analyses are ultimately merely suggestive, there are a few interesting results. First, the point in the vowel at which the F1 onset occurs appears to vary systematically. A closer analysis of this phenomenon would potentially have implications for vowel measurement best practices, as well as studies of reduction. Second, there do not appear to be regular prominent offset steady states, even for longer vowels. This contrasts with the results of Gay (1968), which found offset steady states for long /aɪ/ vowels but not for short /aɪ/ vowels. The difference in results could be due to dialect differences in the sample of speakers for each study; alternatively, Gay (1968) may have found an effect that is present in experimental contexts but not in naturalistic speech. Next, the data suggests regular gender differences in F2 contours for the sample, particularly towards the end of the vowel. Also, onset steady states could be more prominent before voiced alveolars, though preceding environment could have a confounding effect. Lastly, the change in F2 slope near vowel offset for short tokens, as well as the presence of an F2 steady state in the short /aɪb/v/ tokens, suggests a durational effect. While it seems a reach to call this evidence of reduction, the result does suggest that reduction effects might be observable in a vowel's contour, even when they are not observable in a point analysis, such as in Experiment 1.

A few steps could be taken in a future experiment to obtain more interpretable results. Higher token counts could yield more significant results. Controlling or balancing the data set for preceding consonant would make it easier to interpret any onset patterns. Lastly, sampling along more than 5 points of the vowels would result in SSANOVA contours of higher fidelity, with smaller confidence intervals.

6. Conclusions

This study finds evidence of reduction in /aɪ/ production. There is strong evidence for centralization throughout the vowel. The study builds on existing studies of centralization, in two ways: by finding evidence for the phenomenon for a linguistic variable that shows social variation, and by finding evidence for the phenomenon in a corpus of conversational speech.

While the results of this study offer no conclusive evidence for undershoot, the study itself does highlight some of the tenets of undershoot theory that could be clearer. First, the theory implies a durational threshold, such that vowels above a certain duration do not display undershoot. Can such a threshold be identified? Second, under what conditions (phonological, social, typographic) is undershoot avoided? Third, does undershoot vary by speech task? Lastly, does application of undershoot depend on the tenseness of the vowel? Clarification on these issues would result in a theory that has greater predictive power and would necessarily broaden the scope of the theory.

How useful is the current study's methodology in observing centralization and undershoot simultaneously? While there was no evidence for undershoot in the data, the methodology developed in Experiment 1 does appear to have two distinct advantages over the kind of methodology utilized in Moon and Lindblom (1994), for instance. Although there was no evidence of undershoot based on place of articulation, I hope to have made it clear that the regression models employed have the potential to distinguish centralization and undershoot, and to observe both in production. This is the case even when centralization and undershoot predict displacements in opposite directions. Second, the regression analysis employed here for examining reduction allows researchers to account for social variation in the data, rather than simply controlling for speaker variables like gender and ethnicity. Potentially, this methodology allows researchers to observe gender differences in vowel reduction as well, although such differences were not discussed here. The

SSANOVA analysis, while preliminary, suggests that a contour or slope analysis might illuminate reduction effects in the way a point analysis cannot.

While a number of questions remain about the exact nature of reduction in /aɪ/ production, some inroads were made. First, this study does not support the theory that one portion of a diphthong assimilates to the other part at short durations. Second, this study suggests a few differences between centralization and undershoot. There may be a durational threshold for undershoot, above which undershoot does not occur. In contrast, centralization appears to be a continuous duration-based phenomenon. Also, some questions were raised in expanding existing knowledge about vowel reduction in monophthongs to hypothesize about reduction effects on diphthongs. Although there was no concrete evidence for differences in temporal domain between centralization and undershoot, it seems to conform to previous research to hypothesize that, to the extent that diphthongs are subject to undershoot, any undershoot effects are localized in the offset of /aɪ/, while centralization effects appear throughout the vowel. Whether centralization and undershoot are independent phenomena, or if they reflect different explanations for the same phonetic patterns, remains an open question. However, if we can identify different domains for their application in future research, we can establish their independence from one another.

References

Adank, Patti, Smits, Roel, and van Hout, Roeland (2004). A comparison of vowel normalization procedures for language variation research. *JASA* 116(5), 3099-3107.

Andruski, J. E., and Nearey, T. M. (1992). "On the sufficiency of compound target specification of isolated vowels in /bVb/ syllables," J. Acoust. Soc. Am. 91, 390–410.

Assmann, P. F., Nearey, T. M., and Hogan, J. T. (1982). "Vowel identification: Orthographic, perceptual, and acoustic aspects," J. Acoust. Soc. Am. 71, 975–989.

Beckman, Mary E., Kenneth De Jong, Sun-Ah Jun and Sook-Hyang Lee (1992) The interaction of coarticulation and prosody in sound change. *Language and Speech* 35, 45-58.

Beddor, P., Harnsberger, J., and Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. Journal of Phonetics, 30(4):591-627.

Bladon, Anthony (1985). Diphthongs: A Case study of dynamic auditory processing. *Speech Communication* 4, 145-154.

Chomsky, N., & Halle, M. (1968). *Sound Pattern of English*. New York: Harper & Row.

Dailey-O'Cain, Jennifer (1997). Canadian Raising in a midwestern U.S. city. *Language Variation and Change* 9: 107-120.

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. J. Acoust. Soc. Am. 120:407–415.

Delattre, P., Liberman, A,.M., and Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. J. Acoust. Soc. Am. 27,769-773.

Dodsworth, Robin, Christine Mallison, and Lykara Charters (2007). Undershoot in intraspeaker variation. Paper presented at NWAV 36, Philadelphia, PA.

Eberhardt, M (2009). "Back vowel fronting in Pittsburgh AAE." Paper presented at the 38th annual conference on New Ways of Analyzing Variation (NWAV 38), Ottawa, Canada, Oct. 22-25.

Eckert, Penelope (2008). Where do ethnolects stop? *International Journal of Bilingualism* 12, 25-42.

Engstrand, Olle (1992). Systematicity of phonetic variation in natural discourse. *Speech Communication* 11, 337-346.

Farnetani, Edda, & Recasens, Daniel (1999). Coarticulation models in recent speech production theories. In Hardcastle, William J. & Nigel Hewlett, eds., *Coarticulation: Theory, data, and techniques*. New York: Cambridge.

Flemming, Edward (to appear). The phonetics of schwa vowels. To appear in Donka Minova (ed.), *Weak Segments in English*.

Flemming, E., & S. Johnson (2007). Rosa's roses: reduced vowels in American English. Journal of the International Phonetic Association 37, 83-96.

Fought, Carmen (1999). A majority sound change in a minority community: /u/-fronting in Chicano English. *Journal of Sociolinguistics* 3(1), 5-23.

Foulkes, Paul & Docherty, Gerard (2006). The social life of phonetics and phonology. *Journal of Phonetics* 34(4), 409-38.

Fox, R. (1983). Perceptual structure of monophthongs and diphthongs in English. *Language & Speech 26*, 21–49.

Fox, R.A. and Jacewicz, E. (2008). Dialectal differences in dynamic formant patterns in vowels. 156th Meeting of the Acoustical Society of America, Miami, Fl. *Journal of the Acoustical Society of America,* **124**(4): 2436.

Fruehwald, Josef (2008). The spread of raising: Opacity, lexicalization, and diffusion. *University of Pennsylvania Working Papers in Linguistics* 14:2.

Gay, T. (1968). Effects of speaking rate on diphthong formant movements. *Journal of the Acoustical Society of America, 44*, 1570–1573.

Gay, T. (1970). A perceptual study of American English diphthongs. *Language & Speech, 13*, 65–88.

Gendrot, Cedric, & Martine Adda-Decker (2007). Impact of duration and vowel inventory size on formant values of oral vowels: An Automated formant analysis from eight languages. ICPhS XVI, Saarbrucken.

Gottfried, M., Miller, J. D., and Meyer, D. J. (1993). "Three approaches to the classification of American English diphthongs," J. Phonetics 21, 205–229.

Gu, Chong (2002). Smoothing spline ANOVA models. Springer Series in Statistics. New York: Springer-Verlag.

Hall, Kathleen Currie (2005). Defining phonological rules over lexical neighborhoods: Evidence from Canadian Raising. In *Proceedings of the 24th West Coast Conference on Formal Linguistics*, ed. et al. John Alderete, 191-199.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*, 3099–3111.

Hillenbrand, J. M., & Nearey, T. N. (1999). Identification of resynthesized /hVd/ syllables: Effects of formant contour. *Journal of the Acoustical Society of America, 105*, 3509–3523.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. N. (2001). "Effect of consonant environment on vowel formant patterns," J. Acoust. Soc. Am. 109, 748–763.

Holbrook, A., & Fairbanks, G. (1962). Diphthong formants and their movements. *Journal of Speech and Hearing Research, 5*, 38–58.

Jacewicz, E., Fujimura, O., & R. A. Fox (2003). Dynamics in diphthong perception. In . J. Sole, D. Recasens, & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona* (pp. 993–996). Rundle Mall, SA, Australia: Causal Productions.

Johnson, Keith (2003). *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell.

Joos, M. (1948). *Acoustic phonetics* (LSA, Waverly, Baltimore).

Karlsson, I. (1992)."Analysis and Synthesis of Different Voices with Emphasis on Female Speech." Doctoral dissertation, Royal Institute of Technology, Stockholm.

Kewley-Port, Diane (1982). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. *JASA* 72(2), 379-389.

Klatt, D.H. (1987). Review of text-to-speech conversion for English. *J. Acoust. Soc. Amer.* Vol. 82, pp. 737–793.

Koopmans-van Beinum E, J. (1980)." Vowel Contrast Reduction an acoustic and perceptual study of Dutch vowels in various speech conditions." Ph.D. thesis, University of Amsterdam, The Netherlands Academische Pers B. V., Amsterdam.

Koops, Christian (2010). /u/-fronting is not monolithic: Two types of fronted /u/ in Houston Anglos. *University of Pennsylvania Working Papers* 16(2).

Koops, Chris & Nancy Niedzielski (2009). /t, d/-glottalization and vowel variation in Houston AAE. Paper presented at the 38th annual conference on New Ways of Analyzing Variation (NWAV 38), Ottawa, Canada, Oct. 22-25.

Kurath, Hans, and Raven I. McDavid, Jr. 1961. The pronunciation of English in the Atlantic States. Ann Arbor, MI: University of Michigan Press.

Labov, William (1994). Principles of Linguistic Change: Internal Factors. Cambridge, MA: Blackwell.

Labov, William, Malcah Yaeger, and Richard Steiner (1972). A Quantitative Study of Sound Change. National Science Foundation GS-3287.

Labov, William, Ash, Sharon, & Boberg, Charles (2006). *The Atlas of North American English*. New York: Mouton de Gruyter.

Lehiste, I., & Peterson, G. E. (1961). Transitions, glides, and diphthongs. *Journal of the Acoustical Society of America, 33*, 268–277.

Liberman, A., Delattre, P., Cooper, F., and Gerstman, L. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs* 68.

Lindblom, Bjorn (1963) Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773-1781.

Lindblom, Bjorn (1990). Explaining phonetic variation: A sketch of the H&H theory. *Speech production and speech modeling*, ed. by W. J. Hardcastle and A. Marchal, 403-39. Dordrecht: Kluwer.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech, 29*, 3-11.

Moon, Seung-Jae and Bjorn Lindblom (1994) Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America* 96, 40-55.

Moreton, Elliott (2004). Realization of the English postvocalic [voice] contrast in F_1 and F_2. *Journal of Phonetics* 32, 1-33.

Moreton, Elliott and Thomas, Erik. R. (2007) Origins of Canadian Raising in voiceless-coda effects: A case study in phonologization. In Cole, Jennifer and Hualde, José I. (eds.), *Papers in Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 37-64.

Morrison, Geoffrey Stewart (2008). Theories of vowel inherent spectral change: A review. Unpublished ms. Accessed online Sept. 1, 2009. http://geoff-morrison.net/documents/Morrison%20(2007%20unpub)%20Theories%20of%20vowel%20inherent%20spectral%20change%20-%20A%20review%20-%20v30.pdf.

Morrison, G. S., & Nearey, T. M. (2007). Testing theories of vowel inherent spectral change. *Journal of the Acoustical Society of America. 122*, EL15–EL22.

Nearey, T. M. (1978). *Phonetic Feature Systems for Vowels.*Indiana University Linguistics Club, Indiana.

Nord, L. (1986). Acoustic studies of vowel reduction in Swedish. *Quarterly Progress and status report*, 4, 19-36.

Nycz, Jennifer, and Paul De Decker. 2006. A New Way of Analyzing Vowels: Comparing Formant Contours Using Smoothing Spline ANOVA. Poster presented at NWAV 35, Columbus, OH.

Ohala, J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston, & M.Beckman (Eds.), Papers in laboratory phonology I: Between the grammar and the physics of speech (pp. 258–275). Cambridge: Cambridge University Press.

Oxley, Meghan (2009). (ay) Monophthongization in Deer Park, Texas. Masters thesis, University of Washington.

Payne, Arvilla (1980). Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. *Locating language in time and space*, ed. by William Labov, 143-78. New York: Academic Press.

Pearce, Mary (2008). Vowel harmony domains and vowel undershoot. UCLWPL 20: 115-140. http://www.langsci.ucl.ac.uk/linguistics/publications/WPL/uclwpl20.php

Pike, Kenneth Lee (1947). *Phonemics: a technique for reducing languages to writing*. Ann Arbor, MI: University of Michigan Press.

R Development Core Team (2010). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org, ISBN 3-900051-07-0.

Raftery, A.E. (1995). Bayesian model selection in social research (with Discussion). *Sociological Methodology*, 25, 111-196.

Scanlon, Michael, and Wassink, Alicia Beckford (2010). African American English in urban Seattle: Accommodation and intraspeaker variation in the Pacific Northwest. *American Speech* 85(2), 205-224.

Schilling-Estes, Natalie. 1996. "The Linguistic and Sociolinguistic Status of /ay/ in Outer Banks English." Ph.D. diss., Univ. of North Carolina at Chapel Hill.

Shirai, Setsuko (2005). Lexical effects in Japanese vowel reduction. Ph.D. dissertation, University of Washington.

Stalhammar U., Karlsson I., and Fant, G. (1973). "Contextual effects on vowel nuclei." STL-QPSR, No. 4, pp. 1-18.

Stevens, K. N. and House, A.S. (1963). Perturbation of Vowel Articulations by Consonantal Context: An Acoustical Study," J. Speech & Hearing Res. 6, 111-128.

Thomas, E. R. (1991). The origin of Canadian Raising in Ontario. *Canadian Journal of Linguistics* 36, 147-70.

Thomas, Erik R. (1995). Phonetic factors and perceptual reanalyses in sound change. Ph.D. dissertation, The University of Texas at Austin.

Thomas, E. R. (2000). Spectral differences in /ai/ offsets conditioned by voicing of the following consonant. Journal of Phonetics, 28, 1–25.

Thomas, E. R. (2004). Instrumental phonetics. In Chambers, J. K., Peter Trudgill, and Natalie Schilling-Estes, eds., *The Handbook of Language Variation and Change*. Malden, MA: Blackwell, pp. 168-200.

Thomas, E. R. (2007). Phonological and phonetic characteristics of African American Vernacular English. *Language & Linguistics Compass* 1(5), 450-475.

Tiffany, W. R. (1959)."Non-Random Sources of Variation in Vowel Quality." I. Speech Hear. Res. 2, 305-317.

Van Bergem, D.R. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication* 12, 1-23.

Vance, T. J. (1987). Canadian raising in some dialects of the northern United States. American Speech, 62, 195–210.

Wassink, Alicia Beckford (2006). A geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three linguistic varieties. *JASA* 119(4), 2234-2350.

Appendix I: auditory judgments

Following phone	*# tokens judged as diphthongal*	*# tokens judged as reduced*	*% tokens judged as reduced*
b	12	5	29
d	71	11	13
m	25	17	40
n	158	33	17
p	11	3	21
t	81	4	5
v	29	8	22
z	32	13	29

Appendix II: chi-square tests

2-sample tests of equality of proportions (tokens judged as reduced/total tokens):
b/d: χ^2=1.610, p=0.205
m/n: χ^2=10.994, p=0.001
p/t: χ^2=2.887, p=0.089
v/z: χ^2=0.246, p=0.620
b/d/v/z: χ^2=5.337, p=0.149

Appendix III: regression results

Call:
glm(formula = delta.F2.nearey ~ duration + place_f + voice_f +
sex_code + subject_ethnicity + duration:place_f + duration:voice_f +
voice_f:sex_code + place_f:voice_f + duration:place_f:voice_f)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.280559	-0.085825	-0.003848	0.071173	0.520602

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.030724	0.040211	0.764	0.44549
duration	0.570425	0.129141	4.417	1.45e-05 ***
place_flabial	0.079660	0.054575	1.460	0.14555
voice_fvoiceless	0.009853	0.066605	0.148	0.88250
sex_code	-0.000429	0.019487	-0.022	0.98245
subject_ethnicityCauc	0.099003	0.020385	4.857	2.03e-06 ***
duration:place_flabial	-0.074503	0.246240	-0.303	0.76246
duration:voice_fvoiceless	0.458957	0.297736	1.541	0.12437
voice_fvoiceless:sex_code	0.068672	0.032008	2.145	0.03281 *
place_flabial: voice_fvoiceless	-0.388354	0.130454	-2.977	0.00318 **
duration:place_flabial: voice_fvoiceless	1.918359	0.866648	2.214	0.02770 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.01561423)

Null deviance: 6.7962 on 279 degrees of freedom
Residual deviance: 4.2002 on 269 degrees of freedom
AIC: -357.3

Number of Fisher Scoring iterations: 2

Call:
glm(formula = normnearey.F2.80 ~ duration + place_f + voice_f +
sex_code + subject_ethnicity + duration:voice_f + duration:subject_ethnicity +
voice_f:sex_code + place_f:subject_ethnicity)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.3277564	-0.0565010	-0.0004527	0.0570297	0.2173430

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.04657	0.02686	1.734	0.08409 .
duration	0.44763	0.08531	5.247	3.13e-07 ***
place_flabial	-0.04593	0.01456	-3.154	0.00179 **
voice_fvoiceless	-0.08625	0.04414	-1.954	0.05172 .
sex_code	0.01915	0.01397	1.371	0.17160
subject_ethnicityCauc	0.16810	0.03231	5.202	3.90e-07 ***
duration:voice_fvoiceless	0.90953	0.19428	4.682	4.51e-06 ***
duration:subject_ethnicityCauc	-0.31735	0.16742	-1.896	0.05909 .
voice_fvoiceless:sex_code	0.04464	0.02288	1.951	0.05207 .
place_flabial:subject_ethnicityCauc	-0.06017	0.02953	-2.037	0.04258 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.007756835)

Null deviance: 4.1769 on 279 degrees of freedom
Residual deviance: 2.0943 on 270 degrees of freedom
AIC: -554.15

Number of Fisher Scoring iterations: 2

Call:
glm(formula = normnearey.F2.20 ~ duration + place_f + voice_f +
sex_code + subject_ethnicity + duration:place_f + duration:voice_f +
place_f:sex_code + place_f:subject_ethnicity + place_f:voice_f +
duration:place_f:voice_f)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.412289	-0.076139	0.001313	0.064844	0.318614

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.01866	0.03668	0.509	0.61133
duration	-0.17620	0.11981	-1.471	0.14255
place_flabial	-0.03708	0.07652	-0.485	0.62832
voice_fvoiceless	-0.12393	0.04952	-2.503	0.01292 *
sex_code	0.02508	0.01689	1.485	0.13872
subject_ethnicityCauc	0.02576	0.02236	1.152	0.25038
duration:place_flabial	0.04919	0.22888	0.215	0.83000
duration:voice_fvoiceless	0.36280	0.27293	1.329	0.18488
place_flabial:sex_code	-0.05317	0.03655	-1.455	0.14686
place_flabial:subject_ ethnicityCauc	-0.10767	0.04264	-2.525	0.01215 *
place_flabial: voice_fvoiceless	0.33386	0.12185	2.740	0.00656 **
duration:place_flabial: voice_fvoiceless	-1.36661	0.80690	-1.694	0.09149 .

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.01343819)

Null deviance: 4.4254 on 279 degrees of freedom
Residual deviance: 3.6014 on 268 degrees of freedom
AIC: -398.36

Number of Fisher Scoring iterations: 2

Call:
glm(formula = delta.F1.nearey ~ duration + place_f + voice_f +
sex_code + subject_ethnicity + duration:voice_f + duration:sex_code +
voice_f:sex_code + voice_f:subject_ethnicity)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.571529	-0.066259	0.007067	0.070744	0.601986

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.05452	0.08108	-0.672	0.50187
duration	-0.24537	0.35523	-0.691	0.49031
place_flabial	0.03697	0.01983	1.865	0.06332 .
voice_fvoiceless	0.15946	0.07769	2.052	0.04109 *
sex_code	0.07016	0.05192	1.351	0.17769
subject_ethnicityCauc	-0.07666	0.02923	-2.623	0.00922 **
duration:voice_fvoiceless	-1.62054	0.30867	-5.250	3.08e-07 ***
duration:sex_code	-0.36144	0.22331	-1.619	0.10671
voice_fvoiceless:sex_code	-0.06894	0.03947	-1.747	0.08181 .
voice_fvoiceless:subject_ethnicityCauc	0.06604	0.04542	1.454	0.14712

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.01899428)

Null deviance: 9.3234 on 279 degrees of freedom
Residual deviance: 5.1285 on 270 degrees of freedom
AIC: -303.39

Number of Fisher Scoring iterations: 2

Call:
glm(formula = normnearey.F1.80 ~ duration + place_f + voice_f +
sex_code + subject_ethnicity + duration:place_f + duration:voice_f +
duration:sex_code + voice_f:sex_code)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.35747	-0.08871	-0.01330	0.08365	0.33533

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0515676	0.0814582	0.633	0.527233
duration	0.5198288	0.3620637	1.436	0.152234
place_flabial	-0.0009472	0.0481323	-0.020	0.984314
voice_fvoiceless	0.1642773	0.0671256	2.447	0.015030 *
sex_code	0.1785792	0.0504254	3.541	0.000469 ***
subject_ethnicityCauc	-0.0665275	0.0215442	-3.088	0.002225 **
duration:place_flabial	0.3304915	0.2366586	1.396	0.163714
duration:voice_fvoiceless	-1.3786789	0.2933350	-4.700	4.15e-06 ***
duration:sex_code	-0.4750449	0.2174322	-2.185	0.029764 *
voice_fvoiceless:sex_code	-0.0569599	0.0362351	-1.572	0.117133

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.01746275)

Null deviance: 7.4941 on 279 degrees of freedom
Residual deviance: 4.7149 on 270 degrees of freedom
AIC: -326.93

Number of Fisher Scoring iterations: 2

```
Call:
glm(formula = normnearey.F1.20 ~ duration + voice_f + sex_code +
    subject_ethnicity + duration:subject_ethnicity)

Deviance Residuals:
     Min        1Q    Median       3Q       Max
-0.37232  -0.07284   0.00716   0.06707   0.30147

Coefficients:
                          Estimate    Std. Error   t value    Pr(>|t|)
(Intercept)               0.13224     0.03156      4.190      3.77e-05 ***
duration                  0.61249     0.10456      5.858      1.34e-08 ***
voice_fvoiceless          0.04514     0.01557      2.899      0.00404 **
sex_code                  0.09510     0.01497      6.353      8.78e-10 ***
subject_ethnicityCauc     -0.08876    0.04133      -2.148     0.03263 *
duration:subject_
          ethnicityCauc   0.43776     0.21362      2.049      0.04139 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.01356603)

    Null deviance: 5.4239  on 279  degrees of freedom
Residual deviance: 3.7171  on 274  degrees of freedom
AIC: -401.51

Number of Fisher Scoring iterations: 2
```

Appendix IV: distribution of /aɪ/ tokens, by preceding and following phone

		Preceding phone																	
		Labial						Dental		Alveolar						Velar		Glottal	
		b	p	v	f	m	*total*	ð	*total*	d	t	z	s	n	*total*	g	*total*	h	*total*
Following phone	b/v	17	-	-	35	-	*52*	1	*1*	1	-	-	-	-	*1*	-	-	-	-
	d/z	-	-	-	1	1	*2*	-	-	26	2	1	57	-	*86*	36	*36*	2	*2*
	p	-	1	-	-	-	*1*	-	-	13	-	-	-	-	*13*	-	-	-	-

	t	2	2	3	28	15	*50*	-	-	-	6	-	5	24	*35*	-	-	-	-