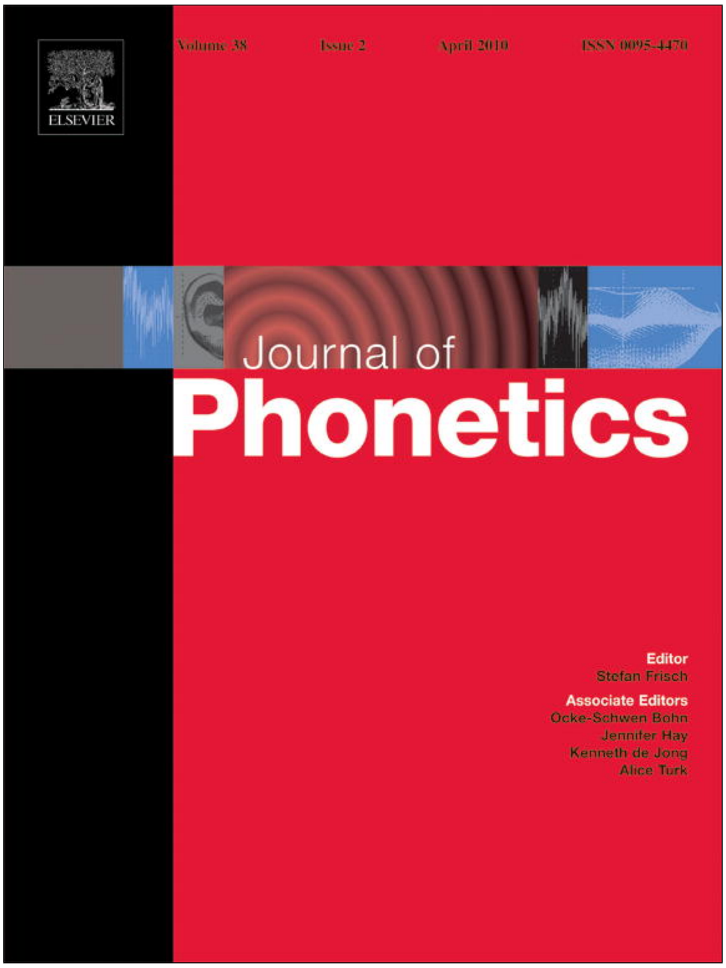
Volume 38
Issue 2
April 2010
ISSN 0095-4470
ELSEVIER
Journal of
Phonetics
Editor
Stefan Frisch
Associate Editors
Ocke-Schwen Bohn
Jennifer Hay
Kenneth de Jong
Alice Turk

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Phonetics 38 (2010) 167–184
Journal of Phonetics
www.elsevier.com/locate/phonetics

Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach

Jennifer Cole[a,*], Gary Linebaugh[b], Cheyenne M. Munson[c], Bob McMurray[c]

[a]*Department of Linguistics, University of Illinois at Urbana-Champaign, 707 South Mathews, Urbana, IL 61801, USA*
[b]*Center for English as a Second Language, Southern Illinois University—Carbondale, Carbondale, IL 62901, USA*
[c]*Department of Psychology and Delta Center, University of Iowa, Iowa City, IA 52242-1407, USA*

Received 25 July 2007; received in revised form 12 August 2009; accepted 22 August 2009

Abstract

Coarticulation is a source of acoustic variability for vowels, but how large is this effect relative to other sources of variance? We investigate acoustic effects of anticipatory V-to-V coarticulation relative to variation due to the following C and individual speaker. We examine F1 and F2 from V1 in 48 V1-C♯V2 contexts produced by 10 speakers of American English. ANOVA reveals significant effects of both V2 and C on F1 and F2 measures of V1. The influence of V2 and C on acoustic variability relative to that of speaker and target vowel identity is evaluated using hierarchical linear regression. Speaker and target vowel account for roughly 80% of the total variance in F1 and F2, but when this variance is partialed out C and V2 account for another 18% (F1) and 63% (F2) of the remaining target vowel variability. Multinomial logistic regression (MLR) models are constructed to test the power of target vowel F1 and F2 for predicting C and V2 of the upcoming context. Prediction accuracy is 58% for C-Place, 76% for C-Voicing and 54% for V2, but only when variance due to other sources is factored out. MLR is discussed as a model of the parsing mechanism in speech perception.

1. Introduction

A fundamental problem in speech perception stems from the fact that during fluent speech production, neighboring segments can profoundly influence one another in their acoustic realization. A result of this coarticulation is that there are few, if any, points in time at which the speech signal can be unambiguously interpreted as a single segment. Vowels are particularly sensitive to coarticulatory influences (Fowler & Brancazio, 2000; Recasens & Pallarès, 2000). The formant structure that indicates their height, backness and roundness is relatively long-lasting, but formant values can be influenced by adjacent context on both sides and the relatively wide separation between phonologically contrastive vowels in phonetic space permits ample within-category variation that would not neutralize contrast.

This paper investigates coarticulation in the production of vowels. Evidence from articulatory studies reveals that the production of a vowel is influenced by the place of articulation of an adjacent vowel (e.g., in diphthong and hiatus contexts), and also by a vowel in an adjacent syllable across an intervening consonant (Alfonso & Baer, 1982; Cho, 2004; Fletcher, 2004; see review in Farnetani, 1997). Vowel-to-vowel (V-to-V) coarticulation introduces variability in the acoustic realization of a vowel as well, resulting in measurable shifts in its formant values relative to a 'neutral', non-coarticulating context (Magen, 1997; Manuel, 1990; Öhman, 1966; Recasens & Pallarès, 2000). As a result, coarticulation increases acoustic variance within a vowel category and reduces the separation between distinct vowel categories in phonetic space, as illustrated in Fig. 1.

Given that coarticulation decreases the phonetic distinctiveness of phonologically contrastive vowels, it might be expected that it results in increased perceptual confusion.

*Corresponding author. Tel.: +1 217 244 3057; fax: +1 217 244 8430.
E-mail addresses: jscole@illinois.edu (J. Cole), linebaug@siu.edu (G. Linebaugh), cheyenne-munson@uiowa.edu (C.M. Munson), bob-mcmurray@uiowa.edu (B. McMurray).

0095-4470/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2009.08.004

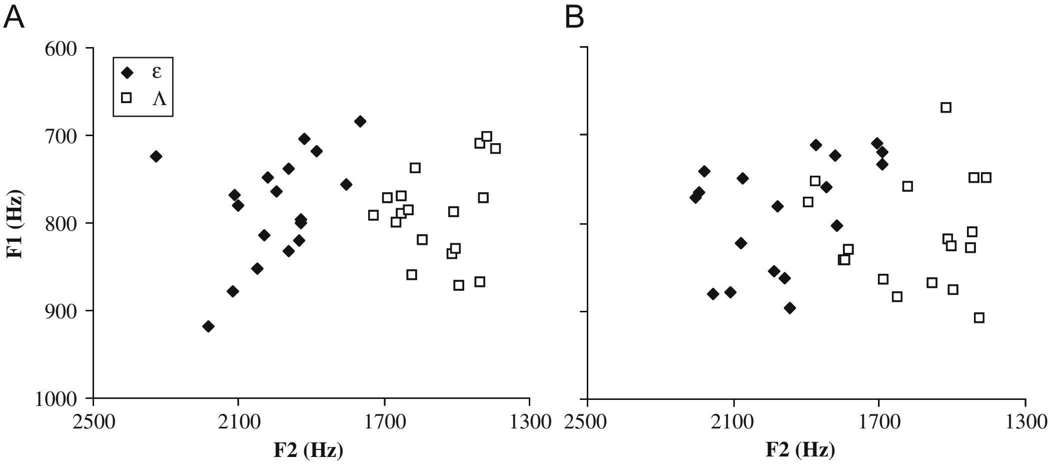

Fig. 1. Representative F1 and F2 frequencies for a sample of /ʌ/'s and /ε/'s for a single speaker. Panel A shows the two vowels spoken in a neutral context. Panel B shows the same vowels across a variety of coarticulatory contexts.

However, a number of studies indicate that the potential confusion in vowel identification is mitigated by a mechanism of perceptual compensation. Fowler (1981, 1984), Fowler and Smith (1986) and Beddor, Harnsberger, and Lindemann (2002), demonstrate that listeners compensate for V-to-V influences on a target vowel, reducing or eliminating the effects of coarticulation on vowel quality when those effects can be attributed to a context (or *trigger*) vowel in a neighboring syllable. Moreover, there is evidence that listeners actually benefit from V-to-V coarticulation in that the altered acoustic form in the target vowel permits predictions about the identity of an upcoming vowel (Fowler, 1984; Martin & Bunnell, 1982; Whalen, 1990).

Fowler (1984) argues that these findings are consistent with a *coproduction* account of coarticulation, in which speech gestures overlap in time such that a given time frame simultaneously provides evidence for multiple segments. This overlap does not result in a complete blending of the segments in perception, though. Fowler's evidence, and that of Beddor et al. (2002), indicates that coarticulatory influences on the target segment are perceived by the listener as evidence for the upcoming context vowel, and are not integrated into the perceptual experience of the target. That is, listeners appear to compensate for the coarticulation, hearing an unmodified target vowel, while simultaneously anticipating an upcoming vowel. Fowler describes both effects as emerging from a single parsing process in which listeners factor or parse the overlapping cues in the acoustic signal into properties deriving from the local segment (i.e., the perceptual target) and those that derive from distal context. After the effects of immediate, local context are identified and parsed out of the current input, the remaining variation can be used to predict upcoming context.

The studies cited above, and others, have demonstrated the acoustic effects of V-to-V coarticulation as a phenomenon of speech production, and reveal how listeners cope with and benefit from V-to-V coarticulation in speech perception. However, we do not yet have a model that bridges these two areas of research. For instance, although we know that listeners cope with the acoustic variability due to V-to-V coarticulation, we also know that V-to-V coarticulation is not the only source of variation in the signal, and it is unclear how it may interact with other sources of variation that could further neutralize vowel contrast (or the V-to-V effects). Moreover, while V-to-V coarticulation offers benefits for anticipating the context segment, it is not clear exactly how much information about upcoming context is available in the acoustic signal, and whether that information is sufficient to actually identify the context sound from other sounds that could occur in the same position.

In order to bridge the gap between our understanding of coarticulation in production and in perception, we must subject V-to-V coarticulation to a broader investigation and ask what challenges and opportunities it affords for speech perception. Such an approach requires simultaneous consideration of the acoustic signal and the many sources of variation influencing phonetic realization, as well as a model of the perceptual process.

The first goal of this paper is to establish that the distinct components of local and distal context are reliably present in the acoustic signal for a variety of coarticulatory contexts in the production of real word sequences, and to evaluate the variability of the acoustic effects of V-to-V coarticulation relative to other sources of variation. We restrict our focus to anticipatory V-to-V coarticulation across a variety of intervening consonants. The second goal is to demonstrate through a statistical simulation of the parsing model that the target and context components of the acoustic signal can be separated from one another and from other sources of acoustic variation, and that this separation is the key to both the successful identification of the target vowel in the face of acoustic variability, and the prediction of the upcoming vowel context.

2. Acoustic effects of V-to-V coarticulation

Our first research question concerns variability in the patterns of coarticulation across target and context vowels, and whether the effects of coarticulation are consistent and similar in magnitude across a range of VCV sequences. In this section we briefly review findings from prior acoustic studies that bear on the variability of coarticulation.

2.1. Coarticulation effects across languages

Coarticulation effects on vowel formants vary across languages. For instance, Öhman (1966) finds effects of V-to-V coarticulation in the F2 measurements of vowels from VCV nonce sequences for both Swedish and English, but not for Russian. Beddor et al. (2002) reports effects of V-to-V coarticulation for both Shona and English, but with differences between the two languages in the direction of the effects (carryover vs. anticipatory) and in the role of stress as a conditioning factor. Similarly, Manuel (1990) finds differences in patterns of V-to-V coarticulation among three Southern Bantu languages, Ndebele, Shona and Sotho, where the magnitude of coarticulatory effects on vowel formants varies inversely with the number of contrastive vowels in the language.

The finding of cross-linguistic variation in coarticulation effects means, of course, that coarticulation is not fully an automatic consequence of the speech production mechanism. Rather, individuals learn the coarticulation patterns of their language, which requires at a minimum that they are sensitive to these patterns at some level in speech perception.

2.2. Coarticulation effects on F1 and F2 measures

The significant finding of Öhman's (1966) seminal study on coarticulation is that the acoustic properties of a vowel are affected by the vowel in the adjacent syllable, across an intervening consonant. Effects of V-to-V coarticulation can be seen across the dominant cues to vowel identity—F1, F2 and to a lesser extent F3 measures—reflecting coarticulation of the tongue body and lip gestures used in the production of vowels in connected speech. Some of the findings from prior studies reveal an asymmetrical influence of V-to-V coarticulation, with larger effects of coarticulation on F2 variation than F1 variation in a number of languages (Catalan: Recasens, 1984; Swedish: Öhman, 1966; English: Alfonso & Baer, 1982; Fowler, 1981, 2005; Fowler & Brancazio, 2000; Huffman, 1986; Martin & Bunnell, 1981, 1982). Manuel (1990) reports a similar asymmetry, with greater F2 variation due to coarticulation for the target vowel /e/ in Ndebele, Shona and Sotho, but when the target vowel is /ɑ/ there are significant coarticulatory effects on both F1 and F2.

The greater and more common effects of coarticulation on F2 measures suggest that coarticulatory effects may afford better predictions about the backness or roundness of a neighboring vowel than about its height. However, a fair comparison of the effects of V-to-V coarticulation on F1 vs. F2 would have to factor out the possible confounding influence from the intervening C (or other sources). For example, F1 may be more influenced by the individual speaker's voice or the voicing of the intervening consonants, factors that may mask a larger V-to-V effect (discussed further in Section 2.4).

2.3. Effects of trigger and target vowel identity

Acoustic effects of V-to-V coarticulation have been reported for a variety of trigger and target vowels. Considering first triggering vowels, nearly all of the studies examining V-to-V coarticulation in English include the two "point vowels" /i/ and /ɑ/ among the trigger vowels, and generally speaking both of these vowels exert a coarticulatory influence on neighboring vowels. Other vowels are also reported to exert a coarticulatory influence, including mid vowels /e,ɛ,ʌ,o/, as well as the two other point vowels, /æ/ and /u/ (Alfonso & Baer, 1982; Beddor et al., 2002; Fowler, 2005).

For the most part, these studies were not designed to make comparisons between different trigger vowels in the magnitude of their influence on the target vowel. However, in some cases asymmetries are apparent. For example, Beddor et al. (2002) presents plots of coarticulated unstressed target vowels that reveal a relatively large fronting effect (increase in F2) with the trigger vowel /i/ compared to the smaller backing effect (decrease in F2) with the trigger vowel /u/, and overall larger effects in F2 than in F1 for all trigger vowels.

Looking next at the target vowels in V-to-V coarticulation, we again find that a wide range of vowels can undergo V-to-V coarticulation. Several studies report coarticulatory effects on a central unstressed target vowel (/ə/ or /ʌ/) (Alfonso & Baer, 1982; Fowler, 1981, 2005; Fowler & Brancazio, 2000), while other studies report effects of coarticulation on full vowels (Beddor et al., 2002; Martin & Bunnell, 1981, 1982; Öhman, 1966; Whalen, 1990), or on both schwa and full vowels (Fletcher, 2004; Magen, 1989, 1997). In English there are greater effects of coarticulatory variation for unstressed or unaccented vowel targets compared to stressed or accented targets (Beddor et al., 2002; Fletcher, 2004; Fowler, 1981; and see Cho, 2004 for parallel findings of prosodic effects on coarticulation based on articulatory [EMA] measures).

The general finding across studies is that coarticulation is widespread, with every vowel a potential trigger and potential target. However, closer consideration suggests that the acoustic effects of coarticulation may vary depending on the identity of the target or trigger vowels. Target vowels that are unstressed, unaccented and central undergo the most coarticulation, and there is suggestive evidence that point vowels /i/ and /ɑ/ may exert a greater coarticulatory influence than other triggering vowels.

2.4. C-to-V coarticulation and its influence on V-to-V coarticulation

In considering other influences on the acoustic realization of a vowel, we need look no further than the adjacent consonant. A voiced consonant, for example, tends to condition lower F1 offset and onset values in the preceding and following vowel, respectively, compared to a voiceless consonant (e.g., Kingston, Diehl, Kirk, & Castleman, 2008; Lisker, 1986; Summerfield, 1981). Similarly, the place of articulation of an adjacent consonant affects vowel formant values, most notably in the centralization of F2 frequencies, with lower F2 of front vowels in the vicinity of labial consonants, and higher F2 of back vowels in the vicinity of alveolar consonants (Hillenbrand, Clark, & Nearey, 2001; Stevens & House, 1963). While these effects are strongest immediately adjacent to the consonant, C-to-V effects can persist further into the vowel, as shown for Catalan by Recasens, Pallarès, and Fontdevila (1997). The effects of C-to-V coarticulation could potentially mask V-to-V effects acoustically (in terms of measurable shifts in vowel formant values) and perceptually (in terms of the listeners ability to attribute effects to upcoming vowel context).

Beyond the direct effects of an adjacent consonant on vowel formant values, we can ask whether C-to-V coarticulation influences the extent or magnitude of V-to-V coarticulation, as a second order effect. Öhman's (1966) seminal study on coarticulation demonstrated that neighboring vowels influence each other; however, that study and studies that follow it demonstrate that the intervening consonant can also mediate V-to-V coarticulatory effects, by interrupting or attenuating them (Fowler, 2005; Fowler & Brancazio, 2000; Martin & Bunnell, 1982; Recasens, 1984, 2002; Recasens & Pallarès, 2000; Recasens et al., 1997). For instance, data from a single speaker in Fowler's (2005) study indicate that the intervening consonant impacts the V-to-V coarticulatory effect in both F1 and F2 values of the target vowel, though with some difference in the magnitude of the effects on the two formant measures. Fowler and Brancazio's (2000) study shows that the interfering effect of the consonant on V-to-V coarticulation is strongest at the edge of the target vowel nearest to the consonant, though the speaker in Fowler's (2005) study demonstrated significant interference effects from the consonant throughout the duration of the first vowel in a [əCV] sequence, with the strongest interference from 'high-resistant' coronal consonants, and considerably lesser interference from 'low-resistant' labials.

These findings from prior studies indicate that the intervening consonant can increase the overall acoustic variability of the target vowel, and mask or highlight the acoustic evidence available in the target vowel for the upcoming context vowel. In addition, the interfering effect of the consonant may impact the effects of V-to-V coarticulation on F1 and F2 differently, which means that the potential for predicting the context vowel based on coarticulatory cues available in the target vowel may vary depending on whether the intervening consonant is high-resistant (e.g., coronal) or low-resistant (e.g., labial).

2.5. Summary

The findings from prior acoustic studies of coarticulation show that coarticulatory effects vary across languages, indicating that coarticulation patterns are learned, presumably on the basis of acoustic evidence. They also establish that variation in both F1 and F2 values of the target vowel is conditioned by the triggering context vowel in a neighboring syllable. Furthermore, coarticulated vowels are the norm, occurring in most if not all VCV contexts. However, the impact of the triggering context vowel on the target vowel can vary significantly depending on the identity (place of articulation) of the target and trigger vowels, the status of the target and trigger vowels as bearing lexical or phrasal stress, and the resistance level of the intervening consonant. Finally, for those studies that report individual speaker data, inter-speaker variability in coarticulation is also observed.

These findings raise several questions for a model that links production and perception. Given that there is some degree of coarticulation in all VCV sequences, the parsing model of speech perception predicts corresponding evidence of perceptual compensation in every VCV context. But how does the perceptual mechanism handle differences in the magnitude of the coarticulatory effect conditioned by the properties of the target and trigger vowels and the intervening consonant? For instance, can the parsing mechanism cope with the combined effects of coarticulation from the context vowel and the intervening consonant, or are the layers of statistical variation inseparable? In particular, can these effects be separated such that parsing affords a prediction of the upcoming context vowel even in the presence of a high-resistant intervening consonant? Is the parsing mechanism sensitive to differences between context vowels in their strength as triggers of coarticulation?

The experiment presented below seeks to answer these questions by examining the magnitude and reliability of anticipatory V-to-V coarticulatory effects on F1 and F2 in English across a range of VCV contexts, and the usefulness of those effects for the prediction of upcoming context. Furthermore, we examine coarticulation in cross-word contexts as a test of the generality of coarticulatory effects in the production of real words and phrases.

Our formant measures reveal variability in the effects of coarticulation on target vowel formants as a function of three factors related to the context: the identity of the context vowel (place of articulation), the identity of the intervening consonant (place and voicing), and the identity of the speaker. Our question is whether there are systematic coarticulatory effects of the context vowel that can be identified from the formant measures of the target vowel, *even in the face of variation due to the intervening consonant and speaker*.

Typical acoustic analyses examine F1 and F2 separately, using tests of significance to determine if each is affected by context vowels. However, this does not really speak to how such information might be used by the perceptual system for a number of reasons. First, such analyses typically assess means across several tokens, ignoring within-speaker, within-condition variability. Second, tests of significance do not map directly onto effect-size measurements, and it is not clear how to translate a significant difference into quantitative predictions about perception. Third, and most importantly, the perceptual system has access to both F1 and F2 simultaneously—any assessment of the utility of V-to-V coarticulation must be predicated on a statistical model that is able to combine sources of information. Thus, one of the most important goals of this work is to reexamine V-to-V coarticulation in this more comprehensive framework.

In short, we are seeking evidence that the expected patterns of coarticulation in F1 and F2 measures persist in a set of materials that introduce additional sources of variation, that they scale up to phonological contexts that span word boundaries, and that these coarticulatory effects will be evident when analyzed using a more comprehensive statistical approach that incorporates multiple sources of variation.

3. Experiment

Variability in V-to-V coarticulation was examined through a production experiment using a variety of VCV contexts and multiple speakers. We assessed the magnitude of anticipatory coarticulation, comparing the size of the shifts in F1 and F2 (in normalized Hz units) for two phonologically contrastive target vowels under conditions of four different context vowels, including a context vowel that was identical to the target and which was assumed to exert a neutral coarticulatory influence. The strength of the V-to-V coarticulation effect was also evaluated in comparison to variation in F1 and F2 due to other sources, such as the voicing or place of articulation of the intervening consonant, or speaker-based variation.

These assessments will reveal how each of the context factors influence F1 and F2 of the target vowel. To determine the size and potential utility of these effects, we conducted analyses in which both F1 and F2 were used to predict the identity of the context (vowel or consonant), as a way of quantifying how much information is available from these two acoustic measures combined.

3.1. Methods

3.1.1. Subjects

Ten native speakers of American English (5 males and 5 females) participated in this experiment. All were either graduate or undergraduate students at the University of Illinois, and all were under 30 years of age. Subjects were administered informed consent in accordance with University guidelines, and were not compensated for their participation.

3.1.2. Materials

Test materials were two-word phrases in which the first word contained the target vowel (either /ʌ/ or /ɛ/) and the second word contained the context vowel (/i/, /æ/ or /ɑ/). For example, in the test phrase *cup occupant*, we examined the influence of the initial /ɑ/ in *occupant* (the *trigger* vowel) on the acoustic realization of the /ʌ/ in *cup* (the *target* vowel). The full set of test phrases used in the experiment are listed in Table 1. We point out here that lexical items in these sometimes unusual phrases were chosen to meet the phonological criteria discussed below, plus the criterion that the phrase denote a picturable entity. The picturability criterion is motivated by our plan to use these test phrases in a future study of coarticulation perception by human listeners, using the visual world paradigm (see e.g., Gow & McMurray, 2007).

We measured the frequencies of the first and second formant for each target vowel. The central vowels /ʌ/ and /ɛ/ were used as target vowels because the location of these vowels in articulatory space allows them to shift in both height and backness under the influence of coarticulation (i.e., these vowels are not at the extreme front/back or height positions of the vowel space). Likewise, context vowels represented either the same vowel as the target (either /ʌ/ or /ɛ/) or three extremes of the American English vowel space (/i/, /ɑ/ or /æ/) that represent combinations of maximal height/lowness with maximal backness/frontness. The fourth extreme, the high back vowel /u/, was not used as it differs from the other context vowels in being rounded, and because /u/ productions by our speakers, as in many current varieties of American English, are often

Table 1
Materials used in the present experiment.

target	*context*	*target*	*context*	*target*	*context*
bed	actor eagle evergreen ostrich	tech	afternoon evening elevator oxygen	web	addict ecologist educator offer
wet	afro Easter Bunny eskimo oxen	deck	alligator easter basket elephant octopus	step	admiral east exit obstacle
mud	apple eater umpire observation	bug	astronaut evil underwear optician	pub	advertisement easel undergrad operator
cut	abdomen evenly onion olive	duck	athlete eating usher officer	cup	appetizer eavesdropping oven occupant

fronted and may therefore fail to condition strong backing effects in V-to-V coarticulation.

The first word in each sequence (containing the target vowel) was always monosyllabic and ended with a plosive consonant. Both voiced and voiceless consonants were used in final position of the first word, and consonants at all three major places of articulation (bilabial, alveolar, velar) were used. The only exception to this was that we excluded the /ɛ/ target vowel in front of /ɡ/ (e.g., as in "leg") as speakers tend to produce a higher, tenser vowel closer to /e/ in the context before /ɡ/ (Hartman, 1985; Kurath & McDavid, 1961, pp. 102, 132–133). Extra words with final /k/ were used in place of final /ɡ/ to yield equal numbers of words ending in labial, alveolar and velar plosives. For each combination of target vowel and final consonant there were two target words (with four words for the /ɛ + k/ combination) for a total of 12 distinct words for each target vowel.

The second word in each sequence was a vowel-initial word with one of four different context vowels as the initial segment. The context vowels /i, æ, ɑ/ were used in combination with both target vowels. The fourth context vowel was either /ɛ/ or /ʌ/, chosen to match the target vowel in order to define an identity context in which coarticulation effects would be null.

Context words were multisyllabic with a full, stressed vowel in the initial syllable. For 44 out of 48 context words, the initial syllable bears primary stress (e.g., *éducàtor*), while four of the context words have secondary stress on the initial syllable (e.g., *òbservátion*). This stress pattern was chosen to create a stressed vowel as the trigger for anticipatory coarticulation, in consideration of the finding from Fowler (1981) that stressed vowels exert a stronger coarticulatory influence on a neighboring vowel than do unstressed vowels. Fowler's study was restricted to within-word coarticulation, but we suppose that a stressed vowel might also be a stronger coarticulation trigger than an unstressed vowel across word boundaries, too.[1] With these restrictions on word structure in place, our coarticulation context is /…ˈVC#ˈV…/ or /…ˈVC#ˌV…/. Follow-up analyses indicate that the speakers in our study produced the target sequences with a syllable break at the word boundary, and did not resyllabify across the word boundary, i.e., they did not syllabify the cross-word VC#V sequence as [CV] [C#VX] in phrases like *bug#underwear*. The resyllabified sequence is predicted to be unlikely since it would leave the lax vowel of the target word stranded in an open syllable, in violation of English phonotactics. Auditory analysis and visual inspection of waveforms and spectrograms support the analysis of the word-final plosives as codas, primarily in finding that these plosives fail to exhibit several characteristics of an onset to a stressed syllable. First, although the final plosives are released (because they are not in pre-pausal, phrase-final position), they have characteristically low-amplitude release bursts. Second, the voiceless plosives have shorter VOT intervals, while the voiced plosives are either voiced throughout the closure, or have gradual devoicing towards the end of the closure. Furthermore, word-final /t/ is often produced with simultaneous glottalization, which is apparent in the presence of irregular pitch periods preceding or following the closure interval. These properties would be atypical for stressed syllable onsets, but are characteristic of coda consonants.

Combining the target words with the context words produced 48 test phrases in total (2 target vowels × 6 intervening consonants × 4 context vowels). The list of test phrases is given in Table 1. Test phrases containing the target vowel /ɛ/ were embedded in the carrier sentence *He said '_______' all the time.* Test phrases containing target vowel /ʌ/ were embedded in the carrier sentence *I love '_______' as a title.* In each case, the vowel in the word that precedes the test phrase was identical to the target (/ɛ/ in *said* and /ʌ/ in *love)* to minimize the influence of carryover articulation from the preceding vowel onto the target vowel.

3.1.3. *Prosodic context*

Prosodic factors can play a role in determining patterns of coarticulation, as noted above, with greater effects of coarticulation on unstressed or unaccented target vowels (Beddor et al., 2002; Cho, 2004; Fletcher, 2004; Fowler, 1981), and diminished coarticulation between vowels across a prosodic phrase boundary (Cho, 2004). On the other hand, accented vowels are not found to be more aggressive triggers of coarticulation (Cho, 2004), despite the fact that they characteristically exhibit more extreme front/back articulation, and in at least some cases, more open aperture. Based on these findings from prior studies of English, we expect coarticulation effects to be greatest in the present study if speakers realize the target word as unaccented, and with no prosodic phrase boundary between target and context words. Conversely, if target words are realized as accented, or if there is a prosodic boundary between target and context words, coarticulation effects may still be evident, but may be weaker.

Speakers were not instructed or coached to produce a particular prosodic realization of the test materials, but the use of a carrier phrase and the nature of the task (reading from a printed list) resulted in strikingly uniform prosody production both within and across speakers. Our auditory impression was that speakers realized both the target and context words as strongly accented (with longer, hyper-articulated stressed syllables and salient pitch movements), and with no intervening prosodic phrase boundary. This impression was confirmed through a follow-up prosodic analysis of 60% of the test utterances, including analysis of

[1] Cho's (2004) articulatory study of coarticulation across word boundaries looks for an effect of phrasal stress (pitch accent) on coarticulation, but does not find evidence that pitch-accented vowels are stronger (more "aggressive") triggers than unaccented vowels. To our knowledge, no study has yet compared the behavior of stressed and unstressed vowels in cross-word coarticulation, in part because few studies have assessed V-to-V coarticulation across word boundaries.

detailed ToBI transcriptions for a randomly selected subset of 120 utterances. Thus, the prosodic realization of the test materials is compatible with the occurrence of coarticulation between target and context words, due to the absence of an intervening phrasal juncture, though the strength of coarticulation may be diminished due to pitch accent assigned to the target word. Overall, these materials provide an appropriate, though somewhat conservative test of V-to-V coarticulation. Details from our prosodic analysis are presented in the Appendix.

3.1.4. Procedure

Stimuli (test phrases embedded in carrier sentences) were presented in text format to subjects using E-Prime presentation software. Subjects were seated at a computer in an isolated booth and were asked to read each sentence as naturally as possible when it appeared on the screen. Each sentence remained on the screen for 4 s. The entire set of 48 sentences was repeated in three blocks for a total of 144 trials (48 test phrases × 3 repetitions).

Test items were presented in quasi-random order within each group. A number of constraints on the sequence were implemented in order to avoid inducing speech errors due to recency effects. Randomization was constrained such that any phrase containing a particular target word was separated from any other phrase containing that same target word by at least two other trials. For example, the trial containing the test phrase *bed actor* was separated from any other phrase with the first word *bed* by a minimum of two trials. This spacing of trials containing the same target word but different context words was intended to discourage a prosodic realization with contrastive focus pitch accent on the context word and an unaccented target word, and no such productions were observed (details in Appendix I). In addition, the set of target words includes three sets of minimal pairs: *wet–web*, *cut–cup* and *duck–deck*. Phrases containing these words were separated from phrases containing the minimal pair counterpart by at least two trials.

The experiment was self-paced in that subjects regulated the amount of time between trials. Subjects were told that it was acceptable to take a break at any time, but none of the subjects took a break for more than 1 or 2 min. Most subjects completed the experiment in about 15 min, and no subject took more than 20 min.

3.1.5. Recording and measurement

The speech data for each speaker was recorded on digital audio tape with a sampling rate of 44.1 kHz, transferred to computer, and then analyzed using Praat speech analysis software (Boersma & Weenink, 2005). Measurements of F1 and F2 at the midpoint of the target vowel were made based on LPC analysis with the Burg formula (the Praat default). The midpoint was selected as a location because V-to-V effects are expected to persist to that point but with a lesser effect of masking by the C-to-V effects that are expected at the vowel offset (Fowler, 2005; Fowler & Brancazio, 2000). Five formants were located within 0–5000 Hz (males) or 0–5500 Hz (females), within a 50 ms Gaussian window (comparable to a Hamming window of 25 ms). Formant values were considered against reference values from Peterson and Barney (1952) and Hillenbrand, Getty, Clark, and Wheeler (1995), and outliers (values that deviated significantly from reference values for speakers of same sex and/or that approximated values for another vowel category) were inspected visually and corrected on the basis of the visual spectrogram and formant displays.[2] Less than 8% of the tokens were visually inspected and adjusted in this procedure.

Formant frequencies were measured in Hz, but converted to Bark for analysis. This was done so that the relative variance in these measurements would be scaled along an approximately psychophysical, not physical, dimension, something that might better reflect the listeners' experience.

Trials that contained speech errors (production of the incorrect vowel or word, or disfluent production) were eliminated from the analysis. Across 1440 trials (10 subjects × 144 trials) there were 22 speech errors. In addition, six of the ten subjects pronounced *ecologist* with an initial schwa rather than the expected /i/. This resulted in an additional 18 trials being eliminated from the study. A total of 1400 target vowels were analyzed (1440—22 speech errors and 18 unexpected pronunciations of *ecologist*).

3.2. Results

Our analyses of these data are reported in two parts. In the first set of analyses we use standard analysis of variance (ANOVA) techniques to establish the presence of V-to-V and C-to-V coarticulation. The ANOVA framework, however, cannot answer our fundamental questions about the relative magnitude of acoustic variability from different sources. Furthermore, although ANOVA can establish reliable differences, it cannot say how useful such differences could be for perception because it collapses data within subjects, cannot treat F1 and F2 simultaneously to predict context, and ignores the inherent temporal sequence of contextual information. Thus, in Section 2, we address these questions with a series of regression analyses that assess the size of the effect from V-to-V coarticulation

[2] Data from the Hillenbrand et al. (1995) study are from speakers in Michigan, and exhibit the characteristic features of the Inland North dialect (Labov, Ash, & Boberg, 2006), including the fronting and raising of /æ/, the backing of /ɛ/ (for females) and the fronting of /ɑ/, relative to the positions of these vowels in the Peterson & Barney dataset. The speakers in our study represent primarily the Midland dialect, and are most similar to the Peterson & Barney speakers in their vowel formants, with two exceptions: the /ɛ/ produced by our female speakers is somewhat backed (lower F2) and closer to the corresponding vowel for females in the Hillenbrand et al. database, and the /æ/ of our male speakers is farther back than the corresponding vowel by male speakers in either of the other two databases, though for all our male speakers the /æ/ distribution is non-overlapping with /ɑ/.

relative to other sources of variation to establish how useful V-to-V coarticulation might be for predicting upcoming context, and whether it is necessary to first account for other sources of variation.

3.2.1. Sources of variance (ANOVA)

There are five sources of variance that potentially affect the measured F1 and F2 values: speaker, target vowel, place and voicing of the intervening consonant, and V-to-V coarticulation (context vowel). These were assessed in a repeated measures ANOVA which treats subject implicitly[3] and assesses the significance of the other four effects, over and above any variance due to subject. Thus, while these analyses do not report the significance of subject as a source of variance, they do take it into account when assessing the others. The regression analysis we present shortly treats subject more explicitly.

An ideal statistical approach would be to use all four experimental factors in a repeated measures ANOVA to address the relative contributions of each and their potential interactions. However, the fact that we excluded the /ɛg/ context makes this impossible (since one cell would be empty). To deal with this, we conducted separate repeated measures ANOVAs, one which included place of the intervening consonant as a factor (averaging across voicing conditions) and one which included voicing (averaging across place). Both included target vowel and context vowel as independent factors. These ANOVAs (place × target × context, and voicing × target × context) were conducted separately for F1 and F2 as dependent variables, yielding four analyses.

Both ANOVAs (one with consonant place as a factor and one with voicing) deal with variance in F1 or F2 due to the target vowel and the intervening consonant—effects that contribute to the perceptual milieu in which V-to-V coarticulation is found, but which are not of central interest to this paper. Thus, while we report the full analyses below, we will not directly report results of planned comparisons designed to explore interactions that do not directly impinge on our findings of V-to-V coarticulation (e.g., a place × target vowel interaction). The complete set of statistical results are reported in an online supplement to this study.

3.2.1.1. First formant frequency. Our first analysis examined the effect of target vowel (/ɛ, ʌ/), place of articulation (labial, coronal, velar) and (most importantly) context vowel (/i, æ, ɑ/ and the neutral context vowel) on the F1 of the target vowel. We did not find a significant effect of place of articulation ($F(2,18)=1.4$, $\eta^2_P=.14$, $p>.2$). Target vowel was significant, however, ($F(1,9)=10.2$, $\eta^2_P=.53$, $p=.011$), with /ʌ/ showing higher F1 values than /ɛ/. Most importantly, context vowel was highly significant ($F(3,27)=23.3$, $\eta^2_P=.72$, $p<.0001$). Planned comparisons revealed that F1 under all three non-neutral context vowels differed significantly from F1 under the neutral context vowel (/æ/: $F(1,9)=8.9$, $p=.015$; /i/: $F(1,9)=16.0$, $p=.0031$; /ɑ/: $F(1,9)=16.6$, $p=.0028$).

The differences between the non-neutral context vowels in their effects on F1 of the target vowel reflected the relative F1 values of the context vowels themselves, as expected. Thus, comparing F1 values of target vowels under different context vowels, F1 under the context vowel /i/ differed from the context vowel /æ/ ($F(1,9)=65.1$, $p<.0001$), while F1 under context vowel /ɑ/ did not differ significantly from F1 under context vowel /æ/ ($F<1$), suggesting that the coarticulatory effect of context vowel on target vowel F1 codes primarily the height of the upcoming vowel. Similarly, the relatively smaller effect of target vowel on F1 variation, as compared to the effect of context vowel, is likely due to the fact that /ʌ/ and /ɛ/ do not differ primarily on height.

There were a number of interactions present as well, which will be discussed only to the extent that they impact our findings of V-to-V coarticulation. Place and target vowel interacted ($F(2,18)=3.6$, $p=.049$). Of more concern was the fact that context vowel and place interacted ($F(6,54)=3.7$, $p=.0039$). This was driven by the fact that context vowel affected F1 of the target vowels when the intervening consonant was coronal ($F(3,57)=3.4$, $p=.023$) or labial ($F(3,57)=3.3$, $p=.025$), but not when the intervening consonant was velar ($F<1$). Finally, the three-way interaction was significant ($F(6,54)=4.4$, $p=.0011$).

We next examined F1 as a function of voicing, target vowel and context vowel. Voicing was significant ($F(1,9)=6.8$, $\eta^2_P=.43$, $p=.028$) with lower F1 values when the intervening consonant was voiced than when it was voiceless (see Kingston et al., 2008). As before, the effect of target vowel was significant ($F(1,9)=12.6$, $\eta^2_P=.58$, $p=.0062$). Most importantly, context vowel was also significant ($F(3,27)=27.5$, $\eta^2_P=.75$, $p<.0001$). None of the interactions were significant (voicing × target: $F<1$; voicing × context: $F(3,27)=2.4$, $p=.087$; target × context: $F<1$; three-way: $F<1$).

3.2.1.2. Second formant frequency. The next analyses examined F2. As before, we start with an ANOVA examining the effect of place of articulation, target vowel, and context vowel on F2 of the target vowel. Place was significant ($F(2,18)=67.0$, $\eta^2_P=.88$, $p<.0001$): the lowest F2 values were found when the intervening consonant was labial, followed by coronals and then velars. There was also a significant effect of target vowel ($F(1,9)=184.9$, $\eta^2_P=.95$, $p<.0001$) with /ɛ/ vowels having much higher F2 frequencies ($M=12.4$ bark) than /ʌ/ vowels ($M=10.8$ bark). Most importantly, context vowel was significant ($F(3,27)=41.4$, $\eta^2_P=.82$, $p<.0001$). Follow-up analyses revealed that all three vowels were different from the neutral condition (/æ/: $F(1,9)=16.3$, $p=.0029$; /i/: $F(1,9)=71.5$, $p<.0001$;

[3] Note that because the repeated measures analysis (and the regressions we use shortly) treats each subject independently, accounting for individual subject effects also accounts for any variance due to between-subject effects like gender.

/ɑ/: $F(1,9)=15.3$, $p=.0036$). In addition, F2 of the target vowel was different for all comparisons of context vowel, reflecting a phonological or phonetic distinction in backness: /æ/ was different from /ɑ/ ($F(1,9)=31.0$, $p=.0003$), and /i/ was different from /æ/ ($F(1,9)=26.1$, $p=.0006$).

All of the interactions were significant. Place interacted with target vowel ($F(2,18)=57.9$, $p<.0001$) and with context vowel ($F(6,54)=4.6$, $p=.0008$). With respect to this latter interaction, follow-up tests revealed that the effect of context vowel held across all three places of articulation (Coronal: $F(3,27)=39.6$, $p<.0001$; Labial: $F(3,27)=13.9$, $p<.0001$; Velar: $F(3,27)=16.2$, $p<.0001$). Context vowel also interacted with target vowel ($F(3,27)=4.8$, $p=.0081$), and as before, separate analyses revealed that the effect of context held up for both /ʌ/ ($F(3,27)=42.2$, $p<.0001$) and /ɛ/ ($F(3,27)=24.6$, $p<.0001$). Finally, the three-way interaction was significant ($F(6,54)=2.6$, $p=.027$).

We next replicated the above analyses using voicing instead of place. This ANOVA examined voicing, target vowel and context vowel effects on F2. Voicing was significant ($F(1,9)=78.8$, $\eta^2_{\mathrm{P}}=.90$, $p<.0001$), with voiced consonants conditioning lower F2 values on the (preceding) target vowel than voiceless consonants. As before, target vowel was also significant ($F(1,9)=168.8$, $\eta^2_{\mathrm{P}}=.95$, $p<.0001$). Finally, and most importantly, context vowel was still significant ($F(3,27)=41.7$, $\eta^2_{\mathrm{P}}=.82$, $p<.0001$).

Of the interactions, only two were significant. The voicing × target vowel interaction was significant ($F(1,9)=25.6$, $p=.0007$), due to the fact that the effect of voicing was larger for /ʌ/ than /ɛ/. Second, target vowel interacted with context ($F(3,27)=3.1$, $p=.04$). Follow-up analyses, however, revealed that the effect of context was significant in both target vowels (/ɛ/: $F(3,27)=20.9$, $p<.0001$; /ʌ/: $F(3,27)=52.2$, $p<.0001$).

3.2.1.3. Summary. These results lay out a robust pattern of coarticulation that is evident even in the centroid frequencies of the first and second formant. Clearly, both formants are strongly affected by the identity of the target vowel as /ɛ/ or /ʌ/, though the effect on F2 ($\eta^2_{\mathrm{P}}=.95$) was greater than that on F1 ($\eta^2_{\mathrm{P}}=.53$). The intervening consonant also has large effects on the vowel, with place affecting F2 ($\eta^2_{\mathrm{P}}=.88$) and voicing affecting both (F1: $\eta^2_{\mathrm{P}}=.43$; F2: $\eta^2_{\mathrm{P}}=.90$).

More importantly for the present purposes, on top of this rich pattern of variation, V-to-V effects are seen to be robust as well. The V-to-V effects appear in both formants, though they may be attenuated in F1 when the intervening consonant is velar. All three context vowels induced coarticulation, showing differences in F1 and F2 from the neutral context. Moreover, V-to-V effects (F1: $\eta^2_{\mathrm{P}}=.72$; F2: $\eta^2_{\mathrm{P}}=.82$) were similar in size to place and voicing effects, particularly for F2, and we find little evidence for a reduced V-to-V effect in F1. Finally, with respect to V-to-V coarticulation, F1 coded predominantly the height of the context vowel, while F2 coded the phonological backness contrast between /æ/ and /ɑ/ (as expected) but also the phonetic backness distinction between /i/ and /æ/, both of which are phonologically [-back].

3.2.2. Statistical modeling of V-to-V effects in context

The foregoing analysis partly replicates prior work, and moreover suggests that the acoustic effects from V-to-V coarticulation (i.e., the effects above due to context vowel) are a potentially powerful source of perceptual information about context vowels. Indeed, this use of fine-grained detail is an important feature of exemplar models (Hawkins, 2003; Johnson, 1997; Pierrehumbert, 2001) which posit that lexical representations of words are specified in terms of fine phonetic and non-phonetic detail. If these models are correct then listeners should be able to harness such information to make robust inferences about context. But is it this simple?

Fig. 2 displays a scatter plot of the all of the data (the F1 and F2 measures from each target vowel), with tokens coded for their context vowel. No immediate pattern of clustering based on context vowel is clearly visible—in fact the two clusters that can barely be discerned correspond to male and female speakers. This seems to suggest that for the listener perceiving a single token of a coarticulated vowel in running speech, differences attributed to vowel context would be of very little use in predicting the upcoming context.

We constructed a statistical model to evaluate whether the acoustic effects of V-to-V coarticulation may provide any useful information about the upcoming context vowel, using a multinomial logistic regression to map the raw formant frequencies onto categories of context vowel (see Nearey, 1997 for a similar approach). This statistical technique works similarly to binary logistic regression in that it maps one or more continuous independent variables onto a discrete category. It differs in that this discrete category need not be binary. In this case, the model was used to predict the category of context vowel (/i, ɑ, æ/ or neutral) from the raw F1 frequency, the raw F2 frequency and an interaction of F1 and F2.

Overall this model barely fit the data ($\chi^2(9)=65.47$, $p=.04$). When used to predict the context vowel (in the given dataset), it averaged only 28.63% correct—barely above chance (25%). It did well at predicting /i/ (51.3%) and /ɑ/ (37.5%) but it did so by simply being biased toward these responses (it responded /i/: 41%, /ɑ/: 32.2%, where only 25% of the stimuli were /i/), rather than making use of the acoustics. In fact, individual likelihood ratio tests of F1, F2 and the interaction revealed that none were significant (F1: $\chi^2(3)=3.47$, $p>.2$; F2: $\chi^2(3)=1.86$, $p>.2$; F1 × F2: $\chi^2(3)=2.7$, $p>.2$). Thus, it does not appear that the sensitivity to raw F1 and F2 values really support the use of context to predict the upcoming vowel.

Given these results, how can we explain the robust effects of context vowel seen in the ANOVA? A crucial consideration is that repeated measures ANOVAs implicitly normalize for subject variation, asking if a given

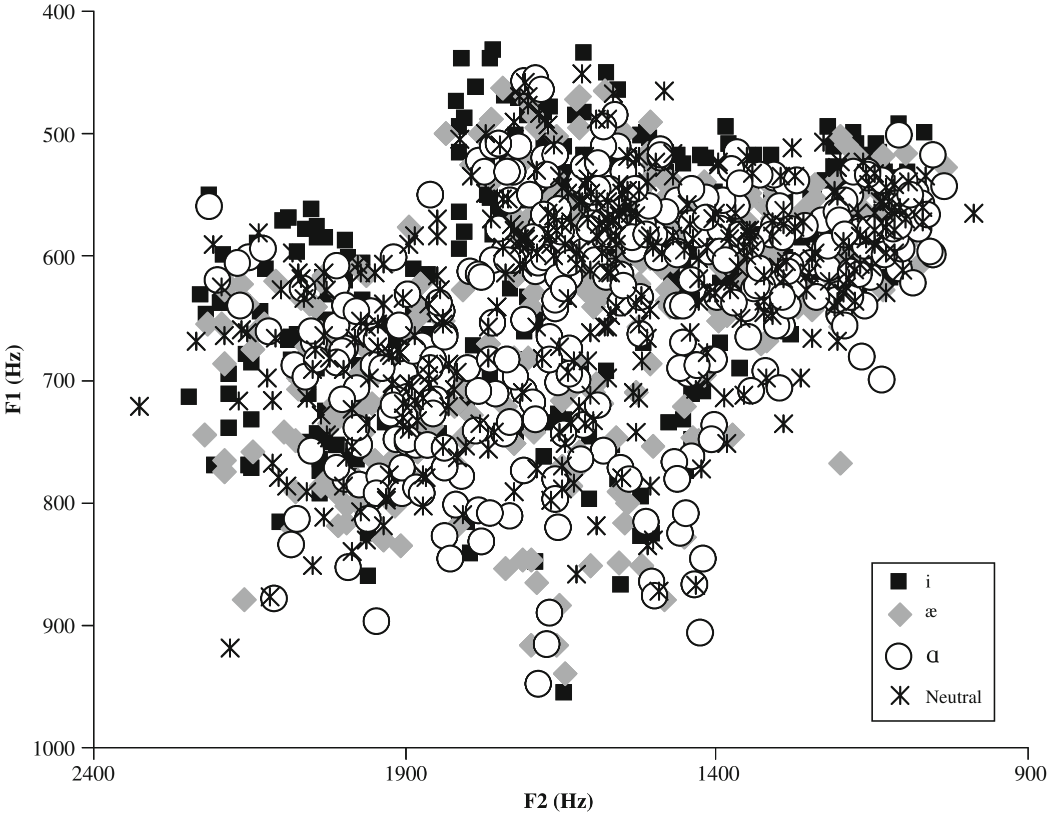

Fig. 2. F1 and F2 frequency of target vowels in each of the four V-to-V coarticulatory contexts (/i, æ, ɑ/). Note that while formant frequencies are presented in Hz, analyses were conducted in Bark scaling.

factor can increase or decrease the dependent measure, relative to the subject's own mean value. For example, it asks if F1 in the context of an upcoming /ɑ/ is higher than in the context of /i/, relative to each subject's own mean. Similarly, ANOVA implicitly accounts for the effects of other factors (e.g., consonant, target vowel), controlling for them while testing for effects of V-to-V coarticulation. As a crude model of speech perception, then, the ANOVA framework considers interactions between factors and features much better than the sort of bottom-up analysis represented by this particular multinomial logistic regression.

This then raises the question of what factors need to be accounted for before the acoustic effects of V-to-V coarticulation are useful for predicting upcoming context. One way to ask this is to use linear regression to partial out the effects of factors like speaker, target vowel and consonant on the acoustic cues to context vowel (F1 and F2) prior to using them in the multinomial logistic regression. Regression thus provides a simple model of the perceptual processes that human listeners use (e.g., processes that underlie compensation effects in speech perception) to take into account multiple sources of variation when interpreting continuous cue values (see McMurray, Cole, & Munson, in review, for a more complete discussion), and the multinomial logistic regression instantiates a categorization process that works from these parsed F1 and F2 values to predict the context vowel.

Moreover, there is an inherent order to the way factors like speaker and consonant are interpreted—information about these factors arrives at different points in time, and psycholinguistic evidence suggests that listeners are exquisitely sensitive to this time course (McMurray, Clayards, Tanenhaus, & Aslin, 2008; Warren & Marslen-Wilson, 1988). ANOVA assumes all factors contribute at the same time when assigning shared variance, yet given the sequential nature of speech, the processing system may give precedence to early arriving factors when assigning shared variance.

Used in this way, linear regression offers a simple model of parsing, a theoretical viewpoint developed by Fowler (1984) and Gow (2003; see McMurray et al., in review for a more complete discussion of the regression approach). Linear regression models use the available variation in the input to identify one contributing factor (e.g., the identity of the target vowel). They then compute a residual (the difference between the prototype value for that target vowel and the current input) and use this residual to identify other factors that influence the signal (e.g., the context vowel). We can control which factors are entered into the model at any given time, allowing us to model the sequential uptake of information.

The next series of analyses use regression to account for variation in the acoustic realization of the target vowels in our database. We will systematically examine each of the four factors manipulated in this study (speaker, target vowel, place and voicing of intervening syllable) as well as interactions between these factors by partialing these effects out of F1 and F2 and then using the residuals to predict the target vowel. We will partial out factors in the order that they are likely to be available to the listener: starting with information about the speaker, then the target vowel, then the intervening consonant. The results of these regressions are reported in Table 2 (for F1) and Table 3 (for F2). The residuals from each step of these analyses will be used in multinomial logistic regressions to determine how well the context vowel can be predicted.

3.2.2.1. Speaker effects. The first analysis examined speaker effects on F1 and F2. Since we expected that speaker gender may play a role, on the first step we entered gender into the model by coding each target vowel token for the gender of the speaker. Next, nine dummy variables were added to account for individual differences between the 10 speakers (e.g., a dummy variable coding *speaker1* will have the value 1 for data from speaker 1, and 0 for data from any other speaker). Gender accounted for 63.2% of the variation in F1 (Table 2, line 1) and 36% of the variance in F2 (Table 3, line 1). Individual speaker differences accounted for an additional 19% of the variance in (Table 2, line 2) and 5% of the variance in F2 (Table 3, line 2). Clearly, speaker effects were seen predominantly in F1.

Table 2
A regression analysis examining F1.

#	Effect	R^2	R^2_{change}	F_{change}	p
1	Gender	.63	.63	$F(1,473)=811.5$	$<.0001$
2	Speaker (10)	.82	.19	$F(8,465)=63.5$	$<.0001$
3	Target vowel	.83	.009	$F(1,464)=25.4$	$<.0001$
4	Cons. voicing	.85	.018	$F(1,463)=56.5$	$<.0001$
5	Cons. place (2)	.85	.003	$F(2,461)=5.1$	.0063
6	Voicing × place (2)	.87	.01	$F(2,459)=22.8$	$<.0001$
7	Voicing × vowel	.87	.00	$F(1,458)<1$	$>.2$
8	Vowel × place (2)	.87	.00	$F(2,456)<1$	$>.2$
9	Three-way	.87	.00	$F(1,455)<1$	$>.2$
10	Context vowel[a]	.88	.012	$F(3,452)=15.3$	$<.0001$

[a] Note residuals from this step were not used in the multinomial logistic regression.

Table 3
A regression analysis examining F2.

#	Effect	R^2	R^2_{change}	F_{change}	p
1	Gender	.36	.36	$F(1,473)=264.7$	$<.0001$
2	Speaker (10)	.41	.05	$F(8,465)=4.8$	$<.0001$
3	Target vowel	.82	.41	$F(1,464)=1066.9$	$<.0001$
4	Cons. voicing	.85	.03	$F(1,463)=107.0$	$<.0001$
5	Cons. place (2)	.90	.05	$F(2,461)=120.9$	$<.0001$
6	Voicing × place (2)	.92	.014	$F(2,459)=39.4$	$<.0001$
7	Voicing × vowel	.92	.004	$F(1,458)=24.7$	$<.0001$
8	Vowel × place (2)	.93	.008	$F(2,456)=26.2$	$<.0001$
9	Three-way	.94	.007	$F(1,455)=49.0$	$<.0001$
10	Context vowel	.94	.007	$F(3,452)=20.4$	$<.0001$

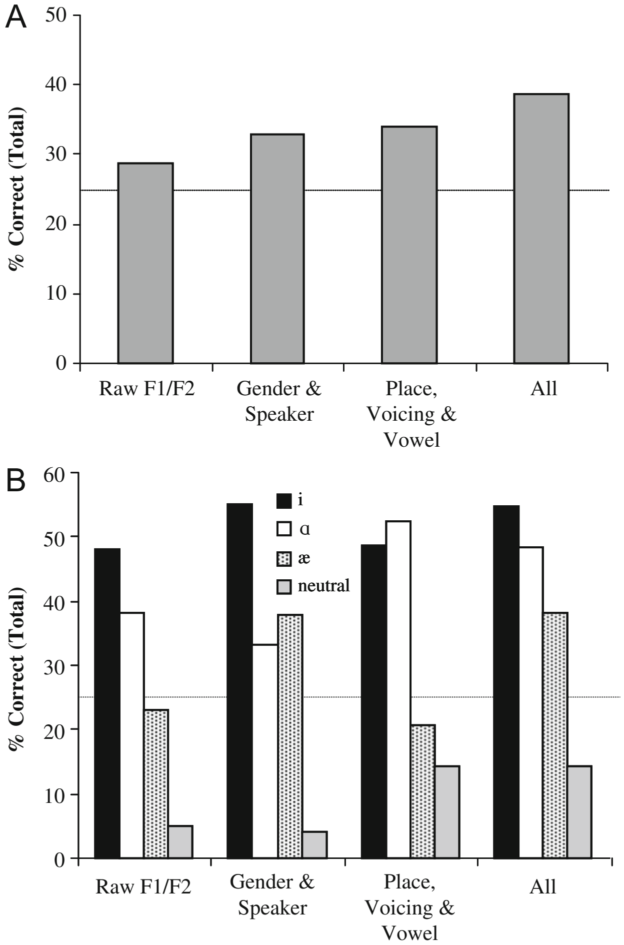

Fig. 3. Mean performance of the multinomial logistic regression models. (A) Overall performance of each model). (B) Performance of the same four models as a function of what context vowel they were predicting. In each panel the horizontal line represents the chance value.

The residuals of these analyses were entered into a multinomial logistic regression (see Fig. 3 for a complete summary). This model was a much better fit to the data than the model using raw values ($\chi^2(9)=35.5$, $p=.00005$) and averaged 32% correct. As before, performance was best for /i/ (49.5% correct), but it was also above chance for /ɑ/ (33.3) and /æ/ (40.0%). Finally, in likelihood ratio tests, F1 was individually significant ($\chi^2(3)=29.25$, $p<.00001$), although F2 and the interaction of F1 and F2 were not (F2: $\chi^2(3)=3.8$, $p>.2$; F1 × F2: $\chi^2(3)=.2$, $p>.2$), establishing the fact that speaker factors were more strongly correlated with F1 than F2.

3.2.2.2. Target vowel effects. In the second analysis, target vowel effects were partialed out of F1 and F2, after the effects of speaker had been removed. Target vowel significantly accounted for an additional 1% of the variance in F1 (Table 2, line 3) and a substantial 41.3% of the variance in F2 (Table 3, line 3)—unsurprising, given that the target vowels contrast phonologically in backness, not in height.

When these effects were partialed out of F1 and F2, the multinomial logistic regression was further improved, providing an excellent fit to the data ($\chi^2(9) = 65.5$, $p < .0001$), and averaging 35% correct (well above chance). At this step, both F1 and F2 were significant by the likelihood ratio tests (F1: $\chi^2(3) = 40.3$, $p < .0001$; F2: $\chi^2(3) = 27.7$, $p < .0001$) although the interaction was not significant ($\chi^2(3) = .59$, $p > .2$). Given that speaker factors primarily affected F1 (which was significant in the prior analysis) but target vowel affected largely F2 (which had a much larger R^2 here), the newly significant effect of F2 for predicting context vowel reinforces the notion that parsing variation out of the signal can improve categorization performance.

3.2.2.3. Consonant effects. The next analysis partialed out the place and voicing of the intervening consonant. First, the effect of voicing was added to the model (which already contained variables for subject and target vowel). Voicing significantly accounted for 1.8% of the variance in F1 (Table 2, line 4) and 3% of the variance in F2 (Table 3, line 4). On the next step, two variables representing place (one variable for ±labial and another for ±velar) were added and significantly accounted for .3% of the variance in F1 (Table 2, line 5) and 5% of the variance in F2 (Table 3, line 5). These results demonstrate a greater effect of the intervening consonant on F2 than F1.

When the residuals from this regression were added to the multinomial logistic regression, the model fit was very good ($\chi^2(9) = 87.80$, $p < .0001$), and averaged 39.3% correct. It was particularly good for /i/ (58%) and /ɑ/ (48.3%), and above chance for /æ/ (36.5%). However, as in the previous models, it achieved this performance with an overall bias against the neutral context (it was less than chance, averaging 14.1% correct). Both F1 and F2 were highly significant components of the model (F1: $\chi^2(3) = 48.0$, $p < .0001$; F2: $\chi^2(3) = 44.9$, $p < .0001$), although as before the interaction was not significant ($\chi^2(3) = 2.15$, $p > .2$).

3.2.2.4. Interactions. In the final analysis, interaction terms were added to the model to account for the two-way interactions between place, voicing and target vowel. First, two variables representing the interaction of voicing and place were added to the model (which already contained speaker, target vowel, place and voicing). They significantly accounted for 1% of the variance in F1 (Table 2, line 6) and 1.4% of the variance in F2 (Table 3, line 6). Next, the interaction of voicing and target vowel was added. It accounted for no new variance in F1, but an additional .4% of the variance in F2 (Tables 2 and 3, line 7). Next, the interaction of target vowel and place was added to the model. As in the prior step, it accounted for no new variation in F1, but significantly accounted for .8% of the variance in F2 (Tables 2 and 3, line 8). Finally, the three-way interaction behaved similarly, accounting for no additional variance in F1, but .7% of the variance in F2 (Tables 2 and 3, line 9). Thus, as a whole the interactions accounted for 1% of the variance in F1 and 3.3% of the variance in F2—interactions between voicing, place and target vowel clearly had more of an effect in this measure.

When the residuals from the analysis including interaction factors were used as independent measures in the multinomial logistic regression, model fit was excellent ($\chi^2(9) = 112.6$, $p < .0001$) and the model performed at 39.8% correct (/i/: 53.9%, /ɑ/: 51.7%, /æ/: 34.2, Neutral: 20.0%). F1 and F2 were both significant (F1: $\chi^2(3) = 41.3$, $p < .0001$; F2: $\chi^2(3) = 59.4$, $p < .0001$) although as before, the interaction was not ($\chi^2(3) = 5.4$, $p = .14$).

3.2.2.5. Coarticulation resistance? Fowler's work on coarticulation resistance (Fowler, 2005; Fowler & Brancazio, 2000) suggests that some consonants (coronals in those studies) may impede V-to-V coarticulation. The ANOVAs provide partial support for this as they showed no effect of V-to-V context on F1 when the intervening consonant was velar. Does this mean that listeners would not be able to take advantage of V-to-V coarticulation in certain consonantal contexts?

This question is difficult to answer by examining a single measure (e.g., F1) as it is not clear whether the loss of one cue could be compensated for by another (F2). However, the complete multinomial logistic regression model developed so far can be used to answer this question by examining the predictability of the context vowel separately for labial, velar and alveolar contexts. Thus, we used the F1 and F2 values for which all of the prior factors had been partialed out in three multinomial logistic regressions looking only at one-third of the data.

The model using only labial contexts did about as well as the complete model. Model fit was good ($\chi^2(9) = 38.8$, $p < .0001$) and the model performed at 39.3% correct (/i/: 51.4%, /ɑ/: 57.5%, /æ/: 32.5, Neutral: 17.5%). F1 and F2 were both significant (F1: $\chi^2(3) = 25.3$, $p < .0001$; F2: $\chi^2(3) = 19.5$, $p < .0001$) although as before, the interaction was not ($\chi^2(3) = 1.2$, $p > .2$).

In contrast, the model looking only at alveolars did much better, showing much better model fit ($\chi^2(9) = 73.3$, $p < .0001$) and performing at 47.5% correct (/i/: 65.0%, /ɑ/: 50.0%, /æ/: 47.5, Neutral: 27.5%)! This model took advantage of both F1 ($\chi^2(3) = 12.9$, $p = .0049$), and F2 ($\chi^2(3) = 33.05$, $p < .0001$) as well as the interaction ($\chi^2(3) = 8.3$, $p = .04$).

Finally, the model using only velar contexts performed worse than either model. Model fit was moderate ($\chi^2(9) = 24.3$, $p = .003$), and performance was only at 35.6% correct, which it achieved largely by predicting /i/ and /ɑ/ and ignoring /æ/ and neutral contexts (/i/: 57.4%, /ɑ/: 60.0%, /æ/: 7.5%, Neutral: 17.5%). Surprisingly, the model made use of both F1 ($\chi^2(3) = 9.4$, $p = .02$) and F2 ($\chi^2(3) = 13.3$, $p = .004$), though the interaction was not significant and ($\chi^2(3) = 3.2$, $p > .2$). Importantly, while it performed worse than models examining labial and alveolar contexts, this model was able to make some accurate predictions (and achieved a significant fit), suggesting that (a) F1 was not totally lost in velar contexts (it was individually significant) and (b) its degradation can be compensated for by F2.

3.2.2.6. Effect of context vowel height and backness. Although the previous models examined V-to-V coarticulation by examining how well the multinomial logistic regression could predict the vowel context, we can also use linear regression to examine how much variance it accounts for. This can help determine the relative sizes of each of these effects. Thus, in the final linear regressions, we added three variables to the model coding the height and backness of the context vowel and whether or not it matched the target vowel. With variance due to speaker, target vowel, consonant place and voicing already partialed out, these new variables coding context vowel accounted for an additional 1.2% of the variance in F1 ($F_{\text{change}}(3,452)=15.3$, $p<.0001$) and .7% of the variance in F2 ($F_{\text{change}}(3,452)=20.4$, $p<.0001$).

3.2.3. Summary and discussion

Clearly, parsing (as instantiated in our simple linear model) is an effective way to remove variance due to one factor, and more importantly *reveal the effects of others.* The unparsed model, taking into account only the raw formant data of the target vowel, averaged 28.6% correct in predicting the context vowel (barely above chance), while the full model that partialed out multiple sources of variance was able to achieve 39.8% correct. The effect of parsing can be seen in the incremental clustering of the formant data at each step in the model in Fig. 4.

Fig. 4 shows the raw F1 and F2 values (panel A) and residuals (panels B–F) used in each of the parsing models. Each data point represents one target vowel token, where different symbols are used to code the context vowel associated with each token. Panel A shows the raw data, and as discussed before, little separation can be seen between target vowels on the basis of the context vowel. However, by Panel E (the complete model), there is significant separation: target vowels produced in the context of an upcoming /i/ (dark squares) generally are to the top left, target vowels in the context of /ɑ/ (open circles) are to the bottom right, and those in the context of /æ/ (gray diamonds) are to the bottom left. Interestingly, when the neutral contexts are added to this plot in Panel F, it is clear why the model consistently struggled to predict the neutral context. Even after parsing out virtually all sources of variation, the formants in the neutral context varied substantially. Evidently, speakers permit more variation in vowel production in a neutral context than in a (non-neutral) coarticulatory one.

The linear regression analyses also let us examine the relative weighting of V-to-V coarticulation to other sources of variation. Not surprisingly, speaker was clearly the most important source of variation for F1, and for a substantial portion of the variation in F2. Also as expected, the choice of target vowel played a huge role in F1 ($R^2=.41$) but a lesser role in F2. Together, these two factors alone accounted for 83% of the variance in F1, and 82% of the variance in F2. Thus, coarticulatory effects due to the adjacent consonant and upcoming vowel context will necessarily be small (since there is less than 20% of the variance remaining).

We saw that for F1 the place and voicing of the intervening consonant and the interactions of these factors accounted for 3.1% of the variance (cumulatively), while context vowel accounted for 1.2%. Thus, the effect of V-to-V coarticulation on F1 is similar in size to that of voicing (1.8%) and about one-third of the size of the combined consonantal interactions. For F2, the story was different. Here, consonantal factors accounted for a cumulative total of 11.3% of the variance, and context vowel only .7%, far smaller than either place (5%) or voicing (3%) individually. However, it is important to point out that by the time information from the context vowel is available (i.e., at the start of the second word), the model has already accounted for 93.7% of the variance in the target vowel, so there is very little ambiguity in the signal left to explain, relative to identification of the target vowel.

The relative effect sizes of context vowel and consonant suggest one final prediction—that a multinomial logistic regression should be quite good at predicting properties of the consonant from the target vowel alone. To test this, we used the F1 and F2 frequencies from which speaker and target vowel had been partialed out in three multinomial logistic regressions predicting either place, voicing, or both properties of the intervening consonant. The analysis examining voicing offered a good fit ($\chi^2(3)=137.7$) and averaged 75.8% correct. Both F1 and F2 were significant predictors of consonant voicing (F1: $\chi^2(1)=40.1$, $p<.0001$; F2: $\chi^2(1)=70.8$, $p<.0001$), although their interaction was not ($\chi^2(1)=1.8$, $p=.17$). The analysis examining place was also quite good ($\chi^2(6)=208.2$, $p<.0001$), and performed at 57.7% correct. All three independent measures were significant under this analysis (F1: $\chi^2(2)=15.2$, $p<.0001$; F2: $\chi^2(2)=163.7$, $p<.0001$; F1 × F2: $\chi^2(2)=25.5$, $p<.0001$). Finally, the model predicting both place and voicing offered an excellent fit ($\chi^2(15)=458.4$, $p<.0001$) and averaged 49.8% correct. All three covariates were significant (F1: $\chi^2(5)=37.0$, $p<.0001$; F2: $\chi^2(5)=291.5$, $p<.0001$; F1 × F2: $\chi^2(5)=22.4$, $p<.0001$).

While the voicing results should be tempered by the fact that chance was only 50%, the place model's task was closer to that of the models predicting the context vowel (estimating three categories instead of four). All three performed better than the context vowel models. Thus, it would appear that based on formant values at the midpoint of the target vowel, the upcoming context vowel is substantially less predictable than are the place and voicing of the right-adjacent consonant.

In summary, and looking broadly over the results from regression analyses, it appears that the acoustic effects

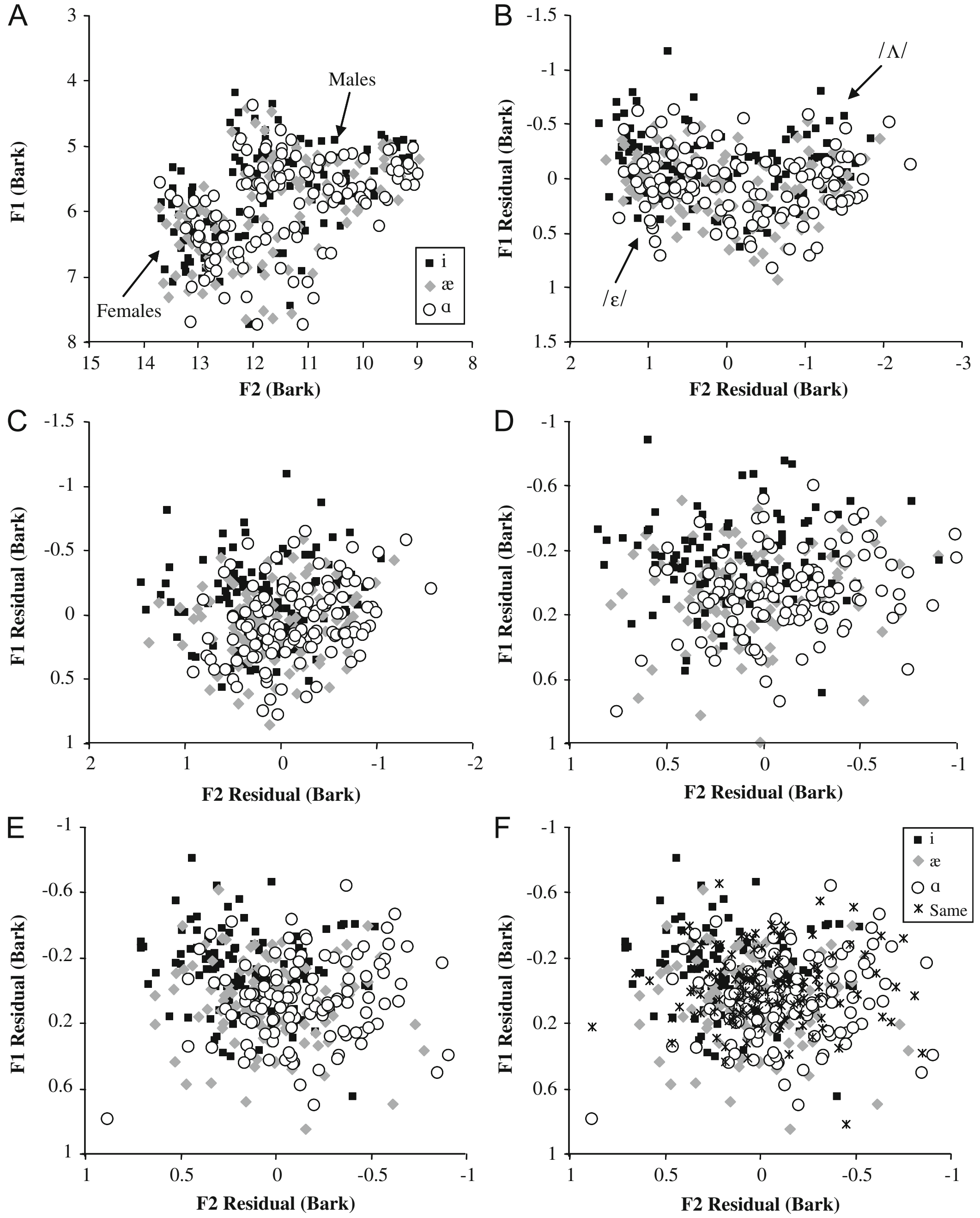

Fig. 4. First and second formant frequencies of target vowels as a function of the context vowel (/i, æ, ɑ/). (A) Raw data. (B) Formant frequency residual values, after the effect of speaker have been removed. (C) Formant frequency residuals after the effect of speaker and target vowel are removed. (D) Results after speaker, target vowel, place and voicing are removed. (E) Results after speaker, target vowel, consonant and interactions are removed. (F) Same as E, but the results of the neutral context are displayed as well.

from V-to-V coarticulation are quite robust and useful for predicting upcoming sounds, but only when multiple other sources of variation for the same acoustic measures are partialed out of the signal. Moreover, when considered as part of a complete model of variability in vowel formant frequencies, it is worth noting how much variance the parsing model accounts for. All together, speaker, target vowel, consonant and context vowel account for 88% of the variance in F1 and 94% of the variance in F2.

4. Discussion

The first research question posed by this study asks if there are robust effects of cross-word, anticipatory V-to-V coarticulation across a variety of VCV contexts. The ANOVA results reported here reveal clear evidence of coarticulatory effects on F1 and F2 for the two target vowels /ɛ/ and /ʌ/. This replicates findings from earlier studies that coarticulation influences vowel production in both the height and backness dimensions. V-to-V coarticulation is triggered by each of the three non-neutral context vowels tested, /i, ɑ, æ/, and in each context the target vowel shifts in the direction of the height and backness of the context vowel: /i/ triggers fronting and raising of the central target vowels, /ɑ/ triggers backing and lowering, and /æ/ triggers fronting and lowering. Target vowels occurring in a neutral coarticulatory context (i.e., where the context vowel is identical to the target vowel) appear to be less constrained in production, exhibiting more variability in both the F1 and F2 measures compared to target vowels in non-neutral coarticulatory contexts.

Another source of variation found to affect target vowels was the right-adjacent consonant. The ANOVA results show effects of consonant place on F2 of the target vowel, and effects of consonant voicing on both F1 and F2. There was also a significant interaction between consonant and context vowel on F1 measures of the target vowel. Specifically, while V-to-V coarticulation effects on F1 are found over an intervening labial or coronal consonant, reduced effects were found in the context of an intervening velar consonant, presumably reflecting the constraint on the vertical position of the tongue body imposed by the velar constriction. These findings are consistent with those of Fowler (2005) and Fowler and Brancazio (2000) in observing different patterns of consonant interference in V-to-V coarticulation for consonants that differ in place of articulation. But while Fowler (2005) finds greater interference from coronals compared to labials, the data reported here show significant interference in V-to-V coarticulation only with velar consonants. There are many differences between the present study and Fowler's in the speech materials used (e.g., the target vowel was unstressed schwa in Fowler's study), which preclude a direct comparison of the results. What is noteworthy is the common finding that in at least some cases the intervening consonant interferes with the acoustic effects of V-to-V coarticulation on the specific measure of F1, with a diminished influence of the context vowel on the target vowel across certain "high-resistant" consonants.

However, it is important to point out that despite this resistance, some inference of the context vowel is still possible. For our purpose, this finding is significant because it underscores the importance of evaluating the acoustic evidence of V-to-V coarticulation in relation to the C-to-V coarticulation context. This finding suggests that in order to make use of the information in the target vowel about V-to-V coarticulation, it is necessary to simultaneously look at the consonantal context and other sources of variability, and also to consider the full range of acoustic cues simultaneously.

Our second research question asks about the relative magnitude of the acoustic effects from V-to-V coarticulation compared to other sources of acoustic variation affecting the target vowel. Using linear regression analyses we compared the variance in F1 and F2 values of target vowels as a function of speaker, target vowel identity (/ɛ/ or /ʌ/), intervening consonant (place and voicing) and context vowel. Variance in F1 was found to be primarily influenced by speaker, while for F2, speaker and target vowel contributed roughly equally to account for the same portion of variance (82%).

In comparison to speaker and target vowel identity, the influence of the upcoming context vowel and the intervening consonant on F1 and F2 variation of the target vowel is relatively small, though significant. The intervening consonant accounts for between .8% (the consonant voicing effect on target vowel F1) to 5% (the consonant place effect on target vowel F2) of total variance in F1 and F2.[4] The effects of V-to-V coarticulation from the context vowel are comparable in size to the consonantal effects, accounting for 1.2% of total variance in F1 and .8% of total variance in F2. While these effects of context are indeed small when considered against the total F1 and F2 variance of the target vowels, when we factor out variation due to sources that occur in the speech stream earlier than the intervening consonant (i.e., speaker and target vowel identity), then the contribution of the intervening consonant and context vowel to the remaining variance is appreciably greater. Furthermore, when the variability due to context vowel is considered as a portion of the variance that remains after speaker, target *and* consonant effects are removed, all of which are signaled prior to the onset of the context vowel, then context vowel accounts for fully 9.2% (F1) and 11.6% (F2) of the remaining variance in the target vowel.

These results from linear regression analyses provide positive evidence that even in the presence of interfering effects from anticipatory C-to-V coarticulation, the acoustic effects of anticipatory V-to-V coarticulation make a distinct contribution to the overall acoustic variance of the target vowel. Furthermore, the effect of V-to-V coarticulation on within-category acoustic variation of the target vowel is a direct reflection of the identity of the context vowel—target vowels are raised, lowered, fronted or backed in relation to the height and backness of the context vowel in phonetic. This finding, which holds for target vowels in non-neutral coarticulatory contexts, suggests that information about coarticulation could be used to predict the height and backness of the context vowel.

To test the strength of this prediction, we built a series of regression models by adding individual sources of target

[4] Our findings on the effect of C-to-V coarticulation can be compared with the findings from Hillenbrand et al.'s (2001) study of the effect of consonantal environment on vowel formants. The data for that study include vowel measures taken from CVC syllables produced in isolation. Though the two studies differ in materials (and notably in the prosodic context of the target word) and in statistical methods, they both show relatively small coarticulatory effects due to consonantal context (accounting for no more than 5% of the total variance in target vowel F1 and F2 measures) in relation to the much larger effects of target vowel identity and (for our study only) speaker.

vowel variance as predictors, in the order in which they appear in the real-time speech signal. The first model included factors related to speaker voice, the second model added the factor of the target vowel identity, and the third model added the intervening consonant. Collectively, the results from these models show that the non-neutral context vowel can be predicted from the F1 and F2 values of the target vowel, with accuracy significantly above chance levels, *but only when the variance due to other sources is first factored out.* This finding addresses our second research question about the effects of parsing in separating proximal and distal sources of variation. In our statistical model, parsing distinct sources of variance not only affords a prediction of the upcoming phonological context, which was shown to rest at or below chance levels without parsing, but it also results in improved identification of the target vowel.

The findings from this production study have implications for the parsing model of speech perception. As proposed by Fowler (1984, 2005; Fowler & Smith, 1986), the parsing mechanism facilitates speech perception by allowing listeners to identify in the acoustic signal a component that signals the identity of the target segment, in our case a vowel, and other components that are sources of variation, including the upcoming context vowel. Through parsing, listeners compensate for the effects of coarticulation and at the same time make predictions about upcoming context. The acoustic evidence from our study establishes the viability of the parsing model of V-to-V coarticulation in English by demonstrating that information about the upcoming vowel is reliably present in the interval of the target vowel and that V-to-V coarticulatory effects index both the height and backness of the context vowel.

At the same time, our data point to limitations and challenges for the parsing model. One limitation relates to the failure of our logistic regression simulations to accurately predict the neutral context vowel. It appears that in our data the distribution of vowels in a neutral coarticulatory context spans the collective distributions of the distinctly (non-neutrally) coarticulated target vowels, as can be seen in the comparison of panels E and F from Fig. 4. Given this distributional pattern, any vowel token will be ambiguous between two interpretations based on its location in F1 × F2 space. Taking our data for example, a token of /ε/ that is high and front in the /ε/ distribution can be interpreted either as a token that is coarticulated with an upcoming /i/, or as a token from the neutral coarticulating context (i.e., with an upcoming /ε/ or perhaps with no following vowel). The implication of this finding for speech perception is that information about V-to-V coarticulation may afford non-unique predictions about upcoming context, increasing the likelihood of some sounds and decreasing the likelihood of others, but the prediction may fall short of identifying a single segment as most probable. Parsing models that use other sources of information, or do not rely on a veridical mapping of the statistics (as these regressions do), may achieve better performance.

A second ramification for parsing models concerns the nature of what is being parsed. Fowler (1984) argues that parsing is primarily a gestural process, geared to unpack the variance caused by overlapping gestures. Gow (2003), on the other hand, sees it as a grouping process which groups together similar phonetic features. Both propose similar operations and can account for similar results. Without taking a strong theoretical stand, our regression model suggests that parsing gestures may not be enough. There are tremendous benefits to be gained by parsing out a speaker's mean F1 and F2 values, distinctly non-gestural sources of information (as they account for the bulk of the variance in both measures). Thus, parsing may be better situated as a general approach to information processing, rather than something geared to a specific type of information.

A final challenge for the parsing model comes from the finding that parsing variance due to the upcoming context vowel requires first identifying the intervening consonant. We found, in concert with Fowler (2005) and Fowler and Brancazio (2000), that consonants differ in the extent to which they interfere in the realization of V-to-V coarticulatory effects on the target vowel. Based on their place of articulation, some consonants resist coarticulation with vowels, and in turn these consonants impede the continuous expression of V-to-V coarticulation across the full span of the target vowel. Given the variability of C-to-V effects on V-to-V coarticulation, the entire VC context must be considered for accurate identification of the target vowel, and for accurate predictions about the identity of the upcoming context vowel based on target vowel formant values. The fact that consonantal effects interact with context vowel effects means more work for the parsing mechanism. The F1 and F2 values of the target vowel must be decomposed to separate out effects from the consonant, effects from the context vowel, and their interaction in order to achieve the most accurate identification of the target vowel and of the context segments themselves.

The logistic regression model that includes contextual factors related to speaker, target vowel and intervening consonant can be considered as a statistical model of speech perception by a human listener. The dramatic improvement that is obtained when contextual factors are partialed out of F1 and F2 suggests that a 3-step parsing process, illustrated in Fig. 5, may be the basis for the facilitation and compensation effects reported by Fowler (1981, 1984, 2005), Beddor et al. (2002) and others. In the first step, evaluating the target vowel in the VCV sequences studied here, the effects of speaker and target vowel are accounted for (permitting a fairly accurate prediction of the upcoming consonant); in the second step, the subsequent consonant is heard and regressive compensation mechanisms account for its coarticulatory effects; in the third and final step, the residuals of this process are used to predict the upcoming vowel.

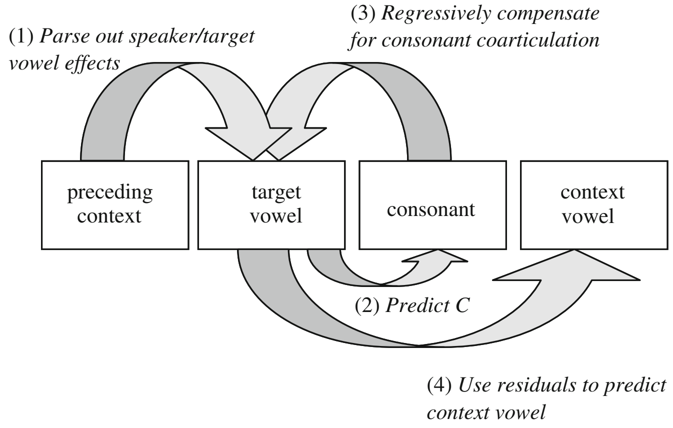
(1) *Parse out speaker/target vowel effects*
(3) *Regressively compensate for consonant coarticulation*
preceding context
target vowel
consonant
context vowel
(2) *Predict C*
(4) *Use residuals to predict context vowel*

Fig. 5. A multi-stage parsing process. (1) As the target vowel is heard, it is identified, and its effects as well as those of speaker are partialed out of F1 and F2. (2) This can be used to make inferences about the upcoming consonant. (3) When the consonant is heard and identified, its effects on F1 and F2 of the target vowel are partialed out. (4) This permits an accurate prediction of the context-vowel.

In all of the multinomial logistic regression models, percent correct identification for /i/ as the context vowel is higher than for any of the other context vowels. In the full model, identification of /i/ context vowels is an impressive 54.8% correct (and reaches 65% when alveolar contexts are considered alone), with /ɑ/ identification not far behind at 48.3% correct. This suggests that a listener who is able to detect the patterned variation in F1 and F2 in the target vowel could make use of this information to make an early prediction about these context vowels. And, as already noted, when the target vowel is neutral, the multinomial logistic regression models yield poor results. It would be interesting to see how these findings relate to the perception behavior of human listeners: do human listeners exhibit a bias to predicting a *different* (non-identical) context vowel over predicting a neutral (identical to target) context vowel? Clearly, lexical statistics might influence this response for real words, but the question of an intrinsic perceptual bias is an interesting matter, which we leave to future research.

Finally, we note that the target vowels in the present study do not coarticulate with the upcoming vowel fully enough to completely and reliably identify that vowel in advance of its position in the word, but the hypothesis space can be substantially narrowed on the basis of coarticulatory cues. The current study sets expectations for a future perception study: to the extent that listeners perceive the fine-grained patterns of variation in F1/F2 due to upcoming context, they should benefit in earlier and more accurate identifications of that context.

5. Conclusion

Methodologically, this study represents a novel approach. Our analyses considered multiple sources of variation simultaneously, both with respect to the dependent variable (e.g., the effect of voicing of F1) and with respect to their effects on each other (e.g., the effect of V-to-V after voicing has been partialed from the consonant). We also advocate for the use of analyses that mirror potential theoretical accounts of perception (e.g., the analogy between parsing and linear regression), and analyses that can be cast directly in terms of perceptual benefits (the multinomial logistic regression approach). Taken together, such an approach can yield new insight on classic problems. In particular, the problem of acoustic invariance does not seem so large when a hierarchical regression can account for 94.1% of the variance using a handful of well understood factors.

With respect to formal models of speech production, we demonstrate that anticipatory V-to-V coarticulation across word boundaries is robust and provides sufficient information to afford an effective prediction about upcoming material (in our study, including the intervening consonant and the context vowel), which at a minimum reduces the hypothesis space of possible upcoming sounds. The acoustic effect of V-to-V coarticulation is quite small, and does not go beyond the area of normal variation in neutral vowels. Thus, at least in this dataset, V-to-V coarticulation does not represent a neutralizing form of variation. This restricted pattern of variation may be key to maintaining coarticulation as a non-neutralizing feature of phonetic realization, suggesting that a first step towards phonologization, where assimilation processes develop from coarticulation, may involve patterns that place the coarticulated vowel token at the periphery of the vowel's 'neutral' distribution, or beyond.

Acknowledgements

This work was supported in part through National Institutes of Health Grant DC-008089 (Bob McMurray). We thank Joe Toscano and José Hualde for provocative questions and helpful discussion related to the ideas presented here.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.wocn.2009.08.004.

References

Alfonso, P. J., & Baer, T. (1982). Dynamics of vowel articulation. *Language and Speech*, *25*(2), 151–173.

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

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer (Version 4.4.04) [Computer program]. Retrieved January 7, 2006, from ⟨http://www.praat.org/⟩.

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*, *32*, 141–176.

Farnetani,, Edda (1997). Coarticulation and connected speech processes. In William J. Hardcastle, & John Laver (Eds.), *The handbook of phonetic sciences* (pp. 371–404). Cambridge, MA: Blackwell.

Fletcher, J. (2004). An EMA/EPG study of vowel-to-vowel articulation across velars in Southern British English. *Clinical Linguistics & Phonetics*, *18*(6–8), 577–592.

Fowler, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing*, *24*, 127–139.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, *36*(4), 359–368.

Fowler, C. A. (2005). Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics*, *33*, 199–213.

Fowler, C. A., & Brancazio, L. (2000). Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation. *Language and Speech*, *43*(1), 1–41.

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell, & D. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 123–136). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, *65*(4), 575–590.

Gow, D. W., & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. In J. Cole, & J. I. Hualde (Eds.), *Laboratory phonology*, Vol. 9 (pp. 173–200). New York: Mouton de Gruyter.

Hartman, J. (1985). Guide to pronunciation. In Frederic G. Cassidy (Ed.), *Dictionary of American regional English* (pp. xli–lxi). Cambridge, MA: Belknap Press.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*, 373–405.

Hillenbrand, J., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America*, *109*(2), 748–763.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

Huffman, M. K. (1986). Patterns of coarticulation in English. In: *UCLA working papers in phonetics*, Vol. 63, pp. 26–47.

Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson, & J. W. Mullenix (Eds.), *Talker variability in speech processing* (pp. 145–166). San Diego: Academic Press.

Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*, *36*, 28–54.

Kurath, H., & McDavid, R. I. (1961). *Pronunciation of English in the Atlantic States*. Ann Arbor: University of Michigan Press.

Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus/p/ in trochees. *Language and Speech*, *19*, 3–11.

Magen, H. S. (1989). *An acoustic study of vowel-to-vowel coarticulation in English*. Ph.D. dissertation, Yale University, New Haven, CT.

Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics*, *25*, 187–205.

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America*, *88*(3), 1286–1298.

Martin, J. G., & Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects. *Journal of the Acoustical Society of America*, *69*(2), 559–567.

Martin, J. G., & Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel–stop consonant–vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(3), 473–488.

McMurray, B., Clayards, M., Tanenhaus, M., & Aslin, R. (2008). Tracking the timecourse of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin and Review*, *15*(6), 1064–1071.

McMurray, B., Cole, J., & Munson, C. (in review). Features as an emergent product of perceptual parsing: Evidence from vowel-to-vowel coarticulation.

Nearey, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America*, *101*(6), 3241–3254.

Öhman, S. E.G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, *39*, 151–168.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175–184.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of linguistics structure* (pp. 137–157). Amsterdam: John Benjamins.

Recasens, D. (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America*, *76*(6), 1624–1635.

Recasens, D. (2002). An EMA study of coarticulatory direction. *Journal of the Acoustical Society of America*, *111*(6), 2828–2841.

Recasens, D., & Pallarès, M. D. (2000). A study of F1 coarticulation in VCV sequences. *Journal of Speech, Language and Hearing Research*, *43*, 501–512.

Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America*, *102*(1), 544–561.

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research*, *6*, 111–128.

Summerfield, A. Q. (1981). On articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1074–1095 (Reprinted by the Acoustical Society of America in *Papers in Speech Communication: Speech Perception*, 1991).

Warren, P., & Marslen-Wilson, W. D. (1988). Cues to lexical choice: Discriminating place and voice. *Perception and Psychophysics*, *43*, 21–30.

Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, *18*, 3–35.