Available online at www.sciencedirect.com
ScienceDirect
Journal of Phonetics 37 (2009) 173–188
Journal of Phonetics
www.elsevier.com/locate/phonetics

Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation

Michael Grosvald[a,b,*]

[a]Department of Linguistics, University of California at Davis, 267 Cousteau Place, Davis, CA 95616, USA
[b]Center for Mind and Brain, University of California at Davis, 267 Cousteau Place, Davis, CA 95616, USA

Received 7 March 2008; received in revised form 20 January 2009; accepted 20 January 2009

Abstract

The phenomenon of coarticulation is relevant for issues as varied as lexical processing and language change, but research to date has not determined with certainty how far such effects can extend. This study investigated the production and perception of anticipatory vowel-to-vowel (VV) coarticulation. First, 20 native speakers of English were recorded saying sentences containing multiple consecutive schwas followed by [i] or [a]. The resulting acoustic data showed significant VV coarticulatory influence up to three vowels before the context vowel, a greater distance than has been seen in previous studies. However, there was substantial variability among speakers in this regard. The perceptibility of these effects was then tested using behavioral methodology; even long-distance effects were perceptible to some listeners. Subjects' coarticulatory production strength and perceptual sensitivity were positively, but only weakly, correlated. Although the very slowest speakers tended to coarticulate less than the rest, speech rate and coarticulatory strength were not significantly correlated for the group as a whole.

1. Introduction

Following Öhman's (1966) groundbreaking work on transconsonantal vowel-to-vowel (VV) coarticulation in Swedish, English and Russian, researchers have sought to understand how different factors influence these effects, such factors being as varied as the specific consonants and vowels involved (Butcher & Weiher, 1976; Öhman, 1966; Recasens, 1984), prosodic context (Cho, 2004; Fletcher, 2004), and the vowel inventory of the language in question (Manuel, 1990; Manuel & Krakow, 1984). Instances of long-distance coarticulation, involving effects crossing two or more intervening segments, have been found for phenomena such as lip protrusion (Benguerel & Cowan, 1974), velum movement (Moll & Daniloff, 1971), and liquid resonances (Heid & Hawkins, 2000; West, 1999), but the possible range of VV coarticulatory effects is not yet known. This study investigates the extent and the perceptibility of long-distance VV coarticulation, with a particular focus on variation among speakers and listeners.

Despite early indications that VV coarticulation might be a relatively local phenomenon (e.g. Gay, 1977), subsequent work has shown that this is not always the case (Magen, 1997; Recasens, 1989). For example, Magen (1997) analyzed [bVbəbVb] sequences produced by four English speakers and found evidence of coarticulatory effects between the first and final vowel, meaning that such effects can cross foot boundaries and multiple syllable boundaries. However, this was not so for all four speakers in the study. This suggests that in order to determine how prevalent such effects are among speakers and how far they can extend, researchers may find it necessary to recruit greater numbers of speakers than has generally been done before, since this may be the only way to get around the statistical problem of large interspeaker variation. The present study's use of a relatively large group of speakers (20) is an attempt to move in such a direction.

An additional goal of this study was to look for coarticulatory effects in natural-language utterances.

*Corresponding author at: Center for Mind and Brain, University of California at Davis, 267 Cousteau Place, Davis, CA 95618, USA.
Tel.: +1 530 297 4427; fax: +1 530 297 4400.
E-mail address: mgrosvald@ucdavis.edu

0095-4470/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2009.01.002

Although the use of nonsense words can often not be avoided when all permutations of even a limited set of consonants, vowels, and prosodic contexts are considered, it seems reasonable to suppose that this may result in study participants speaking less fluently, thus lessening the potential for long-distance effects to occur. Therefore, rather than analyzing the outcome of a small number of speakers saying a large number of items consisting of or containing nonsense words (cf. the Gay (1977), study mentioned above, which analyzed two speakers' production of all VCV combinations of V = {i, a, u} and C = {p, t, k}), the present study investigated a large number of speakers each saying a small number of natural-language sentences, with the corresponding trade-off that the set of contrasts considered was limited.

Although it seems evident from earlier work that some speakers do not coarticulate much or at all over long distances (Gay, 1977; Magen, 1997), the fact that some speakers do must be accounted for in any viable model of speech production (see Farnetani and Recasens (1999) for an overview of relevant approaches). The existence of long-distance coarticulatory effects has similar implications for theories of speech perception, since research has shown that such effects are sometimes perceptible to listeners. For example, Martin and Bunnell (1982) cross-spliced recordings of words in such a way as to create stimuli which varied in the consistency of their VV coarticulatory patterns, and then had listeners perform recognition tasks. Stimuli that were consistent with naturally occurring coarticulation patterns tended to be associated with fewer false alarms and faster reaction times (see also Scarborough, 2003).

The present study provided the opportunity to explore the perception of VV effects from a different, though related, perspective. Here, the ability of study participants to distinguish vowels which had been differently "colored" by coarticulatory effects *at various distances* was examined. Since longer-distance effects may be expected to be more subtle than nearer-distance effects (all else being equal), this provided an opportunity to explore variation among listeners in terms of their sensitivity to such effects. If long-distance coarticulation is sometimes perceptible, this would be particularly relevant to the study of lexical processing, particularly in light of the fact that both the production and the perception of coarticulation appear to be heavily influenced by the coarticulatory patterns of users' native language (Beddor, Harnsberger, & Lindemann, 2002). If listeners are able to "hear ahead" a few segments in the flow of spoken language, it could help them narrow down the possible range of upcoming words more effectively as they process the incoming speech stream. From this standpoint, all else being equal, anticipatory coarticulation would probably be more useful to listeners than carryover coarticulation. Therefore, the current study focused on anticipatory VV coarticulation.

Finally, the present study sought to address one aspect of production–perception interplay discussed by Ohala (1981, 1994) in his examination of the role of the listener in language change. Ohala has argued that over time, perceptible VV coarticulation can become grammaticalized, leading to vowel harmony; if so, this suggests that the production and perception of coarticulation may be correlated. Since some participants in the present study provided both production and perception data, this issue was also investigated.

This paper continues in Section 2 with a presentation of the production experiment. Long-distance VV effects were seen over a greater distance than has been found in previous studies like Magen's (1997): two of the 20 speakers tested showed significant anticipatory VV effects across at least three vowels' distance. A great deal of variation among speakers was also seen, however, and some speakers showed no or only weak effects. Recordings made of some of the production-study speakers were next used as stimuli for the perception study, which is presented in Section 3. All ten listeners were sensitive to nearer-distance effects, while at further distances much more variation between listeners was seen. Two listeners of the ten who participated were sensitive even to effects which had occurred over three vowels' distance. Somewhat unexpectedly, strength of coarticulatory effects was not significantly correlated with either speaking rate or perceptual sensitivity to such effects. Some implications of this study's findings are presented in Section 4.

2. Production experiment

2.1. Methodology

2.1.1. Speakers

Twenty participants (11 females, 9 males) took part in the production experiment. Seven subjects were known personally to the author and agreed freely to take part; the other thirteen were undergraduate students at the University of California at Davis who received course credit for participating. The subjects' ages ranged from 18 to 62 [mean age = 25.2; SD = 13.3]. All participants were native speakers of American English with no history of speech or hearing problems. All participants were uninformed as to the purpose of the study.

The first seven participants took part only in the production experiment. A subset of their vowel recordings was then used as stimuli for subsequent subjects, who took part in both the production and perception experiments (see Section 3 for the perception study).

While every effort was made to recruit only monolingual speakers, most subjects had been exposed to at least one other language as students in a university with a foreign language course requirement. Of these, four felt that they had acquired substantial knowledge of another language. However, including these four participants does not change the results of the statistical analyses of group data, for either the production or perception studies, and in any case, the two speakers who showed the longest-distance

effects were both monolingual. Therefore, all subjects' data were included.

2.1.2. Speech samples

First, it was necessary to create sentences containing plentiful opportunities for VV coarticulation to occur. The vowels [i] and [a] were chosen as context vowels because of their distance in vowel space. Consecutive vowels likely to be produced as schwas, or at least substantially reduced, were used as targets; schwa was chosen because of its susceptibility to coarticulatory influence from neighboring vowels, a property that has emerged in previous studies, both acoustic and articulatory (e.g. Alfonso & Baer, 1982; Fowler, 1981, respectively). The sentences used were as follows:

"It's fun to look up at a key."

"It's fun to look up at a car."

These items were the outcome of the following set of preferences: (1) sentences containing only real words, for the reasons discussed earlier; (2) sentences not differing prior to the context ([i] or [a]) vowel, which was to be the sentence-final vowel; (3) monosyllabic words containing the target (schwa) and context ([i] and [a]) vowels, so that coarticulatory effects would be "spontaneous" and not well-practiced within particular lexical items; (4) function words containing the target vowels and content words containing the context vowels, to encourage reduction of target vowels and full pronunciation of context vowels; (5) context-vowel content words having relatively high (and similar) frequency of use and (6) voiceless intervening consonants, so that vowel boundaries would be as clear as possible when making formant-frequency measurements and when creating stimuli for the perception study (see Section 3).

To minimize interference from intervening consonants' demands on the tongue body (cf. Recasens, 1984), it might have been preferable if the exclusive use of bilabial intervening consonants had been feasible, such as in Magen (1997), where nonsense words of the form [bVbəbVb] were used. However, in the multi-word context necessary for this investigation and given the above constraints, issues like lexical gaps and morphological/syntactic constraints quickly asserted themselves. In addition, excessive alliteration would raise issues of articulatory ease and naturalness (e.g. "Peter Piper picked a peck of pickled peppers"). In the end, the full array of voiceless stops in English was used, with those making fewer demands on the tongue body located further away from the context vowels, to "assist" longer-distance effects at the possible expense of shorter-distance ones.[1] The fact that any distance-3 effects would have crossed consonants articulated at all three major places of articulation of English stops would be an added point of interest (see Section 4).

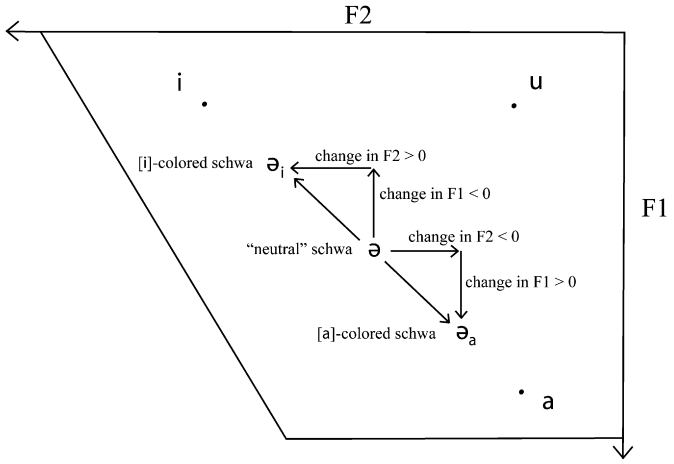

Fig. 1. The diagram shows the expected influence on schwa of coarticulatory effects of nearby [i] or [a]. Near [i], F1 is lowered and F2 is raised, while the reverse holds near [a]. The magnitude of coarticulatory effects may be exaggerated in this figure.

A randomized list containing six copies of each sentence was provided to each speaker. Before recording, speakers were told about list effects and the need for their avoidance. In order to encourage consistent prosodic patterning among these utterances, speakers were asked to say the sentences as if they were being spoken in a normal conversation in response to the question, "What's it fun to look up at?", with the intended effect of obtaining utterances with primary stress on the final word, with some emphasis also expected on the word "fun." Speakers were given a chance to rehearse until any substantial deviations from this pattern, such as "It's fun to look *up* at a car" (i.e. as if meaning "not *down* at a car") were corrected.

The final vowel, either [i] or [a], served as context vowel, while the preceding vowels in the words "up," "at," and "a," were the target vowels. In this paper, these will be referred to as "distance-3," "distance-2" and "distance-1" vowels, respectively. Because of the intervening consonants, these distance conditions correspond, respectively, to a total of 5, 3 and 1 total intervening segments between the context and target vowels. Fig. 1 illustrates the expected coarticulatory effects of [i] and [a] on schwa in two-dimensional formant space.[2]

Since it is difficult to create natural-sounding sentences in English containing multiple consecutive unstressed syllables, the vowel [ʌ] was used as distance-3 vowel because of its acoustic similarity to schwa. In careful

[1] Pilot testing had indicated that VCV effects on schwa were quite common, even across [k].

[2] It was expected that the rhotic occurring after the low vowel in "car" would not significantly color that context vowel, at least not to the point of creating a long-distance coarticulatory [i]–[r] contrast rather than an [i]–[a] contrast. While Heid and Hawkins (2000) and West (1999) have found evidence of long-distance resonance effects of liquids [l] and [r], these were in contexts in which the liquids were in syllable-initial position. Nevertheless, a check was performed against the possibility, as will be explained in Section 2.2.

speech, the vowels in "at" [æ] and "a" [eɪ] are not schwas either, of course, but it was expected that in the casual speech speakers were encouraged to produce here, these vowels would be realized as schwa or at least be substantially reduced. During the rehearsal preceding the recording process, speakers who did not seem to be speaking naturally—presumably because of the perceived formality of participating in a scientific experiment—were gently coached until their production became more relaxed.

Some subjects did occasionally exhibit a slightly [æ]-like quality in "at" or a slightly [eɪ]-like pronunciation of "a" even in casual speech. It should be noted that the research question being investigated (how far VV effects can extend) does not strictly require that schwas be the target vowel, although of course that was the general intention. Overall, the great majority of the "at" and "a" vowels were in fact produced with the expected schwa-like quality, and for convenience, the vowels in "up," "at" and "a" will be referred to here as schwas.

In order to obtain baseline formant values for the context vowels [a] and [i], each speaker was also recorded repeating each of the following sentences three times; in these sentences the context vowels in the final word are preceded by, and therefore coarticulated with, themselves[3]:

"It's fun to see keys."

"It's fun to saw cars."

2.1.3. Recording and digitizing

Participants were seated comfortably at a table in a laboratory environment (a quiet room measuring approximately 15 ft by 18 ft). The recording equipment consisted of a Shure SM48 microphone attached to a Marantz professional CDR300 digital recorder; these digital recordings were made at 16-bit resolution with a 48 kHz sampling rate. The participant was given a randomized list containing the appropriate number of copies of the sentences discussed above (i.e. six copies of both sentences containing the consecutive schwas, and three copies of both sentences containing the repeated context vowels). After the rehearsal process described earlier, the subject was handed the microphone and held it several inches to one side of (not directly in front of) his/her mouth, while saying the sentences in the order indicated on the list. If a disfluency or other unwanted event occurred, that repetition of that sentence was repeated, until all of the sentences had been successfully recorded the required number of times.

2.1.4. Editing, measurements and analysis

Editing of the digital sound files and formant-frequency measurements were performed onscreen using the Sound Edit function in Praat (Boersma & Weenink, 2005) for each sound file, with the following settings: (for spectrogram) analysis window length 5 ms, dynamic range 30 dB; (for formant) maximum formant 5000 Hz [for male speakers] or 5500 Hz [for female speakers], number of formants 5, analysis window length 25 ms, dynamic range 30 dB, and pre-emphasis from 50 Hz, using the Burg algorithm to calculate LPC coefficients.

Each target vowel was excised from the whole-sentence recording and saved as a separate sound file for purposes of data analysis, and also for possible use later as a stimulus in the perception experiment.[4] The starting boundary for each vowel was defined as the formant track marking just prior to the onset of voicing, while the end boundary was the corresponding location after voicing offset, as shown in Fig. 2. Because of the intervening consonants, determining the boundary points of each vowel was generally unproblematic. In less straightforward cases, such as some in which speakers flapped or otherwise reduced the [t] in "at a," visual inspection of the amplitude trajectory usually showed a rapid change of slope at a particular point, which was taken as marking the vowel boundary. In the few cases where this boundary was not so clear, a special notation was made so that measurements made for those tokens could be given further attention if they were clear outliers. In the end, there were only a few such outliers, and their inclusion or exclusion from the analysis did not change the results.[5]

The effects of anticipatory VV coarticulation were expected to be strongest in the later portion of the target vowels (cf. Beddor et al., 2002), where influence of the immediately following consonant was also likely to be greatest. As a compromise between seeking the former while minimizing the latter, measurements of target vowels' F1 and F2 were made at 25 ms before vowel offset. For target vowels with duration under 50 ms, measurements were made at vowel midpoint. The aim was to fill the 25-ms LPC analysis window with only vocalic information, as late in the vowel as was feasible, but without acquiring acoustic information directly from the following consonant (though coarticulatory effects of the consonant on the vowel were sometimes evident, as the results section will show). Since all the target vowels were over 25 ms in length, this

[3]N.B.: The vowels in "saw" and "car" have merged in the variety of American English spoken by these study participants.

[4]Measurements were made after excision instead of before mostly for convenience, because some of these excised vowels were to be used for the perception study, for which repeated measurements (and some alterations; see Section 3) would be necessary, these being more easily performed on the shorter sound files. More importantly, it made essentially no difference whether measurements were made before or after excision (confirmed through pilot testing), because of where the measurements were made and the width of the LPC analysis window.

[5]The issue of outliers was dealt with as follows. Other than those that were the result of genuine errors (such as Praat clearly missing an F1 value and giving an F2 value as F1 instead), I decided I should either keep all of them or omit all of them, rather than making such decisions on a case-by-case basis, which could lead to bias. I did not want any reported significant outcomes to be contingent on any such case-by-case decision-making. In the final analysis, keeping the outliers instead of omitting them made little difference in general, but resulted in more conservative outcomes in a few cases (for example, an additional speaker would have had a significant distance-3 outcome if a particular outlier were removed). Therefore the outliers were included.

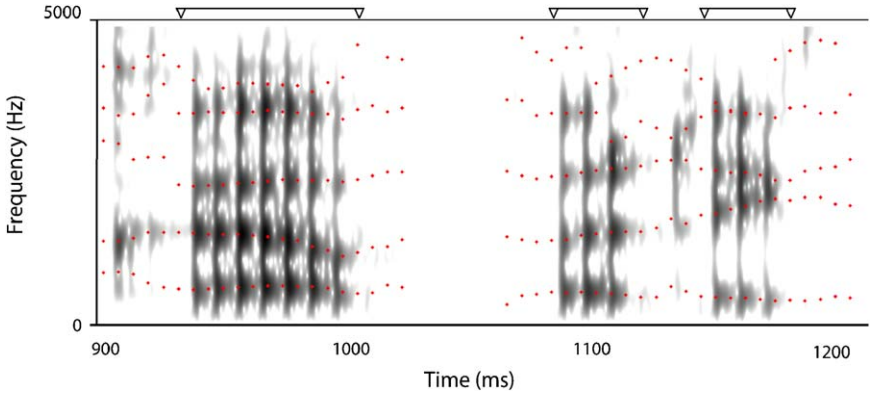

Fig. 2. The acoustic representation of part of one utterance containing the sequence "up at a." The editing points are indicated above each vowel. The starting boundary is taken as the formant track marking just prior to the onset of voicing, while the end boundary is taken as the corresponding location after voicing offset.

was always possible. VV coarticulatory effects were investigated at each distance through a statistical comparison, between the [i] and [a] contexts, of the formant values of the target vowels articulated at that distance from the context vowel.

For all group analyses reported in this paper, a normalization procedure based on Gerstman (1968) was applied to each speaker's nth raw formant values for $n = 1$ and 2. Starting with a given speaker's average first and second formant values for full vowels [a] and [i] and with the raw formant value $F_{n\ \mathrm{raw}}$ (for $n = 1$ or 2), the corresponding normalized value is given by the formula

$$F_{n\ \mathrm{norm}} = 999(F_{n\ \mathrm{raw}} - F_{n\ \mathrm{min}})/(F_{n\ \mathrm{max}} - F_{n\ \mathrm{min}}),$$

where $F_{n\ \mathrm{max}}$ and $F_{n\ \mathrm{min}}$ are the largest and the smallest nth formant values among that speaker's full vowels; in other words, $F_{1\ \mathrm{max}}$ and $F_{1\ \mathrm{min}}$ are given by the speaker's average F1 for [a] and [i], respectively, with the reverse order for F2 values. The procedure has the effect of scaling each speaker's formant values relative to the width and height of his or her own vowel space, as defined by [a] and [i]. Both F1 and F2 are scaled to a range 0–999 with the context [a] at (999, 0) and the context [i] at (0, 999). This makes comparisons between speakers more reasonable (though not unproblematic; see Section 4).[6]

2.2. Results and discussion

2.2.1. Group results

Fig. 3 is a normalized vowel-space plot showing formant frequencies relative to the extremes of the context [a] and [i] for each distance condition and in each context, averaged over all 20 speakers. As one might expect, increased distance from the context vowel is associated with progressively reduced formant differences between the

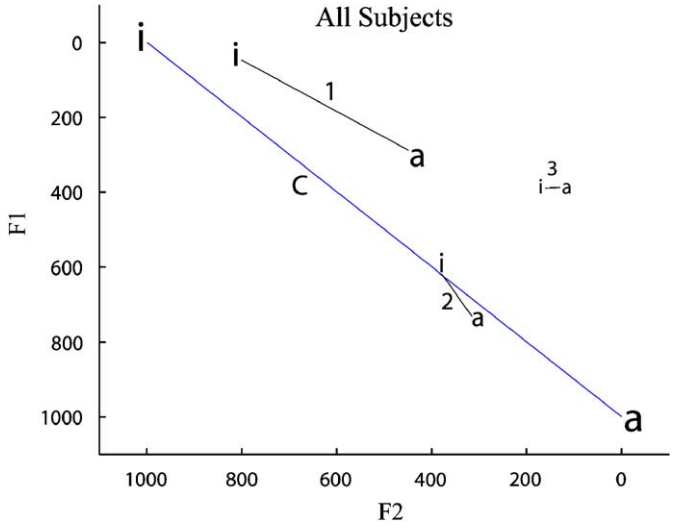

Fig. 3. Target-vowel positions in normalized vowel space, relative to the context [a] and [i], averaged over all 20 speakers; these averaged values are marked by the line segment endpoints, not the labels. Context and distance-1, -2 and -3 vowel pairs are labeled with progressively smaller text size and an adjacent "C," "1," "2" or "3," respectively. Context-related differences are significant at the $p < 0.05$ level or greater for both formants in all 3 distance conditions, except for F1 at distance 3.

[i] and [a] contexts. The mean normalized (F1, F2) for the [a] and [i] contexts are (288, 449) and (46, 801) at distance 1; (731, 315) and (617, 379) at distance 2; and (388, 130) and (389, 160) at distance 3.

It should be noted that the measurements illustrated in Fig. 3 were made near the end of the target vowels, where coarticulatory influence of the context vowel was expected to be strongest. Therefore, these formant values do not always correspond closely to the values one would obtain in the steady-state portion of a schwa vowel produced in isolation. This is especially true considering that the influence of the immediately following consonant in each distance condition appears to be in play as well; for example, as place of articulation changes from labial to alveolar to velar at distances 3 ("up"), 2 ("at") and 1

[6] Gerstman's (1968) original formula specified that F_{max} and F_{min} were to be taken over nine (Dutch) vowels, not just [a] and [i], but given the positions of those two vowels in vowel space, it seems reasonable to take their F1 and F2 as providing the desired maximum and minimum values. Gerstman alluded to this himself (p. 80).

("a" [k]), F2 of the target pairs increases accordingly. Similarly, F1 values are quite low for the target vowels overall, particularly at distances 3 and 1; these vowels immediately preceded [p] and [k], respectively, both of which generally reached fuller closure than the [t] in "at," often realized as a flap.

To determine whether the differences illustrated in Fig. 3 are significant, repeated-measures ANOVAs with context vowel as factor were performed on the group dataset at each distance and for each formant, using normalized formant values as discussed above. For both F1 and F2, there was a highly significant main effect of vowel at both distance 1 (first formant: $F(1,19) = 83.7$, $p < 0.001$; second formant: $F(1,19) = 101.7$, $p < 0.001$) and distance 2 (F1: $F(1,19) = 13.6$, $p < 0.01$; F2: $F(1,19) = 22.9$, $p < 0.001$). At distance 3, these effects appear to taper off, as the non-significant outcome for F1 ($F(1,19) = 0.052$, $p = .82$) and significant but weaker outcome for F2 ($F(1,19) = 5.04$, $p < 0.05$) show. These outcomes provide strong evidence of coarticulatory effects at all three distances, for at least some speakers. The existence of distance-3 effects is particularly noteworthy given that the distance-3 vowel was the target vowel which would be expected to undergo the least amount of reduction based on its [ʌ] quality.

2.2.2. Individual results

In order to explore these results further, the coarticulatory tendencies of individual speakers were next examined. For each speaker, one-tailed heteroscedastic *t*-tests were run for F1 and F2 for each distance condition (1, 2 or 3) to determine if formant values differed significantly between the [i] and [a] contexts at that distance. Raw formant values were used here since formant values were not being directly compared between speakers. One-tailed tests were appropriate because it was predicted that [i]-colored vowels would have lower F1 and higher F2 than [a]-colored vowels. The significance results for all 20 speakers are summarized in Table 1. Significance is given without Bonferroni correction for multiple tests in order to provide a better picture of the differences between individuals and the gradience of the effects over distance. Numerical data for each speaker are given in Appendix A. To address the possibility that speaking rate might be a relevant factor, this was also measured for each speaker; these values are shown in the rightmost column of the table. Speech rate for a given speaker was calculated by averaging, over that speaker's utterances, the time elapsing between the start of the distance-3 vowel and the start of the context vowel, a span of six segments.

As expected, most speakers showed a substantial amount of VV coarticulation, though a great deal of interspeaker variation was evident as well. While two participants had significant results in all three distance conditions, several others showed only distance-1 effects or no significant effects at all. The majority (13 speakers; 65% of the group) produced significant effects as far as distance-2 but no further (at Bonferroni corrected $p < 0.00042$, 10 of 20 speakers had significant effects at distance-1 for F1 or F2 or both, and 4 of 20 speakers had significant effects at distance-2). This confirms and extends Magen's (1997) results, in which she found high variability between speakers in the production of long-distance VV coarticulation. When formant differences did not reach significance, they still tended to pattern the way one would expect, with [i]-colored vowels showing lower F1 and higher F2 with respect to [a]-colored vowels. This pattern held with few exceptions in the distance-1 and distance-2 conditions, but was much less evident in the distance-3 condition, where only two speakers (3 and 7) showed a difference which reached significance. The results for Speaker 7 are pictured in Fig. 4 (with non-normalized values).

Table 1
For each speaker, the significance testing outcomes of six *t*-tests are shown, comparing formant frequency values of that speaker's target vowels for the [i] vs. [a] contexts, for each of F1 and F2 and for each distance condition.

Speaker	Distance 3 ("**u**p")		Distance 2 ("**a**t")		Distance 1 ("**a**")		Speech rate (seg/s)
	F1	F2	F1	F2	F1	F2	
1						*	13.8
2			*	*	***	***	15.5
3		*	***	***		***	13.9
4						*	11.2
5			*	**	***	***	15.2
6				*	***	**	12.7
7		**		***	***	***	15.3
8				**		***	13.6
9			*	*	**	**	12.1
10			**	***		**	15.2
11				*			11.8
12			*		*	***	14.2
13					**	***	11.0
14			*	**	***	**	12.2
15				**	***	**	17.6
16							11.2
17			*	***	**	***	14.4
18			*		*		16.5
19					***	**	12.1
20				*	***	***	12.0

Significant results are noted, with * = $p < 0.05$, ** = $p < 0.01$ and *** = $p < 0.001$ (no Bonferroni correction). The rightmost column shows each speaker's rate of speech in segments per second.

2.2.3. Follow-up tests

At this point, some issues that may be raised about the outcomes reported thus far will be addressed. First, given the number of *t*-tests performed, one may ask whether the significant outcomes for Speakers 3 and 7 at distance 3 may be spurious, since the possibility of a Type I error increases along with the number of such tests performed. Second, how do we know that the presence of the rhotic in the context word "car" was itself not the cause of the coarticulatory contrast with [i], given the resonance effects of liquids reported by Heid and Hawkins (2000) and West

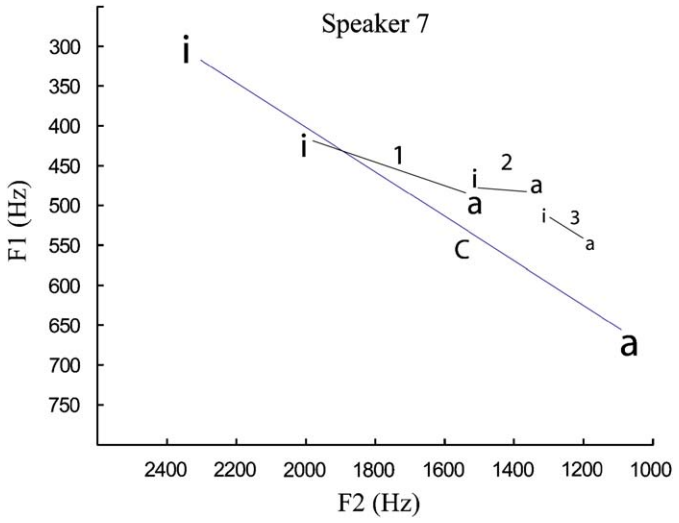

Fig. 4. Subject 7 coarticulated more strongly than any other speaker. The graph shows this speaker's average context and distance-1, -2 and -3 target-vowel positions in vowel space, labeled with progressively smaller text size and an adjacent "C," "1," "2" or "3," respectively. Formant values are not normalized. Values are marked by line segment endpoints, not the labels.

(1999)? Finally, is it not possible that some intervening consonant(s) might act as triggers themselves? In particular, the [k] preceding [i] in "key" is expected to be fronted, so one might suspect that the different [k]s in "key" and "car" were the real trigger for the effects on the preceding vowels.[7]

To answer these concerns, a small follow-up was conducted with the two subjects (Speakers 3 and 7) who had shown significant distance-3 effects. First, they were recorded saying sentences similar to the ones used earlier, but with [r]-free context words:

"It's fun to look up at a keep."

"It's fun to look up at a cop."

Speaker 3 was also recorded saying sentences with [k]-free context words:

"It's fun to look up at a peep."

"It's fun to look up at a pop."

Measurements and significance testing were performed as before; the results are shown in Table 2, along with the original "key/car" results for comparison. Although there are some differences in outcome associated with the different context word pairs, some important similarities are clear: in addition to strong distance-1 and -2 effects, we see significant distance-3 effects in all cases. For these speakers, then, the distance-3 effects seem relatively robust,

[7]It should be noted that in many if not most studies of long-distance coarticulation, it may be impossible to avoid the possibility that long-distance effects are actually shorter-distance effects that are transmitted successively via intermediate segments.

Table 2
For each speaker, the significance testing outcomes of six *t*-tests are shown for each contrast, comparing (non-normalized) formant frequency values of that speaker's target vowels in the [i] vs. [a] contexts, for each of F1 and F2 and for each distance condition.

Speaker	Distance 3 ("**u**p")		Distance 2 ("**a**t")		Distance 1 ("**a**")		Contrast
	F1	F2	F1	F2	F1	F2	
3		*	***	***		***	key/car
3		*	**	**	***	***	keep/cop
3		*	*	***	**	***	peep/pop
7		**		***	***	***	key/car
7		*		***	***	***	keep/cop

Significant results are noted, where * = $p<0.05$, ** = $p<0.01$ and *** = $p<0.001$ (no Bonferroni correction).

and given the variety of contexts in which they are maintained, cannot be due solely to the presence of any particular segments in the context words other than the context vowels [i] and [a].

An additional question is whether these significant distance-3 results might simply be due to some reduction process like segment deletion having occurred somewhere between the context and distance-3 vowels. A closer inspection of the data shows that this was not the case, as exemplified by Fig. 2 earlier. The speaker whose data were used in producing that figure was in fact Speaker 7. Clear transitioning between the consonants and vowels is apparent, with no segment deletion; this was so for all of Speaker 3 and Speaker 7's utterances.

One pattern that might be expected and which Table 1 seems to show is that speakers who coarticulated more strongly at a closer distance were also more likely to show significant results at greater distances. Statistical tests confirm that average vowel-space differences (taken as Euclidean distance in normalized F1–F2 space) between speakers' [i]- and [a]-colored schwas were significantly correlated ($r = 0.48$, $p<0.05$) between distances 1 and 2. However, such correlation was much weaker between distances 2 and 3 ($r = 0.34$, $p = 0.14$), and even more so between distances 1 and 3 ($r = 0.0006$, $p = 0.998$), presumably because of the absence of distance-3 effects for most speakers.

A related but more surprising outcome also seen in Table 1 is that Speakers 3, 10 and 11 all show apparently discontinuous coarticulatory effects; each of these three speakers had a significant distance-2 outcome for one formant without a corresponding distance-1 effect. Recasens (2002) examined "interruption events" such as these in VCV sequences and following Fowler and Saltzman (1993), suggested that such occurrences may be the result of a "fixed, long-term" planning strategy for the second vowel already being executed by speakers during the first vowel. It should also be noted that in all three of these cases, the trends were in the expected direction (see the raw data in Appendix A), indicating that coarticulatory forces may

Table 3
Possible very-long-distance coarticulatory effects for Speakers 3 and 7.

Distance	7 (“**it**’s”)		6 (“f**un**”)		5 (“**to**”)		4 (“l**oo**k”)		3 (“**up**”)		2 (“**at**”)		1 (“**a**”)		Contrast
Speaker	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	
3						*	*			*	***	***		***	key/car
3										*	**	**	***	***	keep/cop
3								*		*	*	***	**	***	peep/pop
7					*					**		***	***	***	key/car
7										*		***	***	***	keep/cop

Results are shown for significance testing between contexts for target vowels at each distance from 1 to 7 before the context vowel, with * = $p<0.05$, ** = $p<0.01$ and *** = $p<0.001$.

have been at work during the distance-2 vowel but were too weak to yield a statistically significant result.

Although some researchers (e.g. Hussein, 1990) have suggested that speaking rate may be related to coarticulatory tendency, an inspection of Table 1 shows that the speakers in this study who coarticulated the most were not the fastest talkers, nor vice versa. Although the slowest speakers (Speakers 4, 13 and 16) showed an absence of significant effects at distances greater than 1, statistical testing for correlation between speakers' speech rate and normalized vowel-space distance between [i]- and [a]-colored schwas found no significant effects in any of the three distance conditions.[8] This result complements work of Hertrich and Ackermann (1995), who found that slower speech was associated with less carryover coarticulation, but no significant difference in anticipatory coarticulation, relative to more rapid speech.

Fowler and Saltzman (1993) have suggested that coarticulatory effects can be considered “long-distance” only in terms of the number of intervening segments, in that the time span across which such effects can occur is relatively narrow. This may be the case, but if so, the upper limit they suggest (approximately 200–250 m s) seems low in light of the fact that the speakers who coarticulated the most in this study (Speakers 3 and 7) showed significant effects across spans of well over 300 m s. The temporal distance between Speaker 7's distance-3 vowel offset and context-vowel onset over his 12 “key/car” utterances ranged from 298 to 377 m s and averaged 333 m s; for Speaker 3 the distances are even greater (range = [301, 472]; mean = 384 m s).

2.2.4. Longer-distance effects

Remaining open is the question of whether VV coarticulatory effects at even greater distances can occur with any substantial frequency. In related work, Heid and Hawkins (2000) and West (1999) have found evidence of different resonances for [r] compared with [l] across several syllables, manifested as lowered formants (F3 for West, F2 + F3 + F4 for Heid and Hawkins), increased lip rounding, and high or back tongue position for [r] contexts compared with [l] contexts. To investigate the possibility of such extreme long-distance effects here, all of the vowels in the utterances of Speakers 3 and 7, who had already shown significant coarticulatory effects as far back as distance 3, were analyzed and compared between the [i] and [a] contexts.

The results are shown in Table 3, and appear to show significant outcomes at distances 4 and 5. However, the magnitude of the effects is consistently small. There are also a number of discontinuities like those seen earlier for Speakers 3, 10 and 11, but occurring over wider ranges and, unlike in those cases, with trends at closer distances not always in the expected direction. To the extent that coarticulatory effects may have occurred over such distances, they are clearly less robust than those reported for these speakers at closer distances.

Despite the uncertainty associated with these longer-distance effects, the production results discussed earlier provide good evidence that many if not most speakers produce a significant amount of anticipatory vowel-to-vowel coarticulation in at least some contexts, and that for a majority of speakers, such effects can extend over at least three intervening segments. The next part of the study investigated the question of how perceptible such effects are.

3. Perception experiment

In an early perception study, Lehiste and Shockey (1972) performed an experiment in which listeners heard recordings of VCV sequences where one of the vowels, together with the adjacent half of the consonant, had been excised, and were asked what they thought the missing vowel was. Results were no better than chance. However, a number of studies using other methods (e.g. Beddor et al., 2002; Fowler & Smith, 1986) have found that VV coarticulatory effects are in fact perceptible to listeners. In the current

[8] A reviewer points out that this may be a case of a threshold effect, rather than an absence of an effect altogether. In other words, such an effect might be present at slow speaking rates, but not be apparent above a certain threshold rate, “perhaps having to do with the fact that one must enunciate to at least some extent to be understandable.”

study, a modified version of Lehiste and Shockey's technique is used, with the following reasoning.

Ohala (1981, 1994) has suggested that acoustic byproducts of physiological linguistic processes may sometimes be perceived by listeners as grammatically important information, and that this may ultimately lead to language change. For example, vowel harmony may sometimes be the outcome of perceptible VV coarticulation that has become grammaticalized. Przezdziecki (2000) found evidence for this in his study of three Yoruba dialects, one of which has vowel harmony and the other two of which do not. Przezdziecki suspected that coarticulation patterns in the two dialects without vowel harmony might be similar in kind but different in degree from those in the third dialect, in which he hypothesized that such patterns had originated in a similar way but had now become grammaticalized. Analysis of vowel formant data from the three dialects offered support for this premise.

For the vowel harmony hypothesis to hold, VV coarticulation would have to be perceptible to some listeners, at least in some environments. An important point here is that the hypothesis does not require that *all* speakers coarticulate heavily, or that *all* listeners perceive such effects readily. Instead, the spread of coarticulation-related change could occur because some listeners who happen to be particularly sensitive even to relatively weak coarticulatory signals would in turn retransmit those signals in a stronger form. This also leads to the intriguing question, addressed later in the present study, of whether listeners who are sensitive to coarticulation might also tend to coarticulate more. The present research was inspired by the idea that in cases where coarticulatory effects are particularly strong, a different result from Lehiste and Shockey's (1972) might be obtained, and that this result would still have important implications, as just discussed. This approach led to positive results in Grosvald (2006), where many listeners were able to determine the final vowel in VCV sequences, where C was /k/ or /p/ and V was /a/ or /i/, when hearing only the initial vowel. Therefore, the current perception study used only recordings from speakers whose VV coarticulation patterns were particularly strong.

3.1. Methodology

3.1.1. Listeners

Out of the 20 subjects who participated in the production experiment, the first seven were those whose recordings provided the raw material for this perception study; no perception data were obtained from them. Of the remaining thirteen subjects, the final three took part in a pilot perception study for which no responses were necessary (see the discussion on ERP methodology in Section 4). Therefore, a total of ten subjects—Speakers 8–17 of the production study—provided data for the perception experiment to be described here. All of these subjects were undergraduate students at the University of California at Davis who received course credit for participating. Six were females and four were males; their ages ranged from 18 to 21 [mean = 19.2; SD = 1.1].

3.1.2. Creation of stimuli

Recordings obtained from the first seven speakers in the production experiment were used to create stimuli for subsequent subjects to respond to in this perception study. The listeners' task was to distinguish [i]-colored schwas from [a]-colored schwas, for each distance condition 1, 2 and 3. Although it might seem that synthesized stimuli would be preferable, their creation would require specific decisions about the kinds of coarticulation that can or cannot occur at various distances, something that is not yet well-established.

While the aim was to determine the perceptibility of long-distance coarticulation, not all speakers in the initial production experiment had produced such effects, so an appropriate subset of the recordings had to be chosen. This should not render irrelevant the obtained results, since the language-change hypothesis described earlier requires only *some* listeners to be sensitive to *some* speakers' coarticulatory effects. The basic approach taken here was to use typical tokens from speakers who had coarticulated more strongly than average. The participants whose schwa tokens were chosen for use here, Speakers 3 and 5, were two whose results were among the strongest from the seven subjects who were initially recorded. In order to err somewhat on the conservative side, recordings from the speaker who coarticulated the most, Speaker 7, were not used, in case the results for that individual were truly exceptional.

Since individual recordings might have quirks which could be used by listeners to determine their distinctiveness independent of vowel quality, four recordings of schwa in each distance condition and vowel context were selected, to be presented interchangeably during each presentation of that vowel type. Four recordings from Speaker 5 were used for each context ([i] and [a]) for each distance 1 and 2, while four of Speaker 3's recordings were taken for each context for the distance-3 condition. Therefore, the total number of tokens used was

$$2(\text{[i] vs. [a] context}) \times 3(\text{distance-1, -2 or -3 condition}) \times 4 \text{ copies of each} = 24.$$

Because speakers had repeated each sentence six times, six tokens for each context and distance condition were available, of which four were needed as stimuli. The choice of which four to use was made in a principled manner. First, for each context and distance condition, average F1 and F2 of the corresponding six tokens were computed, defining a center point in vowel space for that group of six tokens. Next, the Euclidean distance from that center point to each token was computed, and the two outliers—those whose distance from the center point was greatest—were rejected; the remaining four tokens were used in the

Table 4
The duration, amplitude and f0 values used to standardize the tokens used in each of distance conditions 1, 2 and 3.

Distance condition	Duration (m s)	Amplitude (dB)	f0 (Hz)
1	65–70	70	120
2	55–60	70	150
3	75–80	70	200

experiment. Consequently, it was not the most extremely coarticulated, but rather most typically coarticulated, tokens that were used. These tokens were then standardized (re-synthesized) in Praat for duration, amplitude and f0, according to the values shown in Table 4.[9]

3.1.3. Perception task

All perception study subjects began by performing the production task discussed in Section 2. Afterwards, they were given a brief introduction to the purpose of this research. They were told that language sounds can affect each other over some distance, that people can sometimes detect this, and that they were about to begin a task in which their own perceptual abilities were to be tested: they would hear vowel sounds taken from sentences like the ones they had just been saying, with some of these vowels sounding more like [i] than the others. So that they would not be discouraged by the more difficult contrasts, they were told that subjects in such experiments sometimes say afterwards that they felt they were answering mostly at random when the contrasts were very subtle, but often turn out to have performed at significantly better-than-chance levels. (This turned out to be the case in the present study as well.)

Subjects were then seated in a small (approximately 10 ft × 12 ft) sound-attenuated room in a comfortable chair facing a high-quality loudspeaker (Epos, Model ELS-3C) placed 36 in away on a table 26 in high. The stimuli (stored as .wav files) were delivered using a program created by the author using Presentation software (Neurobehavioral Systems), which also recorded subject responses. The tokens had been standardized in Praat at 70 dB (as discussed earlier) and were delivered at this amplitude, as verified by measurement on a sound level meter (Radio Shack, Model 33-2055). To make their responses, subjects used a keyboard that was placed on their lap or in front of them on the table, whichever they felt was more comfortable. Free-field presentation was used because data were also being collected from some subjects for a related study involving event-related potentials (ERPs, see Section 4).

All subjects were given a very easy warm-up task about 1 min long, during which easily distinguishable [i] and [a] sounds (not schwas) were played at the same rate and ratio as their counterparts in the actual task. Subjects were told to hit a response button when they heard a sound like "[i]." After completing this warm-up, they were told that the actual task would be the same in terms of pacing and goal (responding to the [i]-like sounds), but more challenging. Impressionistically, [i]-colored schwas do have a noticeably [i]-like quality to them, particularly for distance 1 and to some extent for distance 2; participants' feedback as well as the results to be presented here indicate that subjects understood the task once they completed the warm-up.

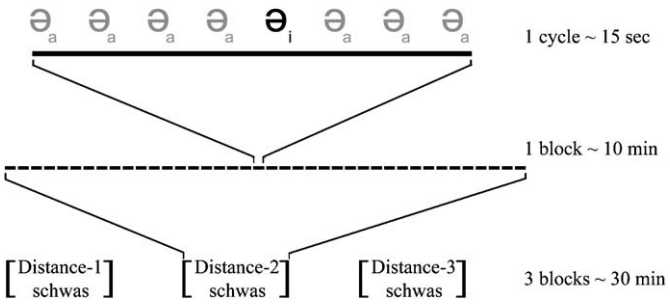

Fig. 5. Organization of the perception task. The experiment consisted of three blocks total, one for each distance condition. Each block consisted of 40 consecutive cycles of eight vowels each, with one randomly placed [i]-colored schwa per cycle.

Fig. 5 illustrates the organization of the perception experiment, which consisted of three blocks, with one block for each distance condition 1, 2 and 3, in that order. This sequence (as opposed to random order) was chosen so that each subject would begin with relatively easy discriminations, which it was hoped would keep them from being discouraged as they then progressed to the more difficult contrasts. Each block consisted of 40 cycles, each of which consisted in turn of eight consecutive schwa tokens, one of which was [i]-colored and the other seven of which were [a]-colored. Therefore, to perform with 100% accuracy, a subject would need to respond 40 times per block, by correctly responding to the one [i]-colored schwa in each of the 40 cycles in that block. The [i]-colored tokens were randomly positioned between the second and eighth slot in each cycle, so that such tokens never occurred consecutively.

The interstimulus interval (ISI) varied randomly between 1200 and 1400 m s, which provided a reasonable amount of time for subjects to respond when they thought they had just heard an [i]-colored vowel.[10] The ISI between each cycle of eight stimuli was somewhat longer, being set randomly between 2800 and 3100 m s. (These varying ISIs were required for ERP data collection.) Each cycle of eight

[9]Statistical testing on the formant values of these standardized tokens confirmed that in terms of their distribution in formant space in the [i] vs. [a] contexts, the standardized token sets were not more widely spaced—hence more easily distinguishable by listeners—than the originals.

[10]Since both stimulus and response timing were recorded by the stimulus delivery software, it was straightforward to assign each response to a vowel stimulus—namely, the immediately preceding one. While it is likely that subjects made occasional "late" responses (i.e. a response intended for one stimulus delayed until after presentation of the subsequent stimulus), participants' feedback indicated that they adjusted readily to the rhythm of the task.

vowels therefore lasted approximately 15 s, and each block of 40 cycles lasted about 10 min. Subjects were not told about the structure of cycles within blocks, but having performed the warm-up task, they had a sense of how often the [i]-colored schwas would tend to occur. Participants could choose to take short breaks of one to two minutes between blocks if they wished, or could proceed straight through the whole experiment.

3.2. Results and discussion

3.2.1. Perception measure

For an analysis of the results of the perception experiment, the raw scores are not an appropriate measure, and a statistic from signal detection theory called d' ("d-prime") will be used instead (see Macmillan & Creelman, 1991). The reasoning behind the idea that raw scores are not the best measure can be illustrated with a simple example. If a subject were to answer "i" for all items, the overall score would be 12.5% since only 1/8 of the tokens were [i]-colored. On the other hand, answering "a" for all items would be equally uninspired but would now result in fully $7/8 = 87.5\%$ correct overall. In less extreme situations, more subtle problems associated with guessing or bias could also be overlooked (see Appendix B for more information on d' and the inferential statistics used to determine statistical significance for d').

3.2.2. Individual results

The values of d' obtained for each listener in each distance condition are shown in Table 5, together with their associated significance levels. Possible d' scores range from −4.65 to +4.65, with higher values reflecting greater sensitivity. The results show that respondents were well able to distinguish [i]- and [a]-colored schwas in the distance-1 condition; many respondents had near-perfect results here. The distance-2 condition was clearly more challenging, and seemed to represent a threshold of sorts, inasmuch as five respondents' scores did not reach significance while four others' did (the remaining subject provided no data here). The distance-3 condition was by far the most challenging, and respondents appear to have answered mostly at random, although one subject did attain the remarkable score of 2.28. It is worth noting that completely random guessing should result in d' scores centered near 0, but all respondents' scores for all conditions are positive, with the sole exception of Subject 9's distance-3 block.

3.2.3. Correlation between production and perception

In a study investigating the possibility of a link between linguistic production and perception, Perkell et al. (2004) found that speakers who articulated phonemically contrasting vowels more distinctly also showed a greater ability to distinguish vowel contrasts. Although the contrasts explored in the present study are sub-phonemic in nature, findings like those of the Perkell et al. study raise a question relevant to the language-change scenario discussed earlier: is there a correlation between individuals' ability to detect coarticulatory effects and tendency to coarticulate?

Table 5
Each subject's d' scores, one for each of the three distance conditions.

Subject	Distance 3	Distance 2	Distance 1
8	2.28***	1.19*	3.97***
9	−0.65	No data	4.65***
10	0.16	0.75	4.65***
11	0.20	0.52	4.18***
12	0.52	1.84***	3.40***
13	0.67*	1.71***	3.45***
14	0.77	1.84**	3.54***
15	0.16	1.04	4.65***
16	0.27	1.16	4.29***
17	0.55	0.89	4.15***
Average	0.49	1.21	4.09

Significant results are noted, where * = $p<0.05$, ** = $p<0.01$ and *** = $p<0.001$. Subject 9 provided no responses for the distance-2 condition.

This possibility was investigated for the ten study participants who provided both production and perception data. As a quantitative measure of perceptual ability for a given subject, the average of that subject's three d' scores was used. Each subject's production measure was the average over the three distance conditions of the Euclidean distance in normalized vowel space between that subject's [i]- and [a]-colored schwas. The perception and production measures so obtained were found to be positively correlated, with $r = 0.52$, but this outcome is not statistically significant ($p = 0.13$). As shown in Fig. 6, the relationship depends on data from the two participants with the highest and lowest d' average and does not hold up if they are excluded. In case some other measures of perception or production might reveal a stronger relationship, other candidate measures were also tested, such as different relative weightings among the distance conditions and the use of logarithmic vowel-space measures instead of raw formant numbers (cf. Johnson, 2003), but none of these led to substantially different results. Evidently, more study will be needed before any strong claims can be made concerning a possible perception–production relationship.

4. General discussion

A number of models of spoken-language coarticulation have emerged in the last three or four decades (see Farnetani & Recasens, 1999, and more generally Hardcastle & Hewlett, 1999), two of the most dominant being coproduction models (Fowler, 1983) and Keating's Window model (1988, 1990a, 1990b). In a coproduction model, the articulatory gesture(s) associated with a given speech segment have a more-or-less fixed temporal duration which may overlap with those of neighboring segments, the

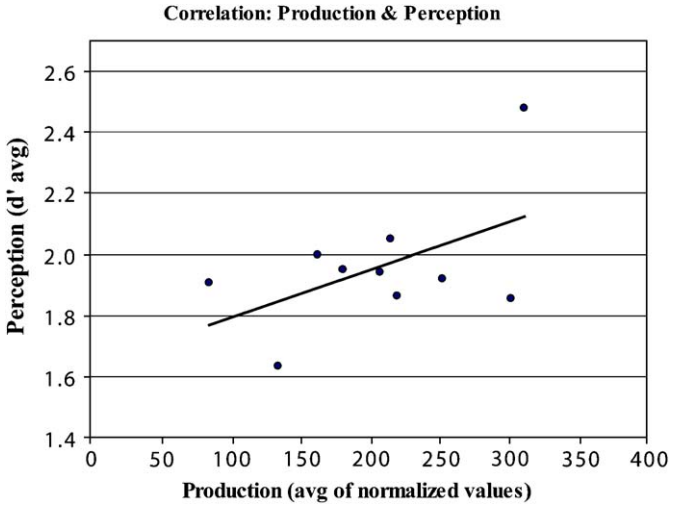

Fig. 6. Correlation between the averaged production and perception measures for the ten subjects who provided data for both ($r = 0.52$, $p = 0.13$).

resulting output at a given moment being an interpolation or averaging of the gestures associated with the segments in play at that time. A key prediction of this theory is that since each gesture's temporal duration is limited, its temporal range of influence on its neighbors should have a rather small upper bound. As was noted earlier, long-distance production results such as those seen in the current study appear inconsistent with this last assertion. It has been seen here that from both a production and a perception standpoint, VV coarticulatory processes may be relevant over at least three vowels' distance, and hence across as many as five intervening segments. These results would seem to pose a problem for any model of coarticulation not allowing for considerable range of influence of segments on one another.

In contrast, the Window model (Keating, 1988, 1990a, 1990b) uses an approach more akin to feature spreading; gestural targets associated with linguistic segments are "windows" (ranges, not points), through which paths are traversed over the course of an utterance. The gestural paths between successive windows are determined through interpolation, and such interpolation may stretch over long distances in cases where intermediate segments are underspecified for a feature. Since this model makes no specific predictions about the limits of long-distance coarticulation, it may be more compatible with the kinds of long-distance results found in this study than the coproduction model is. However, since the production data showed coarticulation occurring almost universally across [k] and frequently across [t], the idea raised by Öhman (1966) and implicit in the Window model—that VV coarticulation should be largely blocked by consonants making more constraints on the tongue body—may be overstated (for relevant work see, for example, Modarresi, Sussman, Lindblom, & Burlingame, 2004; Recasens, Pallarès, & Fontdevila, 1997). The crucial factor here appears to have been the especially large susceptibility of schwa to coarticulatory influence, regardless of the consonant context.

Much of the difficulty in this area of research stems from the fact that coarticulatory patterns seem to a large extent to be idiosyncratic, varying greatly from speaker to speaker. Some of this probably has to do with the relative freedom speakers have in producing a given speech sound, exemplified in studies in which speakers successfully articulate particular vowels in spite of physical obstacles like bite blocks (Fowler & Turvey, 1980; Lindblom, Lubker, & Gay, 1979). A more complete analysis will almost certainly be dependent on recruiting sizable groups of subjects, and in addition may require more sensitive measures for production and perception. For instance, an examination of Speaker 7's numerical production data (see Appendix A) shows that the standard deviations associated with his formant values tended to be smaller than those of other speakers; even if two speakers have similarly sized vowel spaces, one speaker may "hit the targets" more accurately than the other. A good coarticulation measure may need to take this kind of variation into account (see Adank, Smits, & van Hout, 2004). A better production measure would lead to a better quantification of the production/perception relationship as well.

Flemming (1997) mentions VV coarticulation in a discussion in which he argues that phonological representations by necessity contain more phonetic information than has traditionally been assumed; his goal is a "unified account of coarticulation and assimilation." Since it seems evident that coarticulatory effects at various distances are perceptible in some or many cases, a complete account of this phenomenon will be a complicated undertaking indeed, given the variation we see among speakers and listeners. As an example of the subtleties involved, consider the symbols [ə$_i$] and [ə$_a$] that were used in Fig. 1 to represent [i]- and [a]-colored schwas regardless of the distance involved. Suppose that part of the accommodation that listeners make when exposed to a particular speaker includes becoming sensitive to that speaker's coarticulation patterns, including coarticulatory tendency at various distances. Such possibilities might require models recognizing differences among items such as [$_i$a] (carryover [i]-coloring of [a]) vs. [a$_i$] (anticipatory [i]-coloring of [a]), or more generally [ə$_i^1$] (distance-1 anticipatory [i]-coloring of schwa), [$_a^2$o] (distance-2 carryover influence of [a] on [o]), and so on, even if only a subset of listeners is sensitive to such subtleties. The implications of multiple simultaneous effects of different neighboring segments on a single segment might also need to be considered; recall the production results in the present study which appeared to show simultaneous VV and C-to-V effects.

One promising approach in investigating perception is the use of non-behavioral methodology, one example of which is the event-related potential (ERP) technique, which involves the recording of brain-wave (electroencephalogram) data and can provide insight into mental processes

which occur whether or not subjects are consciously aware of them (see Luck, 2005). Groups of neurons firing in response to particular types of stimuli produce positive or negative electrical potentials at characteristic scalp locations during particular timeframes. When averaged over many trials, the background noise tends to zero and the response pattern consistently associated with the stimulus present in each trial remains; such patterns are called ERP components, and many are associated with linguistic processes. For example, Frenck-Mestre, Meunier, Espesser, Daffner, and Holcomb (2005) investigated the ability of French listeners to perceive phonemic contrasts existing in English but absent in French. While data in the current perception study were gathered with behavioral methodology, pilot ERP data using the same paradigm have also been collected and based on the promising outcome, a full study is now underway.

5. Conclusion

This study examined the extent and perceptibility of long-distance vowel-to-vowel coarticulatory effects, and the degree to which these vary among speakers and listeners. It was found that anticipatory VV coarticulation can occur over at least three vowels' distance in natural discourse, and that even such long-distance effects can be perceived by some listeners. Both coarticulatory strength and perceptual sensitivity varied greatly among study participants. The possibility of interplay between coarticulatory production and perception was also investigated, but with inconclusive results; correlation was found to be positive but weak. Speaking rate and coarticulation strength were found not to be correlated. An examination of more contexts and more languages, perhaps complemented by other methodological approaches, should prove useful as we seek a better understanding of these complex issues.

Acknowledgements

I would like to thank Orhan Orgun, David Corina, Carol Fowler, Patricia Keating, Harriet Magen, Daniel Recasens and Keith Johnson for invaluable feedback as this project was being planned and carried out. Additional thanks are due to three anonymous reviewers and to audiences at the UC Berkeley Phonetics/Phonology Phorum, the 2008 Linguistics Society of America meeting in Chicago, and the 2007 coarticulation workshop of the Association Francophone de la Communication Parlée in Montpellier, all of whose comments were also extremely helpful.

Appendix A. Raw formant values for all subjects

The following tables show the average target-vowel F1 and F2 values for each speaker in each distance condition and vowel context obtained in the main production experiment, together with the associated standard deviations. Note that the "a" and "i" labeling refers to context, not the measured vowels themselves, which were always schwa or [ʌ]. The first, second and third tables show results for the distance-1, -2 and -3 conditions, respectively. Significant results are noted, where * = $p<0.05$, ** = $p<0.01$ and *** = $p<0.001$.

Also note that these measurements were made near the end of the target vowels, where coarticulatory influence of the context vowel was expected to be strongest, and where effects of the immediately following consonant appear to be seen as well. Therefore, these formant values should not be expected to correspond too closely to the values one would obtain in the steady-state portion of a schwa vowel.

Formant values of distance-1 target vowels in [a] vs. [i] context

Speaker	F1 [a]		F1 [i]			F2 [a]		F2 [i]		
	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
1 (f)	431	34	405	43		1991	200	2274	121	*
2 (m)	408	14	282	36	***	1442	28	1992	78	***
3 (f)	359	96	315	57		1729	72	2644	165	***
4 (f)	527	104	471	57		2140	239	2524	292	*
5 (f)	585	42	423	55	***	1698	70	2360	100	***
6 (m)	534	48	393	33	***	1622	62	1893	138	**
7 (m)	484	8.6	419	4.4	***	1538	43	1980	21	***
8 (f)	452	110	379	62		1814	88	2605	228	***
9 (m)	355	22	300	32	**	1837	43	2025	90	**
10 (f)	439	77	377	110		2145	147	2731	297	**
11 (f)	458	93	418	97		2089	207	2299	347	
12 (m)	309	93	195	109	*	1562	63	1843	93	***
13 (m)	445	51	331	53	**	1727	33	2123	76	***
14 (f)	432	35	300	45	***	1914	129	2437	305	**
15 (f)	465	29	345	52	***	1826	60	2235	252	**

16 (f)	359	87	307	51		2138	480	2499	354	
17 (m)	447	58	339	53	**	1577	18	1932	122	***
18 (m)	353	68	251	39	*	1808	51	1979	221	
19 (m)	429	49	308	36	***	1634	114	1854	83	**
20 (f)	428	55	305	33	***	1896	93	2862	106	***

Formant values of distance-2 target vowels in [a] vs. [i] context

Speaker	F1 [a]		F1 [i]			F2 [a]		F2 [i]		
	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
1 (f)	572	31	583	21		1622	73	1659	56	
2 (m)	461	7.6	400	57	*	1440	38	1515	56	*
3 (f)	633	82	404	100	***	1731	61	2009	47	***
4 (f)	757	14	746	35		1979	72	1978	72	
5 (f)	594	9.6	558	26	**	1780	79	1914	39	**
6 (m)	551	42	515	52		1311	68	1411	64	*
7 (m)	483	9.0	478	22		1363	14	1504	16	***
8 (f)	870	20	832	68		2052	50	2167	82	**
9 (m)	528	30	487	26	*	1418	43	1481	45	*
10 (f)	617	42	536	50	**	1964	35	2069	29	***
11 (f)	798	34	803	48		1993	42	2058	55	*
12 (m)	539	43	439	78	*	1611	59	1632	32	
13 (m)	631	12	617	19		1439	113	1523	42	
14 (f)	721	38	664	55	*	1866	43	1973	50	**
15 (f)	612	22	595	41		1519	48	1620	39	**
16 (f)	794	37	802	37		1919	105	1735	382	
17 (m)	621	18	587	26	*	1391	27	1523	15	***
18 (m)	459	32	360	83	*	1422	43	1404	92	
19 (m)	506	19	489	17		1491	70	1548	33	
20 (f)	588	29	563	53		1706	34	1841	102	*

Formant values of distance-3 target vowels in [a] vs. [i] context

Speaker	F1 [a]		F1 [i]		F2 [a]		F2 [i]		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
1 (f)	587	27	635	45	1503	251	1527	117	
2 (m)	455	89	479	10	1127	35	1116	99	
3 (f)	718	88	704	95	1412	235	1648	49	*
4 (f)	617	105	592	87	1802	139	1767	213	
5 (f)	692	48	722	67	1497	63	1577	96	
6 (m)	462	210	530	37	1220	294	1290	254	
7 (m)	542	40	515	35	1201	69	1297	31	**
8 (f)	489	199	367	258	1655	103	1709	138	
9 (m)	310	87	287	103	1294	109	1319	59	
10 (f)	395	116	290	158	1655	134	1679	114	
11 (f)	525	78	434	149	1788	157	1706	149	
12 (m)	452	91	487	94	1184	130	1257	218	
13 (m)	214	31	169	64	1201	52	1196	140	
14 (f)	507	124	468	79	1502	58	1543	100	
15 (f)	375	64	416	94	1356	211	1380	197	
16 (f)	318	99	546	180	1680	137	1641	62	

17 (m)	462	63	500	57	1248	85	1314	61
18 (m)	390	58	318	82	1269	86	1208	88
19 (m)	509	36	526	19	1261	73	1312	47
20 (f)	416	117	443	158	1501	275	1690	115

Appendix B. The d' statistic and its use in this study

To analyze the perception study correctly, we can consider the subjects' task to be a signal-detection effort. All of the stimuli were schwa sounds, but in 1/8 of them, there was coarticulatory [i] coloration; this is the "signal" that the subject was trying to detect. In this context, a correctly reported [i] item is referred to as a "hit," an appropriately ignored [a] item is a "correct rejection," a wrong "i" answer is a "false alarm," and a wrong "a" answer is a "miss." The d' statistic is the difference between a subject's normalized hit and false-alarm rates:

$$d' = z(\text{hits/trials with signal present}) - z(\text{false alarms/trials with signal absent}).$$

For rates of 0 or 1, z is not defined, so the values 0.01 and 0.99 are substituted. Given this, possible values of d' range from -4.65 to $+4.65$, with an expected value of 0 if the subject has zero sensitivity to the contrast in question. Scores in the vicinity of 1 or higher generally turn out to be significant at the $p<0.05$ level. The variance of d' is given by the formula

$$H(1-H)/N_{\mathrm{H}}[\varphi(H)]^2 + F(1-F)/N_{\mathrm{F}}[\varphi(F)]^2,$$

where H, F, N_{H}, N_{F} and φ are the hit rate, false-alarm rate, number of "i" trials, number of "a" trials, and probability density function of the normal distribution, respectively (Gourevitch & Galanter, 1967). Using this, a confidence interval for d' can be determined for each subject. If the lower endpoint of the interval is greater than zero, one can be confident (to the chosen degree of significance) that the subject has some sensitivity to the contrast. Note that similar d' scores can be obtained with different distributions of correct and incorrect answers, both of which are used to compute the variance of d', which means that d' scores in the same range can have different significance testing outcomes.

References

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, *116*, 3099–3107.

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

Benguerel, A.-P., & Cowan, H. A. (1974). Coarticulation of upper lip protrusion in French. *Phonetica*, *30*, 41–55.

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer [computer program]. Available from: ⟨http://www.praat.org⟩.

Butcher, A., & Weiher, E. (1976). An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics*, *4*, 59–74.

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

Farnetani, E., & Recasens, D. (1999). Coarticulation models in recent speech production theories. In W. J. Hardcastle, & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 31–65). Cambridge: Cambridge University Press.

Flemming, E. (1997). Phonetic detail in phonology: Towards a unified account of assimilation and coarticulation. In K. Suzuki, & D. Elzinga (Eds.), *Proceeding volume of the 1995 Southwestern workshop in optimality theory (SWOT)*. Tucson, AZ: University of Arizona.

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

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

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms in speech: Cyclic productions of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General*, *112*, 386–412.

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, *36*, 171–195.

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

Fowler, C. A., & Turvey, M. T. (1980). Immediate compensation in bite-block speech. *Phonetica*, *37*, 306–326.

Frenck-Mestre, C., Meunier, C., Espesser, R., Daffner, K., & Holcomb, P. (2005). Perceiving nonnative vowels: The effect of context on perception as evidenced by event-related brain potentials. *Journal of Speech, Language, and Hearing Research*, *48*, 1–15.

Gay, T. (1977). Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America*, *62*, 183–193.

Gerstman, L. H. (1968). Classification of self-normalized vowels. *IEEE Transactions on Audio and Electroacoustics*, *ACC-16*, 78–80.

Grosvald, M. (2006). *Vowel-to-vowel coarticulation: Length and palatalization effects and perceptibility*. Davis, CA: University of California at Davis unpublished manuscript.

Gourevitch, V., & Galanter, E. (1967). A significance test for one-parameter isosensitivity functions. *Psychometrika*, *32*, 25–33.

Hardcastle, W. J., & Hewlett, N. (1999). *Coarticulation: Theory, data and techniques*. Cambridge: Cambridge University Press.

Heid,, S., & Hawkins,, S. (2000). An acoustical study of long-domain /r/ and /l/ coarticulation. In *Proceedings of the fifth seminar on speech production: Models and data* (pp. 77–80). Bavaria, Germany: Kloster Seeon.

Hertrich, I., & Ackermann, H. (1995). Coarticulation in slow speech: Durational and spectral analysis. *Language and Speech*, *38*, 159–187.

Hussein, L. (1990). VCV coarticulation in Arabic. *Ohio State University Working Papers in Linguistics*, *38*, 88–104.

Johnson, K. (2003). *Acoustic and auditory phonetics*. Malden, MA: Blackwell.

Keating, P. (1988). Underspecification in phonetics. *Phonology*, *5*, 275–292.

Keating, P. (1990a). Phonetic representations in a generative grammar. *Journal of Phonetics*, *18*, 321–334.

Keating, P. (1990b). The window model of coarticulation: Articulatory evidence. In J. Kingston, & M. E. Beckman (Eds.), *Papers in Laboratory Phonetics I: Between the grammar and the physics of speech* (pp. 451–470). Cambridge: Cambridge University Press.

Lehiste, I., & Shockey, L. (1972). On the perception of coarticulation effects in English VCV syllables. *Journal of Speech and Hearing Research*, *15*, 500–506.

Lindblom, B., Lubker, J., & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech-motor programming by predictive simulation. *Journal of Phonetics*, *7*, 147–161.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

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. *Haskins Laboratories Status Report on Speech Research*, *103–104*, 1–20.

Manuel, S. Y., & Krakow, R. A. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research*, *77–78*, 69–78.

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*, 473–488.

Modarresi, G., Sussman, H., Lindblom, B., & Burlingame, E. (2004). An acoustic analysis of the directionality of coarticulation in VCV utterances. *Journal of Phonetics*, *32*, 291–312.

Moll, K. L., & Daniloff, R. G. (1971). Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America*, *50*, 678–684.

Ohala, J. (1981). The listener as a source of sound change. In M. F. Miller (Ed.), *Papers from the parasession on language behavior*. Chicago: Chicago Linguistic Association.

Ohala, J. (1994). Towards a universal, phonetically-based, theory of vowel harmony. *ICSLP-1994*, 491–494.

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

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., et al. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America*, *116*, 2338–2344.

Przezdziecki, M. (2000). Vowel harmony and vowel-to-vowel coarticulation in three dialects of Yoruba. *Working Papers of the Cornell Phonetics Laboratory*, *13*, 105–124.

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

Recasens, D. (1989). Long range coarticulation effects for tongue dorsum contact in VCVCV sequences. *Speech Communication*, *8*, 293–307.

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

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

Scarborough, R. A. (2003). Lexical confusability and degree of coarticulation. In *Proceedings of the 29th meeting of the Berkeley linguistics society* (pp. 367–378). Berkeley, CA: Berkeley Linguistics Scoiety.

West, P. (1999). The extent of coarticulation of English liquids: An acoustic and articulatory study. In *Proceedings of the international conference of phonetic sciences*, (pp. 1901–1904) San Francisco.