ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Phonetics 33 (2005) 199–213

Journal of Phonetics

www.elsevier.com/locate/phonetics

Parsing coarticulated speech in perception: effects of coarticulation resistance

Carol A. Fowler*

Haskins Laboratories, 270 Crown Street, New Haven, CT 06511, USA

Received 29 August 2003; received in revised form 4 June 2004; accepted 15 October 2004

Abstract

A speaker produced schwa-CV disyllables in which consonants were low or high in coarticulation resistance. Articulatory and acoustic measurements verified that the magnitude, but not the extent, of anticipatory coarticulation from the stressed vowel to the schwa was modulated by coarticulation resistance. In a perception experiment, listeners heard schwa-CV disyllables that ended in stressed /i/ or /a/ and made speeded identifications of the stressed vowels. Some disyllables had been cross-spliced so that a schwa vowel that had originally been produced in the context of C/i/ was spliced onto a C/a/ syllable or vice versa. Other disyllables were spliced so that a schwa originally from a C/i/ (or C/a/) context was spliced onto a different C/i/ (C/a/) token. Listeners' response latencies were slower to cross spliced than to spliced disyllables but only in the context of low rather than high resistant consonants. The outcome is generally consistent with a hypothesis that listeners to speech "parse" the acoustic signal along coarticulatory or phonetic gestural lines and that success in parsing varies with the amount of acoustic evidence talkers provide. However, findings did not suggest "perfect" parsing. Correlations between the amount of acoustic information provided by the speaker and the extent to which listeners were disrupted by cross splicing were nonsignificant.

*Tel.: +1 203 865 6163; fax: +1 203 865 8963.
E-mail address: carol.fowler@haskins.yale.edu (C.A. Fowler).

0095-4470/$ - see front matter
doi:10.1016/j.wocn.2004.10.003

1. Introduction

Speakers coarticulate the gestures of the consonants and vowels of words. However, they modulate their degree of coarticulation so that it does not interfere with achievement of gestural goals. Following Bladon and Al-Bamerni (1976), Recasens (1984a, b, 1985) has used the term "coarticulation resistance" to reflect this modulation. One of his hypotheses has been that consonants resist coarticulation by vowels to the extent that the two segment types place conflicting demands on an articulator, generally the tongue body. Excepting /g/ in English, consonants with constriction gestures that use the tongue resist coarticulation by vowels more than consonants, such as /b/ and /v/, that do not (e.g., Fowler & Brancazio, 2000). As for English /g/, its place of articulation moves continuously with changes in the constriction location of coarticulating vowels (Dembowski, Lindstrom, & Westbury, 1998).

The present investigation explores the implications of coarticulation resistance for speech perception and production.

As for perception, many studies have shown that listeners are sensitive to information about coarticulation, and that they use information in the domain of one segment about a segment coarticulating with it as such. For example, in pioneering research Martin and Bunnell (1982; see also Martin & Bunnell, 1981) cross-spliced initial vowels from VCV disyllables so that anticipatory coarticulatory information about the final vowel was sometimes misleading. Listeners identified the final vowels in a speeded task. Response times were slower and errors higher when coarticulatory information in the initial vowel was misleading than when it was valid. This result has been replicated many times (e.g., Whalen, 1984) and has been extended to coarticulatory information about vowels in the domain of preceding consonants (e.g., Fowler, 1984). The issue in the present investigation concerns the closeness with which listeners track coarticulatory information. Vowel-to-vowel coarticulation is less in magnitude (although, as above, not necessarily in extent) across a high than a low resistant consonant (Recasens, 1984b). This means that acoustic information for the final vowel of a VCV present in the initial vowel should be more salient in the context of low than of high resistant consonants. If it is and if listeners use coarticulatory information to the extent that it is available, then the consequences of cross splicing the initial vowel across CV contexts should be greater when consonants are low in coarticulation resistance.

Fowler and Brancazio (2000) used a mid-sagittal electromagnetometer to track articulation by two speakers who produced schwa-CVs where the Cs were the low resistant /b/, /v/ and /g/ and the high resistant /ð/, /d/, /z/ and /ʒ/, and Vs were /a/, /ʌ/ and /i/.[1] Both speakers showed less anticipatory vowel-to-consonant coarticulation across high than low resistant consonants. Interestingly, however, as Recasens (1984b, 1989) had found for some, but not all, speakers, reduction of coarticulation in the context of high as compared to low resistant consonants was reflected in the magnitude of coarticulation, but not in its extent. Moreover, Fowler and

[1] /b/ and /v/ resist coarticulation by vowels very little, because their constriction gestures do not involve the tongue. Therefore, tongue gestures for a vowel interfere little with their production. The other consonants all use the tongue. The reason why /g/ is low resistant despite requiring the same part of the tongue for its production that vowels require may be that, in English, there are no stop consonants with places of articulation close to that of /g/ (and /k/). Therefore, when, due to coarticulation, the constriction location for /g/ shifts, it does not become confusable with any other stop of English.

Brancazio found little evidence of effects of coarticulation resistance of the consonant on the magnitude of vowel-to-vowel coarticulation. They subtracted tongue height and tongue fronting measures obtained from four points in the initial schwa vowel of /ə/C/ʌ/ utterances from corresponding points of /ə/C/i/ utterances. They found positive differences, reflecting coarticulatory effects of the stressed vowels in all but one of 112 comparisons (four measurement points in schwa × seven consonant contexts × height vs. fronting measures × two speakers). This is an interesting finding, because it shows that coarticulation resistance does not prevent vowel-to-vowel coarticulation, but, at least for the speakers in the study, only clamps down on its implementation in the vicinity of the consonant. This may mean that planning for vowel-to-vowel coarticulation does not need to be modulated due to demands of the consonant. That is, speakers may begin producing a vowel at a fixed phase in the production of a preceding vowel or consonant regardless of the coarticulation resistance of the consonant. They do not plan differently for vowels in the context of high vs. low resistant consonants. As for the consonant, its implementation interferes more (for high resistant consonants) or less (low resistant consonants) with the continuation of vowel production in its own domain.

In the present investigation, I collected articulatory data from a third speaker, who produced six of the same consonants as those of Fowler and Brancazio's speakers (excepting /ʒ/), also in a schwa-CV context, now with six stressed vowels, /i/, /ei/, /ɛ/, /ʌ/, /u/, and /a/. Although the main purpose was to explore perception of vowel-to-vowel coarticulation, in the interests of comparison with the speakers of the study by Fowler and Brancazio, and to describe the stimuli to which listeners will be responding, production findings on the present speaker will be reported first.

2. Experiment 1

2.1. Method

2.1.1. Participants

The participant was a female adult native speaker of American English, who reported normal speech and hearing. She was unaware of the purpose of the investigation, but had some phonetics training.

2.1.2. Stimulus materials

Target utterances were disyllables beginning with the unstressed schwa vowel followed by one of six consonants and one of six stressed vowels. Consonants were /b/, /v/, /g/, /ð/, /d/ and /z/. Vowels were /i/, /ei/, /ɛ/, /ʌ/, /u/, and /a/. Target disyllables were produced in a carrier sentence: "Enough ____ Bub" designed so that consonants flanking the target disyllable were not produced with a tongue gesture. Each utterance was produced 15 times for a total of 540 utterances. The order of disyllables in the script from which the participant read was random.

2.1.3. Procedure

Articulatory data were collected using an electromagnetic mid-sagittal articulometer (EMMA) designed at MIT (Perkell et al., 1992). EMMA tracks the motions of receivers attached to the articulators in the front-back and height dimensions.

In the experiment, receiver coils were placed on the vermillion border of the upper and lower lips, on four evenly spaced locations on the tongue mid-line, on the bridge of the nose, the maxillary gum line and the mandibular gum line. On the tongue mid-line, a coil was placed at the tongue tip and one as far back on the tongue as the speaker could tolerate. The remaining tongue coils were placed at equal distances from the two extreme coils. Present analyses are based on the third coil on the tongue mid-line, approximately on the tongue body. Receiver coil voltages were sampled at 625 Hz (12 bit resolution) after being low pass filtered at 200 Hz by a hardware filter. Software was used to convert receiver coil voltages to Cartesian coordinates for each receiver. Data were rotated to the occlusal plane and, using the information from receivers on the nose and the maxilla, were corrected for head movement. The acoustic signals from the speaker's productions were sampled at 22 kHz.

The speaker sat in a dental chair facing a computer monitor. She was cued to produce a particular disyllable in the carrier phrase when it was printed in IPA notation on the computer screen. On each trial, an experimenter said "Ready?". If the participant assented, the experiment initiated a trial by button press. A disyllable was presented on the monitor facing the speaker, and data collection was initiated.

Articulatory measures (tongue body fronting and height) were taken at nine points in the target disyllabic utterances. Measurement points were identified from the acoustic waveform: schwa onset and offset, and two points equidistant from them within the schwa, mid-closure of the consonant, final vowel onset, offset and two points equidistant from them within the vowel. Acoustic measures of the first two formants were taken in mid-schwa and mid-stressed vowel, that is, between articulatory measurement points 2 and 3.

2.2. Results and discussion

Fig. 1 shows tongue body fronting at the four measurement points (during schwa (y axes labeled S1–S4; read panels left to right and top to bottom), during consonant closure (C) and at the third measurement point during the stressed vowel (V3)) for all six vowels and for the six consonants. Fig. 2 shows tongue body height at the same measurement points. The sixth panel in each figure shows the differences in tongue body position within the stressed vowels. Coarticulation in the four schwa panels can be gauged by comparing the curves with those in the sixth panel.

Panel 5 verifies the classification of /b/, /v/ and /g/ (open symbols) as low resistant consonants and the remaining consonants (filled symbols) as high resistant consonants. That is, the tongue shows far less shaping by the following stressed vowel if the intervening consonant is high vs. low resistant.

In both figures, during schwa, data points cluster into two groups by curve shape, one for the low resistant consonant contexts (/b/, /v/ and /g/) that shows considerable influence of the frontness and height of the stressed vowel, and the other, a much flatter shape for the high resistant /d/, /ð/ and /z/ contexts. As for tongue fronting, the tongue is farther back during schwa due to coarticulation with back vowels /a/, /u/, and /ʌ/ than in the context of front vowels /i/, /ei/, and /ɛ/. However, the difference in fronting is greater in the context of /b/, /v/ and /g/ than in the context of /d/, /ð/ and /z/. Fig. 2 shows an analogous outcome for tongue height. That is, each of the panels showing tongue height during schwa exhibits an upside down V-shape reflecting

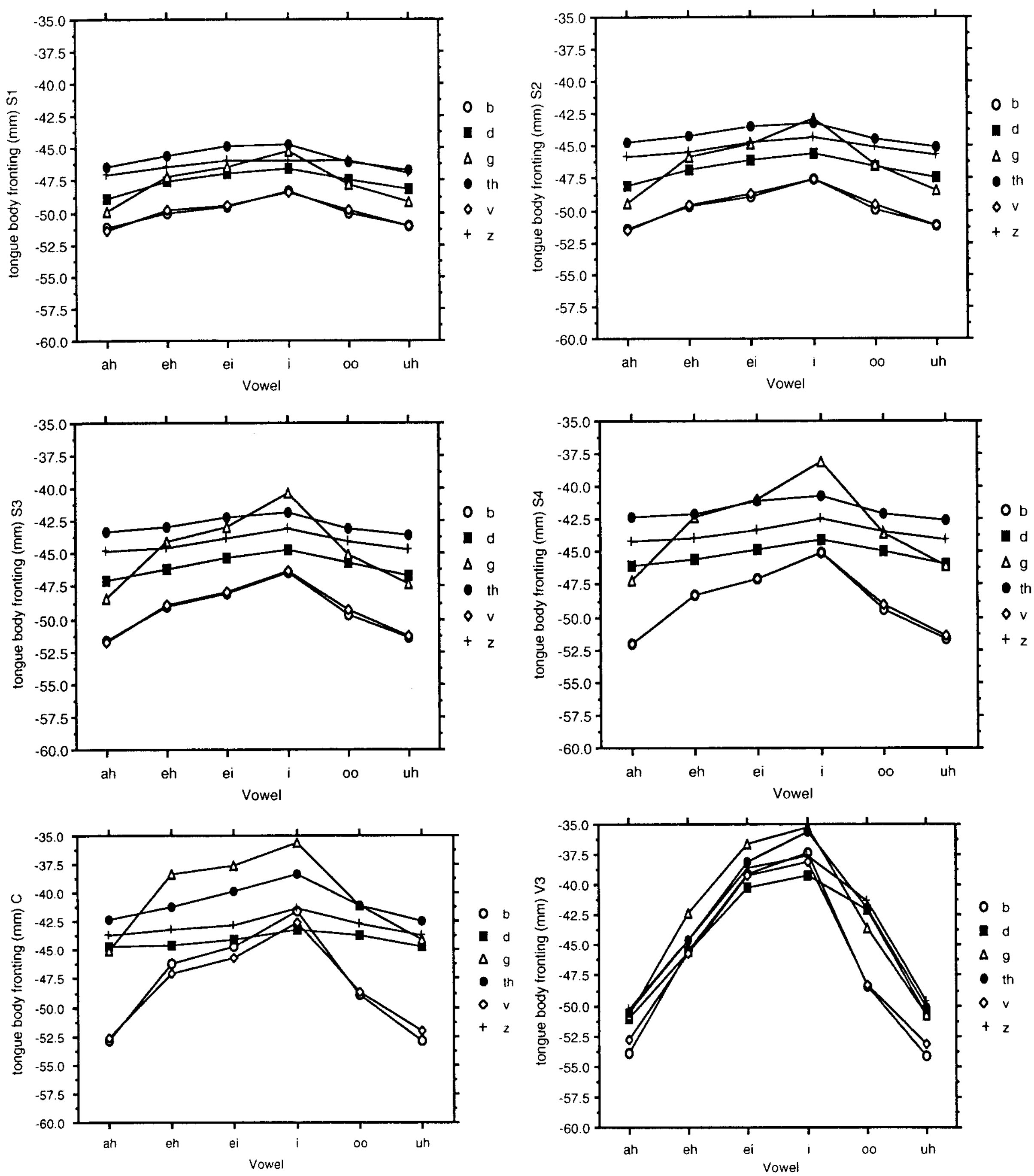

Fig. 1. Tongue body fronting measures at six points in time during schwa-CV disyllables: at four equally spaced intervals during schwa (S1–S4), during consonant closure (C), and at the third articulatory measurement point during V (V3). Less negative values are associated with more front vowels.

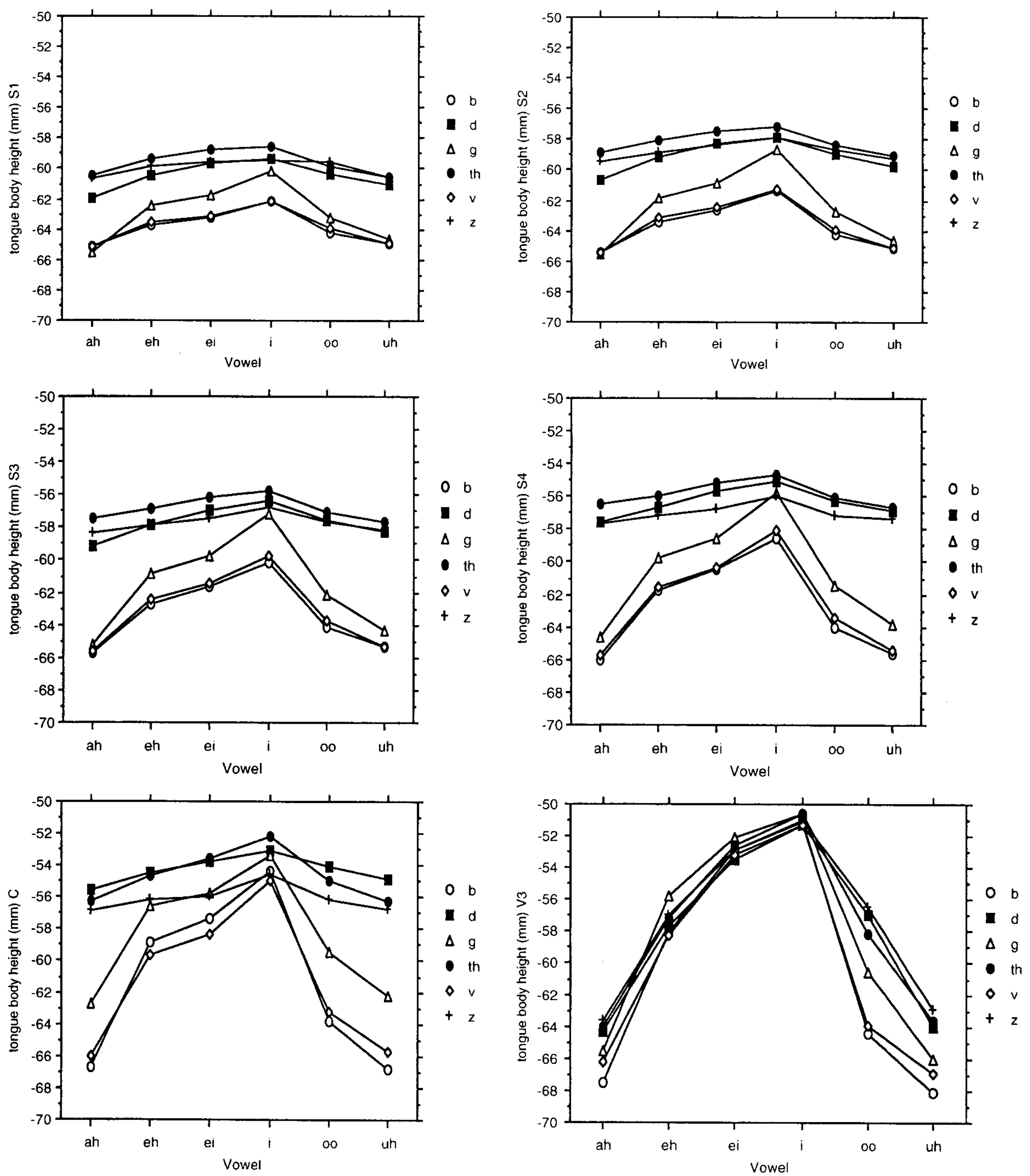

Fig. 2. Tongue body height measures at six points in time during schwa-CV disyllables: At four equally spaced intervals during schwa (S1–S4), during consonant closure (C), and at the third articulatory measurement point during V (V3). Less negative values are associated with a higher positioning of the tongue body.

changes in vowel height for the vowel contexts arrayed along the x-axis. However, the V-shape is only marked for disyllables in which the consonants were /b/, /v/ or /g/.

In analyses of variance with vowel and consonant as factors all main effects and interactions were highly significant. (See Appendix A for the relevant inferential statistics.) This speaker shows clear evidence of the effect of the coarticulation resistance of the consonant on the magnitude of vowel-to-vowel coarticulation.

A question that guided the design of the experiment, however, was whether coarticulation resistance of the consonant affected the temporal extent of vowel-to-vowel coarticulation. I addressed this question by asking whether there was significant vowel-to-vowel coarticulation at the beginning of the schwa vowel in disyllables that included high resistant consonants. There was. In both the measure of tongue fronting and in the measure of tongue body height, the effect of the identity of the stressed vowel was highly significant at schwa onset (fronting; $F(5,252) = 26.43, p<.001$; height ($F(5,252) = 23.34, p<.001$), with analyses restricted to disyllables including /d/, /ð/ and /z/. In this respect, the speaker replicates findings of Fowler and Brancazio (2000).

Acoustic measures were largely consistent with the articulatory findings. Measures of F1 and F2 were obtained during schwa at a point intermediate between the two middle articulatory measurement points in the schwa—that is, approximately in the middle of the vowel. They were analyzed using ANOVA with consonant and stressed vowel as factors. Expectations were that both formants would show effects of coarticulation by the stressed vowel, but to an extent that was modulated by the coarticulation resistance of the intervening consonant. This was the finding consistently for F2, but not for F1.

In the analysis of F1 the main effects of consonant ($F(5,464) = 21.89$, $p<.001$), and vowel ($F(5,464) = 37.58$, $p<.0001$) were highly significant. The interaction did not approach significance. Characteristically, the largest difference in F1 across vowels in the context of a given consonant was between schwa in the context of /a/ vs. /i/ or /u/. This difference averaged 77, 88 and 54 Hz for /b/, /g/ and /v/, respectively. It was 89, 72 and 52 Hz for /ð/, /d/ and /z/, respectively. Accordingly, although F1 showed significant effects of coarticulation, they were not smaller in the context of high than low resistant consonants. Thus, the small acoustic effects on F1 of coarticulation by the stressed vowel do not reflect variation in the coarticulation resistance of the consonant.

In the analysis of F2, both main effects (consonant: $F(5,464) = 441.32$, $p<.001$; vowel: $F(5, 464) = 137.95, p<.0001$) and the interaction ($F(25, 464) = 12.10, p<.0001$) were significant. Here the interaction reflected the predicted effects. The vowels showing the largest difference in F2 were /a/ vs. /i/. These differences averaged 252, 445 and 250 Hz for /b/, /g/ and /v/, respectively. They averaged 116, 152 and 121 for /ð/, /d/ and /z/, respectively.

The two talkers studied by Fowler and Brancazio (2000) and the present talker showed effects of coarticulation resistance. In the study of Fowler and Brancazio, analyses were done on /i/-/ʌ/ difference scores. There, both on measures of tongue body height and on measures of tongue body fronting, /i/-/ʌ/ differences were larger in the contexts of /b/ and /v/ than in the contexts of other consonants. For one talker, /g/ grouped with these low resistance consonants in the measure of tongue body height only. For the other talker, /g/ grouped with the high resistant consonants on both measures. The present analyses were not performed on difference scores, because there were six vowels. However, Figs. 1 and 2 show that /b/, /v/ and /g/ form a group with respect to the magnitude of the coarticulatory influence of the stressed vowel; /d/, /z/ and /ð/ form a separate group. The major difference in outcome between the pair of talkers studied by Fowler and Brancazio and the present talker was that, in the present data, coarticulation resistance of the consonant had a marked effect not only in the domain of the consonant but also throughout the preceding schwa.

In the present study, our speaker showed clear evidence that, throughout the schwa vowel, coarticulation resistance affected the magnitude, but not the extent, of vowel-to-vowel coarticulation. That is, the speaker showed evidence of anticipatory coarticulation from the stressed vowel throughout the schwa that was modulated throughout by the coarticulation resistance of the consonant. One of the two speakers in the study by Fowler and Brancazio showed a similar outcome but only for tongue body fronting, and only at the measurement points closest to the consonant. Although the second talker, like the others, showed anticipatory coarticulation throughout the schwa, her coarticulatory effects were not modulated by coarticulation resistance in either the fronting or the height dimension. As is typical in studies of coarticulation in speech production, the talker in the present and the earlier study show broad similarities in how they speak, but they also show individual differences. The present speaker provides the articulatory and acoustic outcomes required to ask whether, when coarticulation resistance affects the magnitude of vowel-to-vowel coarticulation, listeners use coarticulatory information to the extent that it is available to be used.

In Experiment 2, we ask how closely listeners track coarticulatory information in the schwa for the stressed vowel. The speaker of Experiment 1 provided some schwa vowels in which there is substantial acoustic evidence for the coarticulating stressed vowel, and some, those in the context of high resistant consonants, in which there is much less substantial evidence.

Because differences in F2 were largest between schwa vowels in the context of /a/ and /i/, these vowels were used for the perceptual test of Experiment 2. The magnitude of /a/–/i/ differences was also modulated by the coarticulation resistance of the intervening consonants, allowing us to ask our question whether listeners, closely tracking information about coarticulation, will extract more information about the coarticulating stressed vowel in the context of low than high resistant consonants.

3. Experiment 2

3.1. Method

3.1.1. Participants

Participants were 16 native speakers of English, who reported normal speech and hearing. Some received course credit for their participation; others were paid.

3.1.2. Stimulus materials

The stimuli for the present experiment were disyllables ending in /i/ and /a/. Generally, disyllables were the first and second tokens of our speaker's productions. In six cases in which either the first or second token contained an error or other flaw, a later token was substituted. To make cross-spliced disyllables, the first syllable of tokens 1 and 2 of a disyllable ending in /i/ was spliced onto the second syllable of a disyllable (tokens 1 and 2, respectively) having the same consonant, but ending in /a/. The splice point was the onset of closure for the consonant. For comparison, spliced disyllables were constructed by splicing the first and second tokens of disyllables ending in /i/ onto the second syllable of tokens 2 and 1, respectively, of the same disyllables. Corresponding spliced disyllables were constructed ending in /a/. In this way, within a consonant context, spliced and cross-spliced disyllables included the same schwa tokens and the same stressed syllable tokens; however, the schwas and the stressed syllables were paired differently in spliced and cross-spliced disyllables. For example, the first and second tokens of /əbi/ and /əba/ were used to make spliced and cross-spliced disyllables. Let us call the schwas from the first two tokens of /əbi/, /ə/$_{i1}$ and /ə/$_{i2}$ respectively; call the two /bi/ syllables, /bi/$_1$ and /bi/$_2$, and give the schwas and stressed syllables from the /əba/ tokens corresponding subscripts. Then the cross-spliced disyllables were /ə/$_{i1}$–/ba/$_1$, /ə/$_{i2}$–/ba/$_2$, and /ə/$_{a1}$-/bi/$_1$, /ə/$_{a2}$–/bi/$_2$. The corresponding spliced disyllables were /ə/$_{a1}$–/ba/$_2$, /ə/$_{a2}$-/ba/$_1$, and /ə/$_{i1}$-/bi/$_2$, /ə/$_{i2}$–/bi/$_1$. Because the disyllables were constructed in this way, and because both spliced and cross-spliced stimuli involved a splicing operation, any differences in responding to the two classes of stimuli can almost certainly be ascribed to the pairing of the schwas with the stressed syllables. Whereas cross-spliced disyllables had initial schwa vowels that provided misleading information for the forthcoming stressed vowels, the schwa vowels of spliced disyllables provided accurate information.

Four cross-spliced and four spliced disyllables were constructed for each of the consonantal contexts. This made 48 unique stimuli (6 consonantal contexts $\times$ 8 spliced and cross-spliced disyllables). These were presented five times each in a test order consisting of 240 test stimuli. Stimuli were randomized within each block of 48 trials. They were presented using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993). There was a 1000 ms intertrial interval following each button press response.

Procedure: Participants wore headphones and rested their left and right forefingers on the left and right buttons of the PsycScope button box. They were instructed to identify the final vowel of each disyllable as /i/ or /a/ by making the appropriate button press response as quickly and as accurately as possible.

3.2. *Results and discussion*

Errors (2.08% of responses) and extreme response times (greater than 2.5 standard deviations from each participant's mean; 2.8% of responses) were eliminated from the response time data. In a two-way repeated measures ANOVA with factors splice (cross splice, splice) and consonant, the main effect of consonant ($F(5,75) = 20.72$, $p < .0001$) and the interaction ($F(5,75) = 12.34$, $p < .0001$) were significant. The significant interaction enabled me to test the main prediction of the experiment. I predicted that the deleterious effect of cross splicing as compared to splicing would be greater in disyllables with low as compared to high resistant consonants.

In order to evaluate that prediction, a second ANOVA with factors splice (cross splice, splice) and coarticulation resistance (low, high) was performed, collapsing the consonant factor into groups by coarticulation resistance. The main effect of coarticulation resistance ($F(1,15) = 8.79$, $p < .01$) and the interaction ($F(1,15) = 37.59$, $p < .0001$) were both significant. The factor coarticulation resistance was significant because response times were slower in the context of high than low resistance consonants as shown in Fig. 3. The figure also shows that, as expected,

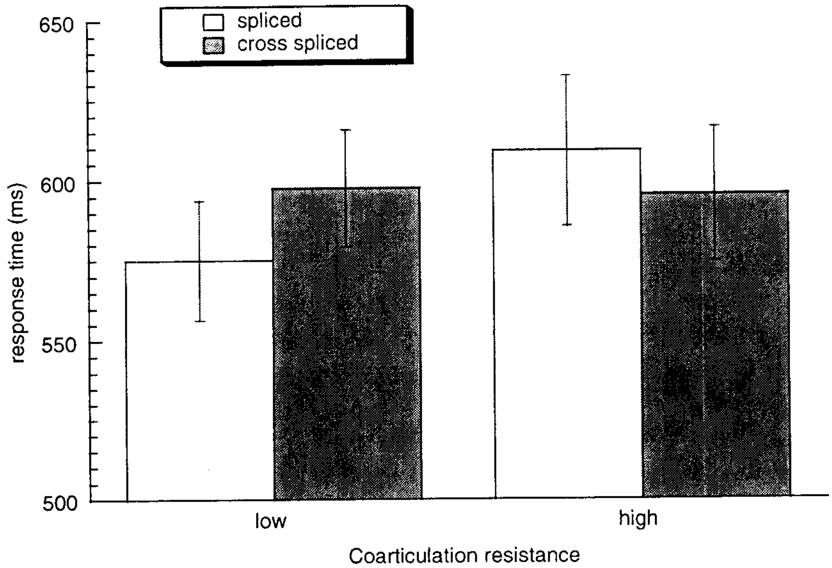

Fig. 3. The interaction of coarticulation resistance and splicing condition in Experiment 2. Response latency means and standard error bars are shown.

response times to cross spliced disyllables were slower than those to spliced disyllables in the context of the low resistance consonants ($t(15) = 5.75$, $p < .0001$). Unexpectedly, responses to vowels in the high resistant context showed a significant difference favoring cross-spliced trials ($t(15) = 3.16$, $p < .01$). An examination of the outcomes separately for the three high resistant consonants revealed that the unexpected outcome was entirely due to the consonant /z/; responses in the spliced condition in disyllables including /z/ were slower than in the cross-spliced condition for 14 of the 16 participants, and averaged 53 ms. Responses in the spliced and cross-spliced conditions were nonsignificant for the other two consonantal contexts. The difference was close to 0 for the /d/ context, and a nonsignificant 16 ms favoring the spliced condition in the /ð/ context.

The reason for the significant, but reversed, outcome for /z/ is not clear. Recall that the same tokens of schwa and the same tokens of /zi/ and /za/ contributed to the spliced and cross-spliced disyllables. Therefore, only the differential pairing of these syllables could have given rise to this outcome. Moreover, the only way that the differential pairing of schwas and stressed syllables could have given rise to a reversed outcome would be if schwa vowels having been produced originally in the context of /a/ provided better information for /i/ than for /a/ and vice versa for schwas having been produced originally in the context of /i/. An examination of the acoustic properties of the four schwas verified that that was not the case. In both tokens of schwa originally produced in the context of /a/, F1 was higher and F2 lower than in both tokens of schwa produced in the context of /i/. The original measurements of F1 and F2 in schwa had been taken at schwa mid-point. I looked at these measures throughout all four schwa vowels and verified that the mid-point measures were representative.

A different possibility is that, somehow, the splicing operation led to disyllables with more artifactual clicks or other infelicities in the spliced than in the cross-spliced disyllables. I could hear no such difference, however.

Finally, I looked to see whether one particular spliced token among the four gave rise to especially long response times or whether just one of the four cross spliced disyllables gave rise to especially short response times. However, every spliced response time, in comparison to the response time to a cross-spliced token having the same schwa vowel, was longer than the cross-spliced response time. If the comparison is made between spliced and cross-spliced disyllables sharing the stressed syllable rather than the schwa vowel, the outcome is the same; in each of the four pairwise comparisons, response times to spliced disyllables are longer than to cross-spliced disyllables.

One other outcome of the experiment was unpredicted. Although the splice-cross splice difference was significant and in the predicted direction for disyllables in the /g/ and /v/ contexts, the difference was nonsignificant and close to 0 in the /b/ context. Attempts to track down the reasons for this departure from prediction were no more revealing than attempts to track down the reason for the reversal of the expected outcome for disyllables including /z/.

Correlations between the mean magnitudes of the cross splicing effect and the differences in F1 and F2 in the corresponding schwa vowels (F1: $r = .60$; F2: $r = .62$) were in the expected direction, but, with just six paired values, were nonsignificant. Together, in a multiple regression analysis, they explained 64% of the variance in the cross splicing effect, another nonsignificant outcome, but almost two thirds of the variance nonetheless. Results were similar when articulatory variables were substituted for F1 and F2. Tongue body height and fronting at the third measurement point in schwa correlated nonsignificantly with the cross splicing effect ($r = .65; r = .52$). Together they explained 67% of the variance in the cross splicing effect ($p = .19$).

Analyses of errors revealed no significant effects.

The general outcome of the perceptual experiment was as expected. Listeners were more disrupted when schwa vowels provided misleading information for the forthcoming vowel than when the information was accurate. This effect was larger if the intervening consonants were low than if they were high in coarticulation resistance. The outcome is consistent with the finding that, in measures of F2, schwa vowels provided more evidence of coarticulation from the stressed vowel in the context of low than high resistant consonants. Responses to disyllables including four of the six consonants were consistent with the overall trend in the data. However, the outcome on one low resistant consonant, /b/, and one high resistant consonant, /z/, were not consistent with expectations. Listeners were not affected by the misleading acoustic information in schwa vowels cross spliced onto /bi/ and /ba/ syllables. The outcome in disyllables with /z/ were especially unexpected.

4. General discussion

The present research project was designed to explore the closeness of the link between information provided by coarticulation in speech and information that listeners extract from the acoustic speech signal. Like other research findings, the present pair of experiments reveals considerable sensitivity of listeners to what talkers do, but the correspondence between information made available and information extracted is not exact.

The speaker of Experiment 1 provided very clear evidence that, in her speech, coarticulation resistance of a consonant affects the magnitude of vowel-to-vowel coarticulation across the consonant. Her results provided much stronger evidence of this than did either speaker tested by Fowler and Brancazio (2000). During the schwa vowel of the disyllables in which medial consonants were low in resistance (/b/, /v/ and /g/), both tongue fronting/backing and tongue height were markedly shaped by the disyllable's stressed vowel. There was considerably less shaping in the schwas of disyllables in which medial consonants were high in resistance (/ð/, /d/ and /z/).

Like the two speakers of Fowler and Brancazio (2000), the present speaker nonetheless showed evidence of vowel-to-vowel coarticulation from schwa onset even when consonants were high in coarticulation resistance. This suggests, that speakers initiate production of a vowel at the same point, perhaps some phase of the consonant's or preceding vowel's production, regardless of the coarticulation resistance of the consonant. However, the consonant's implementation interferes with the vowel's production to an extent that varies with the consonant's degree of coarticulation resistance. If this is the case, it simplifies the task of planning for speech production by the speaker. Vowels need not be phased differently with respect to earlier segments if those segments differ in coarticulation resistance.

In Experiment 2, listeners were affected by misleading information in schwa for the forthcoming stressed vowel particularly when the consonants were the low resistant /v/ and /g/, but not when they were the low resistant /b/ or the high resistant /ð/and /d/. Results on disyllables with the high resistant /z/ were unexpected on any grounds: responses to cross-spliced disyllables were faster than to spliced disyllables.

Considerable research has shown that listeners "parse" acoustic speech signals along phonetic-gestural lines (e.g., Fowler, 1981; Martin & Bunnell, 1981; Silverman, 1987). This means that they pull apart information about multiple phonetic gestures in the acoustic consequences of coarticulated speech. This has two notable results. First, as in this experiment, and in many others, they use acoustic information about a coarticulating segment (in Experiment 2, information about the stressed vowel) that is present in domains where other segments predominate in their effects on the acoustic signal (in Experiment 2, information about the stressed vowel in the schwa vowel) as information for the coarticulating segment. Second, they do not hear the effects of coarticulation by one segment in the domain of another as context-sensitivity of that other segment (e.g., Fowler, 1981; Fowler & Smith, 1986).

For example, Silverman (1987) asked listeners to judge accent peak height in sentences with two accents, one carried by an /i/ vowel and one by an /a/. /i/ has a higher intrinsic fundamental frequency (F0) than /a/. That is, F0 during a vowel has at least two sources of influence, the intonation contour and the intrinsic F0 associated with the vowel. Accordingly, ceteris paribus, F0 on equally accented /i/ and /a/ vowels will not be the same, but will be higher on /i/. Silverman found that listeners judged an accented /i/ vowel to have a lower accent than an accented /a/ vowel if F0 on the two vowels was the same. That is, they "parsed" ostensible effects of intrinsic F0 from the F0 contour before judging intonational peak height. Reinholt-Peterson (1986) showed that listeners use F0 as information for vowel height; along an /o/ to /u/ continuum, listeners reported more /u/ vowels (with higher intrinsic F0 than /o/) for vowels with higher F0 s. Together, these studies index the two signatures of parsing along gestural lines. Listeners do not hear effects of intrinsic F0 as part of an intonation contour; they use the effects, rather, as information for vowel height.

However, an attempt to compare quantitatively the intrinsic F0 difference in speech with the amount parsed by listeners failed to find a close match. Fowler and Brown (1997) had speakers produce words, such as *keyed* and *cod* or *beady* and *body* that were matched in phonetic context but differed in whether the stressed vowel was /i/ or /a/. On average, talkers showed a 13 Hz difference in F0 in the words, with F0 on /i/ being higher than on /a/. Using resynthesis, Fowler and Brown modified the speech of one of their talkers to create versions of *keyed* and *cod* that varied in F0. Listeners heard a word pair and judged which word had the higher pitch. They predicted that, when *keyed* and *cod* were matched in F0, *cod* would have the higher pitch, and that is what was found. However, when they estimated the F0 difference between the words that would lead the words to be judged to have the same pitch, the difference was just one-tenth of the 13 Hz difference that characterized their talkers. In the same study, they obtained other measures of intrinsic F0, now of sung vowels. In one task, participants sang /i/ for a few seconds then shifted to /a/ keeping the pitch the same, or they produced the vowels in the opposite order. In another task, they sang either /i/ or /a/ attempting to match the pitch of vowels to a tone. These tasks yielded differences in intrinsic F0 as had the speech task. Attempts to compare the magnitude of listeners' parsing of intrinsic F0 to the produced magnitude with these sung vowels yielded an estimate that listeners were accurate in one experiment, but that they pulled out too much in a second. Accordingly, across the experiments, it appeared that listeners pulled out too little F0 from spoken vowels, pulled out just the right amount or pulled out too much from sung vowels.

Presumably, in the study of Fowler and Brown (1997), one cause of the mismatch between the magnitude of intrinsic F0 in speech and sung vowels vs. the magnitude of F0 parsed from those productions must lie in the methods used to estimate parsing. The behavioral measures will not be perfectly precise. Indeed, if parsing were very inaccurate as suggested by some of the findings of Fowler and Brown, it is not obvious why listeners would do it. However, it remains the case that neither the present findings, nor those of Fowler and Brown, suggest marked accuracy of parsing.

Acknowledgments

The research was supported by NIDCD grant DC-02717 to Haskins Laboratories. I thank Douglas Honorof and Jeffrey Weihing for help with data collection in Experiment 1, Marianne Pouplier and Mark Tiede for help with data analysis in Experiment 1 and Laura Sabadini for making the stimuli for Experiment 2.

Appendix A

Analyses of variance: tongue fronting during schwa

S1	
Consonant	$F(5,504) = 320.62$, $p < .0001$
Vowel	$F(5,504) = 98.06$, $p < .0001$
Consonant × vowel	$F(25,504) = 3.87$, $p < .0001$

S2	
Consonant	$F(5,504) = 528.55$, $p < .0001$
Vowel	$F(5,504) = 155.62$, $p < .0001$
Consonant × vowel	$F(25,504) = 8.00$, $p < .0001$
S3	
Consonant	$F(5,504) = 687.91$, $p < .0001$
Vowel	$F(5,504) = 212.14$, $p < .0001$
Consonant × vowel	$F(25,504) = 14.55$, $p < .0001$
S4	
Consonant	$F(5,504) = 766.17$, $p < .0001$
Vowel	$F(5,504) = 266.54$, $p < .0001$
Consonant × vowel	$F(25,504) = 22.28$, $p < .0001$

Analyses of tongue height during schwa

S1	
Consonant	$F(5,504) = 359.55$, $p < .0001$
Vowel	$F(5,504) = 108.28$, $p < .0001$
Consonant × vowel	$F(25,504) = 4.48$, $p < .0001$
S2	
Consonant	$F(5,504) = 640.06$, $p < .0001$
Vowel	$F(5,504) = 174.09$, $p < .0001$
Consonant × vowel	$F(25,504) = 8.98$, $p < .0001$
S3	
Consonant	$F(5,504) = 803.36$, $p < .0001$
Vowel	$F(5,504) = 239.45$, $p < .0001$
Consonant × vowel	$F(25,504) = 15.87$, $p < .0001$
S4	
Consonant	$F(5,504) = 802.71$, $p < .0001$
Vowel	$F(5,504) = 295.82$, $p < .0001$
Consonant × vowel	$F(25,504) = 23.74$, $p < .001$

References

Bladon, A., & Al-Bamerni, A. (1976). Coarticulation resistance in English /l/. *Journal of Phonetics*, *4*, 137–150.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods and Instrumentation*, *25*, 257–271.

Dembowski, J., Lindstrom, M., & Westbury, J. (1998). Articulator point variability in the production of stop consonants. In M. P. Cannito, K. M. Yorkston, & D. R. Beukelman (Eds.), *Neuromotor speech disorders* (pp. 27–46). Baltimore: Paul H. Brookes Publishing Co.

Fowler, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research*, *46*, 127–149.

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

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*, 1–41.

Fowler, C. A., & Brown, J. (1997). Intrinsic f0 differences in spoken and sung vowels and their perception by listeners. *Perception & Psychophysics*, *59*, 729–738.

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.

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

Martin, J., & Bunnell, T. (1982). Perception of coarticulation effects in vowel-stop syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 473–488.

Perkell, J., Cohen, M. H., Svirsky, M. A., Matthies, M., Garabieta, I., & Jackson, M. T. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America*, *92*, 3078–3096.

Recasens, D. (1984a). V-to-C coarticulation in Catalan VCV sequences: an articulatory and acoustical study. *Journal of Phonetics*, *12*, 61–73.

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

Recasens, D. (1985). Coarticulatory patterns and degrees of coarticulation resistance in Catalan CV sequences. *Language and Speech*, *28*, 97–114.

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

Reinholt-Peterson, N. (1986). Perceptual compensation for segmentally-conditioned fundamental-frequency perturbations. *Phonetica*, *43*, 31–42.

Silverman, K. (1987). *The structure and processing of fundamental frequency contours*. Ph.D. dissertation, Cambridge University.

Whalen, D. H. (1984). Subcategorical mismatches slow phonetic judgments. *Perception & Psychophysics*, *35*, 49–64.