337

PRODUCTION AND PERCEPTION OF COARTICULATION AMONG STRESSED AND UNSTRESSED VOWELS

CAROL A. FOWLER
Dartmouth College, Hanover, New Hampshire

A pair of experiments examines first the coarticulatory relations among certain stressed and unstressed vowels, and next the perception of coarticulated unstressed vowels. The first study finds the acoustic properties of unstressed medial /ə/ and, to a substantially lesser extent, of stressed medial /ʌ/, to be assimilated to the properties of their flanking vocalic contexts. Both initial and final flanking vowels coarticulate with medial /ʌ/, but carryover coarticulatory effects tend to exceed anticipatory effects. In a second experiment, listeners' manners of perceiving the coarticulated unstressed vowels of the first experiment are shown to be coupled to, or to be compatible with, the talkers' coarticulatory strategies. In particular, perceivers hear acoustically identical vowels to be different when the vowels appear in different contexts of flanking vowels. Similarly, instances of /ə/ that are acoustically different due to different coarticulatory influences on them sound the same to listeners as long as each appears in its appropriate context of flanking vowels.

Patterns of coarticulatory influence among the segments of an utterance are indicators of a talker's strategies of articulatory planning or control. As such, they are quite interesting in themselves, but also are intriguing in respect to the complications they seem to offer a perceiver who tries to identify the phonetic categories intended by the talker. The present experiments examine coarticulatory influences among vowels that span a consonant. Both purport to discover something about the talker's strategies for producing multisyllabic stretches of speech, and also to learn how the perceiver handles the acoustic variability to which the strategies give rise.

For the first aim, the possibility is explored in a preliminary way that two largely separate realms of observation in the speech literature may be integrated, and in so doing, that the talker's organizational strategies for producing syllables and longer sequences of speech may be clarified. The "separate realms" are the literature on coarticulation, and that on compensatory shortening.

In respect to the second aim, the consequences for a perceiver of certain of the talker's coarticulatory strategies are examined. Due to coarticulation, the spectral and durational properties of phonetic segments are strongly sensitive to context. Under special experimental conditions (for example, excising a coarticulated segment from its context) the acoustic effects of coarticulation can be shown to be readily audible to perceivers (e.g., Kuehn & Moll, 1972; Ohde & Sharf, 1977). Despite this, perceivers are able to identify the phonetic segments intended by talkers. Whatever their perceptual strategy may be, it enables and even compels them to hear acoustically identical segments as different just when the context specifies different categories of phonetic segments as produced by the talker (Liberman, Delattre, & Cooper, 1952; Schatz, 1954). Similarly, it leads them to identify different acoustic signals as the same segment just when the context specifies articulatory gestures belonging to the same phonetic category (Liberman, Delattre, Cooper, & Gerstman, 1954). Several studies concerning the perception of coarticulated speech confirm that perceivers reliably identify a coarticulated vowel or consonant based on the acoustic signal within the time frame of the segment itself, but evaluated relative to its coarticulatory context (Liberman et al., 1954; Ochiai & Fujimura, 1971).

In addition, the studies show that perceivers can use information within the time frame of a coarticulated segment to identify the neighboring segments of the coarticulation (e.g., Kuehn & Moll, 1972; Ohde & Sharf, 1977). These latter findings reveal a high degree of sensitivity on the part of perceivers to information in the acoustic speech signal that specifies the talker's coarticulatory strategies.

The following series of studies further examine the coupling between a talker's patterns of coarticulatory influence and the listener's perception of a coarticulated sequence. In contrast to most other studies of coarticulation and its perception, the present studies examine coarticulatory relations among non-adjacent segments—in particular, among vowels that span a consonant. The separate rationales for the production and perception studies are given further consideration, respectively, in the introductions to Parts 1 and 2 of the paper.

Carol A. Fowler, Ph.D., is currently an Assistant Professor at Dartmouth College, and a Research Associate with Haskins College, in Hanover, New Hampshire.

0022-4685/81/2401-0127$1.00/0

PART 1: THE PRODUCTION OF STRESSED AND UNSTRESSED SYLLABLES OF A WORD

The production study was motivated in part by the observation of an apparent symmetry between certain coarticulatory influences among segments on the one hand, and so-called "compensatory shortening" on the other. As to coarticulation among the segments of a syllable, the particular interest here are coarticulatory relations between the primary articulators of a vowel—that is, the tongue body and jaw—and gestures for neighboring consonants. Articulatory gestures of the tongue body for a vowel can influence the production (e.g., Gay, 1977; Sussman, MacNeilage, & Hanson, 1973) and acoustic signal (Ohman, 1966), both of preceding and of following consonants. Similarly, a consonant preceding or following a vowel shortens the vowel's measured acoustic duration (e.g., Lindblom & Rapp, 1973).

The bidirectional symmetry of these two sets of findings—one relating to coarticulation and one to shortening—may signify that the two are different measures of the same articulatory phenomenon, namely partially concurrent production ("coproduction") of vowels and neighboring consonants. On that interpretation, vowels exert an influence on the portion of the acoustic signal conventionally assigned to the consonant in a *CV* or *VC* because it is produced partially in concurrence. For its part, the consonant may be said to be superimposed, in whole or in part, on the production of flanking vowels (cf. Carney & Moll, 1971; Kent & Moll, 1972; Ohman, 1966; Perkell, 1969). All else equal (in particular, in the absence of fully compensatory lengthening of the vowel), coproduction of a vowel with a consonant will lead to a measured shortening of the vowel relative to the vowel's duration when produced in isolation. By hypothesis, the vowel is measured to be shorter not necessarily because it is shorter in any articulatory sense, but because most of the durational extent over which it is coproduced with a consonant is conventionally assigned only to the consonant when measurements are made. On this view, compensatory shortening is bidirectional simply because coarticulatory influences of vowels on consonants extend in both directions. This is illustrated in Figure 1.

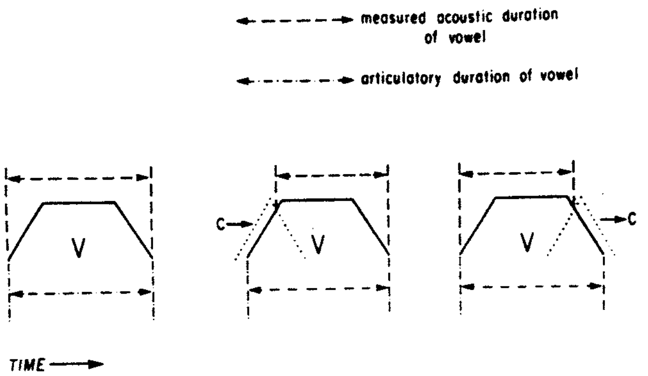

FIGURE 1. Schematic representation of coproduction between vowels and consonants leading to measured shortening of the vowel.

A systematic investigation of the proposal that shortening and coproduction are different measures of the same phenomenon of coproduction would require measures of coarticulation and shortening to be obtained on the same utterances so that the extents of their effects could be correlated. This kind of investigation, though conceptually straightforward, is methodologically very difficult. (This is due in part to the problem of finding measures of the two effects that are approximately equally sensitive to small effects.) Here a different and substantially more convenient test is devised as a first step. The test derives from the observation that shortening effects of contextual units on stressed vowels are tiered. Stressed vowels are shortened not only by preceding and following consonants, but also by preceding and following unstressed syllables (Fowler, 1977; Huggins, 1975; Lehiste, 1972; Lindblom & Rapp, 1973). Moreover, the directional asymmetry of the two levels of shortening effects—segmental and unstressed syllabic—is the same; that is, anticipatory shortening exceeds backward shortening (Fowler, 1977; Lindblom & Rapp, 1973). Clearly the shortening effects due to segments and unstressed syllables are strictly analogous and require analogous explanations. If the explanation of segmental shortening invokes coproduction, then so should the explanation of shortening due to neighboring unstressed syllables. That is, the shortening of a stressed vowel or syllable by neighboring unstressed syllables must be ascribed to something like a superimposition of the production of the unstressed vowels on the stressed vowel (cf. Martin, 1972).

This proposal is intriguing in suggesting that talkers organize the production of relatively long stretches of speech—that is, the intervals between stressed syllables within a phrase—more-or-less as they organize the production of syllables. Their strategy, perhaps, is to subsume the production of segments other than stressed vowels within the domain of production of a stressed vowel. Indeed, this subsumption strategy may account for the linguists' (e.g., Abercrombie, 1964; Pike, 1945) and naive listeners' (Donovan & Darwin, 1979; Lehiste, 1973) impression that English is stress-timed.

If indeed both coarticulation and shortening are due to coproduction, then predictions may be devised based on the reported shortening effects about coarticulation of stressed and unstressed vowels. These predictions are tested in Experiment 1. In particular, neighboring unstressed vowels should be affected acoustically by their stressed vocalic context. Moreover, unstressed vowels following a stressed vowel should be affected more than unstressed vowels preceding a stressed vowel. That is, carryover coarticulatory effects should exceed anticipatory effects.

The literature offers only a small amount of data. Bell-Berti and Harris (1979) looked at acoustic correlates of coarticulation in utterances of the form /pə*CVC*əp/ in which *C* was /p/, /t/, or /k/ and *V* was stressed /i/, /a/, or /u/.

The investigators looked for evidence of anticipatory and carryover coarticulation in the stressed vowel due to the flanking consonants, and in the unstressed schwas due to the consonants or the stressed vowel. In respect to vowel-to-vowel effects, they found no anticipatory influences, but noticeable carryover effects. As noted earlier, stressed vowels in a *CVC* are reported to affect preceding consonants at least when EMG or X-ray measures are taken (MacNeilage & DeClerk, 1969). Thus this implies that anticipatory effects of the stressed vowel are present but are less extensive than carryover effects, and that only carryover effects extend to a transconsonantal unstressed vowel.

Bell-Berti and Harris (1979) also report EMG data on the genioglossus muscle in utterances of the form /əp*VCVC*pə/ in which *C* is /p/ or /k/ and *V* is /i/, /a/, or /u/. The genioglossus is active for the vowels /i/ and /u/. In these utterances, either the first or the second *V* was stressed. The EMG signals showed evidence of anticipatory effects on only 25% of the utterances in which it could be assessed and evidence of carryover effects in 75% of the cases. These findings support the proposed relationship between coarticulation and shortening in showing the expected asymmetry between anticipatory and carryover effects. However, the anticipatory effects are nearly absent. Moreover, data were obtained only from a single talker.

Experiment 1 is designed to uncover any anticipatory and carryover coarticulatory effects of stressed vowels on a neighboring unstressed vowel. A failure to find these effects, or a failure to find stronger carryover than anticipatory effects would tend to falsify the proposal that stressed and unstressed vowels are, in any sense, coproduced. Positive outcomes along these lines would be compatible with this view in showing that measures of coarticulation mirror reported measures of shortening.

By way of contrast, the coarticulatory effects of flanking unstressed vowels on a medial stressed vowel are also examined. Shortening of unstressed vowels due to neighboring stressed vowels has not been examined systematically in the literature, in part, perhaps because it is not expected on the assumption that English is a stress-timed language (e.g., Abercrombie, 1964). Therefore, clear predictions regarding their coarticulatory influence cannot be devised with any confidence from evidence relating to shortening. However, at least one consideration, other than the stress timing just mentioned, suggests that unstressed vowels should shorten very little in any context and hence that their coarticulatory influence on contextual segments should be small. The literature indicates that unstressed vowels are quite short in any context (see, e.g., Klatt, 1975) and probably approach their compression limit (see Klatt, 1976). Thus, the addition of a stressed vowel before or after an unstressed vowel should shorten it very little. If shortening reflects coproduction, it should follow, then, that an unstressed vowel should exert little influence on the acoustic properties of neighboring stressed vowels. There is some evidence that this is the case. Nord (1974) measured F1 and F2 of stressed and unstressed /a/ and /e/ spoken in sets of Swedish words in which the critical vowels appeared in an initial or final syllable in the word. He reports strong transconsonantal influences on F2 of the unstressed vowels and suggests that unstressed reduced vowels are more compliant to coarticulatory influence than are stressed vowels.

Experiment 1 tests the proposals of the two preceding paragraphs. Its design replicates a pilot investigation reported in Fowler (1977).

EXPERIMENT 1

Method

Stimuli. Two lists of trisyllables of the form $V_1bV_2bV_3$ were devised. The initial and final vowels were /i/, /a/, or /u/, and the medial vowel was /ʌ/. All combinations of initial and final vowels were included in the lists to make nine different utterances in each list. The two lists differed in stress pattern. In the first list, V_1 and V_3 were stressed, while the medial vowel, /ʌ/, was unstressed (as in "misinform"). In the second list, the stress pattern was reversed (as in "deliver"). The order of stimuli within each list was random, and the same random ordering was used for each list.

The central vowel /ʌ/ was selected as the vowel on which measures of coarticulation were to be made. It was not selected to be representative of the class of vowels; in contrast, its special property of being a central lax vowel, presumably with substantial room to show coarticulatory effects, led to its selection (cf. Bell-Berti & Harris, 1979). It is possible, but not necessarily the case, that other vowels show coarticulatory effects compatible in pattern to those exhibited by /ʌ/; the effects on these vowels might well be substantially smaller than those on /ʌ/, however, due, for example, to their closer proximity to articulatory floors and ceilings. The point vowels, /i/, /a/, and /u/ were selected as the flanking vowels because their production at the extremes of the oral cavity should maximize their coarticulatory effects on /ʌ/. According to the values reported by Peterson and Barney (1952), /i/ has a substantially higher F2 value than /ʌ/ (for males, 2290 Hz for /i/ and 1190 for /ʌ/) while /a/ and /u/ have somewhat lower values (1090 and 870, respectively). Therefore, a coarticulatory effect of /i/ on /ʌ/ should raise F2, while an effect of /a/ and /u/ should lower F2 for /ʌ/. Although choosing only these vowels limits the ability to generalize these findings, it may maximize the chance that whatever coarticulatory effects do occur are large enough to be measured from a spectrographic display. Thus, the experiment asks: What is the pattern of coarticulatory effects that may be observed under conditions designed to maximize their occurrence?

Procedure. Phonemic transcriptions of each trisyllable were printed with a black felt tip pen, one trisyllable to a 3×5-inch file card. Tested individually, each subject was asked to read both lists of trisyllables. The order of list presentation was partially counterbalanced across the five subjects. For each list, subjects were instructed as to the pronunciation of the phonemic symbols and the appropriate stress patterning of the trisyllables. They were given practice trials on the list until their productions were fluent. During the test, subjects read through the list five times. Their productions were recorded on audiotape and sound spectrograms were made of the middle three repetitions of each trisyllable. If an error of pronunciation occurred on any of these three repetitions, the fifth repetition was substituted for the error. Errors occurred on just 9 of the 270 critical utterances.

Measurements. The measures on each utterance were the center frequencies of the first and second formants of the medial vowel /ʌ/ and of the flanking vowels, /i/, /a/, and /u/ as measured from a wide-band spectrographic display. The short duration of the medial unstressed vowels precluded measurements of other than the center frequency at the middle of the vowels' durational extents.

To check for reliability, each measurement was made twice, separated by an interval of several months. The sets of measurements were close enough that none of the analyses (reports to follow) gave significantly different outcomes depending on which set of measurements was used. Here the second set of measurements is reported.

On the spectrographic displays, each millimeter represents 55 Hz. Measures were rounded to the nearest millimeter giving a (conservative) estimate of measurement error of 28 Hz. It is worth pointing out that this or any other estimate of error on a single measurement is not in itself an estimate of the size of coarticulatory effects that can be detected. That is, it does not imply that coarticulatory effects smaller than 30 Hz are to be discounted as possibly due to measurement error. So long as measurement errors are unbiased as to different levels of the independent variables, their only effect will be to increase the variability around the mean value at each level of the independent variable—that is, measurement error will decrease the sensitivity of analyses to any coarticulatory effects in the data. However, the errors cannot create spurious effects unless they are correlated with levels of the independent variable.

Subjects. Subjects were five undergraduates (three men, two women) at Dartmouth College who participated in the experiment for course credit.

Results and Discussion

For each subject, 108 measurements were made on the productions of the medial vowels of lists 1 and 2 (3 tokens × 9 different utterances × 2 stress patterns × 2 formants). The coarticulatory effects were assessed separately for $F1$ and $F2$ by two three-way analyses of variance, both with these factors: Initial Flanking Vowel (i,a,u), Final Flanking Vowel (i,a,u), and Stress of medial vowel. Tables 1 and 2 show the means compared in these analyses.

Table 1 gives the values of $F1$ for stressed (a) and unstressed (b) medial /ʌ/ in the context of initial (first column) or final (second column) /i/, /a/, /u/. Each mean in the first column averages across the five subjects and the three final-vocalic contexts. Thus, the value 628 Hz in the first column of Table 1 averages $F1$ for /ʌ/ in /ibʌbi/, /ibʌba/, and /ibʌbu/ across the five talkers; 640 Hz in column two averages $F1$ across the five talkers for /ibʌba/, /abʌba/, and /ubʌba/. The values in parentheses are the values of $F1$ of the contextual vowel. Thus, 401 Hz in column one is $F1$ for initial stressed /i/ averaged across the five talkers and across the trisyllables, /ibʌbi/, /ibʌba/, and /ibʌbu/; 662 Hz in column two is $F1$ for final stressed /a/ averaged across the talkers and across the trisyllables, /ibʌba/, /abʌba/, and /ubʌba/. Table 2 shows the comparable measures of $F2$. In these Tables, coarticulatory effects may be gauged by comparing $F1$ or $F2$ for /ʌ/ in different contexts. For example, because $F1$ for /i/ and /u/ are lower than that for /ʌ/, and $F1$ for /a/ is higher, any coarticulatory effects of the flanking vowels on /ʌ/ would lower its $F1$ in the context of /i/ or /u/ and would raise it in the context of /a/. However, on this measure, neither the initial flanking vowel (carryover coarticulation) nor the final flanking vowel (anticipatory coarticulation) had reliable effects on medial /ʌ/. Tables 1a and b confirm the analyses in showing little effect of the flanking vowel

TABLE 1. Mean $F1$ (Hz) of stressed (a) and unstressed (b) /ʌ/ in the context of flanking vowels /i/, /a/, and /u/. Mean $F1$ of flanking unstressed (a) and stressed (b) vowels are given in parenthesis. See text for clarification.

a) stressed medial /ʌ/	*Initial*	*Final*
/i/	628 (401)	643 (444)
/a/	649 (729)	640 (662)
/u/	643 (419)	637 (467)
b) unstressed medial /ʌ/	*Initial*	*Final*
/i/	468 (416)	466 (402)
/a/	485 (879)	479 (818)
/u/	459 (419)	466 (427)

TABLE 2. Mean $F2$ (Hz) of stressed (a) and unstressed (b) /ʌ/ in the context of flanking vowels /i/, /a/, and /u/. Mean $F2$ of flanking unstressed (a) and stressed (b) vowels are given in parenthesis. See text for clarification.

a) stressed medial /ʌ/	*Initial*	*Final*
/i/	1468 (2519)	1465 (2381)
/a/	1416 (1491)	1432 (1456)
/u/	1449 (1123)	1436 (1236)
b) unstressed medial /ʌ/	*Initial*	*Final*
/i/	1435 (2521)	1395 (2405)
/a/	1350 (1577)	1329 (1521)
/u/	1290 (1146)	1352 (1279)

on either the stressed or the unstressed medial vowel. In contrast, stress of medial vowel had a striking effect, with the stressed medial vowel having a substantially higher first formant than the unstressed vowel ($F[1,4] = 29.18$, $p = 0.007$). We return to this finding shortly. No interactions among any of the factors approached significance.

On $F2$, a coarticulatory effect of /i/ would raise $F2$ of /ʌ/, while /a/ and /u/ would lower it. On this measure, both initial and final flanking vowels influenced medial /ʌ/ in the expected directions ($F[2,8] = 6.20$, $p = 0.02$ and $F[2,8] = 4.53$, $p = 0.05$, respectively). The effect of stress on $F2$ of /ʌ/ was also significant ($F[1,4] = 11.09$, $p = 0.03$), $F2$ having higher average values for stressed (1444 Hz) than unstressed (1358 Hz) /ʌ/. Finally, the interaction of initial vowel and stress was significant ($F[2,8] = 5.81$, $p = 0.03$) due to the substantially larger carryover effects on unstressed than on stressed /ʌ/.

For both stressed and unstressed medial /ʌ/, carryover effects on $F2$ were larger than anticipatory effects, although for stressed /ʌ/, the difference was negligible. Stressed /ʌ/ in trisyllables with initial /i/ had an average $F2$ of 1468; with an initial /a/ or /u/, $F2$ for /ʌ/ averaged 1433. The difference between these values, which we will call the "carryover range," is 35 Hz. $F2$ of stressed /ʌ/ in trisyllables ending in /i/ averaged 1465 Hz; for trisyllables ending in /a/ or /u/, it averaged 1434 Hz giving an "anticipatory range" of 31 Hz. Just three of the five subjects showed this direction of difference between carryover and anticipatory coarticulation. The comparable ranges for carryover and anticipatory coarticulation on unstressed /ʌ/ were 115 Hz and 55 Hz, respectively. Four of the five subjects displayed this pattern. As these figures indicate in addition, the coarticulatory effects of stressed flanking vowels on unstressed /ʌ/ were larger than the effects of unstressed flanking vowels on stressed /ʌ/. This effect, computed on the measure of coarticulatory range is highly significant ($F[1,4] = 73.49$, $p = 0.002$). Figure 2 displays the carryover and anticipatory effects separately for stressed and unstressed medial /ʌ/.

One of the strongest effects in the current data is the effect of stress on the formant values for the vowels—that is, the "reduction" of unstressed relative to corresponding stressed vowels (cf. Lindblom, 1963; Nord, 1974; Tiffany, 1959). A simple explanation for this effect in the present data attributes it to the smaller lip opening for /ʌ/ in its unstressed environment. The lip opening is expected to be smaller in this environment because of the consonantal frame of labial consonants and the shorter duration of the unstressed than the stressed vowel.[1]

To summarize the findings of primary interest in this study: Compatible with the literature on compensatory shortening, the present study found coarticulatory effects of a stressed vowel on a preceding or following unstressed transconsonantal vowel. The additional finding that a preceding stressed vowel exerts a more powerful influence than a following stressed vowel (at least for four of five subjects) matches the directional asymmetry both of segmental and of syllabic shortening effects (Lindblom & Rapp, 1973).

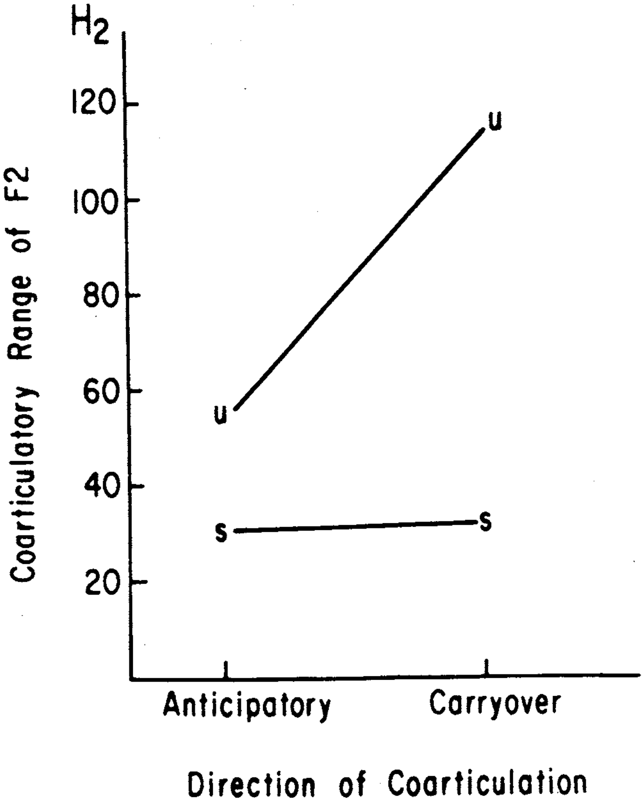

FIGURE 2. Anticipatory and carryover coarticulatory effects on unstressed (u) and stressed (s) /ʌ/. The measure of "coarticulatory range" is obtained by subtracting the mean $F2$ for /ʌ/ in the context of flanking /a/ or /u/ from that for /ʌ/ in the context of flanking /i/.

These findings are compatible with the evidence on coarticulation and shortening at the segmental level, which suggests that the two effects, coarticulation and shortening, are symmetric. However, to demonstrate that coarticulation and shortening, both at the level of the segments in a syllable and at the level of stressed and unstressed syllables, are symmetric is not to confirm that they are in fact two different measures of the same phenomenon as proposed in the introduction. That demonstration requires that measures of coarticulation and shortening be obtained on the same utterances and that the magnitudes and extents of each be correlated. For reasons given in the introduction to Part 1, such a study is methodologically difficult; hence, it is not attempted here. We should point out that despite the difficulties involved, the study is worth pursuing for the information it may offer concerning the talker's strategy of organizing stretches of speech larger than the syllable and shorter than the phrase.

Rather than pursue this line of inquiry, however, Experiments 2a and 2b carry forward the question of the nature and extent of the coupling between the coar-

---

[1] I thank an anonymous reviewer for this account.

ticulatory patterns produced by talkers and the hearers' perceptual strategies. In particular, Experiments 2a and 2b examine the implications for strategies of perceiving vowels of the coarticulatory effects found in Experiment 1.

PART 2: THE PERCEPTION OF COARTICULATED VOWELS

At least two considerations suggest that the effects found in Experiment 1 are unlikely to be irrelevant to the perceiver's strategy for identifying vowels in running speech. First, as noted earlier, many acoustic effects of coarticulation are found to be readily perceptible even when insensitive measures of perceptibility are used. Thus, hearing a coarticulated segment or a sequence of coarticulated segments isolated from its context, a listener can often supply the missing context (see, e.g., Kuehn & Moll, 1972; Lehiste & Shockey, 1972; Ohde & Sharf, 1977; Sharf & Beiter, 1974; Sharf & Hemeyer, 1972). If the vowel-to-vowel effects discovered in Experiment 1 are perceptible, then the listener receives what to him may be different stimuli signaling the same vowel in different contexts, and his strategy for identifying vowels must be able to handle that variability.

Second, one effect of coarticulation is to increase the potential confusability among segments—that is, it generates some overlap in the possible acoustic realizations of different segments across their different contexts. For example, the effect of flanking /i/ on /ʌ/ is to give /ʌ/ the acoustic characteristics of a higher vowel than its characteristics when produced in isolation or between flanking /a/s. Thus if context is not considered, the acoustic properties of /ʌ/ produced between flanking /i/s may be confused with those of a higher vowel it resembles. Consequently, certain strategies of segment perception would seem to be precluded and others promoted for a listener who is able to identify segments as the ones intended by talkers. Precluded are styles of perceiving that identify a vowel based on its acoustic properties considered independently of its surrounding context. Promoted, then, are styles of perception in which a segment is identified relative to its context. The relevant context would be the coproduced segments.

To date, studies of the perception of coarticulated vowels have not examined the perceptibility of vowel-to-vowel coarticulatory effects or their influence on the ability to identify vowels. However, several studies have looked at the influence of the consonantal context on vowel identification (e.g., Kuehn & Moll, 1972; Lehiste & Shockey, 1972; Ochiai & Fujimura, 1971; Ohde & Sharf, 1977). For present purposes the most relevant studies are by Lehiste and Shockey (1972) and by Ochiai and Fujimura (1971).

Lehiste and Shockey (1972) partitioned naturally-produced *VCV* disyllables into their component initial and final syllables. The consonants were the voiceless stops, and the segmentation was made in the consonantal closure interval. This left the initial vowel and the implosive vowel-consonant transitions in one syllable, and the explosive transitions along with the final vowel in the second syllable. Based on Ohman's (1966) findings that the implosive (and explosive) transitions of a *VCV* utterance are influenced by both vowels, Lehiste and Shockey assumed that the implosive transitions in their utterances also carried some information about the identity of the final vowel and that the explosive transitions carried information about the initial vowel.

Their experimental procedure was to present the excised monosyllables to listeners and to ask them to supply the missing vowel. Listeners' performance in supplying the initial or final vowels was at chance. Lehiste and Shockey conclude from this that any coarticulatory effects that a final vowel may have on implosive transitions or that an initial vowel may have on explosive transitions are not perceptible to listeners.

If Lehiste and Shockey's (1972) conclusion holds generally, and perceptible coarticulatory influences of vowels do not extend even throughout an adjacent consonant, then it is unlikely that any coarticulatory effects between transconsonantal vowels will be perceptible. However, several considerations suggest that this conclusion is too strong, or at least is premature.

Most important, the technique used in the experiment is rather insensitive in asking a talker to supply a missing segment based on syllables excised from their original context. Quite possibly, information that can be used in its proper context may not be interpretable in the absence of its contextual support. Because the technique is insensitive, negative findings are difficult to evaluate. (A more sensitive technique might compare the identifiability of the presented vowel and consonant as a function of the identity of the excised segment. For example, compare the identifiability of /ap/ excised from /apa/ and /api/. If the /i/ has effected perceptible acoustic changes in /ap/, then they may impair the ability to identify /ap/ with the /i/ removed.)

Besides the procedural insensitivity of the study, the disyllables used as stimuli are not optimally selected to address the present concern—namely, the perceptibility of coarticulatory effects among vowels—nor, perhaps, even to address the investigators' own concern—namely, the perceptibility of the coarticulatory effects found by Ohman (1966). In the disyllables used, both syllables are stressed. Possibly relations of shortening and of coarticulation are small between stressed vowels relative to those between stressed and unstressed vowels. Moreover, in contrast to Ohman, Lehiste and Shockey (1972) used voiceless stops, which may make a difference both in the perceptual salience of any coarticulatory influence of a vowel on transitions into or out of a consonant, and also, possibly in the occurrence of any such coarticulatory effects. Gay (1977) reports that in *VCV* utterances with voiceless medial stops, and both vowels stressed, movements of the tongue body toward the final vowel cannot be detected until consonantal closure is achieved. If his results are representative, this would preclude any coarticulatory effect of a final stressed vowel on implosive transitions from an initial stressed vowel to a voiceless stop. Hence, there may have been little or no coarticulatory effect to be perceived in the stimuli of Lehiste and Shockey.

A second technique for examining the influence of coarticulatory effects on vowel identification is reported by Ochiai and Fujimura (1971). These investigators extracted 50-msec intervals of vocalic segments that had been produced in real (Japanese) words; the listeners, native speakers of Japanese, were asked to identify the vowels. Thus, in contrast to the more usual procedure, listeners were not asked to supply the missing context, but instead were asked to identify the segments that were presented.

When a segment is extracted from its coarticulatory context, presumably a listener has difficulty recognizing coarticulatory influences as coarticulatory influences. Consider, for example, a 50-msec section of a vowel with formants that would be characteristic of a higher-than-central vowel (say /ɛ/), were the vowel to be produced in isolation. Given a signal like this, a listener may not know whether the presented vowel is a central vowel whose tongue position had been raised due to its missing context, a high vowel whose tongue position had been lowered due to its context, or in fact a vowel whose tongue position is unchanged over its characteristic position when produced in isolation. In the absence of information about the coarticulatory context that would help a listener distinguish among these three possibilities, presumably that listener will assume the last of them, as it approximates a null hypothesis. If so, the errors of identification would be predictable based on coarticulatory context, and that, indeed, is what Ochiai and Fujimura report.

Subjects exceeded chance performance by a substantial margin. This indicates that despite coarticulatory influences, vowels can be identified with some accuracy based on their acoustic signal considered independently of its context. However, substantial numbers of errors were made, especially by naive listeners, and they often patterned according to the consonantal context from which the vowel had been extracted. The effects were both anticipatory and perseverative in direction. Examples are /e/ misidentified as /i/ in /aeta/. Here the raised tongue tip for the /t/ may have changed the acoustic properties of the /e/ so that they tended to resemble those of isolated /i/. Similarly, /u/ was misidentified as /i/ in /yuyusii/. This study indicates, then, that coarticulatory effects of neighboring consonants on vowels are perceived.

The following pair of experiments was designed to examine the influences on unstressed vowel perception of the coarticulatory effects observed in Experiment 1. To maximize the sensitivity of the experiment to any such influences that may obtain, a 4IAX discrimination task was used instead of an identification task (Pisoni, 1971). In brief, the studies attempted to ask whether a pair of unstressed medial syllables produced between flanking stressed vowels (as in Experiment 1) sounds more alike or less alike to a listener when they are acoustically identical, but in different coarticulatory contexts, than when they are acoustically different but each shows coarticulatory influences appropriate to its context.

The first outcome would obtain if listeners perceive unstressed vowel quality based solely on the acoustic speech signal within the time frame of the vowel's syllable—that is, independent of its stressed vocalic context. The second outcome would obtain if unstressed vowels are perceived relative to their flanking vocalic contexts.

There are several reasons for predicting the second outcome. First, although the findings of Ochiai and Fujimura (1971) show that vowels can be identified with better than chance accuracy when excised from their context, the error pattern suggests that context ordinarily does play a role. Second, as noted earlier, a listener presumably aims to recover the phonemes intended by a talker. If he is to do so in the case of segments affected by their segmental context, he must perceive them relative to those contexts. Experiments 2a and 2b test this prediction.

EXPERIMENTS 2A AND B

Two 4IAX discrimination experiments were run. They differed mainly in that the second experiment corrected a potential design flaw in the first experiment. In addition, the second experiment used the trisyllabic productions of a different talker and used a slightly different stimulus set from the first. Because none of these changes made a difference in the outcome of the two studies, Experiment 2b constitutes a replication of Experiment 2a, and the two studies are reported together.

Method

Design. The experiments obtained similarity judgments of pairs of the medial unstressed syllables of Experiment 1. The syllables of a pair were either acoustically identical or acoustically different, and each was embedded in a context of flanking vowels that was either the same as the one in which it had been produced originally, or different from its original context. Thus, the experimental design included two independent variables, each with two levels: Acoustic Similarity of the medial unstressed vowels of a pair of trisyllables (levels: acoustically the same or different), and Context of flanking vowels in which the pair of medial syllables is presented (levels: both appear in their original (appropriate) contexts, or one appears in a different (inappropriate) context from its original one).

Figure 3 shows the design of the experiment and provides an example of a pair of trisyllables for each cell of the design. (In Figure 3, the subscripts surrounding each medial syllable indicate the context of flanking vowels in which it had been produced originally. Instances in which the subscripts are not identical to the flanking vowels are instances in which a medial syllable was spliced out of its original context and into a new one

	Acoustic	
	same	different
appropriate context	A-12 i$_i$ bʌ$_i$ bi i$_i$ bʌ$_i$ bi	B-12 i$_i$ bʌ$_i$ bi a$_a$bʌ$_u$ bu
inappropriate context	B-34 i$_i$ bʌ$_i$ bi a$_i$ bʌ$_i$ bu	A-34 i$_i$ bʌ$_i$ bi i$_a$ bʌ$_u$bi

FIGURE 3. Design of Experiments 2a and 2b. The experiments provide an orthogonal comparison between the effects of acoustic similarity of the medial vowels of a pair of trisyllables (same or different) and the coarticulatory contexts in which the vowels appear (the contexts are appropriate to the acoustic properties of the medial /ʌ/ or one context is not).

with different flanking vowels. The procedure is described in the "Materials" section following. The labels A-12, A-34, B-12, B-34 refer to types of 4IAX trials and are also explained.)

Two dependent measures were obtained in the experiment: similarity judgments of pairs of medial syllables, and judgments of the confidence with which the similarity assessments had been made.

The four trisyllabic pairs shown in Figure 3 allow an orthogonal comparison to be made between the two perceptual strategies intended to be distinguished by the experiments. One strategy judges similarity of two vowels based only on their acoustic similarity; the second judges similarity based on the acoustic properties of each vowel in relation to its coarticulatory context. On the first perceptual strategy, trisyllabic pairs located in the left column of Figure 3 should sound more similar than those in the right column because the medial syllables within a pair are acoustically identical in the left column and different in the right column. On the second strategy, pairs in the top row should be judged more similar than pairs in the bottom row because items in the top row appear in a coarticulatory context compatible with their original ones. Hence, although the medial unstressed vowels in some of these pairs may be acoustically different from each other (as in the cell labeled B-12), both are acoustically characteristic of the unstressed vowel /ʌ/ produced in the particular contexts in which they appear. On a context-sensitive perceptual strategy, both should sound like unstressed /ʌ/ to a perceiver even if they are acoustically discriminable when presented in isolation.

Materials: The nine trisyllables of Experiment 1 in which the medial syllable was unstressed provide 36 pairs of different medial syllables that might have been used in the present experiments. Of these, a subset of 11 pairs in Experiment 2a and 14 pairs in Experiment 2b was selected based on procedures to be described shortly. For each pair of medial syllables used in the experiments, four trisyllables were constructed. For example, for the pair $_i$bʌ$_i$ and $_a$bʌ$_u$, they were /i$_i$bʌ$_i$bi/, /i$_a$bʌ$_u$bi/, /a$_a$bʌ$_u$bu/, and /a$_i$bʌ$_u$bi/. Thus, each syllable was placed into a context compatible with its original one and one compatible with that of its counterpart syllable. The splicing procedure is described below. In Experiment 2a, in trisyllables such as /i$_i$bʌ$_i$bi/, $_i$bʌ$_i$ was replaced into the context in which it had originally been produced; in Experiment 2b, it was replaced into a context that was a different token of the same type as the original context. That is, $_i$bʌ$_i$ from one instance of /ibʌbi/ was replaced into the context /i--bi/ or another instance of /ibʌbi/. This was to control for any artifacts introduced into the trisyllables by the splicing and replacement technique. In Experiment 2a, these artifacts would be restricted to trisyllables in which a syllable was replaced into a new context. In 2b, they would be as likely to occur on reconstructions such as /i$_i$bʌ$_i$bi/ as on those such as /a$_i$bʌ$_i$bu/.

From each set of four trisyllables, sixteen 4IAX trials were constructed, eight of each of two types, here labeled A and B:

	1 2		3 4
A:	i$_i$bʌ$_i$bi---i$_i$bʌ$_i$bi	-------------------------	i$_i$bʌ$_i$bi---i$_a$bʌ$_u$bi
B:	i$_i$bʌ$_i$bi---a$_a$bʌ$_u$bu	-------------------------	i$_i$bʌ$_i$bi---a$_i$bʌ$_i$bu

In trials of both types, one pair of stimuli (A-12, B-34) had acoustically identical medial syllables and one pair (A-34, B-12) had medial vowels that were acoustically different due to coarticulation effects. (Refer to Figure 3.) In trials of type A, when the medial syllables of a pair were identical, they occurred in their appropriate contexts (A-12). When they were different, one was in a different context of flanking vowels from that in which it had been produced originally (A-4). In contrast to A trials, on B trials, when the medial syllables were acoustically identical, one was in a context identical to its original one (B-3) and one was not (B-4). When the medial vowels within a pair were different, both were in their original contexts (B-12). Also in contrast to A trials, on B trials, the flanking vowels within each pair were never identical. Either the initial vowels or the final vowels or both were different.

For trials of both types, if the sequence of four trisyllables is numbered 12-34, the additional orderings 34-12, 21-43, and 43-21 were constructed giving four A and four B trial variants on the example given. In this set of eight trials, one of the flanking vowel contexts (in the example, i--bi) serves as a basis for the variations. Three of the four medial syllables of each trial were extracted from the context of the base. Therefore, in addition, a symmetrical set of eight trials was constructed using the second flanking vocalic context of the pair (in the example ab bu) as the base trisyllable:

	1 2		3 4
A:	a$_a$bʌ$_u$bu---a$_a$bʌ$_u$bu	--------------------	a$_a$bʌ$_u$bu---a$_i$bʌ$_i$bu
B:	a$_a$bʌ$_u$bu---i$_i$bʌ$_i$bi	--------------------	a$_a$bʌ$_u$bu---i$_a$bʌ$_u$bu

In this way, Experiment 2a had 88 A trials and 88 B trials; Experiment 2b had 112 trials of each type. The onset-onset times between members of a pair within a trial was 600 msec. The interval between pairs within a trial was 1000 msec; that between trials was 4 sec.

For the purposes of selecting the stimulus set for Experiment 2a, a pilot study was run. Its aim was to identify pairs of medial unstressed syllables from among those produced in Experiment 1 that were discriminably different due to the different coarticulatory influences on them. The pilot study presented pairs of medial unstressed syllables /bʌ/ each extracted from trisyllables of the form used in Experiment 1. The syllables were extracted only from trisyllables with unstressed medial syllables because coarticulatory effects on those syllables had been largest.

The medial syllables were excised from their coarticulatory context using the pulse-code-modulation system at Haskins Laboratories. Cuts were made at the onsets of the closure intervals of the two /b/s in the nine *VbVbV* utterances.

The talker was not one of the five subjects of this Experiment, but had been a subject in a pilot study for the Experiment reported in Fowler (1977). His coarticulatory effects were similar in all respects to those of the subjects of Experiment 1.

Subjects in the pilot study heard a random ordering of each of the nine syllables paired with itself eight times to make 72 "same" trials, and paired with each of the other syllables once in an AB order and once in a BA order to make 72 "different" judgments. Subjects made "same" or "different" judgments on each trial. Pairs of syllables were selected for use in Experiment 2a if subjects correctly judged them different in every instance in which the pair was presented. Eleven pairs were selected that met this criterion. They were: ${}_{a}\text{bʌ}_{a}$--${}_{a}\text{bʌ}_{i}$, ${}_{u}\text{bʌ}_{i}$--${}_{u}\text{bʌ}_{a}$, ${}_{u}\text{bʌ}_{i}$--${}_{u}\text{bʌ}_{u}$, ${}_{i}\text{bʌ}_{a}$--${}_{a}\text{bʌ}_{a}$, ${}_{i}\text{bʌ}_{i}$--${}_{a}\text{bʌ}_{i}$, ${}_{i}\text{bʌ}_{i}$--${}_{a}\text{bʌ}_{u}$, ${}_{i}\text{bʌ}_{i}$--${}_{u}\text{bʌ}_{a}$, ${}_{i}\text{bʌ}_{i}$--${}_{a}\text{bʌ}_{a}$, ${}_{i}\text{bʌ}_{a}$--${}_{a}\text{bʌ}_{u}$, ${}_{a}\text{bʌ}_{a}$--${}_{u}\text{bʌ}_{i}$, and ${}_{a}\text{bʌ}_{u}$--${}_{u}\text{bʌ}_{i}$. This gives three pairs in which acoustic differences are due only to anticipatory effects, two pairs in which the differences are due only to carryover effects, and six pairs in which both directions of coarticulation distinguish the monosyllables.

A different talker was used in Experiment 2b.[2] The pairs of medial syllables used in this experiment were selected based on their degree of acoustic dissimilarity rather than on perceptual discriminability. Fourteen pairs were selected. They were: ${}_{i}\text{bʌ}_{i}$--${}_{i}\text{bʌ}_{a}$, ${}_{a}\text{bʌ}_{i}$--${}_{a}\text{bʌ}_{a}$, ${}_{a}\text{bʌ}_{i}$--${}_{a}\text{bʌ}_{u}$, ${}_{i}\text{bʌ}_{i}$--${}_{a}\text{bʌ}_{i}$, ${}_{i}\text{bʌ}_{a}$--${}_{a}\text{bʌ}_{a}$, ${}_{a}\text{bʌ}_{u}$--${}_{u}\text{bʌ}_{u}$, ${}_{i}\text{bʌ}_{i}$--${}_{a}\text{bʌ}_{a}$, ${}_{i}\text{bʌ}_{i}$--${}_{u}\text{bʌ}_{u}$, ${}_{i}\text{bʌ}_{i}$--${}_{a}\text{bʌ}_{u}$, ${}_{i}\text{bʌ}_{i}$--${}_{u}\text{bʌ}_{a}$, ${}_{i}\text{bʌ}_{a}$--${}_{a}\text{bʌ}_{u}$, ${}_{a}\text{bʌ}_{a}$--${}_{u}\text{bʌ}_{i}$, ${}_{a}\text{bʌ}_{u}$--${}_{u}\text{bʌ}_{a}$, and ${}_{u}\text{bʌ}_{i}$--${}_{a}\text{bʌ}_{u}$. This gives three anticipatory pairs, three carryover pairs, and eight pairs in which the coarticulatory effects are bidirectional.

Procedure. In both experiments, subjects were asked to judge which pair of the two pairs on each trial sounded most similar. Subjects were told that the only basis for their choices was the medial syllable /bʌ/ of each trisyllable. They were told to indicate their choice by writing a "1" or a "2" on an answer sheet next to the appropriate trial number. "1" specified that the first pair of trisyllables sounded more similar than the second pair; a "2" indicated that the second pair sounded more similar. In addition to selecting the first or the second pair on each trial as more similar, subjects were asked to make confidence judgments: 0 signifying a random guess, 3 signifying certainty, and 1 and 2 indicating intermediate levels of certainty.

Scoring. Judgments that trisyllables with acoustically identical medial syllables were more similar than those with acoustically different medial syllables were given a score of 1. The alternative judgments were given a score of 0. For each subject, A and B trials were scored separately. For each type of trial, the number of scores of 1 was summed and divided by the total number of trials of that type. Thus, a ratio of 0.5 would indicate an inability to choose consistently based on acoustic similarity or based on contextual conditions. A consistent strategy of picking pairs in which the medial syllables are acoustically identical would give a ratio greater than 0.5 on both A and B trials. Moreover, subjects' degree of confidence should reflect their consistency in using this strategy. That is, they should be more confident of trials on which they receive scores of 1 than on trials in which they receive scores of 0.

In contrast, if subjects make their judgments based on an assessment of the medial vowels that is sensitive to the vowels' flanking vocalic context, the outcome should be somewhat different. On A trials, the acoustically identical syllables are in their appropriate coarticulatory contexts, whereas the different syllables are not. Hence, a context-sensitive strategy would lead the identical syllables to be selected just as before. However, on B trials acoustically different syllables are in their appropriate contexts, whereas one of the syllables of an identical pair is not. Hence, on a context-sensitive strategy, subjects should choose the different pair and should obtain overall ratios of less than 0.5. Moreover, whereas on A trials their degree of confidence should be high when they pick an acoustically identical pair, it should be low on these choices on B trials. On B trials, high confidence should be correlated with a score of 0, and low confidence with a score of 1.

Subjects. Eleven subjects were run in Experiment 2a and 14 in Experiment 2b. All were students at Dartmouth College.

Results

The first outcome of interest is the proportion of A and B trials on which subjects obtained a score of 1. In Experiment 2a, 0.84 of A trials received a score of 1. This differs substantially from a chance (i.e., no-preference) score of 0.50 ($t(10) = 15.10, p < 0.001$). In marked contrast to this outcome, 0.44 of B trials received a score of 1 ($t(10) = -3.54, p = 0.005$). Thus, on A trials, significantly more scores of 1 were obtained than chance, and on B trials, significantly fewer scores of 1 were obtained than chance. All subjects received lower B than A scores.

---

[2] I thank Tom Gay for serving as talker in the present experiment.

In Experiment 2b, the corresponding proportions for A and B trials are 0.73 ($t[13] = 9.11$, $p < 0.001$) and 0.44 ($t[13] = -2.94$, $p = 0.01$). Again, all subjects received lower scores on B trials than on A trials.

Reasons for the closer proximity of B than A trials to the chance value of 0.5 will be considered in the Discussion to follow.

The pattern of responses obtained in Experiments 2a and b is compatible with a manner of perceiving coarticulated medial vowels in reference to their flanking vocalic contexts. It is not compatible with a context-free manner of perception.

Figures 4a and b show the degrees of confidence on A and B trials separately for trials on which scores of 1 or 0 were obtained. As indicated earlier, higher confidence on trials with a score of 1 than 0 are expected on both A and B trials if subjects' perceptions of the vowels are such that they selected pairs of acoustically identical vowels as most similar, independently of their coarticulatory context. However, if their perceptions are sensitive to the flanking vocalic context, then subjects should show the foregoing pattern on A trials, but not on B trials. On B trials, they should be more certain when they select the acoustically different pair (each in a coarticulatory context appropriate for the acoustic signal to be /bʌ/) than when they select the acoustically identical pair (one of which is not in a context appropriate for that acoustic signal to be /bʌ/). This would give a lower confidence on B trials with a score of 1 than on trials with a score of 0, because the 1-scored trials would be mistaken guesses, whereas the 0-scored trials would be correct guesses and correct choices.

As Figure 4a shows, the patterning of the confidence judgments indicates that perception of the medial vowels was sensitive to the trisyllables' flanking vocalic context. On A trials, subjects were more confident of 1-scored trials than of 0-scored trials. On B trials, the pattern was reversed. An analysis of variance with factors Score (1,0), and Trial-type (A,B) showed both main effects to be significant ($F[1,10] = 22.55$, $p < 0.001$; $F[1,10] = 10.11$, $p < 0.001$). In general, subjects were more confident on trials with a score of 1 than on trials with a score of 0, and were more confident on A than on B trials. However, the significance of these main effects is mitigated substantially by the highly significant interaction between them ($F[1,10] = 105.43$, $p < 0.0001$). The significant interaction is due to A and B trials showing the reverse confidence patterns as Figure 4a illustrates. The difference between 1- and 0-scored trials is significant on both A and B trials considered separately, according to the results of Scheffé post hoc tests.

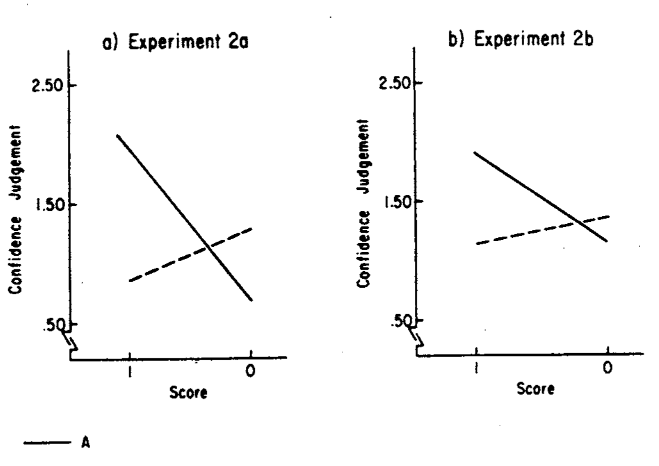

FIGURE 4. Confidence judgments on 1-scored and 0-scored A and B trials in Experiments 2a (a) and 2b (b).

As Figure 4b indicates, Experiment 2b replicated the foregoing outcome in every respect. The main effects of Score and Trial-type and of the interaction are all highly significant ($F[1,13] = 58.62$, $p < 0.001$; $F[1,13] = 15.00$, $p = 0.002$; $F[1,13] = 75.05$, $p < 0.001$). The significant interaction is due to A trials showing higher confidence on 1-scored trials than on 0-scored trials and B trials showing the reverse patterning. Again both differences in confidence are significant beyond the 0.05 level according to Scheffe's tests.

Next the trials of both experiments were partitioned into three categories: *Anticipatory*, *Carryover*, and *Both*, according to the direction of the coarticulatory differences that distinguish the acoustically different medial syllables of a different pair (A-34, B-12). On reanalysis of the data, Experiment 2a shows significant departures of A and B trials from chance in the expected directions on *Both* trials (A trials: $t(10) = 17.93$, $p < 0.001$; B trials: $t(10) = -4.34$, $p = 0.002$). Likewise, subjects exceed chance on A trials both on *Anticipatory* and on *Carryover* trials ($t[10] = 5.53$, $p < 0.001$; $t[10] = 14.76$, $p < 0.0001$). However, subjects are at chance on B trials when coarticulatory effects are unidirectional. In fact, they nearly exceed chance in the positive direction on *Anticipatory* trials of type B ($t[10] = 2.20$, $p > 0.05$).

In Experiment 2b, subjects were at chance on both A and B trials when the direction of coarticulation was *Anticipatory*. But they deviated from chance on both A and B trials when the direction was *Carryover* ($t[13] = 9.47$, $p < 0.001$; $t[13] = -2.96$, $p = 0.01$) or *Both* ($t[13] = 9.54$, $p < 0.001$; $t[13] = -2.68$, $p = 0.02$).

In general, these patterns of performance levels mirror the acoustic differences among the members of an acoustically different pair. That is, as Experiment 1 showed, carryover effects exceed anticipatory effects, and their combined effects tend to exceed either effect alone.

A final analysis examined performance on individual trial-types. Results from sets of four trials were combined and were averaged across the 11 subjects of Experiment 2a, and separately across the 14 subjects of Experiment 2b. The groups of four trials that constituted a "set" had the same coarticulatory contexts and the same medial syllables. An example of a set of four trials that were collapsed in these analyses are those generated from the A trials of the first example in the Methods section above. That is, results were collapsed across the trial given in the example and its variants obtained by permuting the pairs (12-34, 34-12, 21-43, 43-21). In this way, performance expressed as a deviation from the chance level of 0.5 could be examined for each of the 22-A and 22-B trial-types (3 *Anticipatory* trials, 2 *Carryover* trials, and 6 *Both* trials times two base medial syllables) and for the 28-A and 28-B trial-types of Experiment 2b (3 *Anticipatory* trials, 3 *Carryover* trials, and 8 *Both* trials times 2 base medial syllables).

On this measure, subjects exceed an average score of 0.5 on all of the sets of A trials in Experiment 2a and on 25 of the 28 trial-types in Experiment 2b. Of the three reversals, all were trials in which the direction of the coarticulatory differences between the medial syllables was unidirectional, either *Anticipatory* or *Carryover*. On the B trials of Experiment 2a, subjects fall below 0.5 on 11 of 12 sets in which coarticulatory differences are bidirectional (*Both* trials), but on only 2 of 4 and 1 of 6 trials in which its direction is *Carryover* and *Anticipatory*, respectively. In Experiment 2b, they fall below 0.5 on 11 of 16 *Both* trials, on 4 of 6 *Carryover* trials, and on 2 of 6 *Anticipatory* trials.

These analyses indicate that under the conditions of these experiments, bidirectional coarticulatory effects influence the vocalic percept, but unidirectional effects, in particular the smaller anticipatory effects, are less likely to do so.

Discussion

The results of the experiments show very clearly that the way in which unstressed coarticulated vowels "sound" to a perceiver depends not on the acoustic properties of their own time frame considered alone, but instead on those properties in relation to their flanking vocalic context.

In this respect, the perceiver's strategy for recognizing unstressed vowels is, as expected, tightly coupled to, or compatible with, the talker's strategy for producing them. As proposed in Part 1, coarticulatory effects of stressed on unstressed vowels may signify that the talker subsumes the production of unstressed vowels within the domain of the production of stressed vowels. More particularly, he may produce the unstressed vowel relative to the final phases of production of a preceding vowel, and, to a lesser extent, relative to the beginning phases of a following vowel. Compatible with this, a perceiver judges the quality of an unstressed vowel relative to its stressed vocalic context.

Yet to consider is one outcome in the data that was not predicted, and then to ask how the perceiver achieves one's context-sensitive percept. The not-predicted outcome was the closer proximity of performance on B than A trials to the chance value of 0.5. The different relations among the flanking vowels of a pair in A and B trials might explain this. On A trials, the trisyllables of a pair had the same flanking vowels—that is, their initial flanking vowels were the same, as were their final flanking vowels. In contrast, on B trials, either the initial flanking vowels of a pair or the final flanking vowels or both were different. Possibly, judgments of similarity of the unstressed medial vowels of a pair are facilitated when the flanking vocalic contexts of each are the same. On B trials, the listener's attention may be attracted to the large differences in the flanking vowels and away from the degrees of similarity of the medial vowels.

A different interpretation of this outcome cannot be rejected, however. It is that perceivers use two criteria in judging the degree of similarity of a pair of medial syllables. One is the absolute acoustic similarity of the syllables, and the second is the acoustic properties of each evaluated relative to its coarticulatory context. On A trials, judgments based on the two criteria lead to the same decision. On B trials, they lead to different decisions, giving a lower overall departure from chance on B trials. Of course, even if perceivers do use both sources of information in making their judgments, rather than just the second as we prefer to conclude, the second criterion is given more weight because performance on B trials consistently falls below 0.5.

We turn now to consider how the perceiver achieves his context-sensitive percept by considering two accounts. The two are quite different and both have substantial independent support in the literature. One involves "contrast" effects which occur quite generally, and are instances in which the identification of a stimulus is affected by its context in a dissimilative way. That is, the identity of the stimulus is moved away from that of its neighbor. An example outside the realm of speech is provided by the work of Sherif, Taub, and Hovland (1958). They obtained ratings of the weights of six objects under two conditions. In one condition, the objects were presented singly, and subjects provided a number between 1 and 6 according to the apparent weight of the objects. In the other condition, the object to be rated was lifted after the subject first had lifted a comparison weight. The first weight of the pair was always heavier than the weight to be rated, and most typically was heavier than the heaviest of the six objects that the subject was asked to rate. In comparison to the first condition, objects were rated overall lighter in the second condition when the comparison weights were heavier than the heaviest object in the test series. Thus, judgments tended to be displaced away from the comparison stimulus.

Compatible effects to these are reported for speech continua when the stimuli are vowels (Eimas, 1963; Fry, Abramson, Eimas, & Liberman, 1962; Repp, Healy, & Crowder, 1979). Consonants also show contrast effects (e.g., Diehl, Elman, & McCusker, 1978; Diehl, Lang, & Parker, 1980), but are more resistant to them than are vowels (Simon & Studdert-Kennedy, 1978).

In the context of the 4IAX discrimination procedure used in Experiments 2a and b, contrast effects would work in the following way. Due to coarticulation, F2 for /ʌ/ between flanking /i/s is raised relative to its value between flanking /a/s. However, a contrast effect on /ʌ/ due to the flanking /i/s would perceptually lower F2 for /ʌ/, thereby effectively erasing the coarticulatory influence, at least if the magnitudes of the coarticulatory and contrast effects were similar. Likewise, the flanking /a/s would perceptually raise F2 for /ʌ/ erasing its coarticulatory lowering of the acoustic value of the same formant.

This account can explain all of the findings of Experiments 2a and b, but it is unattractive on two counts. Most importantly, if the contrast account is to explain how listeners recognize the vowels intended by talkers, it must be supposed that contrast effects are not only opposite in sign, but also very similar in magnitude to coarticulatory effects. Because the two effects are independent—one articulatory and one psychoacoustic—it would be remarkably fortunate for the listener were the magnitudes consistently the same.

A second unattractive aspect of the contrast account is related to the first: the account treats an observed tight coupling in the patterns of behavior of talkers and listeners as if it were fortuitous. This seems implausible. Nonetheless, the contrast account can not be rejected based on the current findings.

A second way in which context might be expected to affect the perceived quality of a vowel is offered by such views of perception as the motor theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) or Gibson's more general perceptual theory (Gibson, 1966). On the latter view, the stimulus for perception, the acoustic signal in the case of speech, is informative insofar as it has been patterned by a "distal source," the vocal tract in speech. The systematic properties of the source—its structure or its changing structure over time—are reflected in the systematic properties of the proximal (acoustic) stimulus. These systematic properties constitute the stimulus for perception; to the extent that a perceiver registers the structure in the acoustic stimulus, structure that corresponds to articulatory gestures is registered, and perceived phonetic categories will tend to have an articulatory definition.

If perceivers are sensitive to information in the signal that reflects articulation, among other things, they extract information corresponding to the talker's coarticulatory patterns. Thus, for example, when a talker produces a medial unstressed /ʌ/ in /ibʌbi/ relative to a high positioning of the tongue body for /i/, and toward a somewhat lower positioning for /ʌ/, the acoustic signal he generates reflects the ongoing gestures for the stressed vowel with the unstressed vowel produced relative to it. Similarly, in /abʌba/, when the talker produces a medial vocalic segment relative to a low positioning of the tongue for /a/ and toward a higher one, the acoustic signal reflects the complex of gestures. By hypothesis, the listener distinguishes the stressed articulatory frame from the unstressed vowels produced relative to it. Thus, in both examples above, acoustic information corresponding to movement toward the target positioning of the vowel /ʌ/ is extracted. It could not be so readily extracted were the acoustic stimulus for the medial syllable excised from its context and presented in isolation or in an inappropriate context.

This account is no less general than the explanation that invokes contrast effects. In particular, several perceptual phenomena involving speech seem to require accounts that invoke an articulation-referential description of the acoustic signal. These include the perception of certain coarticulated consonants (most notably, the synthetic /d/s in two formant versions of /di/ and /du/), the role of silence as a stimulus for perceiving phonetic segments (see Liberman and Pisoni, 1977, for a review), various trading relations that obtain among acoustically quite disparate "cues" for segments (e.g., Fitch, Halwes, Erickson, & Liberman, 1979; see also, Dorman, Raphael, & Liberman, 1979; Dorman, Studdert-Kennedy, & Raphael, 1977), and, perhaps also the perception of isochrony in speech (Fowler, 1979; Tuller & Fowler, 1980). Moreover, Gibson offers an analogous theoretical account of visual perception, and, indeed, of perception in any modality.

Unfortunately, it is unclear how the contrast and articulatory accounts can be distinguished experimentally. Specifically, the contrast account presumes that coarticulatory influences on the acoustic signal for a vowel will always be erased by perceptual contrast effects. The articulatory account presumes that the acoustic consequences of gestures for the coarticulatory context constitute a frame perceptually, and that the coarticulated segment is perceived relative to the frame. The effect of this is to "subtract" the coarticulatory influence of the frame from the signal for the coarticulated segment.

To distinguish these views would require that the vowels to be compared resist coarticulation—perhaps stressed tense vowels. On a contrast view, these should be subject to contrast effects from a preceding or following context. However, on the articulatory view, their perceived quality should be unaffected by their vocalic context.

ACKNOWLEDGMENTS

The research reported here was supported by NICHD Grant HD01994 and NINCDS Grant NS-13617 to Haskins Laboratories and by Biomedical Research Grant RR-05392 from the Biomedical Research Support Branch, Division of Research Facilities and Resources, National Institutes of Health. I thank George Wolford for critiquing an earlier version of the manuscript. Requests for reprints should be addressed to Carol A. Fowler, Gerry Hall, Dartmouth College, Hanover, N.H. 03755.

REFERENCES

ABERCROMBIE, D. Syllable quantity and enclitics in English. In D. Abercrombie, D. Fry, P. MacCarthy, N. Scott, & J. Trim (Eds.) *In honour of Daniel Jones*. London: Longman, 1964.

BELL-BERTI, F., & HARRIS, K. Anticipatory coarticulation: Some implications from a study of lip rounding. *Journal of the Acoustical Society of America*, 1979, 65, 1268-1270.

CARNEY, P., & MOLL, K. A cinefluorographic investigation of coarticulation. *Phonetica*, 1971, 23, 707-721.

DIEHL, R., ELMAN, J., & MCCUSKER, S. Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, 4, 599-609.

DIEHL, R., LANG, M., & PARKER, E. A further parallel between selective adaptation and contrast. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, 6, 24-44.

Donovan, A., & Darwin, C. The perceived rhythm of speech. *Ninth International Congress of Phonetic Sciences.* Copenhagen: University of Copenhagen, 1979.

Dorman, M., Raphael, L., & Liberman, A. Some experiments on the sound of silence in phonetic perception. *Haskins Laboratories Status Reports on Speech Research,* 1979, *SR-58,* 105-138.

Dorman, M., Studdert-Kennedy, M., & Raphael, L. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent. *Perception and Psychophysics,* 1977, 22, 109-122.

Eimas, P. The relation between identification and discrimination along speech and nonspeech continua. *Language and Speech,* 1963, *6,* 206-217.

Fitch, H., Halwes, T., Erickson, D., & Liberman, A. Perceptual equivalence of two acoustic cues for stop-consonant manner. *Haskins Laboratories Status Reports on Speech Research,* 1979, *SR-57,* 183-200.

Fowler, C. *Timing control in speech production.* Indiana University Linguistics Club, Bloomington, Indiana, 1977.

Fowler, C. "Perceptual centers" in speech production and perception. *Perception and Psychophysics,* 1979, *25,* 375-388.

Fry, D., Abramson, A., Eimas, P., & Liberman, A. The identification and discrimination of synthetic vowels. *Language and Speech,* 1962, 5, 171-189.

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

Gibson, J. *The senses considered as perceptual systems.* Boston: Houghton-Mifflin, 1966.

Huggins, A. On isochrony and syntax. In G. Fant & M. Tatham (Eds.) *Auditory analysis and perception of speech.* London: Academic Press, 1975.

Kent, R., & Moll, K. Tongue-body articulation during vowel and diphthong gestures. *Folia Phoniatrica,* 1972, *24,* 278-300.

Klatt, D. Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics,* 1975, *3,* 129-140.

Klatt, D. The linguistic uses of segment duration in English. *Journal of the Acoustical Society of America,* 1976, *59,* 1208-1221.

Kuehn, D., & Moll, K. Perceptual effects of forward coarticulation. *Journal of Speech and Hearing Research,* 1972, *15,* 654-664.

Lehiste, I. The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America,* 1972, *51,* 2018-2020.

Lehiste, I. Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America,* 1973, *54,* 1228-1234.

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

Liberman, A., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review,* 1967, *74,* 431-461.

Liberman, A., Delattre, P., & Cooper, F. The role of selected stimulus variables in the perception of the unvoiced stop consonants. *American Journal of Psychology,* 1952, *65,* 497-516.

Liberman, A., Delattre, P., Cooper, F., & Gerstman, L. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs,* 1954, *68* (Whole No. 379).

Liberman, A., & Pisoni, D. Evidence for a special speech-perceiving subsystem in the human. In T. Bullock (Ed.), *Recognition of complex acoustic signals.* Berlin: Dahlen Konferenzen, 1977.

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

Lindblom, B., & Rapp, K. Some temporal regularities of spoken Swedish. *Papers in Linguistics from the University of Stockholm,* 1973, *21,* 1-59.

MacNeilage, P., & DeClerk, J. On the motor control of coarticulation in CVC monosyllables. *Journal of the Acoustical Society of America,* 1969, *45,* 1217-1233.

Martin, J. Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review,* 1972, *79,* 487-509.

Nord, L. Vowel reduction—Centralization or contextual assimilation. *Preprints of the Speech Communication Seminar,* 1974, 2, 149-154.

Ochiai, K., & Fujimura, O. Vowel identification and phonetic contexts. *Reports from the University of Electrocommunications* (Tokyo), 1971, 22-2, 103-111.

Ohde, R., & Sharf, D. Order effect of acoustic segments of VC and CV syllables in stop and vowel identification. *Journal of Speech and Hearing Research,* 1977, *20,* 543-554.

Ohman, S. Coarticulation in VCV utterances. *Journal of the Acoustical Society of America,* 1966, *39,* 151-168.

Perkell, J. *Physiology of speech production.* Cambridge, Mass: MIT Press, 1969.

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

Pike, K. *Intonation of American English.* Ann Arbor: University of Michigan Press, 1945.

Pisoni, D. On the nature of categorical perception of speech sounds. Supplement to the *Haskins Laboratories Status Reports on Speech Research,* 1971.

Repp, B., Healy, A., & Crowder, R. Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance,* 1979, 5, 129-145.

Schatz, C. The role of context in the perception of stops. *Language,* 1954, *30,* 47-56.

Sharf, D., & Beiter, R. Identification of consonants from formant transitions presented forward and backward. *Language and Speech,* 1974, *17,* 110-118.

Sharf, D., & Hemeyer, T. Identification of place of consonant articulation from vowel formant transitions. *Journal of the Acoustical Society of America,* 1972, *51,* 652-658.

Sherif, M., Taub, T., & Hovland, C. Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology,* 1958, *55,* 150-156.

Simon, H., & Studdert-Kennedy, M. Selective anchoring and adaptation of phonetic and nonphonetic continua. *Journal of the Acoustical Society of America,* 1978, *64,* 1338-1357.

Sussman, H., MacNeilage, P., & Hanson, R. Labial and mandibular dynamics during the production of bilabial consonants. *Journal of Speech and Hearing Research,* 1973, *16,* 397-420.

Tiffany, W. Nonrandom sources of variation in vowel quality. *Journal of Speech and Hearing Research,* 1959, 2, 305-317.

Tuller, B., & Fowler, C. Some articulatory correlates of perceptual isochrony. *Perception and Psychophysics,* 1980, *27,* 277-283.

Received January 14, 1980.
Accepted February 25, 1980.