Listeners' perception of compensatory shortening

CAROL A. FOWLER

*Haskins Laboratories, New Haven, Connecticut
and University of Connecticut, Storrs, Connecticut*

AND

JAQUELINE M. THOMPSON

*Haskins Laboratories, New Haven, Connecticut
and Yale University, New Haven, Connecticut*

English exhibits compensatory shortening, whereby a stressed syllable followed by an unstressed syllable is measured to be shorter than the same stressed syllable alone. This anticipatory shortening is much greater than backward shortening, whereby an unstressed syllable is measured to shorten a following stressed syllable. We speculated that measured shortening reflects not true shortening, but coarticulatory hiding. Hence, we asked whether listeners are sensitive to parts of stressed syllables hidden by following or preceding unstressed syllables. In two experiments (Experiments 1A and 1B), we found the point of subjective equality—that is, the durational difference between a stressed syllable in isolation and one followed by an unstressed syllable—at which listeners cannot tell which is longer. In a third experiment (Experiment 2), we found the point of subjective equality for stressed monosyllables and disyllables with a weak–strong stress pattern. In all of the experiments, the points of subjective equality occurred when stressed syllables in disyllables were measured to be shorter than those in monosyllables, as if the listeners heard the coarticulatory onset or the continuation of a stressed syllable within unstressed syllables.

Coarticulation has many definitions. For the present purposes, it refers to temporal overlap among gestures for serially nearby phonetic segments; it does not refer, for example, to adjustments in production of one segment to accommodate it to requirements of another. Temporal overlap of gestures ensures that the acoustic speech signal is highly context sensitive, because it is caused nearly everywhere by gestures for more than one consonant or vowel. Our interest in the present study is in the perception of coarticulated speech.

Compensation for Coarticulation and the Companion Finding

Listeners compensate for coarticulation of many kinds. That is, they respond to speech signals as if they have adjusted for the context sensitivity caused by coarticulation. For example, Mann (1980) found that listeners identified ambiguous members of a /da/–/ga/ continuum differently, depending on whether a precursor syllable was /al/ or /ar/. In particular, they identified more continuum members as /ga/ if the precursor was /al/ than if it was /ar/. Mann identified two possible accounts of the finding. One is that listeners track coarticulation. Productions of /l/ and /r/ persist after the onset of /da/ or /ga/ production (carryover coarticulation). A consequence of perseveration of /l/ may be that the persisting tongue-tip constriction or its gradual release causes the tongue dorsum constriction

during /g/ to be pulled forward. In contrast, persistence or gradual release of the pharyngeal constriction for /r/ during /da/ production might pull the alveolar constriction for /da/ back. When listeners hear ambiguous continuum members, they hear a consonant that is rather far forward for /g/ and rather far back for /d/. If they are tracking coarticulation, and the precursor syllable is /al/, they may hear the ambiguous continuum member as /ga/, pulled forward by /l/; if the precursor is /r/, they may hear the continuum member as /da/, pulled back by /r/. An alternative account of the finding is that it reflects spectral contrast. The high-ending *F3* of /l/ causes the onset *F3*s of continuum members to sound lower than they are and so more like /g/; the low-ending *F3* of /r/ has the opposite effect.

These accounts of compensation for coarticulation have been intensely debated (see, e.g., Fowler, 2006; Lotto & Holt, 2006). In our view, the debate is coming to a resolution that clearly favors Mann's (1980) first account (see Viswanathan, Fowler, & Magnuson, 2009; Viswanathan, Magnuson, & Fowler, in press). The debate has some relevance to the research that we report here, in which durational contrast offers an alternative account of our findings, and so we will return to the debate in the General Discussion section. For the present, we focus on the range of speech domains in which compensation for coarticulation occurs and on a companion finding to it that, in our view, sheds light on listeners' perception of coarticulated speech.

Mann's (1980) finding revealed compensation for carryover coarticulation. However, compensation for anticipatory coarticulation—that is, for the onset of gestures of a later segment during production of an earlier segment—also occurs. For example, Mann and Repp (1980) found that listeners reported more /s/ responses along an /s/–/ʃ/ continuum when the following vowel was rounded (/u/) than when it was unrounded (/a/), as if they compensated for the spectrum-lowering effects of anticipatory lip rounding. Accordingly, consonants with lower spectra counted as /s/ in the context of /u/ as contrasted with one of /a/.

There is also compensation for coarticulatory effects that are neither left-to-right as in Mann (1980) nor right-to-left as in Mann and Repp (1980). For example, Silverman (1987) found that judgments of the height of an intonational pitch accent in such sentences as “They always feast before fasting” and “They always fast before feasting” reflected compensation for the higher intrinsic f_0 (Sapir, 1989) of the high vowel /i/ than of the low vowel /a/. This finding is of interest not only because it is a nondirectional coarticulatory effect (and hence not subject to an account in terms of spectral contrast, which requires a context to trigger contrast), but also because it shows compensation for coarticulation that affects a suprasegmental speech variable—namely, intonation.

Findings of compensation for coarticulation reveal that listeners somehow perceptually remove effects of coarticulatory overlap from the domain on which they encroach. As Fowler (2006) pointed out, there is a companion effect to compensation that provides insight into what listeners do to achieve that. For example, companion to the finding that listeners compensate for anticipatory lip rounding in the study of Mann and Repp (1980) cited above is a finding that listeners use information for rounding in a fricative as information for a rounded vowel (e.g., Whalen, 1984, Experiment 2). That is, they behave as if they do not hear spectrum lowering as part of the fricative (Mann & Repp, 1980), but that is not because they fail to perceive it (Whalen, 1984). Rather, they use that information for what it *is* information for—namely, the coarticulating rounded vowel. Likewise, listeners did not hear f_0 raising because of a high vowel as part of a pitch accent in Silverman's (1987) research, but they did use it as information for a high vowel (Reinholt-Peterson, 1986).

This result suggests a perspective on coarticulation and its perception like that shown schematically in Figure 1. In the figure, time is represented along the x-axis and a variable, prominence, along the y-axis. *Prominence* refers both to the extent to which a speech gesture exerts control over the vocal tract and to the extent to which its acoustic consequences predominate in the acoustic signal. Gestures wax then wane in prominence. The figure shows two temporally overlapping speech gestures. Gesture 1 persists after the onset of production of Gesture 2 (carryover coarticulation). Gesture 2 begins during production of Gesture 1 (anticipatory coarticulation). Typically, when an acoustic signal is segmented for the purposes, say, of measuring duration, it is segmented at the vertical line in Figure 1—that is, at a point where the acoustic consequences of Gesture 1 cease to dominate in the acoustic

Figure 1. Schematic depiction of the temporal overlap of coarticulating speech gestures.

signal, and those of Gesture 2 take over. Normally, that is the best that speech scientists can do, but it does not accurately capture the extents of Gestures 1 and 2.

The findings of compensation for coarticulation and its companion suggest that listeners do not segment the acoustic signal in the way that speech scientists do. They do not hear the encroachment of Gesture 1 into the domain of Gesture 2 as information for Gesture 2—that is, they compensate for coarticulation—but, instead, they hear the encroachment as information for Gesture 1 (the companion finding).

Measured Durational Shortening Also Indexes Coarticulatory Overlap

In the present study, we extend the investigation of listeners' tracking of coarticulation to a relatively unstudied domain: the perception of so-called *compensatory shortening*. This is the finding that, in English (e.g., Lehiste, 1972) and some other languages (e.g., Swedish; Lindblom & Rapp, 1973), stressed syllables followed by unstressed syllables are shortened by them (so that, e.g., other things being equal, /beɪ/ in *baby* is shorter than in *bay*), a phenomenon that Lindblom and Rapp called *anticipatory compensation*. (We will use the term *shortening* rather than *compensation*.) In these languages, stressed syllables are also shortened by unstressed syllables that precede them (backward compensation or backward shortening), but the latter shortening is much less than anticipatory shortening.

It is likely that this pattern of shortening reflects the metrical structure of these languages, which are sometimes said to be stress timed (e.g., Abercrombie, 1967; Classe, 1939; Pike, 1945; Ramus, Nespor, & Mehler, 1999) or to have left-dominant metrical feet—that is, canonically, a strong–weak (trochaic) pattern of disyllables. In a stress-timed language, there is said to be a tendency (difficult to document) for intervals between stressed syllable onsets to be isochronous. Accordingly, a stressed syllable is shortened if it is followed by an unstressed syllable before the next stress. The tendency for isochrony—if there is one—is very weak. The shortening of a stressed syllable by an unstressed syllable is far less than the duration of the unstressed syllable itself (so that the total duration of

baby is longer than the duration of *bay*—by 176 msec in the stimulus materials that we used here).

Fowler (1981b) reported a correlation between compensatory shortening and coarticulation that she interpreted as evidence that compensatory shortening is a measurement artifact. Talkers produced reiterant speech versions of such words as *easy* and *asleep* in a carrier sentence. The reiterant versions of *easy* were /sisə/ and /səsə/. Those for *asleep* were /səsi/ and /səs/. Reiterant versions of monosyllables and multisyllabic words with different stress patterns were also produced. The stressed syllable was either /si/ or /sə/; the unstressed syllable(s) always had a schwa vowel. The duration of the stressed vowel in each context was measured. The coarticulatory effects of /i/ in each unstressed schwa vowel were assessed by comparing F_2 in the schwa vowels, depending on whether the stressed vowel was /i/ or /ə/.

Fowler (1981b) found that durational shortening mirrored earlier findings by Lehiste (1972) and, in Swedish, by Lindblom and Rapp (1973). Monosyllables were longer by 99 msec than stressed syllables in trochaic disyllables. Only 8 msec of additional shortening occurred in trisyllables with initial stress. Backward shortening due to one unstressed syllable was much less than anticipatory shortening; the stressed syllable in a weak-strong (iambic) disyllable was just 19 msec shorter than the monosyllable. There was no further shortening in a trisyllabic utterance with final stress. In fact, the duration of the stressed syllable was slightly longer than that in the disyllable.

Coarticulation closely mirrored the durational pattern. Carryover coarticulation in a final schwa of a trochaic disyllable averaged 180 Hz. That is, F_2 in schwa differed by 180 Hz, depending on whether the initial stressed syllable was /i/ or /ə/. In a trisyllable with initial stress, the carryover effect on the final schwa was just 21 Hz. The anticipatory coarticulation of /i/ in a disyllable with an initial schwa was 133 Hz in magnitude. In the first schwa of a trisyllable ending in a stressed /i/ or /ə/, it was just 15 Hz. The pattern mirrored the durational pattern, in that, in both available comparisons, carryover coarticulation was larger than anticipatory coarticulation, and (unfortunately in a terminological sense) correspondingly anticipatory shortening was greater than backward shortening. In addition, both carryover coarticulation and anticipatory shortening decreased with the distance of the stressed from the unstressed syllable. A failure to mirror occurred in one case. Whereas anticipatory coarticulation, like carryover coarticulation, decreased with distance between the stressed and unstressed syllables, backward shortening showed a slight increase.

The correlations between magnitudes of durational shortening and coarticulation magnitude were high (ranging from .53 to .92) and were significant for five of the six talkers in the study. Fowler (1981b) proposed that the correlations reflected the means by which the measured shortening took place. She proposed that the shortening was not true shortening but, rather, reflected coarticulatory overlap, as is shown in Figure 1. Imagine, for example, that Figure 1 displays a trochaic disyllable, such as *baby*. In this case, Gesture 1 is the stressed syllable's vowel, and Gesture 2 is the unstressed vowel. As the figure shows, the

stressed syllable's articulatory waning right edge is hidden by temporal overlap of the unstressed syllable. That its production in fact continues beyond the end of its measured extent is shown by Fowler's (1981b) measures of carryover coarticulation. Likewise, in iambic disyllables, such as *obey*, the stressed syllable begins to be produced during the unstressed initial syllable, as is indexed by measures of anticipatory coarticulation. Its produced extent, then, is longer than its measured extent. The carryover coarticulation in trochaic disyllables is greater than the anticipatory coarticulation in iambic disyllables. Thus, the anticipatory compensatory shortening exceeds backward shortening (perhaps only in languages with left-dominant feet).

The Present Study

In the present study, we asked whether listeners' perceptions of syllable durations respect coarticulatory overlap. That is, we asked whether listeners hear the stressed syllable (Gesture 1 in Figure 1) as beginning at the beginning of the syllable and ending at (or near; see below) its end, despite the fact that its end is overlapped by the following unstressed syllable (Gesture 2).

There is already evidence that compensation for coarticulation occurs in this domain. For example, Fowler (1981a; Fowler & Smith, 1986) gave listeners two pairs of trisyllables such as those in Example 1. The subscripts around the schwa vowel identify the context in which the schwas had been produced originally. (Accordingly, the schwa in the first trisyllable has been cross-spliced from an /a/ context into an /i/ context. Schwas in the other trisyllables were spliced from a different trisyllable with the same phonetic content.) The listeners were asked in which pair—the first or the second—the schwas sounded more alike. The acoustic signals during the schwas were identical in the first pair but, owing to coarticulation, were different in the second. However, the listeners systematically judged the schwas in the second pair as more alike than those in the first. That is, they compensated for coarticulation in failing to hear the effects of /i/ and /a/ production during schwa as properties of the schwa.

$$ib_a\underset{a}{e}bi \quad ab_a\underset{a}{e}ba \quad ib_i\underset{i}{e}bi \quad ab_a\underset{a}{e}ba \quad (1)$$

This finding may be understood in more than one way. On one hand, listeners may track coarticulation. They may not hear the waning edge of the first /i/ or the waxing edge of the final /i/ in /ibəbi/ as part of the schwa vowel. On the other hand, the schwas may be subject to a spectral contrast effect. The schwa in /ibəbi/ is measured to have a higher F_2 than that in /abəba/. A contrast effect from the high F_2 of /i/ would lead to perception of a schwa with a lower F_2 .

Because contrast effects reduce sensitivity to coarticulatory information, it is relevant to ask whether a companion effect occurs as well. That is, if listeners experience contrast, they lose sensitivity to the coarticulatory information in the schwa. But if they track speech gestures, that is the reason that they do not hear /i/ or /a/ information in the schwa as part of the schwa. It is not that they are insensitive to it; they should hear it as part of the flanking vowel, and there is some evidence that they do. Listeners identify the final vowel in trisyllables such as those above

(and others, such as /ab_aə_abi/ or /ib_iə_aba/) faster if the schwa vowel provides accurate coarticulatory information for it than if (as in the first trisyllable in Example 1) it provides misleading information (Fowler & Smith, 1986).

There is another prediction of a view that listeners track speech gestures that refers to compensatory shortening, a companion effect involving duration. In its boldest form, it is that the perceived duration of a gesture will exceed its measured duration by an amount that is the same as its amount of coarticulatory overlap. That is, in Figure 1, Gesture 1's perceived duration should be its produced duration, not its produced duration up to the vertical line in the figure where it is measured to end. This prediction is almost certainly too bold, because, as the gesture wanes, its acoustic effects become quite small and are therefore likely to be inaudible. Accordingly, the prediction to be tested here is that the perceived duration of a stressed syllable in a word such as *baby* will exceed its measured duration, so that the shared /be/ syllables in *bay* and *baby* are perceived to have the same duration when that in *bay* is longer than the measured duration of /be/ in *baby*. However, the point of subjective equality will not be at a duration that entirely offsets the difference in the syllables' measured duration.

Other researchers have explored listeners' sensitivity to the (measured) durational difference between shared syllables of monosyllabic and disyllabic words, either to test whether the duration difference is represented lexically (Shatzman & McQueen, 2006) or to explore the role of prosodic information in segmentation of continuous speech (Davis, Marslen-Wilson, & Gaskell, 2002; Salverda, Dahan, & McQueen, 2003). These studies are consistent in showing listeners' sensitivity to, and use of, the duration difference.

In contrast to these studies, the present study is not about sensitivity to the duration difference. It is complementarily about listeners' relative *insensitivity* to the measured duration difference, because, for example, in trochaic disyllables they hear the stressed syllable continuing to be produced during the unstressed syllable. The present study is about listeners' sensitivity to *coarticulatory information* and their consequent insensitivity to measured durational differences. If listeners were sensitive to all of the carryover coarticulatory information in the signal of trochaic disyllables, the outcomes by Davis et al. (2002), Salverda et al. (2003), and Shatzman and McQueen (2006) would not have occurred. Listeners would hear the shared syllables of monosyllables and trochaic disyllables as the same when they differed by the measured durational amount. That is, they would not hear a duration difference. Yet they did. However, as was noted above, it is highly likely that some of the coarticulatory information about stressed syllables in the domain of unstressed syllables is inaudible. We expect to find a point of subjective durational equality when the duration of the shared syllable in the monosyllable exceeds that in the trochaic disyllable by less than the measured amount. If that occurs, the findings in the preceding studies that listeners are sensitive to compensatory shortening are not inconsistent with a finding that they are not sensitive to all of it.

Experiments 1A and 1B were designed to explore perceived durations of stressed syllables in trochaic disyllables and in monosyllables, and so to explore the percep-

tion of metrical structure in contexts in which anticipatory shortening occurs. Experiment 2 provides a similar exploration of the perception of iambic disyllables and monosyllables and the perception of backward shortening.

In all of the experiments, we used a duration judgment procedure that was certainly unlike ordinary speech perception, and this may have had an undesired impact on response patterning. However, the procedure is the best one that we could think of to address the question of listener's insensitivity to measured durational shortening. We will discuss this issue further in the General Discussion section.

EXPERIMENTS 1A AND 1B

In the first pair of experiments, we explored the shortening that is most robust in English: anticipatory shortening. The main aim was to identify the listeners' point of subjective equality—that is, the duration difference of the common syllables in words such as *bay* and *baby* at which the listeners were unable to tell which syllable was longer. Our prediction was that the point of subjective equality would be one at which the /be/ in *baby* would be shorter than that in *bay* by an amount that was somewhat less than the measured durational difference in natural speech. This would be the outcome if perceivers were sensitive to coarticulatory overlap but could not detect all of the waning production of the stressed syllable in the unstressed syllable.

Method

Participants. There were 25 participants in the experiments—12 in Experiment 1A and 13 in Experiment 1B. They were students at the University of Connecticut who participated for course credit. All were native speakers of English who reported having normal hearing and speech.

Stimuli. We generated 14 pairings of monosyllabic and disyllabic words. The disyllables were stress initial, and the initial syllable was shared between the monosyllable and the disyllable. We recorded a native English speaking female (JB) producing each word in the carrier phrase "The next word to say is ____ because." (The carrier sentence fragment was meant to ensure that the target words were neither initial nor final in the utterance.) The speaker was recorded in a sound-attenuating booth. The 14 pairs are listed in Table 1 with

Table 1
Word Pairs and Their Durations (in Milliseconds)
for the Stimuli of Experiment 1

Monosyllable	Duration	Disyllable	Duration of Shared Syllable
<i>bar</i>	330	<i>barber</i>	190
<i>bay</i>	363	<i>baby</i>	187
<i>bee</i>	310	<i>beacon</i>	130
<i>can</i>	400	<i>candle</i>	300
<i>do</i>	315	<i>duty</i>	175
<i>go</i>	280	<i>gopher</i>	160
<i>low</i>	390	<i>local</i>	190
<i>may</i>	390	<i>maple</i>	210
<i>my</i>	369	<i>miser</i>	305
<i>mow</i>	405	<i>motor</i>	265
<i>new</i>	368	<i>noodle</i>	240
<i>tie</i>	443	<i>tiny</i>	223
<i>toe</i>	375	<i>total</i>	215
<i>tore</i>	385	<i>torso</i>	197
Mean	366	Mean	213

the durations of their shared syllables. The average duration of the monosyllables was 366 msec; that of the shared syllable in the disyllable was 213 msec.

To address a concern that intensity might differ between monosyllables and disyllables and that it would affect perceived duration (e.g., Turk & Sawusch, 1996), we measured both the peak and the mean intensity of monosyllables and shared syllables of disyllables. Both measures differed by less than 1 dB across pairs of items. The results of paired *t* tests did not approach significance ($ps > .42$).

For each monosyllable and disyllable, we created six edited shortened and lengthened versions of each shared syllable. We chose a step size for shortening and lengthening that allowed us to include the original durations of both the monosyllable and the disyllable. The step size, which had to be adjusted to include whole pitch pulses, averaged 39 msec. We lengthened the monosyllable by one step and shortened it by five steps. Likewise, we shortened the disyllable's shared syllable by one step and lengthened it by five steps. Shortening was accomplished by removing pitch pulses that were not adjacent in the waveform. Likewise, lengthening was accomplished by replicating

pitch pulses that were not adjacent in the vowel of the shared syllable. (That is, we might replicate Pitch Pulse 5 and then Pitch Pulse 10, but never Pitch Pulses 5 and 6.) We also included the original version of each word to make seven total unique stimuli per word.¹

In Experiment 1A, to keep the number of trials reasonable, on each trial, we paired the original duration of either the monosyllable or the disyllable with one of the seven durations of the disyllable or monosyllable, respectively. Figure 2A shows the pairings that we used. This gave 182 unique trials: 2 original durations $\times$ 7 variants $\times$ 14 words minus the 14 duplicate pairings in which the original durations of the monosyllable and disyllable were presented. (Duplication occurred because the same trial type, in which both words of a pair had their original durations, was selected twice—once in which the monosyllable was selected to have its original duration and once in which the disyllable was selected to have its original duration.) Each trial type was presented in two orders (monosyllable first, disyllable first), to give 364 trials.

A disadvantage of the pairing scheme used in Experiment 1A was that the monosyllables were longer than the disyllables on most tri-

Figure 2. Durations (in milliseconds) of sample syllable *bar* in monosyllables (top) and disyllables (bottom) in Experiments 1A (A) and 1B (B). Lines linking the rows of durations reflect stimulus pairings. The bolded numbers are the original durations of the monosyllable and the disyllable. Bolded lines are duplicate trials, used only once in each order type—that is, monosyllable first, disyllable first—for each item (see the Method section). The pairings in Experiment 2 were the same as those in Experiment 1B.

als. We anticipated that the effect of this length difference would be offset by the predicted finding that disyllables would be perceived as longer than they were measured to be. Therefore, the proportion of trial types on which the monosyllable *sounded* longer than the disyllable would be reduced, but by an unknown amount. In any case, to address the measured imbalance in trial types, we ran Experiment 1B, in which we paired stimuli so that monosyllables and disyllables were balanced in their durational difference. Now, on each trial, either the disyllable or the monosyllable had the intermediate duration among the seven durations. On one seventh of the trials, the stimuli had the same intermediate duration. On the remaining trials, either the monosyllable or the disyllable was longer by one, two, or three steps. There were 364 trials in this test. Figure 2B shows the trial pairings for Experiment 1B.

For both experiments, test orders were created in which trials were presented in a random order.

Procedure. The participants were run individually. They were told that they would hear two words over headphones—one of them a one-syllable word and the other a two-syllable word. The two words would share a syllable. They were to judge whether the first or second word that they heard had the longer shared syllable. They made their response by hitting a key labeled “1” or “2” on the computer keyboard. They were told that the decision would be difficult on many occasions and that they should guess if necessary. There was an enforced break halfway through the test, but they were invited to take a break at any time by not entering their response. Test trials were presented, and responses were recorded using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993).

Results

The results for the two experiments were similar. In Experiment 1A, in which the monosyllable was longer than the shared syllable of the disyllable on most trials, when the shared syllables in the two words were equal in measured duration, the monosyllable was judged longer on just 13% of the trials.

Figure 3 shows the results. The figure plots the proportion of judgments that the monosyllable was longer than the shared syllable in the disyllable as a function of the

Figure 3. Results of Experiment 1A. The figure shows the proportion of judgments that the monosyllable was longer as a function of relative duration. MS1, the monosyllable was shorter than the shared syllable of the disyllable by one step. ND, there was no difference in duration. ML1–5, the monosyllable was longer by one, two, three, four, or five steps. The data are plotted separately for the trials on which the monosyllable provided the original duration and for those on which the disyllable provided it.

durational difference (relative duration) between them. The figure plots the data separately for trials on which the monosyllable had its original duration and for trials on which the disyllable had its original duration. The figure shows, as was expected, that as the monosyllable grows increasingly longer relative to the disyllable, judgments that it is longer increase. However, it also shows that the location at which the monosyllable is judged longer than the disyllable half of the time occurs when it is measured to be longer by two or three steps.

To estimate the durational difference between the syllables at the point of subjective equality, we collapsed the two curves in Figure 3 and performed simple regression analyses, predicting the proportions of judgments that the monosyllable was longer from the relative durational difference between the disyllable and monosyllable. Linear fits were close for all of the participants, with R s ranging from .95 to .99 (averaging .97). In an items analysis, the outcome was similar (range = .89–.99; average = .97). The point of subjective equality—that is, the duration difference between the monosyllable and disyllable at which listeners judged the monosyllable longer on half of the trials—was determined by setting the y value of the regression equation to .5 and solving for x , the independent variable. The average duration by which the monosyllable exceeded the duration of the disyllable was 99.1 msec (102.4 msec by items). This compares to a measured durational difference of 153 msec, as is shown in Table 1.

We asked whether significant perceptual sensitivity to coarticulatory hiding occurred by testing whether the points of subjective equality for the 12 participants and the 14 items differed from a value of zero. That difference was highly significant in paired t tests [$t_1(11) = 25.49$, $p < .001$; $t_2(13) = 11.37$, $p < .0001$]. We also asked whether the perceptual points of equality differed from the measured duration difference by comparing the points of subjective equality for participants and items with 153 msec. Those results were also significant [$t_1(11) = 8.43$, $p < .001$; $t_2(13) = 5.64$, $p < .0001$].

Another finding reinforces the analyses above in showing that, although listeners compensate for coarticulation, they do not do so fully. When both the monosyllable and the disyllable were at their original durations, the listeners judged the monosyllable as longer 73% of the time. Had the listeners fully compensated for anticipatory shortening (or, as was proposed, for coarticulatory hiding), that value would have been 50%. The difference between 73% and 50% is significant [$t_1(11) = 6.52$, $p < .001$; $t_2(13) = 2.17$, $p = .05$].

The results of Experiment 1B are shown in Figure 4. When the monosyllable and the disyllable had the same duration, the listeners judged the monosyllable as longer just 23% of the time. They judged the monosyllable and disyllable to have the same duration when the monosyllable was just under two steps longer than the disyllable.

Doing the curve-fitting analysis that we did in Experiment 1A on the participants' data gave an estimate of the point of subjective equality of 72.0 msec (95.1 msec in the items analysis). The linear fits were slightly worse than

Figure 4. Results of Experiment 1B. The plot is analogous to that of Figure 1. The x-axis labels differ, because they reflect the different way in which we manipulated relative duration in this experiment. MS1–3, the monosyllable was shorter than the shared syllable of the disyllable by one, two or three steps. ND, there was no difference in duration. ML1–3 means that the monosyllable was longer by one, two, or three steps. The data are plotted separately for the trials on which the monosyllable provided the original duration and for those on which the disyllable provided it.

those in Experiment 1A, but were still quite good (average $R = .95$, range = .92–.99, in the participants analysis; $R = .92$, range = .84–.98, in the items analysis). This result is similar to that of Experiment 1A in showing that the point of subjective equality is not where the monosyllable and the shared syllable of the disyllable are measured to have the same duration but, rather, where the monosyllable exceeds the duration of the disyllable. However, the equality point is at a duration difference less than the measured duration difference between the monosyllables and disyllables produced by our model speaker (and less than the estimate in Experiment 1A).

As for Experiment 1A, we asked whether the 72–(95-) msec points of subjective equality differed from zero; they did [$t_1(12) = 10.08$, $p < .001$; $t_2(13) = 5.05$, $p < .001$]. In addition, we asked whether the point of subjective equality fell below complete compensation by comparing the points of subjective equality with the measured duration difference of 153 msec; this comparison did show a significant difference [$t_1(12) = 7.56$, $p < .001$; $t_2(13) = 31.08$, $p < .001$]. Accordingly, as in Experiment 1A, the participants appear to have heard the stressed syllable in a disyllable as longer than it is measured to be, but not so much so as to make the duration of the stressed syllable in a disyllable perceptually equal to that of the monosyllable.

Discussion

In Experiments 1A and 1B, the stimuli were monosyllables and disyllables with a trochaic stress pattern. The stimuli as they were recorded had the durational patterning that the literature (e.g., Lehiste, 1972) shows is typical of English. That is, shared syllables of monosyllables and trochaic disyllables are measured to be longer in the monosyllables. Experiments 1A and 1B showed that lis-

teners exhibit some insensitivity to this difference. When asked to judge whether a monosyllable or the shared syllable in a disyllable was longer, the listeners' point of subjective equality was located at a duration difference between the shared syllables at which the monosyllable was measured to be longer than the disyllable.

However, as others have shown (Davis et al., 2002; Salverda et al., 2003; Shatzman & McQueen, 2006), the listeners showed some sensitivity to the difference as well. In both experiments, the duration difference at the point of subjective equality was less than the measured duration difference. If the measured shortening reflects coarticulatory hiding, as Fowler (1981b) proposed, one reason that the listeners did not have points of subjective equality at the duration difference measured in speech production in the present experiment may be that the listeners did not hear the hidden portion of the stressed syllable in its entirety. This may occur because production of the stressed syllable wanes (or waxes in an iambic disyllable) during its hidden portion and has a progressively (or initially) less audible impact on the acoustic signal.

In Experiment 2, we explored listeners' tracking of backward shortening—that is, the smaller amount of measured shortening of a stressed syllable when it is produced in a disyllabic word with an iambic stress pattern (e.g., *obey* vs. *bay*).

EXPERIMENT 2

Method

Participants. The participants were 18 students at the University of Connecticut who participated for course credit. They were native speakers of English who reported normal speech and hearing. The data from 2 participants were eliminated, because the participants appeared (and 1 confirmed it during debriefing) to be judging the duration difference between the monosyllable and the entire disyllable rather than the syllable shared with the monosyllable. They judged the disyllable as longer on nearly every trial.

Stimuli. The stimuli were monosyllabic and iambic disyllabic pairs that shared a syllable. They were recorded by a female native speaker of English (author C.A.F.) in a sound-attenuating booth. As in Experiments 1A and 1B, the words were recorded in a carrier sentence fragment: "The next word to say is ____ because." Eighteen word pairs were recorded. From them, we selected the 14 listed with their durations in Table 2. Selection was based on ease of excision of clear tokens from the carrier sentence. Table 2 shows that the monosyllables averaged 451 msec in duration, whereas the shared syllable of the disyllables averaged 433 msec, a difference much smaller than the anticipatory shortening in Experiments 1A and 1B. Moreover, not all items did show shortening. Three showed numerical lengthening, and 1 showed no difference (starred items in Table 2). Despite the reversals, the overall amount of shortening was reliable across the items [$t(13) = 2.16$, $p < .05$]. However, the reversals, as well as the overall considerably weaker backward than anticipatory shortening, were consistent with the previous findings summarized in the introduction.

As in Experiment 1B, we edited the monosyllables and disyllables to create three additional tokens that were progressively longer than the produced duration and three tokens that were shorter. Because the durational shortening in this stimulus set was so small, we chose to use the same intermediate (original) duration for both the disyllable and the monosyllable. For seven of the stimuli, the original duration was that of the monosyllable. For seven, it was that of the disyllable. That monosyllable, or else the shared syllable of the disyllable, was

Table 2
Word Pairs and Their Durations (in Milliseconds)
for the Stimuli of Experiment 2

Monosyllable	Duration	Disyllable	Duration of Shared Syllable
<i>bout</i>	366	<i>about</i>	364
<i>claim</i>	483	<i>proclaim</i>	454
<i>deign</i>	436	<i>mundane</i>	424
<i>feet*</i>	455	<i>defeat</i>	460
<i>go</i>	430	<i>forego</i>	354
<i>moat</i>	440	<i>promote</i>	429
<i>note*</i>	427	<i>denote</i>	461
<i>pain</i>	480	<i>champagne</i>	443
<i>pass</i>	518	<i>surpass</i>	481
<i>rate</i>	387	<i>irate</i>	371
<i>send</i>	490	<i>transcend</i>	477
<i>tail*</i>	442	<i>curtail</i>	442
<i>turn*</i>	463	<i>return</i>	480
<i>veil</i>	497	<i>prevail</i>	422
Mean	451	Mean	433

*Item did not show shortening.

then edited by shortening it or lengthening it to make the remainder of the continuum of durations. If the monosyllable contributed its original duration, after editing, we spliced the unstressed syllable onto the seven versions of the monosyllable to create the disyllabic tokens. If the disyllable provided the original duration, before editing, we excised the unstressed syllables and edited the remaining stressed syllable. Those edited syllables served as our monosyllabic duration continuum. With the unstressed syllable spliced back on, they served as our disyllable continuum. In this experiment, the step size averaged 34 msec. The procedure for adding or subtracting pitch pulses was that of Experiments 1A and 1B.

The stimuli were paired in the same manner as those in Experiment 1B, so that, on each trial, either the monosyllable or the disyllable had the intermediate duration, and the other word had one of its seven durations. As in Experiment 1B, there were 364 trials, in which the monosyllable was presented first on half the trials. Neither the peak nor the mean intensity of the monosyllables and the shared syllables of disyllables differed in paired *t* tests.

Figure 5. Results of Experiment 2. MS1–3, the monosyllable was shorter than the shared syllable of the disyllable by one, two, or three steps. ND, there was no difference in duration. ML1–3, the monosyllable was longer by one, two, or three steps. The data are plotted separately for the trials on which the monosyllable provided the original duration and for those on which the disyllable provided it.

Procedure. The procedure was the same as those of Experiments 1A and 1B.

Results and Discussion

Figure 5 shows the data plotted in the same manner as those in Experiments 1A and 1B. The curves are like those in Figures 3 and 4 in being very similar one to the other; that is, it mattered little whether the monosyllable or the disyllable provided their original duration. In addition, the curves resemble those in Figures 3 and 4 in suggesting a point of subjective equality to the right of the ND location on the *x*-axis.

We performed the curve-fitting procedure of Experiments 1A and 1B and, again, found that linear curves fit the data well (participants' average $R = .98$, range = .95–.99; items' average $R = .98$, range = .93–.99). Predicting the duration difference between monosyllables and their counterparts in disyllables at which the participants' judgments averaged .5 gave an estimate of 12.4 msec (10.8 msec in the items analysis). That is, the monosyllable and the shared syllable in the disyllable had subjectively the same duration when the monosyllable was measured to be longer by about 12 msec. As in the previous experiments, this result underestimates the average measured duration difference of 18 msec, as is shown in Table 2. This result closely mirrors those of Experiments 1A and 1B. The listeners appear to have heard the stressed syllable in a disyllable as having a longer duration than its measured duration.

In both the participant and item analyses, the point of subjective equality significantly exceeded a value of zero [$t_1(15) = 3.77$, $p = .002$; $t_2(13) = 4.35$, $p < .001$]. That is, the participants behaved as if they were sensitive to coarticulatory hiding. In addition, nonsignificantly in the participants analysis and significantly in the items analysis, it fell below the measured shortening value of 18 msec [$t_1(15) = 1.70$, $p = .11$; $t_2(13) = 2.92$, $p = .012$].

GENERAL DISCUSSION

In three experiments, listeners showed that they are sensitive to the durational patterning of stressed syllables as they are produced in a stress-timed language—English. English has left-dominant, maximally binary (strong[–weak]) stress feet, and a possible correlate of that—phonetically—is that measured shortening occurs in a stressed syllable when an unstressed syllable follows. Substantially less shortening occurs when an unstressed syllable precedes a stressed syllable. Metrically, the stressed syllable and a following unstressed syllable are members of the same foot. A preceding unstressed syllable is not a member of a canonical stress foot. The asymmetry in shortening may reflect that difference in constituent membership.

Fowler (1981b) found a correlation between the magnitudes of coarticulation and of compensatory shortening in stressed syllables preceded or followed by one or more unstressed syllables. She proposed that the correlation reflects coarticulatory hiding rather than true shortening. That is, the stressed syllable in *baby* is measured to shorten relative to *bay*, not because it is reduced articula-

torily, but because its waning articulatory edge is hidden by the articulation of the following unstressed syllable. Acoustic measures of anticipatory and carryover coarticulation index that hiding. In this context, carryover coarticulation is more extensive than anticipatory coarticulation (i.e., a talker beginning production of /be/ of *obey* during production of the initial unstressed syllable), perhaps because of the different metrical relations between the stressed syllable and the following versus preceding unstressed syllable.

In the present investigation, we explored listeners' perception of compensatory shortening. A general hypothesis was that the listeners would achieve parity with talkers (Lieberman & Whalen, 2000) in accurately perceiving what the talkers provide as information. Such a finding would be consistent with many others in the study of compensation for coarticulation and the companion finding to it. A more specific hypothesis was that listeners' perceptions can be interpreted as their hearing the stressed syllable as starting during a preceding unstressed syllable and persisting during a following unstressed syllable.

The findings are consistent with the general hypothesis that listeners closely track what talkers do and with the more specific hypothesis that they are sensitive to coarticulatory hiding. In the test for compensation for anticipatory shortening, the listeners' points of subjective equality between the durations of monosyllables and trochaic disyllables occurred when the monosyllable exceeded the duration of the shared syllable of the disyllable by about 84–100 msec (averaging over the participant and item analyses). Likewise, consistent with the smaller effect of backward shortening, the listeners' point of subjective equality between monosyllables and iambic disyllables occurred when the duration of the monosyllables exceeded that of the shared syllable of the disyllables by about 12 msec.

The amount of compensation was less than the amount of measured shortening by a factor of about two (Experiment 1B) or less (about 1.5 in Experiments 1A and 2) in the experiments in which we assessed compensation for anticipatory shortening. One reason for why we saw less than full compensation in Experiments 1A and 1B, as was noted earlier, may be perceptual. As production of *ba* in *baby* winds down during the unstressed syllable, it may come to have inaudible acoustic effects. This interpretation is also consistent with findings cited earlier (Davis et al., 2002; Salverda et al., 2003; Shatzman & McQueen, 2006) that listeners are aware of and can use the durational difference. An analogous account may explain the undercompensation in Experiment 2.

In this study, we extended the findings of compensation for segmental coarticulation (Mann, 1980; Mann & Repp, 1980; and many follow-ups) and compensation for coarticulation involving *f*0 (e.g., Fowler & Brown, 1997; Pardo & Fowler, 1997; Silverman, 1986, 1987) to compensation involving coarticulatory effects that affect measured syllable duration. More than one theoretical account might explain these findings. We have proposed that listeners track speech gestures and that, therefore, they detect coarticulatory overlap as such. An alternative ac-

count might invoke durational contrast (e.g., Kluender, Diehl, & Wright, 1988; but see Fowler, 1992). That is, the short unstressed syllable of *baby* might make the longer stressed syllable sound even longer than it is, leading to a perceived duration closer to that of *bay* than what is actually measured. Correspondingly, the short unstressed syllable of *obey* may make the following stressed syllable sound closer in duration to that of the isolated *bay*.

There are two findings that, although compatible with the contrast account, are not predicted by it. That is, the contrast account is equally compatible with other imaginable configurations of the data in these domains that did not occur. One finding is that compensation is consistently partial. The other is that the effects on trochaic disyllables exceed those on iambic disyllables.

The durational contrast account can accommodate both findings. The contrastive effect is possibly smaller than the shortening effect; perhaps the context (weak-strong) that leads to less shortening also leads to less contrast. The point, however, is that the contrast account is equally compatible with outcomes that did not occur. If it had turned out that compensation consistently overshoot shortening instead of undershooting it, the account could have been that contrast exceeds shortening. If it had turned out that compensation exactly matched shortening, the account could have been that contrast just matches shortening, perhaps because we learn to talk in ways that anticipate auditory interactions with acoustic inputs. As for the directional difference, had it turned out that compensatory effects on iambic disyllables exceeded those on trochaic disyllables, an account could be that contrast has the same directionality.

In any case, the major purpose of the present research was not to distinguish contrast accounts from accounts in which listeners track speech gestures. The purpose was to look for sensitivity to coarticulation in a prosodic domain in which it has infrequently been sought. Effects were found. Like other findings of compensation for coarticulation and its companion finding, the present findings appear to expose listeners' close tracking of what speakers do.

Two further considerations may suggest caution in accepting this interpretation, however. One is that our task was offline. The other, related observation is that it required the participants to make a metalinguistic judgment.

That the task was offline means, among other things, that responses occurred long after the listeners' tracking of gestures (if the listeners tracked gestures) was completed. By the time the listeners made their judgment, for example, lexical access had occurred. However, lexical information is not, in the present account, involved in perceiving the duration of speech gestures. It is logically possible, therefore, that processes taking place after gestural tracking has taken place account for the present findings.

Indeed, the problem may be more severe than that. Because the task was offline, there was room for processes that play no role at all in speech perception outside of the laboratory to play a role in our participants' responding. For example, demand characteristics of the task may have an impact on task performance. That is, listeners may take

a guess, accurate or not, about what the predicted findings are, and that may affect how they respond. Or they may reinterpret our request to judge duration as a request to judge something else.

Finally, that the task called for metalinguistic judgments means that the responses were based on kinds of attention to the speech stimuli that do not characterize speech perception outside the laboratory. How can we know that findings have any bearing on how speech is *perceived* outside the laboratory rather than on how it is *cognized* in this particular laboratory setting?

These concerns cannot be dismissed. Nor, however, should they be used to dismiss the present findings' relevance to speech perception outside the laboratory, unless there is some way to connect these dangers of the artificial nature of the laboratory task to the findings. Our findings were these: (1) The listeners' points of subjective equality suggest that listeners hear the stressed syllables of disyllabic words as longer than they are measured to be; (2) the effect is greater in magnitude for words with a trochaic stress pattern than for words with an iambic pattern; and (3) in both instances, the effects were partial, in that stressed syllables in the disyllables were nonetheless judged shorter than those in the monosyllables when each had its unmodified duration.

What about the fact that the responses were late relative to the gestural tracking that we posit would create this set of outcomes? For example, how might lexical access taking place before the responses occurred create them? It is not at all obvious to us. Imagine that lexical representations are of the sort classically supposed. They are representations in which lexical items are represented as sequences of (timeless) ordered symbols. In that case, the representations should have no biasing effect on responses to our stimuli. Imagine instead that representations are more dynamic in nature, so that gestural overlap is represented in the entries as Browman and Goldstein (1992) might propose. Or suppose, as others might, that gestures are not represented lexically, but the duration differences are. *Bay* has a relatively longer duration lexically than the shared syllable in *baby*. What about that would create our pattern of findings? It is not obvious to us. If, for example, listeners were in that case to use the stored durational differences rather than the differences in the stimuli, two findings are unaccounted for in the present experiments. First, judgments differ depending on shared durational differences in the stimuli. Second, judgments undershoot measured durational differences.

Now consider demand characteristics. On each trial, our participants heard either a monosyllable followed by a disyllable or a disyllable followed by a monosyllable. The items shared a syllable. They were asked whether the first or the second word that they had heard had the longer shared syllable. What might the participants think that we were expecting to find so that they might bias their responding in that direction? It is very doubtful that they guessed that we expected the stressed syllables of disyllables to be perceived as longer than they are measured to be.

Alternatively, how might the participants have reinterpreted our instructions? Perhaps they thought that we

meant for them to judge speaking rate. We doubt this, because we asked them to judge duration, but we cannot rule out that they reinterpreted our instructions as instructions about rate. However, would judgments of rate have led to Findings 1–3 above? It appears not. The syllable that was shared between the monosyllables and the disyllables in the present study were measured to be shorter in the disyllable than in the monosyllable when the two words were produced at the *same rate*. When the monosyllable was lengthened, it might have been perceived as being produced more slowly than the disyllable at its original rate. Accordingly, the participants choosing to judge rate would choose the monosyllable as being produced more slowly than the disyllable when the lengthening was sufficient to be noticed. When it was shortened, it might have been perceived as being produced more quickly than the disyllable at its original duration. But how does this lead to Findings 1–3 above? In particular, consider Finding 2, that the effects were much larger in trochaic disyllable–monosyllable than in iambic disyllable–monosyllable pairings. The step sizes of shortening were comparable in the two experiments (39 msec in Experiments 1A and 1B, 34 msec in Experiment 2); yet the compensation ranged from about 80 to 100 msec in Experiments 1A and 1B, but was just 12 msec in Experiment 2.

In short, although our task was offline and called for a metalinguistic judgment, we do not see how those features of the task could have given rise to the pattern of findings that we obtained. We predicted the pattern that occurred from our theoretical account that listeners closely track gestural overlap in speech perception, a perceptual strategy that optimizes achievement of talker–listener parity. We conclude in favor of that account.

AUTHOR NOTE

The research was conducted with support from NICHD Grant HD-01994 and NIDCD Grant DC-02717 to Haskins Laboratories. Correspondence concerning this article should be addressed to C. A. Fowler, Haskins Laboratories, 300 George St., Ste. 900, New Haven, CT 06511 (e-mail: carol.fowler@haskins.yale.edu).

REFERENCES

- ABERCROMBIE, D. (1967). *Elements of general phonetics*. Chicago: Aldine.
- BROWMAN, C., & GOLDSTEIN, L. (1992). Articulatory phonology: An overview. *Phonetica*, **49**, 155–180.
- CLASSE, A. (1935). *The rhythm of English prose*. Oxford: Blackwell.
- COHEN, J., MACWHINNEY, B., FLATT, M., & PROVOST, J. (1993). PsychoScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, **25**, 257–271.
- DAVIS, M. H., MARSLER-WILSON, W. D., & GASKELL, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 218–244.
- FOWLER, C. A. (1981a). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech & Hearing Research*, **46**, 127–139.
- FOWLER, C. A. (1981b). A relationship between coarticulation and compensatory shortening. *Phonetica*, **38**, 35–50.
- FOWLER, C. A. (1992). Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *Journal of Phonetics*, **20**, 143–165.
- FOWLER, C. A. (2006). Compensation for coarticulation reflects gesture

- perception, not spectral contrast. *Perception & Psychophysics*, **68**, 161-177.
- FOWLER, C. A., & BROWN, J. M. (1997). Intrinsic f_0 differences in spoken and sung vowels and their perception by listeners. *Perception & Psychophysics*, **59**, 729-738.
- FOWLER, C. A., & SMITH, M. R. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. S. Perkell & D. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 123-136). Hillsdale, NJ: Erlbaum.
- KLUENDER, K. R., DIEHL, R. L., & WRIGHT, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, **16**, 153-169.
- LEHISTE, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, **51**, 2018-2024.
- LIBERMAN, A. M., & WHALEN, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, **4**, 187-196.
- LINDBLOM, B., & RAPP, K. (1973). Some temporal regularities of spoken Swedish. *Papers in Linguistics From the University of Stockholm*, **21**, 1-59.
- LOTTO, A. J., & HOLT, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics*, **68**, 178-183.
- MANN, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, **28**, 407-412.
- MANN, V. A., & REPP, B. H. (1980). Influence of vocalic context on perception of the [j]-[s] distinction. *Perception & Psychophysics*, **28**, 213-228.
- PARDO, J. S., & FOWLER, C. A. (1997). Perceiving the causes of coarticulatory acoustic variation: Consonant voicing and vowel pitch. *Perception & Psychophysics*, **59**, 1141-1152.
- PIKE, K. (1945). *Intonation of American English*. Ann Arbor: University of Michigan Press.
- RAMUS, F., NESPOR, M., & MEHLER, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, **73**, 265-292.
- REINHOLT-PETERSON, N. (1986). Perceptual compensation for segmentally-conditioned fundamental-frequency perturbations. *Phonetica*, **43**, 31-42.
- SALVERDA, A. P., DAHAN, D., & MCQUEEN, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, **90**, 51-89.
- SAPIR, S. (1989). The intrinsic pitch of vowels: Theoretical, physiological, and clinical observations. *Journal of Voice*, **3**, 44-51.
- SHATZMAN, K. B., & MCQUEEN, J. M. (2006). Prosodic knowledge affects the recognition of newly-acquired words. *Psychological Science*, **17**, 372-377.
- SILVERMAN, K. (1986). F_0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica*, **43**, 76-92.
- SILVERMAN, K. (1987). *The structure and processing of fundamental frequency contours*. Unpublished doctoral dissertation, Cambridge University, Cambridge, England.
- TURK, A. E., & SAWUSCH, J. R. (1996). The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America*, **99**, 3782-3790.
- VISWANATHAN, N., FOWLER, C. A., & MAGNUSON, J. S. (2009). A critical examination of the spectral contrast account of compensation for coarticulation. *Psychonomic Bulletin & Review*, **16**, 74-79.
- VISWANATHAN, N., MAGNUSON, J. S., & FOWLER, C. A. (in press). Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception & Performance*.
- WHALEN, D. H. (1984). Subcategorical mismatches slow phonetic judgments. *Perception & Psychophysics*, **35**, 49-64.

NOTE

1. For example, the original duration of *bar* was 330 msec; that of *barber* was 190 msec, a difference of 140 msec. That difference was divided by 4 (35 msec) to choose a step size. We lengthened the monosyllable by one step (365 msec) and shortened it by five steps (so the durations were 365, **330**, 295, 260, 225, **190**, and 155 msec; bolded numbers are the original durations of the monosyllable and the disyllable). We shortened the disyllable by one step and lengthened it by five steps to get the same durations as those just listed for the monosyllable (because we made cuts at zero-crossings, step sizes and the durations above are approximate). On each trial, either the monosyllable or the disyllable (and sometimes both) had their original duration. The duration of the other was one of the seven values.

(Manuscript received January 15, 2009;
revision accepted for publication September 26, 2009.)