Some effects of later-occurring information on the perception of stop consonant and semivowel

JOANNE L. MILLER
Northeastern University, Boston, Massachusetts 02115

and

ALVIN M. LIBERMAN
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

In three experiments, we determined how perception of the syllable-initial distinction between the stop consonant [b] and the semivowel [w], when cued by duration of formant transitions, is affected by parts of the sound pattern that occur later in time. For the first experiment, we constructed four series of syllables, similar in that each had initial formant transitions ranging from one short enough for [ba] to one long enough for [wa], but different in overall syllable duration. The consequence in perception was that, as syllable duration increased, the [b-w] boundary moved toward transitions of longer duration. Then, in the second experiment, we increased the duration of the sound by adding a second syllable, [da], (thus creating [bada-wada]), and observed that lengthening the second syllable also shifted the perceived [b-w] boundary in the first syllable toward transitions of longer duration; however, this effect was small by comparison with that produced when the first syllable was lengthened equivalently. In the third experiment, we found that altering the structure of the syllable had an effect that is not to be accounted for by the concomitant change in syllable duration: lengthening the syllable by adding syllable-final transitions appropriate for the stop consonant [d] (thus creating [bad-wad]) caused the perceived [b-w] boundary to shift toward transitions of shorter duration, an effect precisely opposite to that produced when the syllable was lengthened to the same extent by adding steady-state vowel. We suggest that, in all these cases, the later-occurring information specifies rate of articulation and that the effect on the earlier-occurring cue reflects an appropriate perceptual normalization.

In exploratory work with synthetic speech, we chanced on a phenomenon that seemed to hold promise for the study of two related perceptual effects: most directly, how later-occurring aspects of the speech signal modify the perception of an earlier-occurring cue and, by implication, how duration of the syllable specifies, inter alia, the articulatory rate to which the listener must adjust. Following the early findings of Liberman, Delattre, Gerstman, and Cooper (1956), we had used the duration of the initial consonant-vowel transitions to produce the perceived distinction in manner between stop-vowel ([ba]) and semivowel-vowel ([wa]) syllables. Then we varied the duration of the syllable—by extending the steady-state portion of the vowel—and observed that perception

of the transition cue was, in consequence, quite markedly affected: when we made the syllable longer, the manner boundary, as we perceived it, was displaced toward a longer duration of transition.

We were reminded, then, of a series of studies by Summerfield (1975, Note 1) that dealt with the effect of rate of articulation on the perception of the voicing distinction in stop-vowel syllables. Having put his attention on the location of the perceptual boundary along a continuum of voice onset times, a major cue for the distinction in question, Summerfield found that variations in the articulatory rate of the sentence frame caused the perceptual boundary for the target phone to be displaced. More to the point of our interest here, he also found that the rate effect can be quite local, so local, indeed, that it was observed when the syllable containing the target (syllable-initial) phone was isolated and the rate information was conveyed only by variations in the duration of that syllable. If is, of course, just there that Summerfield's results with voicing anticipate ours with manner.

0031-5117/79/060457-09$01.15/0

It is, perhaps, to be expected that a cue like voice onset time should be affected by rate of articulation, for the cue is temporal in nature. Indeed, other temporal cues have been shown to be perceived in relation to speech rate. These include, for example, the following: silence duration as a cue for voicing in intervocalic stop consonants (Port, 1976, Note 2) and as a cue for single vs. double consonants (Pickett & Decker, 1960); frication duration and silence duration as cues for the fricative-affricate manner distinction (Repp, Liberman, Eccardt, & Pesetsky, 1978; Dorman, Raphael, & Liberman, Note 3); and vowel duration as a cue for vowel quality (Ainsworth, 1972, 1974; Verbrugge, Strange, Shankweiler, & Edman, 1976; Verbrugge & Shankweiler, Note 4). Since the transition cue for the [b-w] distinction is essentially temporal, it, too, should be perceived in relation to articulatory rate. That it is, in fact, so perceived is indicated by a study carried out concurrently with those we are reporting here. That study (Minifie, Kuhl, & Stecher, Note 5) found that varying the articulatory rate of a sentence frame did affect whether a word within the sentence was heard as beginning with [b] or [w]. Moreover, investigations into the production of syllable-initial consonants at various rates of articulation have revealed changes in the speed with which the relevant gestures are made and also in the durations of the resulting acoustic transitions (Gay, 1978; Gay & Hirose, 1973; Gay, Ushijima, Hirose, & Cooper, 1974). Given that changes in transition duration occur with changes in rate, we should expect that the listener would make the appropriate normalization when using transition duration to cue a phonetic distinction.

We will here report three experiments designed to enlighten us further about the phenomenon described in our opening paragraph, namely, that the perceived distinction between syllable-initial stop and semivowel, as cued by duration of transitions, is affected by acoustic information that occurs later in time. In the first experiment, we are concerned primarily to establish the phenomenon more securely than our preliminary observations can have done. In the second experiment, we ask whether the effect of later-occuring information is confined to a single syllable. The point of the third experiment is to see if the effect of the later-occurring aspects of the signal is only by changes in syllable duration, or whether it is also influenced by the structure of the syllable.

EXPERIMENT 1

That duration of the steady-state portion of the syllable-final vowel affects the perceived boundary between syllable-initial stop and semivowel was, as we have pointed out, reasonably apparent from our initial observations. But those were based on stimulus variations that were not so systematic as they might have been, and the judgments were made only by us. It is appropriate, then, that we do the experiment properly, the more so in order to delimit the range of durations over which the effect can be found.

Method

Stimuli. The stimuli for all experiments we report in this paper were synthetic speech patterns generated on the Haskins Laboratories parallel-resonance synthesizer. For Experiment 1, we synthesized four series of syllables that ranged from [ba] to [wa]. The syllables were three-formant patterns, consisting of a fixed initial 20 msec of prevoicing (first formant only), a variable duration of formant transition appropriate for [b] or [w], and a subsequent period of steady-state formants. To create each series, we varied the transition duration from 16 to 64 msec in 4-msec steps, for a total of 13 stimuli per series. Syllables with relatively short transitions were perceived as [ba] and those with longer transitions as [wa]. As we increased the duration of transitions in 4-msec steps, we decreased the duration of the steady-state formants by the same amount; thus, within a given series, all stimuli had the same overall duration. The four [ba-wa] series differed from each other in the overall duration of the syllables, specifically, in the duration of the steady-state formants. These syllable durations were: 80, 152, 224, and 296 msec.

For all stimuli, the first formant (F1) started at 234 Hz and rose linearly to a steady-state value of 769 Hz, while the second formant (F2) began at 616 Hz and rose linearly to a final value of 1,232 Hz. The third formant (F3) remained constant at 2,862 Hz. The overall amplitude of each syllable had a gradual onset, increasing by 28 dB over the course of the prevoicing and the formant transitions, and remained constant thereafter.[1]

Each of the 52 stimuli (4 series × 13 stimuli per series) was digitized at a 10-kHz sampling rate by the pulse code modulation (PCM) system at the Haskins Laboratories. These stimuli were then used to make three randomized test orders, each containing eight instances of each stimulus. All stimuli were recorded with an interstimulus interval of 3 sec.

Procedure. All subjects listened to the three test orders over the course of two sessions. They were informed that they would hear computer-generated syllables, [ba]-like or [wa]-like, and of variable duration. They were asked to decide about each syllable whether it was [ba] or [wa], guessing if necessary, and to indicate the decision by marking an appropriately formatted response sheet. The stimuli were presented to the subjects through earphones at approximately 78 dB SPL, measured for the peak intensity of the vowel.

Subjects. The subjects were eight college students who were paid for their participation in the experiment. None reported a history of a speech or hearing disorder.

Results

As is apparent from Figure 1, there was a systematic effect of syllable duration on the identification of the stop consonant [b] and the semivowel [w]: as syllable duration increased, an increasingly longer transition was required to perceive [w]. To obtain a summary account of the effect, we calculated for each subject the location of the [b-w] phonetic boundary for each of the four syllable durations, using the procedure introduced by Eimas, Cooper, and Corbit (1973). That procedure calls for transforming the percentages of [b] responses to z scores, fitting a regression line to the transformed data by finding a
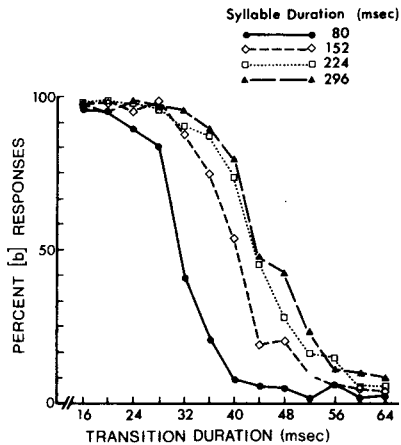
Syllable Duration (msec)
- • 80
- ◇ 152
- □ 224
- ▲ 296

Figure 1. Effect of syllable duration on the [b-w] distinction as cued by transition duration.

Syllable Duration (msec)
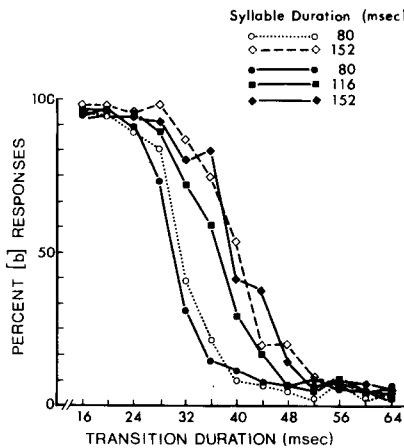- ○ 80
- ◇ 152
- • 80
- ■ 116
- ◆ 152

Figure 2. Replication and extension of the effect of syllable duration on the [b-w] distinction. The data from the main study are shown by open symbols and the data from the auxiliary study by filled symbols.

least-mean-squares solution and then taking the boundary to be the stimulus value corresponding to a z score of zero. Averaged across subjects, the mean boundary locations for the 80-, 152-, 224-, and 296-msec series proved to be at transition durations of 31.9, 41.3, 44.7, and 46.6 msec, respectively. Thus, over the range of syllable durations tested, we obtained a shift of about 15 msec in the location of the boundary. As one can see either by examining the calculated boundaries or by inspecting the identification functions of Figure 1, the largest boundary shifts were at the shortest syllable durations, with increasingly smaller shifts occurring as syllable duration increased. A consequence is that the gap between the functions for the 80- and 152-msec series—about 9 msec of difference in the location of the boundary—is quite large.

In order to fill that gap, and at the same time to test the replicability of our initial findings, we conducted an auxiliary study in which we included the two shortest syllable durations (80 and 152 msec) of the initial study, together with a new condition of syllable duration (116 msec) lying midway between them. (These new syllables were constructed by extending the steady-state vowel of each syllable in the 80-msec series by 36 msec.) The subjects for this auxiliary experiment were the same eight students who had participated in the initial one. In Figure 2, we see the results of the three syllable-duration conditions of the auxiliary experiment and, for comparison, the results of the two corresponding conditions of the initial study. It is plain that the two corresponding conditions of the two studies did indeed produce essentially the same results and that the new, intermediate condition produced an intermediate result.

The combined data from the two studies, shown in Figure 3, clearly indicate that, as syllable duration increases, there is a perfectly regular change in the way our subjects identified the patterns as [ba] and [wa]: the longer the duration of the syllable, the longer the duration of transition needed in order to hear [wa]. Using the same combined data, we calculated, for each subject, the location of the [b-w] boundary for each of the five syllable durations. Those data are presented in Table 1, where we see that the effect of syllable duration, shown in Figure 3, occurred for every subject in nearly every syllable-duration condition. The mean boundary values are shown, as a function of syllable duration, in Figure 4. There we see, as we might have inferred from Figure 3, that the transition duration at the boundary is a smooth and negatively accelerated function of the duration of the syllable.

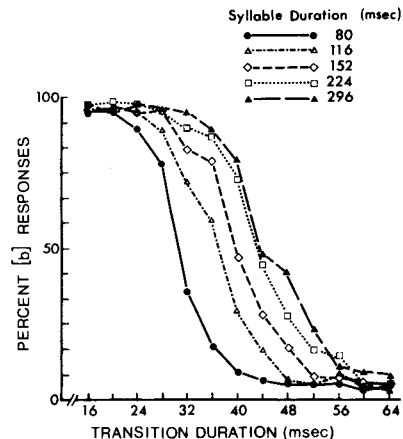Syllable Duration (msec)
- • 80
- ▲ 116
- ◇ 152
- □ 224
- ▲ 296

Figure 3. Combined results of the two studies of the effect of syllable duration on the [b-w] distinction. (The identification functions for the 80- and 152-msec series are based on data from both studies of Experiment 1).
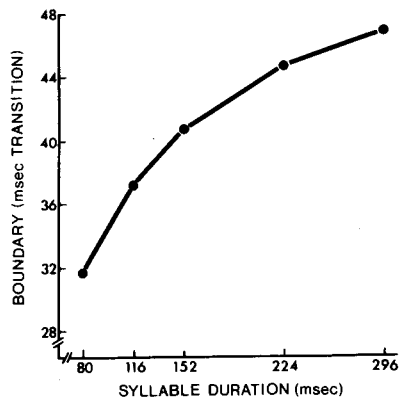
Figure 4. Effect of syllable duration on the location of the [b-w] boundary. (The boundary values for the 80- and 152-msec series are based on data from both studies of Experiment 1).

EXPERIMENT 2

In Experiment 1, we found that later-occurring information in the syllable affected the perception of an earlier-occurring cue: the duration of the syllable determined whether the initial formant transitions of different durations were perceived as [b] or [w]. Now we mean to find out whether such effects are contained within syllable boundaries. What is the effect, if any, of adding a second syllable of variable duration to the one containing the transition cue?

There is reason to believe that the effect of adding a second syllable will be small. We have in mind the experiments by Summerfield (1975, Note 1) described in the introduction to this paper. As the reader may recall, Summerfield found that perception of an important temporal cue for voicing (voice onset time) was affected by variations in the rate at which the utterance was articulated and, further, that the effect was quite local. That is, rate information closer to the target phone had more effect on its perception than rate information that occurred farther away. A similar result has been reported by Port (Note 2) for the intervocalic voicing distinction as cued by duration of intersyllabic silence. If we assume about the results of Experiment 1 that syllable duration had its effect because it specified the rate of articulation, then we might suppose that, following the Summerfield and Port studies, the most important duration would be that of the syllable containing the target phone.

Method

Stimuli. For this experiment, we used stimuli that were identical to some of those used in Experiment 1, except that a [da] was added to each syllable. To create these disyllables, we selected from the stimuli of Experiment 1 just those that had durations of 80 and 224 msec. The reader will recall that stimuli of both durations—let us call them the 80-msec series and the 224-msec series—had initial formant transitions that ranged from 16 to 64 msec in steps of 4 msec, and were perceived, depending on the duration

of the formant transitions, as [ba] or [wa]. For the purposes of this experiment, we added to each of those patterns a synthetic syllable, [da], 72 msec in duration. Thus, we had one series containing stimuli composed of an 80-msec [ba] or [wa] followed by a 72-msec [da], which we will refer to as the 80-72-msec series, and one series with stimuli consisting of a 224-msec [ba] or [wa] followed by a 72-msec [da], which we will refer to as the 224-72-msec series. Next, we created two additional [bada-wada] series by extending the steady-state formants of the [da] so as to make the syllable 216 (instead of 72) msec long. These will be referred to as the 80-216-msec series and the 224-216-msec series. Note that a comparison of performance on the two series containing a short [ba] or [wa] (the 80-72- and 80-216-msec series) with performance on the two series containing a long [ba] or [wa] (the 224-72- and 224-216-msec series) will allow us to assess the effect of lengthening the syllable containing the transition cue for [b-w]. On the other hand, comparing performance on the two series containing the short [da] (the 80-72- and 224-72-msec series) with that on the series containing the long [da] (the 80-216- and 224-216-msec series) will show the effect of lengthening not the target syllable, but the one following.

Each of the two [da]s contained an initial 24 msec of transition followed by steady-state formants, 48 msec in length for the 72-msec [da] and 192 msec in length for the 216-msec [da]. The starting frequency values for the [da] transitions were 234 Hz (F1), 1,541 Hz (F2), and 3,195 Hz(F3), and the steady-state formant frequency values for the [da] were the same as those for the [ba] or [wa], namely, 769 Hz (F1), 1,232 Hz (F2), and 2,862 Hz (F3). As for the amplitude of the [da], it increased by 10.5 dB over the initial transition segment, reaching and maintaining a level equal to that of the first syllable ([ba]) or ([wa]). Fundamental frequency for [da] was also set equal to that of the first syllable.

Each of the 52 stimuli (4 series × 13 stimuli per series) was digitized at a 10-kHz sampling rate, using the PCM system. We then generated three randomized test orders, each containing eight instances of each of these 52 tokens. These orders were recorded on audio tape with an interstimulus interval of 3 sec.

Procedure. The subjects were presented with the three test orders over the course of two sessions. They were informed that they would hear computer-generated disyllables, [bada] or [wada], and that the durations of both syllables would vary. They were asked to decide whether the first syllable of each stimulus was [ba] or [wa], guessing if necessary, and to indicate their choice by writing B or W on an answer sheet. All subjects heard the stimuli through earphones at approximately 78 dB SPL, measured for the peak intensity of the vowel in the first syllable.

Subjects. Fourteen paid listeners participated in this experiment, including three subjects who served as listeners in the first experiment. All were college students or staff who reported no history of a speech or hearing disorder.

Results

In Figure 5 are plotted the data from the four [bada-wada] series. We should first examine the effect on the [b-w] boundary of changing the duration of the first syllable—that is, the one containing the target [b] or [w] phone—while holding constant the duration of the second syllable ([da]). Clearly, lengthening the first syllable (from 80 to 224 msec) shifted the [ba-wa] boundary to a longer duration of transition, and this was true whether the duration of the second syllable was 72 or 216 msec. Calculating the boundary values by the method described in Experiment 1, we obtained values (in milliseconds of transition) for the four disyllables as follows: 80-72 = 33.0, 80-216 = 35.9, 224-72 = 40.7, and 224-216 = 41.9. The average

boundary value for the two series with a short [ba] or [wa] (80-72 and 80-216) is 34.4 msec, and that for the two series with a long [ba] or [wa] (224-72 and 224-216) is 41.3 msec. Thus the boundary shift attributable to the difference in duration of the first syllable is about 7 msec. The reliability of this difference was confirmed by an analysis of variance, First Syllable by Second Syllable by Subject, performed on the boundary scores, that showed a significant effect of first syllable (p < .001) but no significant interaction. Thus, the magnitude of shift resulting from lengthening the first syllable did not depend on the duration of the second syllable.

Consider, next, the effect on the [b-w] boundary of changing the duration of the second syllable ([da]) from 72 to 216 msec. Making the appropriate comparisons in Figure 5, we see that there was, indeed, an effect and, further, that the effect was in the same direction as that produced by variation in the duration of the first syllable. Moreover, the analysis of variance on the boundary scores, mentioned above, showed that the effect of lengthening the second syllable was reliable (p < .01). Again, the lack of a reliable interaction between First Syllable and Second Syllable indicated that the effect of lengthening [da] did not differ as a function of the duration of the first syllable. Of particular relevance to our purposes, however, is the fact that the boundary shift produced by changing the duration of the second syllable, when averaged over the two durations of the first syllable, proved to be only 2 msec, smaller by fair margin than the 7-msec boundary shift produced by varying the duration of the first syllable. A difference in that direction was, in fact, found with 13 of the 14 subjects; the remaining subject showed no difference.

We move now to comparisons that permit us to see more directly whether the location of the [b-w] boundary is determined primarily by the duration of the first syllable, as the results so far suggest, or, alternatively, by the duration of the entire disyllable. For that purpose we reproduce some of the results of Experiment 1 and set them alongside some of the results of Experiment 2, with which we are now concerned. That is done in Figure 6. Now we are able to take advantage of an aspect of our experimental design. It is that the durations of two of the [bada-wada] patterns were so chosen as to create disyllables having the following characteristics: (1) their overall durations (152 and 296 msec) are equal to the durations of two of the four monosyllables of Experiment 1, while (2) the duration of the first syllable of each of these disyllables (80 msec) is equal to another of the monosyllables of Experiment 1. If the boundary is determined primarily by the duration of the first syllable, then the results obtained with both of the disyllables, whether the durations are 80-plus-72 or 80-plus-216, would be like those for the 80-msec condition of Experiment 1. But if it is the duration
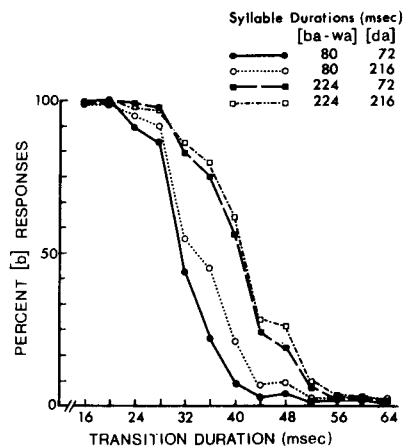

Figure 5. Effect of duration of a first and second syllable on perception of the [b-w] distinction in initial position in the first syllable.
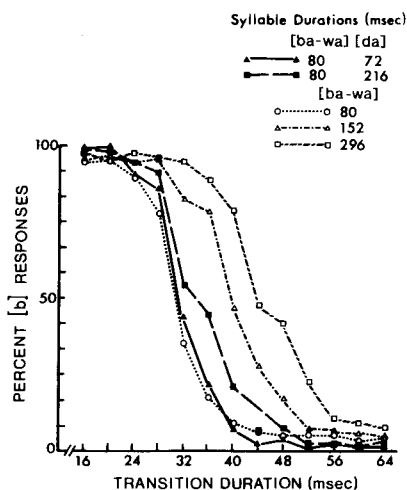

Figure 6. Comparison of the effect of varying the duration of a first and second syllable on perception of the [b-w] distinction in initial position in the first syllable.

of the disyllable that is important, then the results with the disyllables should compare with those obtained in Experiment 1 when the monosyllables were 152 and 296 msec. We see from Figure 6 that, in fact, the results with the disyllables lie quite close to those of the 80-msec monosyllables, closer certainly than to the monosyllables whose durations (152 and 296 msec) are equal to the overall durations of the disyllables. Plainly, then, the location of the boundary for a syllable-initial [b-w] contrast is primarily determined by the duration of the syllable that contains the target phone.

EXPERIMENT 3

In Experiments 1 and 2, we found that perception of a syllable-initial transition-duration cue for the

Syllable Duration (msec)
●———● [ba-wa] 80
○········○ [ba-wa] 116
▲——▲ [bad-wad] 116
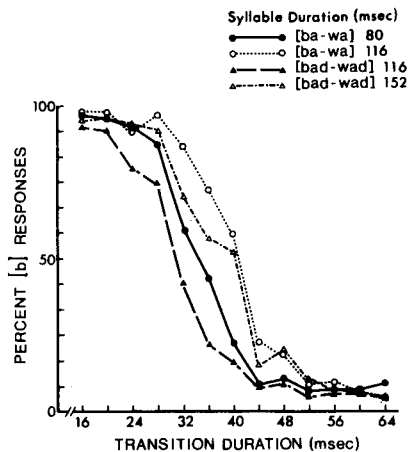▲····-··▲ [bad-wad] 152

Figure 7. Comparison of the effect on the [b-w] distinction of lengthening the syllable by extending the steady-state formants and by adding transitions that cue a final consonant.

distinction between [b] and [w] was affected by the duration of the syllable containing the target phone and, to a lesser extent, by the duration of a following syllable. Putting our attention now again on the syllable containing the target phone, we ask whether its internal structure, as well as its duration, has an effect. To find out, we have compared two ways of changing syllable duration: by extending the steady-state formants, as we had done in the earlier experiments, and by adding transitions appropriate for a syllable-final stop consonant.

If we assume, as we have before, that the effect of syllable duration is via an adjustment (by the listener) for the rate of articulation it specifies, we might suppose that the internal structure of the syllable would have an effect independently of its duration. Thus, for example, two syllables that have the same overall duration but different internal structures—one ending in a voiced stop and the other ending in a vowel—would presumably have been produced at different rates of articulation. For such different syllables, then, the [b-w] boundary should be located at different durations of transitions.

Method

Stimuli. The stimuli for this study comprised four series of synthetic syllables. Two of these, identical with some that were used in Experiment 1, consisted of prevoicing and initial formant transitions (appropriate for [b] or [w]) of variable duration (16 to 64 msec in steps of 4 msec), followed by steady-state formants (appropriate for [a]) of variable duration, to yield CV syllables with total stimulus durations of 80 msec in the case of one series and 116 msec in the other. We will refer to these "old" series as the [ba-wa]-80 and the [ba-wa]-116 series. The two new series were formed by simply adding to the end of each of the "old" patterns 36 msec of formant transitions appropriate for a syllable-final [d]. Consequently, these syllables sounded like [bad] or [wad] and had a duration of 116 msec in the case of the one series ([bad-wad]-116) and 152 msec in the other ([bad-wad]-152). Across the final [d] transition, the first formant fell linearly from its

steady-state value of 769 to 234 Hz, as the second and third formants rose linearly from their steady-state levels of 1,232 and 2,862 Hz to 1,541 and 3,363 Hz, respectively. During this period, the overall amplitude fell 4.5 dB.

Using the PCM system, we digitized the stimuli (10-kHz sampling rate) and then created three test orders, each containing random arrangements of eight tokens of each of the 52 stimulus types (4 series × 13 stimuli per series). The interstimulus interval was 3 sec.

Procedure. The three test orders were presented to the subjects in two sessions. The subjects were told that they would hear one of four syllables—[ba], [wa], [bad], or [wad]—and that these would vary in duration. They were asked to indicate for each syllable whether it began with [b] or [w], and to guess if necessary. The syllables were presented through earphones at approximately 78 dB SPL, measured for the peak intensity of the vowel.

Subjects. The subjects were 10 paid college students who reported no speech or hearing disorders. Two of the listeners had participated in one or more of our previous experiments with [b-w].

Results

Having in mind that half of this experiment was an exact repetition of Experiment 1, we should first make the appropriate comparison of results. For that purpose we look, in Figure 7, at the functions that were obtained with the two [ba-wa] series. We observe that, as in Experiment 1, lengthening the steady-state vowel caused the [b-w] boundary to shift toward a longer duration of transition. As determined by the method referred to in Experiment 1, the phonetic boundaries were found to lie at 36.1 msec for the syllables with overall durations of 80 msec ([ba-wa]-80) and at 41.8 msec for those with durations of 116 msec ([ba-wa]-116). These are to be compared with boundaries of 31.8 and 37.2 msec that were obtained for the same conditions in Experiment 1 (see Table 1). Thus, the magnitude of the boundary shift owing to the 36-msec change in syllable duration was approximately equal in the two experiments, but in Experiment 3 all boundaries fell at longer transition durations. We do not know why. The only differences

Table 1
Individual and Mean [b-w] Boundary Values, in Milliseconds of Transition Duration, for the Several Syllable Durations of Experiment 1

Subject	Syllable Duration				
	80	116	152	224	296
1	28.8	35.2	40.0	47.2	48.0
2	33.6	36.0	40.8	43.2	46.4
3	38.4	45.6	46.4	48.8	48.0
4	28.8	32.0	37.6	40.8	41.6
5	28.8	36.0	39.2	45.6	46.4
6	33.6	36.0	42.4	48.8	50.4
7	28.8	35.2	39.2	44.0	44.8
8	33.6	41.6	41.6	39.2	47.2
Mean	31.8	37.2	40.9	44.7	46.6

Note—The scores for the 80- and 152-msec durations are the averages of the scores from the main and auxiliary studies.

were in the subjects and in the overall contexts in which the stimuli were presented.

In all of the experimental results so far presented, the target syllable—that is, the one containing the syllable-initial [b] or [w]—had the structure CV. It is of interest, then, to examine the effect of varying the duration of the steady-state vowel in a CVC syllable. To see that effect, we look at the two remaining functions in Figure 7—that is, those for the [bad-wad] series having durations of 116 msec ([bad-wad]-116) and 152 msec ([bad-wad]-152). We observe that with the CVC syllable, as with the CV, increasing the duration of the steady-state vowel causes the [b-w] boundary to shift toward longer durations of transition. Calculating these boundaries as we have the others, we obtained locations of 32.5 and 39.4 msec for the [bad-wad]-116 and [bad-wad]-152 series, respectively.

Having now seen the effect of increasing the duration of the syllable by adding steady-state vowel (in both CV and CVC structures), we can turn to the question that is of greatest interest to us in this experiment—namely, how does this effect compare with that which is obtained when the same increases in syllable duration are produced by adding, not steady-state vowel, but syllable-final formant transitions appropriate for a stop? To answer that question, we should use the functions shown in Figure 7 to make two comparisons. The first is between [ba-wa]-80 and [bad-wad]-116. We see, then, that the phonetic boundary for [bad-wad]-116 is at a *shorter* transition duration than that for [ba-wa]-80. Recall, now, that the effect of increasing syllable duration by adding 36 msec of steady-state vowel was to shift the phonetic boundary toward a *longer* duration of syllable-initial transition. Thus, the two ways of increasing syllable duration—adding steady-state vowel or adding syllable-final (stop) transitions—have exactly opposite effects.

The second comparison we want to make is between [ba-wa]-116 and [bad-wad]-152. Here we see the same effect that we observed in the comparison we just made between [ba-wa]-80 and [bad-wad]-116—namely, that lengthening the syllable by adding 36 msec of syllable-final transitions shifted the phonetic boundary for the [b-w] distinction toward shorter durations of syllable-initial transitions. Since we do not have, in this experiment, a condition of [ba-wa]-152, we cannot directly compare the effect of lengthening the syllable by the two different means, as we did above for the cases of [ba-wa]-80, [ba-wa]-116, and [bad-wad]-116. However, we know from the results of Experiment 1 that changing syllable duration from 116 to 152 msec by adding 36 msec of steady-state vowel did, in fact, shift the boundary toward longer durations of transitions. Thus, we have further evidence that adding 36 msec of steady-state vowel and

the same amount of syllable-final transition have opposite effects.[2] This conclusion is supported by an analysis of variance performed on the individual boundary scores, Adding Steady-State Vowel by Adding Formant Transitions by Subject. The opposite effects of Adding Steady-State Vowel and Adding Formant Transitions were both significant (p < .001 and p < .05, respectively), and the interaction between these two effects was not significant (p > .10).

DISCUSSION

Our three experiments show that information occurring later in the speech stream affects the perception of an earlier-occurring cue. When considered, most generally, as an after-going effect, our finding is one of a class that is common in speech perception. For example, it has been found by several investigators that perception of cues for stop and fricative consonants depends in some instances on the nature of the following vowel (e.g., Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Dorman, Studdert-Kennedy, & Raphael, 1977; Fischer-Jorgensen, 1954; Kunisaki & Fujisaki, Note 6; Repp & Mann, Note 7). Moreover, the effect is not confined within a single syllable, since information present in a following syllable can affect the perception of phonetic segments that belong to the preceding one (e.g., Repp, 1978; Repp et.al., 1978; Dorman, Raphael, Liberman, & Repp, Note 8). Such after-going effects are important, because they imply that phonetic perception is not accomplished, phone by phone, in a simple progression through the acoustic signal. Apparently, the perceiver operates over relatively long stretches of sound, integrating into unitary phonetic percepts a numerous variety of acoustic cues that are quite widely distributed in time and thoroughly overlapped with cues for other phones (cf. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

Of the after-going effects reported in the literature, most are associated with the way articulatory and co-articulatory maneuvers smear the acoustic information for the discrete and successive segments of the phonetic message. Thus, the cues for a single phone may be spread through several acoustic segments as, for example, when stop-consonant closure and opening into a following vowel produce a period of silence, a transient burst of sound, a period of frication, some aspiration, and, finally, the onset of voicing, usually at some point during the formant transitions into the vowel (cf. Fant, 1973; Fischer-Jorgensen, 1954; Halle, Hughes, & Radley, 1957). Or, in apparently opposite fashion, coarticulation may cause cues for successive phonetic segments to be collapsed into a single segment of sound and conveyed simultaneously on the same acoustic para-

meter as, for example, when the initial formant transitions convey information both about the initial consonant and the following vowel of a CV syllable (Liberman et al., 1967).

The after-going effects we have reported in this paper are, however, of a different sort, in that they are apparently owing to a different cause: normalization for the consequences of changes in rate of articulation. As we pointed out in the introduction to this paper, there is reason to believe that different rates of articulation produce different durations of formant transition in syllable-initial consonants. There is, moreover, considerable basis for supposing that information about rate of articulation is provided by syllable duration, one of the variables of our experiment (cf. Gay, 1978; Klatt, 1976; Peterson & Lehiste, 1960; Gaitenby, Note 9). We conclude, therefore, that the result of the first two experiments—namely, that the perceived [b-w] boundary shifted as a function of syllable-duration—is to be interpreted as an appropriate adjustment by the listener for changes in articulatory rate.

Having in mind that one of the rate-specifying variables in our experiments was simply duration of the syllable, we should take note here of those cases in which duration is a cue in its own right. There is, for example, the distinction between voiced and voiceless stops in syllable-final position (e.g., [εd] vs [εt]). In that case, the syllable is longer, other things being equal, when the final stop is voiced, and there is evidence that listeners use the duration appropriately in identifying the voicing value of the final segment (e.g., Denes, 1955; Raphael, 1972). Or, as is well known, a similar situation exists for certain vowel distinctions. Thus, [æ] is inherently longer than [ε] and here, too, duration is, per se, a cue for the phonetic identity of the vowel (e.g., Peterson & Lehiste, 1960; Mermelstein, Liberman, & Fowler, Note 10; cf. Verbrugge & Shankweiler, Note 4).

Thus duration can, in fact, specify duration and not, as in our experiment, rate of articulation. Accordingly, it is of interest to know what happens in such cases when, as in our third experiment, duration is increased, not by extending the steady-state vowel but by adding formant transitions appropriate for a stop consonant. Such experiments have shown that adding (stop-consonant) formant transitions has the same effect as adding a certain duration of steady-state vowel (Mermelstein et al., Note 10; Raphael, Dorman, and Liberman, Note 11; cf. Mermelstein, 1978). That is, when duration, qua duration, is being specified, formant transitions contribute to it, just as we should expect, given the impossibility of dividing a syllable into acoustic segments (transitions and steady-state) that correspond one-to-one with the phonetic segments. In contrast, as the reader will recall, we found in our third study that increasing the duration of the sylla-

ble by adding formant transitions had an effect precisely opposite to that produced by adding the same amount of steady state.[3] This suggests that, in our experiments, duration itself was not the cue; its effect was presumably owing instead to its role in specifying rate of articulation. In that case, we should have expected that the structure of the syllable as well as its duration would be important.

The experiments reported here have demonstrated · that the effect of transition duration as a cue for the [b-w] distinction is influenced by the duration and structure of the syllable containing the cue and, to a lesser extent, by the duration of a subsequent syllable. In our view, this after-going effect reflects an adjustment by the listener to the articulatory rate of the speaker: the duration and structure of the syllable provide information about rate, and the listener uses this information when making a phonetic judgment of [b] vs. [w].

REFERENCE NOTES

1. Summerfield, A. Q. *On articulatory rate and perceptual constancy in phonetic perception.* Unpublished manuscript, 1978.

2. Port, R. F. *Effects of word-internal versus word-external tempo on the voicing boundary for medial-stop closure.* Paper presented at the 95th meeting of the Acoustical Society of America, Providence, R.I., 1978.

3. Dorman, M. F., Raphael, L. J., & Liberman, A. M. *Further observations on the role of silence in the perception of stop consonants.* Paper presented at the 91st meeting of the Acoustical Society of America, Washington, D.C., 1976.

4. Verbrugge, R. R., & Shankweiler, D. P. *Prosodic information for vowel identity.* Paper presented at the 93rd meeting of the Acoustical Society of America, University Park, 1977.

5. Minifie, F., Kuhl, P., & Stecher, B. *Categorical perception of [b] and [w] during changes in rate of utterance.* Paper presented at the 94th meeting of the Acoustical Society of America, Miami, 1977.

6. Kunisaki, O., & Fujisaki, H. *On the influence of context upon perception of voiceless fricative consonants.* Research Institute of Logopedics and Phoniatrics, University of Tokyo, Annual Bulletin No. 11, 1977, 85-91.

7. Repp, B. H., & Mann, V. A. *Influence of vocalic context on perception of the [s]-[ʃ] distinction.* Paper presented at the 96th meeting of the Acoustical Society of America, Honolulu, 1978.

8. Dorman, M. F., Raphael, L. J. Liberman, A. M., & Repp, B. *Some maskinglike phenomena in speech perception.* Paper presented at the 89th meeting of the Acoustical Society of America, Austin, Tex., 1975.

9. Gaitenby, J. *The elastic word.* In Haskins Laboratories Status Report on Speech Research, SR-2, 1965.

10. Mermelstein, P., Liberman, A. M., & Fowler, A. *Perceptual assessment of vowel duration in consonantal context and its application to vowel identification.* Paper presented at the 94th meeting of the Acoustical Society of America, Miami, 1977.

11. Raphael, L. J., Dorman, M. F., & Liberman, A. M. *The perception of vowel duration in VC and CVC syllables.* Paper presented at the 89th meeting of the Acoustical Society of America, Austin, Tex., 1975.

12. Suzuki, H. *Mutually complementary effect of rate and amount of formant transition in distinguishing vowel, semi-vowel, and stop-consonant.* Research Laboratory of Electronics, Quarterly Progress Report No. 96, 1970, M.I.T., 164-172.

REFERENCES

AINSWORTH, W. A. Duration as a cue in the recognition of synthetic vowels. *Journal of the Acoustical Society of America*, 1972, **51**, 648-651.

AINSWORTH, W. A. The influence of precursive sequences on the perception of synthesized vowels. *Language and Speech*, 1974, **17**, 103-109.

COOPER, F. S., DELATTRE, P. C., LIBERMAN, A. M., BORST, J. M., & GERSTMAN, L. J. Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, 1952, **24**, 597-606.

DENES, P. Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, 1955, **27**, 761-764.

DORMAN, M. F., STUDDERT-KENNEDY, M., & RAPHAEL, L. J. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, 1977, **22**, 109-122.

EIMAS, P. D., COOPER, W. E., & CORBIT, J. D. Some properties of linguistic feature detectors. *Perception & Psychophysics*, 1973, **13**, 247-252.

FANT, G. *Speech sounds and features*. Cambridge: MIT Press, 1973.

FISCHER-JORGENSEN, E. Acoustic analysis of stop consonants. *Miscellanea Phonetica*, 1954, **2**, 42-59.

GAY, T. Effect of speaking rate on vowel formant transitions. *Journal of the Acoustical Society of America*, 1978, **63**, 223-230.

GAY, T., & HIROSE, H. Effect of speaking rate on labial consonant production. *Phonetica*, 1973, **27**, 44-56.

GAY, T., USHIJIMA, T., HIROSE, H., & COOPER, F. S. Effect of speaking rate on labial consonant-vowel articulation. *Journal of Phonetics*, 1974, **2**, 47-63.

HALLE, M., HUGHES, G. W., & RADLEY, J. P. A. Acoustic properties of stop consonants. *Journal of the Acoustical Society of America*, 1957, **29**, 107-116.

KLATT, D. H. Linguistic uses of segmental duration in English. *Journal of the Acoustical Society of America*, 1976, **59**, 1208-1221.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431-461.

LIBERMAN, A. M., DELATTRE, P. C., GERSTMAN, L. J., & COOPER, F. S. Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology*, 1956, **52**, 127-137.

MERMELSTEIN, P. On the relationship between vowel and consonant identification when cued by the same acoustic information. *Perception & Psychophysics*, 1978, **23**, 331-336.

PETERSON, G. E., & LEHISTE, I. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 1960, **32**, 693-703.

PICKETT, J. M., & DECKER, L. R. Time factors in perception of a double consonant. *Language and Speech*, 1960, **3**, 11-17.

PORT, R. F. *The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words*. Unpublished PhD thesis, University of Connecticut, 1976.

RAPHAEL, L. J. Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America*, 1972, **51**, 1296-1303.

REPP, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*, 1978, **24**, 471-485.

REPP, B. H., LIBERMAN, A. M., ECCARDT, T., & PESETSKY, D. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, **4**, 621-637.

SUMMERFIELD, A. Q. *Information processing analysis of perceptual adjustments to source and context variables in speech*. Unpublished PhD thesis, Queen's University of Belfast, 1975.

VERBRUGGE, R. R., STRANGE, W., SHANKWEILER, D. P., & EDMAN, T. R. What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America*, 1976, **60**, 198-212.

NOTES

1. Since the starting and terminating formant-frequency values of the transition segment were kept constant as its duration was changed, its rate was necessarily changed as well. That is, as transition duration varied from short ([ba]) to long ([wa]), transition rate varied from fast to slow. Given that Liberman et al. (1956) have shown that transition duration, and not rate, appears to be the effective cue for the [b-w] contrast, we will refer to the stimulus manipulation in our experiments as one of duration (cf. Suzuki, Note 12).

2. Summerfield (Note 1) has reported a similar finding for the syllable-initial voiced-voiceless boundary as cued by voice onset time (VOT). Specifically, he found that the boundary was shifted toward a longer VOT value as the syllable was lengthened by extending the steady-state vowel, but that it was shifted toward a shorter value when the syllable was lengthened by adding a final fricative.

3. We should point out a difference between the experiments of Mermelstein et al. (Note 10) and Raphael et al. (Note 11), on the one hand, and those conducted by us, on the other. In their experiments, the transitional information was added to the beginning of the syllable (thus, for example, changing [ɛd] vs. [ɛt] to [dɛd] vs. [dɛt]), whereas we added the consonantal transitions to the end of the syllable (so that [ba] vs. [wa] became [bad] vs. [wad]). Although unlikely, it may be that the added transitions functioned differently in their experiments and ours because of this difference in where in the syllable they were added.