Limitations of context conditioned effects in the perception of [b] and [w]

P. C. SHINN, S. E. BLUMSTEIN, and A. JONGMAN Brown University, Providence, Rhode Island

Two series of experiments were performed to test the effect of varying vowel length (reflecting different speaking rates) on the crossover boundary of a [b-w] continuum. Earlier research (Miller & Liberman, 1979) indicated context-dependent perception of [b] and [w] as a function of syllable duration. The object of our study was to determine whether the syllable duration effects shown for [b-w] would be lost if the parameter values of the test stimuli were more similar to natural speech. Different combinations of acoustic parameters were manipulated, as was the method of stimulus presentation. Results showed that the magnitude of the syllable-duration effect was greatly reduced and that the effect was eventually lost as the parameters increasingly approximated natural speech. In addition, the syllable-duration effect was differentially affected depending on the experimental design being used: in most experimental conditions the syllable-duration effect was more affected when the stimuli were presented in a blocked, as compared to a mixed, design. Nevertheless, when parameters similar to natural speech were used, the syllable-duration effect was completely lost, regardless of which experimental design (mixed vs. blocked) was used. These results suggest that the syllable-length effect for the perception of the [b-w] contrast may not be as robust as originally believed, and, furthermore, that the role of context-dependent cues in the perception of speech may have been overestimated.

Research in speech perception has emphasized the importance of context-dependent cues for the perception of the phonetic dimensions of speech. While there is little reason to doubt the importance of these cues, our own research strategy has emphasized the possible role of invariant, context-independent acoustic properties in speech processing. As part of our research methodology, we have conducted a number of detailed acoustic analyses of the characteristics of natural speech (Blumstein & Stevens, 1979; Lahiri, Gewirth, & Blumstein, 1984; Mack & Blumstein, 1983; Stevens & Blumstein, 1981). These analyses have revealed that acoustic characteristics of natural speech are often quite different from those of the synthetic stimuli used in speech perception experiments.

The parameters of the synthetic stimuli seem to vary in at least two ways from natural speech. In some cases, they are simplified versions of natural speech. For example, a place of articulation continuum for stop consonants might contain only formant transitions and no burst. In other cases, however, the synthetic stimuli compromise the acoustic characteristics of natural speech. As a consequence, some parameters used are, in essence, inappropriate for the particular phonetic dimensions being explored. For example, [b-w] continua endpoints usually have identical formant frequencies, and yet natural-speech [b] and [w] differ along a number of acoustic dimensions, including formant frequencies (Dalston, 1975; Fant, 1960; Fischer-Jorgensen, 1954; Lehiste, 1964; Lehiste & Peterson, 1961; Mack & Blumstein, 1983; Pickett, 1980). On the basis of these observations, we question whether a number of the context-dependent effects shown for the perception of the phonetic dimensions of speech are as robust as has been suggested, and, in particular, whether these effects would still emerge if the synthetic stimuli contained parameters more similar to those found in natural speech.

The [b-w] contrast is a good case in point. Speech perception research using synthetic stimuli has shown that the perception of [b] and [w] is significantly affected by phonetic context. In particular, the locus of the phonetic boundary for a [b-w] continuum varies across vowel context as a function of different transition slopes and rates intrinsic to these phonetic contexts (Cooper, Ebert, & Cole, 1976; Diehl, 1976; Godfrey & Millay, 1981; Liberman, Delattre, Gerstman, & Cooper, 1956; O'Connor, Gerstman, Liberman, Delattre, & Cooper, 1957; Schwab, Sawusch, & Nusbaum, 1981). Nevertheless, the synthetic speech stimuli used to investigate this effect did not systematically vary at least one critical acoustic parameter-namely, the amplitude envelope in the vicinity of the consonant release. Acoustic analyses of natural speech have revealed stable, context-independent acoustic properties for the [b-w] contrast related to the transient versus gradual onset of energy at the release of stops and glides, respectively (Mack & Blumstein, 1983). Further research has shown that these properties are perceptually salient and can override the context-dependent cues

We would like to thank John Mertus for development of the synthesizer and wave-editing program and Molly Mack for measurements of natural speech and for synthesizing the first set of stimuli. This research was supported in part by Grant NS15123 to Brown University. Address correspondence to Sheila Blumstein, Department of Linguistics, Brown University, Providence, RI 02912.

of formant transition rate and duration (Shinn & Blumstein, 1984).

Since a number of studies have indicated that speaker rate affects syllable and segmental duration (Gay, 1978; Gay & Hirose, 1973; Gay, Ushijima, Hirose, & Cooper, 1974: Kozhevnikov & Chistovich, 1965: Peterson & Lehiste, 1960), it was natural to inquire whether speaker rate would affect the segmental boundary of a [b-w] continuum. Miller and Liberman (1979) demonstrated context-dependent perception for the [b-w] contrast when they showed that the locus of the phonetic boundary of a [b-w] continuum in which the stimuli varied in duration and slope of the formant transitions changed as a function of the syllable length, and in particular, of the vowel duration. As the vowel steady-state decreased, listeners required a shorter formant transition duration to perceive a [w]. Miller and Liberman interpreted their results as evidence for speaker-rate normalization in ongoing speech.

This effect has proven to be a stable perceptual phenomenon, having been shown in a number of different experimental manipulations (Miller, 1980a, 1980b, 1981), not only in adults, but also in infants (Eimas & Miller, 1980; Miller & Eimas, 1983), and not only for formant-driven speech stimuli but also for sine-wave stimuli (Jusczyk, Pisoni, Reed, Fernald, & Myers, 1983; Pisoni, Carrell, & Gans, 1983).

In this study, we report the results of two sets of experiments designed to explore whether rate-dependent effects for the perception of [b-w] would be affected if the synthetic speech stimuli had acoustic characteristics more similar to natural speech (see also Maxwell & Landahl, 1983). In the first set of experiments, stimuli were presented in a blocked design, and in the second set of experiments, they were presented in a mixed design.

SERIES 1

In the first set of experiments, four stimulus conditions were employed, in which the acoustic parameters of the synthetic stimuli were systematically varied. In Condition 1, a set of stimuli were created similar to those used by Miller and Liberman, in order to replicate the syllableduration effect. The stimuli contained no bursts, shared formant frequency characteristics, and had the same transition lengths for Formants 1–3, but varied in duration and rate of the formant transitions.

The remaining three stimulus conditions were designed to systematically increase the naturalness of the test stimuli. The stimuli used in Condition 2 were designed to investigate the perceptual effects of varying amplitude characteristics in the vicinity of the consonantal release. Recent research on acoustic invariance for the stop-glide contrast has emphasized the importance of this parameter (Mack & Blumstein, 1983; Shinn & Blumstein, 1984). For stops, there is an abrupt amplitude increase at the release of the closure, whereas for glides, there is a more gradual increase. One of the acoustic cues contributing to this amplitude difference is the presence of a burst at the release of a stop consonant and the absence of a burst at the release of a glide. In order to increase the naturalness of the stimuli, in Condition 2 a burst was added to the end-point [b] stimulus and was systematically decreased along the continuum until it was absent at the [w] end of the continuum.

The acoustic analysis of natural speech shows that, in contrast to the stimulus parameters used in Conditions 1 and 2, the formant frequency characteristics of [b] and [w] are not the same (Mack & Blumstein, 1983). Consequently, in Condition 3, we used all of the stimulus parameters of Condition 1, except that the formant frequencies of the transitions varied along the continua.

Finally, in Condition 4, the stimulus parameters were derived directly from natural speech measures. Stimuli along the continuum varied in onset formant frequencies, duration and rate of the formant transitions, and amplitude characteristics at consonantal release.

Method

Subjects. A total of 40 college students were randomly assigned to one of the four stimulus conditions, resulting in 10 subjects per stimulus condition. All subjects were paid for their participation. None reported any history of a speech or hearing disorder.

Stimuli. Stimuli for all conditions were five formant patterns generated at the Brown University Phonetics Laboratory using a computer simulation of a terminal analog speech-synthesizer in which the tuned circuits for the vowel-generating portion were connected in cascade (Klatt, 1980). All stimuli in this set of experiments were generated at 10 kHz, and the output was low-pass filtered with a cutoff frequency of 4800 Hz. The stimuli for Condition 1 were patterned after those used by Miller and Liberman (1979). Parameters for formant frequencies, amplitude of voicing, fundamental frequency, and the bandwidths of the formants were identical for the stimuli across the continuum, whereas transition durations varied. Formant 1 (F1) began at 234 Hz, stayed flat for 40 msec, then rose to 769 Hz, and stayed flat until the end of the syllable. Similarly, for both endpoints, F2 began at 616 Hz for the first 40 msec, then rose to 1232 Hz and stayed flat until the end of the syllable. F3, F4, and F5 were kept constant at 2862, 3500,

Table 1 Synthesizer Input Parameters for Condition 1 Endpoint Stimuli Over Time (in msec)

	F1		F2		F3		F	4	F5	
Time	[b]	[w]	[b]	[w]	[b]	[w]	[b]	[w]	[b]	[w]
. 0	234	234	616	616	2862	2862	3500	3500	4500	4500
40	234	234	616	616						
45	412	275	821	663						
50	591	316	1027	711						
55	769	357	1232	758						
60		399		806						
65		440		853						
70		481		900						
75		522		948						
80		563		995						
85		604		1042						
90		646		1090						
95		687		1137						
100		728		1185						
105		769		1232						
110										
317	769	769	1232	1232	2862	2862	3500	3500	4500	4500

Note-F1, F2, F3, F4, and F5 are the frequencies of the first five formants.

Table 2 Synthesizer Input Parameters for Condition 2 Endpoint Stimuli Over Time (in msec)											
	Α	V	A	F	A	.2	A3		А	4	
Time	[b]	[w]									
0	40	40	0	0	0	0	0	0	0	0	
25	51	51	0	0	0	0	0	0	0	0	
30	0	53	50	0	55	0	45	0	30	0	
35	50	55	0	0	0	0	0	0	0	0	
40	55	55									

Note-AV is amplitude of voicing, AF is amplitude of frication, and A2-A4 are burst amplitudes. Missing time values (or blanks) are predicted by linear interpolation. Synthesizer input values updated every 5 msec.

0 0

0 0

0

0 0

445

55 55

 Table 3

 Synthesizer Input Parameters for Condition 3

 Endpoint Stimuli Over Time (in msec)

	F	71	F	2	F3	⁷ 3
Time	[b]	[w]	[b]	[w]	[b]	[w]
0	274	346	500	500	1500	1800
40	554	424	500	717	1500	2029
55	653		1139		2351	
105		653		1139		2351
445	653	653	1139	1139	2351	2351

Note-F1, F2, and F3 are the frequencies of the first three formants.

and 4500 Hz, respectively (see Table 1). The bandwidths of F2 and F3 were widened for the first 35 msec of the stimulus to simulate prevoicing. For all stimuli, the fundamental frequency was flat at 115 Hz. F1 and F2 transition durations varied across the continuum from 15 msec (Stimulus 1, [ba]) to 65 msec (Stimulus 11, [wa]) in 5-msec steps.

The 11-member basic continuum was edited to produce five syllable-length conditions. The syllable durations were 102, 136, 171, 248, and 317 msec. These were chosen to be similar to those used by Miller and Liberman (1979), but with 20 msec added, since the prevoicing of our stimuli was 20 msec longer than that used by Miller and Liberman.

For Condition 2, the stimulus parameters were the same as those used in the first condition, except that a 5-msec burst was appended 10 msec before the beginning of the formant transitions for the [b] exemplar stimulus. In addition, the amplitude of voicing (AV) parameter was changed for the exemplar stimulus to simulate closure and release. These parameters are shown in Table 2. The 11 stimuli on the continuum were generated by interpolating between the exemplar parameter values. The various syllable-length conditions were the same as those used in Condition 1.

In Condition 3, the stimulus parameters were exactly the same as those used in Condition 1, except that the formant transitions for F1-F3 were different for the endpoint stimuli. These values are given in Table 3. The formant frequency values used were schematized so that the changes in formant frequency were consistent between endpoints. To this end, the prevoicing values and steadystates were taken from the natural speech measures, and the formant frequency transitions were interpolated between these values. Although it provided naturalistic endpoints, this procedure did, in the final analysis, compromise the F2 and F3 frequency values used in the endpoints. Although both endpoints had different values for formant transition onset, contrary to natural speech, the glide had higher transition-onset frequencies for F2 and F3 than did the stop. The intermediate stimuli on the 11-member continuum were created by interpolating between the endpoint values. The construction of the five syllable-length continua was the same as in the previous two conditions.

Finally, for Condition 4, stimuli were constructed by first mak-

ing [ba] and [wa] endpoints using parameters derived from measurements of natural speech (Mack & Blumstein, 1983). Intermediate stimuli for the basic continuum were made by interpolating between these endpoints. The synthesizer input parameters for the basic endpoints are given in Table 4.

The formant frequency, formant duration, burst frequency, and burst duration values for the stimuli were taken from Mack & Blumstein (1983). As can be seen in Table 4, the formant frequency values changed throughout the stimuli. For both [ba] and [wa] stimuli, prevoicing was 40 msec. To simulate prevoicing, the bandwidths of F2 and F3 (B2 and B3) were widened for the first 35 msec of the stimulus. For the stop [ba], a 5-msec burst was appended to the transitions. This burst was synthesized by turning on the amplitude of frication for 5 msec at 40 msec. The AV was shut off at 40 msec, and then was abruptly turned back on 10 msec later to simulate closure and release. Transitions began at 50 msec for both endpoints. For [ba], the F1 transition was 40 msec long, whereas F2 and F3 transitions were 15 msec long. The endpoint syllable [wa] had 40 msec of prevoicing, 10 msec of glide, a 65msec F1 transition, and 95-msec F2 and F3 transitions. The AV for the glide endpoint rose gradually. The frequency at the end of the transitions was identical for both syllables: F1 of both [ba] and

Table 4 Synthesizer Input Parameters for Condition 4 Endpoint Stimuli Over Time (in msec)

								•		·		
	F 1		F2			F3				F5		
Time	[b]	[w]	[Ь]	[w]]	[b]	[w]	[b]	[w	_] [b]	[w]
0	274	346	500	50	0 1	500	1300	3500	250	0 45	500	4500
35	252	412	500)	1	500			278	0		
40	466	421	1248	5	2	366			282	0		
45	466		1248	;	2	366			286	0		
50	554		1100)	2	342			290	0		
65			1139)	2	351			302	0		
70		431		80	4		2321		306	0		
90	653								322	0		
135		653							350	0		
165		•		113	9		2351					
445	675	675	1073	107	3 2	400	2400	3500	350	0 45	500	4500
	Ā	AV Fo		Fo	_		B1			B3		
Time	[b]	[w]	[b]	[w]]	[b]	[w]	[b]	[w] [b]	[w]
0	40	35	95	; 9	5	90	90	500	500 5		00	500
35	38	48	95	; 9	9	90	90	500	500 5		00	500
40	0	50	125	5 10	0	90	90	110	11	01	70	170
45	0	51										
50	52	52	120) 12	0							
60			115	5 11	5							
65	57	54										
80	57	57										
445	57	57	100) 10	0	90	90	110	110	0 1	70	170
	A	F	Α	A1		A2 A		.3A4		.4	A5	
Time	[b]	[w]	[b]	[w]	[b]	[w]	[b]	[w]	[b]	[w]	[b]	[w]
0	0	0	0	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0
40	48	0	30	0	55	0	47	0	38	0	0	0
45	0	0	30	0	55	0	47	0	38	0	0	0
50	0	0	0	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0	0	0
65	0	0	0	0	0	0	0	0	0	0	0	0
80	0	0	0	0	0	0	0	0	0	0	0	0
445	0	0	0	0	0	0	0	0	0	0	0	0

Note-F1, F2, F3, F4, and F5 are the first five formant frequencies; AV is amplitude of voicing; F0 is fundamental frequency; B1, B2, and B3 are the bandwidths of F1, F2, and F3; AF is amplitude of frication; A1-A5 are burst amplitudes. Missing time values (or blanks) are predicted by linear interpolation. Synthesizer input values updated every 5 msec. [wa] ended its transition at 653 Hz, F2 at 1139 Hz, and F3 at 2351 Hz. The steady-state values were extended to produce [ba] and [wa] syllables of 445 msec. F4 was held constant at 3500 Hz for [ba]; for [wa], F4 started at 2500 Hz and rose gradually to reach a steady-state value of 3500 Hz at 125 msec. F5 was held constant at 4500 Hz. Intermediate stimuli were made by interpolating the endpoint parameter values to produce a continuum of 11 stimuli. For this continuum, the transition duration thus varied from 15 to 65 msec in 11 5-msec steps. The steady-state values of this 11-member continuum were then edited to produce four other continua which differed by syllable length. In total, five continua were generated, with syllable lengths of 85, 175, 265, 355, and 445 msec. These syllable lengths encompass those of Miller and Liberman (1979), who used continua with syllable durations ranging from 80 to 296 msec.

For each of the four stimulus conditions, five separate tapes were made, one per syllable-length condition. Each tape contained 10 repetitions of each stimulus, resulting in 110 stimuli per tape. Each tape had a different stimulus randomization. The interstimulus interval was 2 sec, with a 6-sec pause after every 10 stimuli. Two repetitions of each stimulus preceded each test tape, to familiarize subjects with presentation rate and stimuli.

Procedure. Ten subjects were randomly assigned to each stimulus condition. For each condition, they were tested in two groups of 5, with one group receiving the test tapes in the order of shortest to longest syllable duration, and the other group the reverse order. Subjects were told they would hear [ba]-like and [wa]-like syllables, and were to circle the appropriate letter on the answer sheets provided. If unsure, they were to guess. Each tape took about 8 min, and with breaks during tape changing, the entire session lasted about



Figure 1. Total categorization results from Series 1, showing the number of [b] responses for the four stimulus conditions as a function of stimulus number and syllable length. The top left panel shows the results for Condition 1, the top right panel for Condition 2, the bottom left panel for Condition 3, and the bottom right panel for Condition 4 (see text). The abscissa of each panel shows the stimulus number, with Stimulus 1 at the stop end of the continuum and Stimulus 11 at the glide end. The five separate categorization functions on each panel represent the five syllable-length conditions, with 1 being the shortest and 5 the longest. Syllable-length (SL) Condition 1 is represented by the line of dots with 1s, Condition 2 by dots and dashes and 2s, Condition 3 by a line of short dashes and 3s, Condition 4 by longer dashes and 4s, and Condition 5 by a solid line and 5s.

I h per stimulus condition. Tapes were played on an MCI tape recorder through a NAD amplifier and an MPC signal splitter to five AKG headphones. Each headphone channel was balanced, and stimuli were played at a comfortable listening level.

Results

Figure 1 shows the total number of [b] responses for the five syllable-length (SL) conditions as a function of stimulus number across the four stimulus conditions. The syllable-length conditions are numbered 1 to 5, with the shortest (85 msec) being SL Condition 1 and the longest (445 msec) being SL Condition 5. As the figure shows, syllable-length effects occurred in some stimulus conditions and not in others.

To explore the statistical significance of these results, crossover boundaries were derived by transforming the results to z-scores and performing a linear regression analysis. A two-way repeated measures analysis of variance (ANOVA) with one grouping factor (stimulus condition) and one within factor (syllable length) was then conducted. Results showed significant main effects for both stimulus condition [F(3,36) = 4.09, p < .02] and syllable length [F(4,144) = 11.15, p < .0001], as well as a stimulus condition × syllable length interaction [F(12,144) = 3.67, p < .001], confirming that the crossover boundaries for the [b-w] continua did indeed vary as a function of stimulus condition and syllable length. To explore these effects in more detail, each stimulus condition was then analyzed separately.

In Condition 1, as the top left panel of Figure 1 shows, responses to SL Conditions 1 through 4 occurred in the predicted direction, with fewer [b] responses in the shorter syllable conditions and increasingly more [b] responses as syllable length increased. However, contrary to expectations, the longest syllable length (SL Condition 5) did not produce the greatest number of [b] responses. A one-way repeated measures ANOVA across the five SL conditions yielded a significant effect [F(4,36) = 11.73, p < .001]. Post hoc Newman-Keuls tests revealed that there were significantly fewer [b] responses in the shortest SL condition than in the other SL conditions (p < .01). All other comparisons were nonsignificant, although comparisons of SL Conditions 2 and 4 and SL Conditions 2 and 5 approached significance.

Nevertheless, analysis of individual labeling functions indicated that there was a test-order effect. Figure 2 shows these results. The top panel shows the scores of the group that received SL Condition 1 first, and the bottom panel shows the scores of the subjects who received SL Condition 5 first. As the figure shows, the unexpected pattern in the group data for SL Condition 5 is due to the test order effects, and occurred solely for those subjects who received the tapes in the order of longest to shortest. A two-way repeated measures ANOVA with one grouping factor (test order) and one within factor (syllable length) yielded a significant effect for syllable length [F(4,32) =25.33, p < .0001 and a significant test order \times syllable length interaction effect [F(4,32) = 11.37], p < .0001, indicating that the test-order effect was statistically reliable.



Figure 2. Categorization results from Condition 1, by group. The top functions are for the group receiving the tapes in order of shortest to longest syllable-length condition, the bottom functions are the opposite order.

The pattern of results obtained in Condition 1 was similar to that obtained by Miller and Liberman (1979). Perception of a [ba-wa] continuum varying only in transition slope and duration was significantly affected by the duration of the syllable. Although there was a significant difference only between the shortest SL condition and the other SL conditions, there was a trend in the expected direction for the longer syllables to have more [b] responses. Thus, these results seem to replicate the results of Miller and Liberman. However, it is important to note that although Miller and Liberman showed a difference across SL conditions, they did not conduct statistical tests to see if the differences between the stimulus conditions were significant.

The finding of a test-order effect in Condition 1 is surprising, as there have been no similar reports in other studies exploring the effects of syllable length on [b-w] perception. It is worth noting, however, that in only one study could we find a methodology similar to our own (Miller, 1980a). In all other experiments, subjects received the stimuli in a mixed, rather than blocked, design. We will explore whether this effect appears in the other stimulus conditions before considering what the implications of these findings might be.

The top right panel of Figure 1 shows the results for Condition 2, in which a burst was added to the stimuli. A one-way repeated measures ANOVA showed a significant effect for syllable length [F(4,36) = 3.62, p < .03]. Post hoc tests revealed that SL Condition 1 was significantly different from SL Conditions 2 (p < .05), 4 (p < .01), and 5 (p < .05). All other comparisons were nonsignificant.

A test-order effect was again revealed, as shown in Figure 3. A two-way repeated measures ANOVA yielded a significant effect both for test order [F(1,8) = 5.47, p < .05] and for syllable length [F(4,32) = 8.50, p < .0002], as well as a significant test order × syllable length interaction [F(4,32) = 13.08, p < .0001]. As in Condition 1, when subjects received the shortest SL condition first, there was a stronger syllable-length effect than when



Figure 3. Categorization results from Condition 2, by group. See Figure 1 for description of symbols.

the SL conditions were presented in the reverse order. In the latter case, the stimuli in the longest SL condition were unexpectedly displaced toward the [b] end of the continuum.

The addition of a burst to the [b-w] continua slightly reduced the syllable-length effect, in that, unlike Condition 1, the locus of the boundary for the shortest stimulus condition was not significantly different from all other SL conditions. Nevertheless, there was clearly a trend in the data consistent with the syllable-length effect, particularly for those subjects who received the test order longest to shortest.

In some ways, it is not surprising that the addition of a burst did not play a greater role in reducing the syllablelength effect, because the burst was strongly present only in the first few stimuli at the [b] end of the continuum. Although the burst amplitude was gradually reduced over the 11 stimuli, its perceptual effect was probably minimal by Stimulus 4 or 5, at which point the amplitude of frication was reduced to 30 dB. Thus, the likelihood of the burst's affecting the stimuli lying in the vicinity of the locus of the [b-w] boundary (which occurred between Stimulus 4 and Stimulus 8 for the individual subjects across the various SL conditions) was negligible.

With respect to the test-order effect, it seems to be a reliable phenomenon. We will consider its implications after exploring whether it occurs across the remainder of the stimulus conditions.

The results for Condition 3, in which the formant frequencies varied, are shown in the bottom left panel of Figure 1. As can be seen, there is no evidence for a syllable-length effect. The labeling functions are virtually identical for SL conditions 2-5, although the shortest SL condition (1) was shifted slightly in the expected direction toward fewer [b] responses. Nevertheless, a oneway repeated measures ANOVA showed no effect of syllable length [F(4,36) = 1.68, p = .17].

Figure 4 shows the results as a function of test order. The top panel shows the scores of the group that received SL Condition 1 first, and the bottom panel depicts the scores of the subjects who received SL Condition 5 first. Although the top panel shows fewer [b] responses in the shortest SL condition, it is clear that the syllable-length effect is virtually gone, as SL Conditions 2, 4, and 5 have overlapping functions, and SL Condition 3 has fewer [b] responses than SL Condition 2. For those subjects who received the longest syllables first, there is no evidence of a syllable-length effect. A two-way repeated measures ANOVA yielded no significance either for test order [F(1,8) = .60, p = .46] or for syllable length [F(4,32)]= 1.75, p = .16]; there was no significant test order \times syllable length interaction [F(4,32) = 1.52, p = .221

The results for Condition 3 indicate that by changing the formant frequencies for stimuli along a [b-w] continuum, the syllable-length effect is lost. As for test-order effects, there is a slight performance difference as a function of order of presentation. However, unlike Condi-



Figure 4. Categorization results from Condition 3, by group. See Figure 1 for description of symbols.

tions 1 and 2, in which presentation of the longest SL condition first showed the strongest and most consistent syllable-length effect, in Condition 3 this test order did not produce such an effect.

The bottom right panel of Figure 1 shows the results for Condition 4, in which the stimulus parameters were patterned after natural speech. As in the previous condition, there does not seem to be a syllable-length effect. A one-way repeated measures ANOVA was done for the 10 subjects across the five SL conditions; the results confirmed that there was no significant difference between SL conditions [F(4,36) = .57, p = .68]. Figure 5 shows these same results plotted as a function of the order in which the test groups received the stimulus tapes. It is clear that for both groups, the expected change in [b] responses as a function of syllable length was not obtained.¹

The results for Condition 4 show that the locus of the phonetic boundary for a [b-w] continuum did not vary as a function of syllable length. In other words, the syllableduration effect was not replicated for a [ba-wa] continuum in which the stimuli were patterned after natural speech.

Discussion

Contrary to all of the other studies exploring the influence of syllable duration on the perception of speech and nonspeech stimuli, we found test-order effects for two of the four stimulus conditions. In particular, we presented the test stimuli in a blocked design, and found a reduction in the syllable-duration effect in Conditions 1 and 2, especially when the subjects received the longest SL condition first. With one exception (Miller, 1980a), all other studies discussed in the literature presented the test stimuli in a mixed, not blocked, design. It is possible that a blocked design acts to reduce any potential context effects. Our blocked design involved five separate test conditions (one for each syllable-length condition). Given that the subjects heard each condition separately, they could have established five different standards or strategies for perceiving the test stimuli. As a result, the amount of varia-



Figure 5. Categorization results from Condition 4, by group. See Figure 1 for description of symbols.

bility could have potentially increased, thus reducing or suppressing the presence of the syllable-length effect. In contrast, in a mixed design, subjects cannot apply different standards across the various SL conditions, because the stimuli are randomly presented within the same test. Thus, the amount of variability may well be reduced, and the syllable-length effect may still emerge.

In order to determine whether our failure to replicate the syllable-duration effects in all stimulus conditions was a function of our methodology, of the changes in the acoustic parameters in the test stimuli, or of both, we decided to run a second series of experiments, using a mixed, rather than blocked, design. In this series of experiments the stimuli used in Conditions 1, 3, and 4 of the previous series were presented separately in a mixed design. It will be recalled that in Condition 1, stimuli similar to those of Miller and Liberman (1979), varying only in formant rate and transition duration, were used, and a syllable-length effect was found. In Condition 3, in which stimuli with varying formant transitions were used. and in Condition 4, in which stimuli were patterned after natural speech parameters, no syllable-length effect was found.

SERIES 2

Method

Subjects. Thirty new volunteers participated, 10 for each stimulus condition. None reported any speech or hearing disorder.

Stimuli and Procedure. The stimulus parameters for the three stimulus conditions of Series 2 were identical to those used in Conditions 1, 3, and 4 of Series 1. Three test tapes were made, one for Condition 1 (Miller and Liberman replication), a second for Condition 2 (formant frequencies varied), and a third for Condition 3 (patterned after natural speech parameters). Each tape consisted of a random presentation of the stimuli in the various syllable-length conditions, yielding 11 stimuli \times 5 length conditions \times 10 repetitions = 550 stimuli per test tape. The interstimulus interval was 2 sec, with a 6-sec pause after every 10 stimuli. Preceding each tape, the five 11-stimuli continua were presented to the subjects to familiarize them with presentation rate and stimuli.

Results

Figure 6 shows the mean results for the three stimulus conditions. As the figure shows, the syllable-length effect clearly seems to vary as a function of the stimulus condition. A two-way repeated measures ANOVA with one grouping factor (stimulus condition) and one within factor (syllable length) confirmed that the boundary values varied as a function of stimulus condition and syllable length. There was a significant main effect for both stimulus condition [F(2,27) = 9.47, p < .001] and syllable length [F(4,108) = 54.31, p < .001], as well as a significant stimulus condition \times syllable length interaction [F(8,108) = 15.14, p < .001].

To explore these effects in more detail, each stimulus condition was then analyzed separately. As the top panel of Figure 6 shows, for those stimuli based on the Miller and Liberman stimulus parameters, a syllable-length ef-



Figure 6. Categorization results for the three stimulus conditions of Series 2. The top panel shows the results for Condition 1, the middle panel for Condition 2, and the bottom panel for Condition 3 (see text). See caption of Figure 1 for description of symbols.

fect is clearly present, with fewer [b] responses in the shorter syllable condition and increasingly more [b] responses as syllable length increases. A one-way repeated measures ANOVA showed a significant effect for syllable length [F(4,36) = 64.48, p < .0001]. Post hoc tests revealed that all comparisons between five SL conditions were significant, with the exception of the difference between SL Conditions 4 and 5, which was nonsignificant. The pattern of results obtained in this condition clearly replicates those obtained by Miller and Liberman (1979).

The middle panel of Figure 6 shows the results for the stimulus condition in which formant frequencies varied over the 10 subjects. Again, the results show a syllable-length effect. A one-way repeated measures ANOVA showed a significant effect for syllable length [F(4,36) = 18.47, p < .0001]. Post hoc tests revealed that the two shortest SL conditions, 1 and 2, are significantly different from each other and from the remaining (longer) SL conditions, 3, 4, and 5. No significant differences were found among SL Conditions 3, 4, and 5.

The bottom panel of Figure 6 shows the results for Condition 3, in which the test stimuli were patterned after natural speech. As is apparent, there is no evidence for a syllable-length effect. The labeling functions are virtually identical for SL Conditions 1–4, with SL Condition 5 yielding slightly more [b] responses. Nevertheless, a oneway repeated measures ANOVA showed no effect of syllable length [F(4,36) = 1.47, p = .23], indicating that the syllable-length effect is absent in this stimulus condition. Thus, regardless of the experimental design being used, the syllable-length effect does not occur when parameters are closely matched to those of natural speech.

Discussion

When stimuli were presented to subjects in a mixed design, Miller and Liberman's (1979) syllable-length effect was replicated. Thus, the loss of the syllable-length effect in Condition 3 of Series 1, in which the stimuli varied in formant frequency, rate, and duration, seemed to be due to the blocked experimental design. Nevertheless, although the syllable-length effect emerged in the mixed design, the size of the effect was greatly reduced when the stimulus parameters were more like those of natural speech, and the effect disappeared completely when the stimulus parameters were modeled after those of natural speech. In particular, in Condition 1 of Series 2, in which the stimulus parameters were similar to those of Miller and Liberman, the change in boundary value between the shortest SL condition (102 msec) and the longest SL condition (317 msec) was approximately 13.3 msec. In other words, a roughly threefold change in syllable duration produced a boundary shift of approximately 13.3 msec. In contrast, in Condition 2 of Series 2, in which the stimulus parameters varied in formant frequency as well as in rate and duration, the change in boundary value between the shortest and longest SL condition was only 5.5 msec. That is, in this condition, a threefold change in syllable duration produced a boundary shift of only 5.5 msec. It

is apparent, then, that although a syllable-length effect did show up in Condition 2 of Series 2, the use of more natural parameters reduced this effect by a magnitude of approximately 2.5. In Condition 3 of Series 2, in which the stimulus parameters were based on natural speech values, the syllable-length effect disappeared.

GENERAL DISCUSSION

The results of these two series of experiments suggest that context-dependent effects of syllable duration on the perception of the [b-w] contrast are strongly reduced or even completely suppressed when the stimuli are patterned after natural speech. In particular, results of the first series of experiments revealed an absence of the syllablelength effect for a [ba-wa] continuum in which the endpoint stimuli contrasted in a number of acoustic dimensions, including presence/absence of a burst, amplitude envelope, transition rate and duration, and formant frequencies. Nevertheless, the Miller and Liberman (1979) effect was replicated in this series of experiments, which showed that the perception of a [ba-wa] continuum varying solely in formant transition duration and rate was affected by syllable duration. Smaller effects were shown, however, when a burst was added to the [b] end of the continuum, and the effect was completely lost when the formant frequency parameters were varied. Thus, even though the frequencies of the transition onsets were somewhat different from measured values of natural speech, the loss of the syllable-duration effect was a function of having different formant-frequency values at transition onset. The syllable-duration effect was found only in those conditions in which the transition-onset frequencies were identical throughout the continuum, and was absent when these frequencies were different.

However, the results of the second series of experiments showed that the syllable-length effect was also influenced by the experimental design used. In our attempt to replicate Miller and Liberman's (1979) study, we found a syllable-length effect whether we used a blocked design or a mixed design. However, when stimuli more like natural speech were used, we did not find a syllable-length effect when we used a blocked design, but did find such an effect when we used a mixed design. The fact that in this experimental condition the syllable-length effect occurred as a function of the experimental design (a finding which, to our knowledge, has never been reported before) suggests that the effect is not particularly strong, especially when the test stimuli have been modeled more like natural speech.

The occurrence of the test-order effect was surprising and puzzling, and does bring into question the overall robustness of the syllable-duration phenomenon. We have no ready explanation for the changes in patterns of perception as a function of test order. Obviously, subjects are sensitive to the stimulus set they are receiving, and are also affected by the stimulus sets presented earlier in the testing session.

The second series of experiments shows a gradual reduction in the syllable-length effect as the test stimuli more nearly approximate natural speech. In Condition 1 (Miller and Liberman replication), the boundary change as a function of syllable length was approximately 2.5 msec per SL condition. In Condition 2 (in which formant frequencies changed), the magnitude of the effect was very small, averaging a change of only 1 msec per SL condition. Although no study has reported the just noticeable difference (JND) for a change in duration for either stops or glides, the available literature on JNDs for both speech (Klatt & Cooper, 1975) and nonspeech stimuli (Abel, 1972; Creelman, 1962) suggests that listeners are not sensitive to differences in duration of the magnitude that has been reported either in Miller and Liberman (1979) or in Conditions 1 and 2 of Series 2. Finally, in Condition 3, in which the synthetic stimuli were modeled after natural-speech parameters, the syllable-length effect was completely suppressed. The use of increasingly natural-speech-like stimuli thus reduces the syllable-length effect until the effect has completely disappeared when stimuli have been modeled very closely after natural speech. The present study thus suggests that stimuli that closely resemble natural speech contain contextindependent acoustic properties, in the sense that they are not affected by changes in syllable length (speaking rate) or experimental design (blocked vs. mixed).

In conclusion, it has been argued that the perception of the phonetic dimensions of speech is context-dependent and is influenced by neighboring segments (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Schatz, 1954). Although there is little doubt that contextdependent effects play some role in the speech perception process, we suggest that the strength of this dependence may have been greatly overestimated, in large part due to the methodologies used over the past 25 years in speech perception research. Most studies have used highly schematized stimuli which, in many cases, contain inappropriate rather than simplified characteristics of natural speech. Although this method has been useful in isolating particular variables and exploring their effects on phonetic identification and discrimination, it has also suggested a greater reliance on these parameters in the perception of the phonetic categories of speech than may be warranted. Because in the present study some context effects were shown to disappear as the result of the use of more natural-speech-like stimuli, further research on the strength of these context effects need to be done by running experiments using stimuli that more nearly reflect the acoustic characteristics of natural speech.

REFERENCES

- ABEL, S. M. (1972). Duration discrimination of noise and tone bursts. Journal of the Acoustical Society of America, 51, 1219-1223.
- BLUMSTEIN, S. E., & STEVENS, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, 66, 1001-1017.

- COOPER, W. E., EBERT, R. R., & COLE, R. A. (1976). Perceptual analysis of stop consonants and glides. *Journal of Experimental Psychol*ogy: Human Perception & Performance, 2, 92-104.
- CREELMAN, C. D. (1962). Human discrimination of auditory duration. Journal of the Acoustical Society of America, 34, 582-593.
- DALSTON, R. M. (1975). Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults. *Journal of the Acousti*cal Society of America, **57**, 462-469.
- DIEHL, R. (1976). Feature analyzers for the phonetic dimension stop vs. continuant. *Perception & Psychophysics*, **19**, 267-272.
- EIMAS, P. D., & MILLER, J. L. (1980). Contextual effects in infant speech perception. Science, 209, 1140-1141.
- FANT, G. (1960). Acoustic theory of speech production. The Hague: Mouton.
- FISCHER-JORGENSEN, E. (1954). Acoustic analysis of stop consonants. Miscellaneous Phonetics, ii, 42-59.
- GAY, T. (1978). Effect of speaking rate on vowel formant transitions. Journal of the Acoustical Society of America, 63, 223-230.
- GAY, T., & HIROSE, H. (1973). Effect of speaking rate on labial consonant production. *Phonetica*, 27, 44-56.
- GAY, T., USHIJIMA, T., HIROSE, H., & COOPER, F. S. (1974). Effect of speaking rate on labial consonant-vowel articulation. *Phonetica*, 2, 47-63.
- GODFREY, J. J., & MILLAY, K. K. (1981). Discrimination of the "tempo of frequency change" cue. Journal of the Acoustical Society of America, 69, 1446-1448.
- JUSCZYK, P. W., PISONI, D. B., REED, M. A., FERNALD, A., & MYERS, M. (1983). Infants' discrimination of the duration of rapid spectrum changes in nonspeech sounds. *Science*, **222**, 175-177.
- KLATT, D. H. (1980). Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America, 67, 971-995.
- KLATT, D. H., & COOPER, W. E. (1975). Perception of segment duration in sentence contexts. In A. Cohen & S. G. Nooteboom (Eds.), Structure and process in speech perception. Berlin: Springer-Verlag.
- KOZHEVNIKOV, V., & CHISTOVICH, L. (1965). Speech: Articulation and perception. (Federal Technical and Scientific Information JPRS 30,543). Washington, DC: U.S. Department of Commerce.
- LAHIRI, A., GEWIRTH, L., & BLUMSTEIN, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study. *Journal of the Acoustical Society of America*, 76, 391-405.
- LEHISTE, I. (1964). Acoustic characteristics of selected English consonants. International Journal of American Linguistics, 30, 1-197.
- LEHISTE, I., & PETERSON, G. E. (1961). Transitions, glides and diphthongs. Journal of the Acoustical Society of America, 33, 268-277.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461.
- LIBERMAN, A. M., DELATTRE, P. C., GERSTMAN, L. J., & COOPER, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology*, **52**, 127-137.
- MACK, M., & BLUMSTEIN, S. E. (1983). Further evidence of acoustic invariance in speech production: The stop-glide contrast. *Journal of* the Acoustical Society of America, **73**, 1739-1750.
- MAXWELL, E., & LANDAHL, K. (1983). The b/w contrast and considerations of syllable duration. Journal of the Acoustical Society of America, 73, S53.
- MILLER, J. L. (1980a). Contextual effects in the discrimination of stop consonant and semivowel. *Perception & Psychophysics*, 28, 93-95.
- MILLER, J. L. (1980b). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study* of speech. Hillsdale, NJ: Erlbaum.
- MILLER, J. L. (1981). Some effects of speaking rate on phonetic perception. *Phonetica*, 38, 159-180.
- MILLER, J. L., & EIMAS, P. D. (1983). Studies on the categorization of speech by infants. Cognition, 13, 135-165.
- MILLER, J. L., & LIBERMAN, A. M. (1979). Some effects of lateroccurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, 25, 457-465.

- O'CONNOR, J. O., GERSTMAN, L. J., LIBERMAN, A. M., DELATTRE, P. L., & COOPER, F. S. (1957). Acoustic cues for the perception of initial /w,j,r,l/ in English. Word, 13, 24-43.
- PETERSON, G., & LEHISTE, I. (1960). Duration of syllabic nuclei in English. Journal of the Acoustical Society of America, 32, 693-703.
- PICKETT, J. M. (1980). The sounds of speech communication. Baltimore: University Park Press.
- PISONI, D. B., CARRELL, T. D., & GANS, S. J. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Perception & Psychophysics*, 34, 314-322.
- SCHATZ, C. D. (1954). The role of context in the perception of stops. Language, **30**, 47-56.
- SCHWAB, E., SAWUSCH, J., & NUSBAUM, H. C. (1981). The role of second formant transitions in the stop-semivowel distinction. *Percep*tion & Psychophysics, 29, 121-128.
- SHINN, P. C., & BLUMSTEIN, S. E. (1984). On the role of the amplitude envelope for the perception of [b] and [w]: Further support for a theory of acoustic invariance. *Journal of the Acoustical Society of America*, **75**, 1243-1252.

STEVENS, K. N., & BLUMSTEIN, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, NJ: Erlbaum.

NOTES

1. An additional experiment was conducted, based on natural speech parameters for [bi-wi]. These results also failed to replicate the syllable length effect. However, a test-order effect was revealed. When subjects received the stimulus tapes in the order of shortest syllable length to longest syllable length, there was no differential effect of syllable length; however, when subjects received the longest syllable length condition first, some differential effects emerged. Specifically, consistent with Miller and Liberman's (1979) findings, SL Condition 2 (175 msec) produced fewer [b] responses than the longer SL Conditions 3 (265 msec), 4 (355 msec), and 5 (445 msec). Contrary to their findings, however, the shortest SL condition (85 msec) produced the greatest number of [b] responses.

(Manuscript received March 5, 1984; revision accepted for publication October 17, 1985).