Published in final edited form as:
Percept Psychophys. 1983 October ; 34(4): 314–322.

Perception of the duration of rapid spectrum changes in speech and nonspeech signals

D. B. PISONI, **T. D. CARRELL**, and **S. J. GANS**
Indiana University, Bloomington, Indiana

Abstract

For a number of years, investigators have studied the effects one acoustic segment has on the perception of other acoustic segments. In one recent study, Miller and Liberman (*Perception & Psychophysics,* 1979, 25, 457–465) reported that overall syllable duration influences the location of the labeling boundary between the stop [b] and the semivowel [w]. They interpreted this "context effect" as reflecting a form of perceptual normalization whereby the listener readjusts his perceptual apparatus to take account of the differences in rate of articulation of the talker. In the present paper, we report the results of several comparisons between speech and nonspeech control signals. We observed comparable context effects for perception of the duration of rapid spectrum changes as a function of overall duration of the stimulus with both speech and nonspeech signals. The results with nonspeech control signals therefore call into question the earlier claims of Miller and Liberman by demonstrating clearly that context effects are not peculiar to the perception of speech signals or to normalization of speaking rate. Rather, such context effects may simply reflect general psychophysical principles that influence the perceptual categorization and discrimination of all acoustic signals, whether speech or nonspeech.

For many years, investigators have been interested in how one phonetic segment affects the perception of adjacent segments in the speech signal. This form of "context conditioned variability" has been of major theoretical interest in the past and, despite some 30 years of research, it still continues to occupy the attention of many researchers even today (Liberman, 1982; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Studdert-Kennedy, 1978). While some formal attempts have been made to deal with this problem by offering theoretical accounts of speech perception framed in terms of motor theory, analysis-by-synthesis, or feature detector models of the early stages of recognition, there is still no satisfactory solution to the problem of how listeners compensate for the extensive amount of context-conditioned variability observed in the speech signal. These various forms of acoustic variability in speech arise from several sources, including: the effects of the immediately surrounding phonetic context, differences in speaking rate, differences in talkers, and the variability of segmental durations that is conditioned by the syntactic and semantic structure of sentences and passages of fluent speech (see, e.g., Liberman & Studdert-Kennedy, 1978, Miller, 1980b, and Summerfield, 1981).

The problem of context-conditioned variability in speech has been so elusive that some investigators, such as Klatt, have proposed to solve it by completely denying that the problem exists at all. Instead, Klatt (1979) has proposed a set of context-sensitive spectral

D. B. Pisoni's mailing address is: Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, Indiana 47405.

An earlier version of this paper was presented at the 100th meeting of the Acoustical Society of America, November 19, 1980, Los Angeles, California.

templates that directly encode the acoustic-phonetic variability of phonemes in different phonetic environments. Words are recognized directly from sequences of these spectral templates without the need for an intermediate level corresponding to a phonetic segmental representation.

We became interested in the problem of context-conditioned variation after reading a report by Miller and Liberman (1979). They found that the perceptual boundary between [b] and [w] was influenced by the nature of the context immediately following the critical acoustic cues to the stop vs. semivowel distinction—namely, the duration of the formant transitions at stimulus onset. Miller and Liberman (1979) reported that the duration of the vowel in a CV syllable systematically influences the perception of the formant transition cues for the stop-semivowel distinction. With short syllables, subjects required shorter transition durations to perceive a [w] than they did with longer syllables. Miller and Liberman interpreted these results as a clear demonstration of perceptual normalization for speaking rate—listeners adjusted their decision criteria to compensate for the differences in vowel length that are conditioned by the talker's speaking rate. It is well known, for example, that vowels become much shorter when speaking rate increases. According to Miller and Liberman's account, the listener apparently interprets a particular set of acoustic cues, such as the duration of a transition for a [b] or a [w], *in relation* to the talker's speaking rate rather than by reference to some absolute set of context-invariant acoustic attributes in the stimulus pattern itself. In their study, the location of the boundary for a syllable-initial [b-w] contrast was determined by the overall duration of the syllable containing the target phoneme (see, also, Miller, 1980a, 1981; Miller & Grosjean, 1981).

The particular claims surrounding perceptual compensation and adjustment for speaking rate have recently taken on even more significance with the findings reported by Eimas and Miller (1980) on young infants. They found that young prelinguistic infants also discriminate these same [b-w] stimuli in a "relational manner" that is similar to that found earlier by Miller and Liberman with adult listeners. Moreover, these results, like the ones obtained with adults, were interpreted as support for the argument that speech is not processed in a strictly left-to-right linear fashion one phoneme at a time; rather, phonetic perception requires the integration of numerous widely distributed acoustic cues in the speech signal (see Miller & Eimas, 1983, for a review).

The experiments reported in this paper were concerned with the extent to which the context effects reported by Miller and Liberman (1979) were a consequence of perceptual mechanisms that are specific to processing speech signals. Miller and Liberman suggest that the context effects caused by later-occurring information reflect "an appropriate adjustment by the listener for changes in articulatory rate" of the talker. Although not stated explicitly, Miller and Liberman implied that this form of perceptual normalization for speaking rate was unique to processing speech signals and the mechanisms used in phonetic categorization (see, also, Eimas & Miller, 1980, Miller, 1981, and Miller & Eimas, 1983). We wanted to know if these findings with adults and infants were a result of mechanisms involved in phonetic categorization per se or whether they were due to somewhat more general factors related to auditory perception of both speech and nonspeech signals (Pisoni, 1979). In order to answer these questions, we first generated several sets of new synthetic speech stimuli that were modelled as closely as possible after the parameter specifications provided by Miller and Liberman to see if we could replicate their initial findings with adults in our own laboratory. We then generated several sets of nonspeech "control" stimuli using three time-varying sinusoids that followed the formant patterns of the speech stimuli. These nonspeech stimuli preserved all of the temporal and durational cues present in the original speech stimuli, although they did not sound like speech to our subjects.

EXPERIMENT 1

The purpose of this first experiment was twofold. First, we wanted to replicate, as closely as possible, the major findings reported by Miller and Liberman (1979) on the perception of [b] and [w]. In particular, we wanted to replicate their finding that the precise location of the labeling boundary between [b] and [w] was a function of the overall syllable duration; Miller and Liberman found that as syllable duration increased, the [b-w] boundary moved toward formant transitions of longer duration. We also wanted to replicate an additional finding reported by Miller and Liberman, that altering the internal structure of the syllable by adding syllable-final transitions would cause the [b-w] labeling boundary to shift toward transitions of shorter duration—an effect that was precisely the opposite of that produced when the syllable was lengthened by simply increasing the overall duration of the steady-state vowel.

The second purpose of this experiment was to make available a set of precisely specified speech stimuli that could be used as models to generate several sets of matched nonspeech control stimuli. The results obtained with the nonspeech stimuli could then be compared directly with the data obtained with the speech stimuli from which they were derived. Such a comparison in perception between speech and nonspeech would permit us to assess whether the context effects found by Miller and Liberman were due to phonetic categorization of speech signals or to more general auditory processes common to perception of both speech and nonspeech signals.

Method

Subjects—Thirteen naive undergraduate students at Indiana University were recruited as paid volunteer subjects from a laboratory subject pool used for perceptual experiments. They were all right-handed monolingual speakers of midwestern American English and reported no history of a hearing or speech disorder on a pretest questionnaire given at the time of testing. The subjects were paid at the rate of $3.50/h for each testing session.

Stimuli—The stimuli for this experiment consisted of four sets of synthetically produced speech signals: two sets of CV syllables and two sets of CVC syllables. Each set contained 11 test stimuli that formed a continuum between [b] and [w]. Schematic representations of the formant motions of the endpoint stimuli for each set are shown in Figure 1.

The top panel in this figure shows the long and short CV stimuli, the bottom panel shows the long and short CVC stimuli. The 11 stimuli in each set contained identical initial formant transitions and differed only in the overall duration of the syllable. The construction of the stimuli was based on parameter values provided in Miller and Liberman's (1979) earlier report. The long CV stimuli were 295 msec in duration; the short CV stimuli were 80 msec long. For each set, the duration of the formant transitions was varied in 5-msec steps from 15 msec, a value appropriate for a [b], to 65 msec, a value appropriate for a [w]. As the duration of the transitions was increased, the duration of the steady-state vowel was decreased so as to hold the overall syllable duration constant. The parallel sets of CVC stimuli were constructed by adding 35 msec of formant transitions appropriate for a [d] in syllable final position to the long and short CV stimuli, respectively.

Digital waveforms of all stimuli were generated on a version of the cascade-parallel software synthesizer designed by Klatt (1980). They were then output at 10 kHz via a 12-bit D/A converter controlled by a PDP-11/34 and low-pass filtered at 4.8 kHz. All stimuli contained five formants and consisted of a 20-msec period of low-frequency, low-amplitude prevoicing, variable-length formant transitions, and a variable-length steady-state vowel. The first formant (F1) transition started at 234 Hz and rose linearly to a steady-state value of

769 Hz. The second formant (F2) transition started at 616 Hz and rose linearly to a steady-state value of 1232 Hz. The third, fourth, and fifth formants (F3, F4, and F5) were set at 2862, 3600, and 3850 Hz, respectively, and remained constant for the entire duration of the steady-state vowel. For the two sets of CVC syllables, the final 35 msec consisted of an F1 transition that fell linearly from 769 to 234 Hz, and F2 and F3 transitions that rose linearly from 1232 and 2862 Hz to 1541 and 3363 Hz, respectively. The steady-state values of F4 and F5 remained constant during the final transitions.

Procedure

All experimental events involving presentation of the stimuli, feedback, and collection of the subjects' responses were controlled on-line in real time by a PDP-11/34 computer. The subjects participated in small groups in a quiet room equipped with individual cubicles interfaced to the computer. The test stimuli were converted to analog form and presented to the subjects binaurally through Telephonics TDH-39 matched and calibrated headphones. All stimuli were presented at a comfortable listening level of about 80 dB SPL for the steady-state portion of the vowel. The same voltage levels were maintained across sessions with a VTVM.

The present experiment consisted of one session that lasted about 1 h. The session was divided into three phases. In the first phase, the subjects heard each of the eight endpoint stimuli presented three times each in a random order. They were told to identify each stimulus as beginning with a [b] or a [w] by pressing one of two buttons located directly in front of them on a response box. Feedback was presented after each trial, indicating the correct response by illumination of a light above the appropriate response button. In the second phase, another block of 24 trials (8 stimuli × 3 repetitions) using endpoint stimuli was presented. However, no feedback was in effect. Finally, in the third phase, all 44 test stimuli (4 continua × 11 stimuli per set) were presented in two randomized blocks containing five repetitions of each stimulus for a total of 440 trials. In all three phases, the stimuli were presented one at a time with a warning light preceding the onset of the signal by 1 sec. Timing and sequencing of trials was paced to the slowest subject in a given session.

Results and Discussion

The group labeling functions for the four stimulus series are shown in Figure 2.

The data for the CV stimuli are shown in the left panel; the data for the CVC stimuli are shown in the right panel. The parameter in each panel is syllable duration; the filled circles represent the short syllables, and the open triangles, the long syllables in each panel. It can be seen very clearly in this figure that there is an effect of syllable duration on identification of [b] and [w]. The labeling function for the short syllables (CVS) in each panel is displaced toward shorter transition durations relative to the labeling function for long syllables (CVL).

To quantify these observations in more precise terms, we fitted a normal ogive to each subject's data with the procedures outlined by Woodworth (1938), and obtained means and standard deviations for the four conditions for each subject. This procedure provided us with numerical values of the location of the phonetic boundary in terms of the 50% crossover point (means) and their respective slopes (standard deviations). Means and standard deviations for the individual subjects in these conditions are given in Table 1.

The mean phonetic boundaries for the group labeling functions shown in Figure 2 were 24.3 vs. 39.3 msec for the CVS and CVL stimuli and 30.3 vs. 39.5 msec for the CVCS and CVCL stimuli, respectively. The differences in the phonetic boundaries due to syllable duration were, in each case, highly significant by correlated t tests [for the CV syllables,

$t(12) = 8.43$, $p < .001$; for the CVC syllables, $t(12) = 9.95$, $p < .001$]. On the other hand, analyses of the slopes of the labeling functions indicated that they did not change when syllable duration was shortened ($p > .05$).

The results of this experiment replicate the earlier findings reported by Miller and Liberman (1979). Syllable duration strongly influences the location of the phonetic boundary between [b] and [w] in synthetically produced CV and CVC syllables differing in the duration of their formant transitions. Every subject in our experiment showed the predicted effect across both syllable types. Thus, the effect of syllable duration is not only reliable, but also very consistent across subjects.

Miller and Liberman (1979) also found that the internal structure and organization of the syllable had an effect on the location of the phonetic boundary between [b] and [w]. When the syllable was lengthened by adding final transitions appropriate for the stop consonant [d], the labeling function shifted towards transitions of shorter duration, a finding that was precisely the reverse of the earlier results that had shown that lengthening the steady-state vowel caused the boundary to shift towards longer transition values. Since Miller and Liberman's findings demonstrate that both the physical duration of the syllable *and* its internal structure and organization affect the perceptual analysis of the relevant acoustic cues, we thought it would be appropriate to attempt a replication of this important result as well. In order to do this, we generated a new set of CVC stimuli that were specifically matched to the overall duration of the long CV stimuli. These were called the CVCL (new) stimuli. Using the same procedures as described above, we ran three new groups of subjects. Group A heard the CVS-CVL series, Group B heard the CVCS-CVCL series, and Group C heard the CVCL (new)-CVL series. During identification testing, each stimulus was presented 20 times, each in a random order, for identification as [b] or [w].

The results of the additional experimental conditions are shown in Figure 3. The labeling data in Panels A and B replicate the findings obtained in the previous experiment. For both CV and CVC syllables, increasing the duration of the syllable shifts the locus of the labeling boundary between [b] and [w]. The differences shown in panels A and B are highly significant [$t(11) = 8.27$, $p < .001$, and $t(14) = 9.00$, $p < .001$, for the CV and CVC comparisons, respectively]. The labeling data shown in Panel C are for the two test series which were matched for overall duration but which differed in their internal structure. The results for this condition also replicated the previous findings reported by Miller and Liberman (1979), although only weakly. The effect of adding 35-msec transitions appropriate for a syllable final [d] was to shift the labeling boundary for the CVCL (new) stimuli toward transitions of shorter duration. The difference in the location of the phonetic boundaries between the two conditions was very small and only marginally significant ($p < .05$), as compared with those of the other conditions. Nevertheless, we did replicate the second effect reported by Miller and Liberman: "later occurring" information in the speech signal related to the internal structure of a syllable modifies the perception of earlier-occurring information used to make a phonetic decision.

Having replicated the main findings reported by Miller and Liberman, we can now turn to the major purpose of the present investigation, namely, to determine if these later occurring contextual effects are due to perceptual processes and mechanisms that are specific to the phonetic coding of speech signals. The results reported by Miller and Liberman and replicated in Experiment 1 clearly demonstrate that the perception of [b] and [w] is *relational* rather than invariant and is strongly dependent on the surrounding context; in particular, the precise location of the labeling boundary is determined by the duration and internal structure of the syllable itself. These results were interpreted by Miller and Liberman (1979) as evidence that the listener somehow normalizes for speaking rate in

perceiving segmental phonetic distinctions such as [b] and [w]. The assumption underlying this claim is that syllable duration varies inversely with rate of articulation and therefore can be used via an appropriate adjustment by the listener to specify the talker's speaking rate. As Miller and Liberman (1979) have put it: "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]" (p. 464).

EXPERIMENT 2

In this experiment, we obtained labeling functions for several sets of nonspeech control stimuli that were carefully modeled after the synthetic speech stimuli used in the previous experiment. Our goal in using nonspeech control stimuli was to determine if we could obtain contextual effects that were comparable to those found with the speech stimuli. Specifically, would the overall duration and internal structural organization of nonspeech signals affect the labeling of their onset characteristics? Evidence of both effects in the perception of nonspeech signals would argue strongly, in our view, against the interpretations proposed by Miller and Liberman (1979), that the duration of the syllables is used to specify the articulatory rate of the talker and that this after-occurring information is then used by the listener in making a phonetic judgment of [b] vs. [w]. Since nonspeech signals do not receive, by definition, a phonetic interpretation, it follows that they cannot therefore be perceived in relation to the articulatory rate of the talker. Thus, comparable effects with nonspeech signals would raise important questions about the basis for the claim that the shifts in the perceived [b-w] boundary are due to an adjustment (via normalization) by the listener to changes in the talker's articulatory rate. Of course, it could be argued that the same basic perceptual mechanisms are employed, although differently, in processing speech and nonspeech signals. In either case, a demonstration of the same type of context effects for nonspeech control signals would effectively weaken the strong claims made recently by Eimas and Miller (1980) and Miller and Eimas (1983) about the specialized speech processing capabilities of young infants and the presumed basis of these abilities in terms of a phonetic mode of processing that somehow uses knowledge of articulation to rationalize contextual variation and lack of acoustic-phonetic in variance.

Method

Subjects—Forty-six additional naive subjects who met the same requirements as in Experiment 1 were recruited for this experiment. None of the subjects had been in the previous experiment, and nor had they participated in any other perceptual experiments in our laboratory before the present tests.

Stimuli—Five sets of nonspeech control stimuli were generated using a program that permits independent control over the frequencies, amplitudes, and temporal characteristics of three sinusoids (Kewley-Port, Note 1). Each set contained 11 stimuli that differed in the duration of a rapid spectrum change at onset. The parameter specifications for the first three formants of the synthetic speech stimuli used in Experiment 1 were used as the input values to generate the nonspeech stimuli. The parameter values for the formant frequencies and durations of the speech stimuli were translated directly into values that controlled the frequency and durations of three sinusoids, corresponding to values of the first three formants of the speech stimuli. The relative amplitudes of the component sinusoids were set to constants based on measurements of the formant amplitudes in the steady-state portion of the vowel. The second (T2) and third (T3) components, corresponding to the second and third formants, were set 3 dB lower than the first component tone (T1). The five sets of

nonspeech stimuli used here corresponded to the CVS-CVL, CVCS-CVCL, and CVCL (new)-CVL speech continua used to collect the labeling data shown in Figure 3.

Procedure—As in the previous experiment, the presentation of stimuli, feedback, and collection of responses was controlled online by a PDP-11/34 minicomputer. The present experiment consisted of two 1-h sessions conducted on separate days. In order to obtain reliable and consistent labeling functions with the nonspeech signals, it was first necessary to train and then screen subjects on identification of the endpoint stimuli from each continua. Thus, on Day 1, the subjects were trained to identify the endpoint stimuli of two test continua (i.e., four stimuli) as beginning with either an "abrupt onset" or a "gradual onset," using a disjunctive conditioning procedure (Lane, 1965; Pisoni, 1977). On Day 2, all the stimuli from the entire continuum were presented for identification testing.

Training on Day 1 consisted of two phases. The first phase simply involved familiarization with the test stimuli. The four endpoint stimuli were presented in alternating fashion (i.e., 1, 11, 12, 22), 10 times each. The subjects were told to listen carefully to the beginning of each sound and to try to determine if it had an "abrupt" or a "gradual" onset. Each trial consisted of a cue light, presentation of one of the four endpoint stimuli, and then feedback indicating the correct response. The subjects were not required to make any overt responses during this part of the training procedure; they simply listened to the stimuli to familiarize themselves with the signals.

The second phase involved identification training with the end-points. The four stimuli were presented 40 times each in a random order with feedback. The subjects were required to identify each stimulus as "abrupt" or "gradual" by pressing one of two response buttons. Feedback indicating the correct response was provided immediately after a response was entered.

The third phase of the training procedure involved criterion testing. The four endpoint stimuli were again presented 40 times each in a random order for 160 trials. However, no feedback was provided after each response. The subjects were told that their performance on this testing phase would determine if they would be invited to return for the second day of the experiment. Subjects who performed at or above 90% correct for each of the four stimuli were asked to return the next day.

On Day 2, subjects were first given a brief warm-up session before identification testing began. The warm-up consisted of two blocks of 40 trials. In the first block, the four endpoints were presented 10 times each in a random order with feedback in effect. In the second block, feedback was removed. Identification testing consisted of the presentation of two blocks of 220 trials. Each block contained 10 repetitions of each of the 11 stimuli from two continua presented in random order. No feedback was presented during the final phase of identification testing. Subjects in Group A were assigned to the CVS-CVL condition, subjects in Group B were assigned to the CVCS-CVCL condition, and subjects in Group C were assigned to the CVCL (new)-CVL condition. The design of this experiment was exactly parallel to that used for the speech data shown in Figure 3.

Results and Discussion

The labeling functions for the three groups of subjects are shown in Figure 4. Panel A on the left displays the data for the CV stimuli, Panel B displays the data for the CVC stimuli, and Panel C displays the data for the cross-series comparisons for CVs and CVCs that were matched for overall stimulus duration. Examination of Panels A and B reveals a very substantial shift in the identification functions for "abrupt" and "gradual" onsets as the overall duration of the nonspeech stimulus pattern is increased from 80 to 295 msec. The

means and standard deviations for individual subjects in these three conditions are presented in Table 2. The shifts in location of the identification functions were statistically reliable and, as in the previous experiment with speech stimuli, quite consistent over subjects [$t(15) = 4.89$, $p < .001$, and $t(15)=1.97$, $p < .05$, for the differences shown in Panels A and B, respectively]. Thus, these two sets of identification data obtained with nonspeech control stimuli demonstrate context effects that are due to overall stimulus duration. These effects appear to be quite comparable to those found earlier with speech stimuli by Miller and Liberman (1979) and the results of our replication in the previous experiment.[1]

Of special interest, however, are the nonspeech results, shown in Panel C of Figure 4, which involved comparisons of nonspeech stimuli having equal overall durations but different internal structures. The result shown here also replicates the effect reported by Miller and Liberman (1979) for speech stimuli containing formant transitions in syllable-final position. In particular, the labeling boundary for the CVCL (new) nonspeech condition is shifted toward the left to shorter transition durations, as if these stimuli were perceived as having a shorter overall stimulus duration [$t(13) = 3.59$, $p < .005$]. This context effect occurred despite the fact that both sets of nonspeech stimuli were of equal physical duration; they differed only in terms of their structural organization. Thus, the identification of the onsets of nonspeech control stimuli are affected by later occurring information in the stimulus configuration.

GENERAL DISCUSSION

Taken together, all three nonspeech stimulus conditions replicate the context effects reported by Miller and Liberman (1979) with speech stimuli. Our nonspeech stimuli were designed to model, as closely as possible, the four sets of speech stimuli that differed in the acoustic cues for the [b-w] distinction. As we found in the first experiment, using speech stimuli, the major effects of syllable duration on identification of [b] and [w] could be replicated quite easily with newly created synthetic speech stimuli that differed in the duration of their formant transitions. More importantly, however, the present results also demonstrate that comparable context effects can also be obtained with nonspeech stimuli.

We believe that our findings with nonspeech stimuli undermine the major conclusions of Miller and Liberman, who have argued that such context effects in speech perception reflect a form of “perceptual compensation” that is due to the specification of articulatory rate of the talker by overall stimulus duration. It seems very unlikely to us that listeners in the nonspeech control conditions carried out an appropriate perceptual “adjustment” for changes in articulatory rate, since the nonspeech stimuli were not perceived as speech signals at all in our experiments (see, however, Grunke & Pisoni, 1982; Schwab, 1981). Moreover, the present findings demonstrate quite clearly that the perceptual categorization of stimulus onsets as either “abrupt” or “gradual” is also influenced by later occurring events in the stimulus configuration itself and that nonspeech signals may also be processed in a “relational” and nonlinear fashion; that is, in a manner that is comparable to the perception of speech signals. The present nonspeech results are particularly striking because we replicated not only the contextual effects reported by Miller and Liberman (1979) for syllable duration, but also the effects observed when simulated formant transitions were

[1]It could be argued that our subjects perceived these nonspeech patterns as speech stimuli and that the context effects observed here were actually due to phonetic coding or covert labeling processes (see, e.g., Grunke & Pisoni, 1982; Remez, Rubin, Pisoni, & Carrell, 1981; Schwab, 1981). To insure that this did not occur, we administered a posttest questionnaire to ascertain whether subjects thought the sounds they heard resembled speech in any way. An examination of the responses to these questions indicated that none of the subjects perceived the stimuli as speech or speech-like. Indeed, in the debriefing session, the subjects were surprised to learn that the stimuli were actually based on speech replicas of the syllables [ba] and [wa].

added to the ends of the sinusoidal replicas of CV syllables, thus changing the internal structure of the stimulus pattern.

Finally, the present findings with nonspeech signals raise serious questions regarding the conclusions Eimas and Miller (1980) have drawn from their recent study with young infants which demonstrated comparable context effects for discrimination of the duration of formant transitions for the [b-w] distinction. While 2–4-month-old infants may show a form of context-conditioned sensitivity in discriminating differences in formant transition duration, it seems very unlikely to us that these context effects are brought about through the operation of perceptual mechanisms that are involved in the phonetic coding of speech signals or the interpretation of speech signals as linguistic entities such as phonemes, phonetic segments, or distinctive features. The finding that infants demonstrate sensitivity to context effects in discrimination of [b] and [w] *and* that they discriminate these differences in a relational manner, as do adults, in no way implies that they are in fact coding the speech signals phonetically, since comparable relational effects have now been obtained with nonspeech signals. As in the case of the discrimination of voice onset time, we suspect that infants are responding to the basic psychophysical or sensory properties of these acoustic signals without reference to coding them phonetically as speech (see Aslin & Pisoni, 1980; Aslin, Pisoni, Hennessy, & Perey, 1981; Jusczyk, Pisoni, Walley, & Murray, 1980; Pisoni, 1977, 1979).

The results of recent experiments using the same nonspeech signals with infants have also provided further evidence against the hypothesis that context effects in speech perception are due to phonetic coding or a specialized mode of processing. Jusczyk, Pisoni, Reed, Fernald, and Myers (1983) found that infants' discrimination of nonspeech contrasts differing in the duration of a rapid spectrum change was affected by overall stimulus duration in a manner that was identical to that reported by Eimas and Miller (1980) for speech stimuli. Thus, the context effects Eimas and Miller found with speech stimuli, suggesting a relational or nonlinear mode of processing, are not limited specifically to the perception of speech signals or to a distinctive phonetic mode of response. From these recent findings with infants, it would appear that such context effects in discrimination may simply reflect the operation of fairly general auditory processing capacities that infants use to perceive speech as well as other acoustic signals in their environment (see Jusczyk, 1982, for further discussion).

In summary, we have carried out several critical comparisons in perception between speech and comparable nonspeech control signals that differed in terms of the duration of a rapid spectrum change at stimulus onset, an acoustic cue that has been shown to be sufficient to distinguish between the stop [b] and the semivowel [w]. Our findings, using new synthetic speech stimuli, replicated the earlier context effects reported by Miller and Liberman (1979). Overall syllable duration clearly affected the location of the [b-w] labeling boundary. However, we also found similar context effects in the perception of nonspeech stimuli that varied in terms of the duration of a rapid spectrum change at onset. We interpret these results with nonspeech control stimuli as evidence that context effects in speech perception may not be peculiar to the perception of speech signals or to the listener's perceptual normalization or adjustment to the talker's speaking rate. Our findings with nonspeech control signals therefore call into question the major conclusions of Miller and Liberman that listeners somehow monitor or extract estimates of the talker's articulatory rate and then subsequently carry out normalization operations that adjust their perceptual criteria for interpreting a particular set of acoustic cues as a specific phonetic segment.

Our findings with nonspeech stimuli also raise important questions concerning Eimas and Miller's (1980) recent interpretations of infant discrimination data. These authors have

argued that the presence of context effects in the discrimination of [b-w] contrasts by young infants implies that they are perceiving speech signals in a “relational manner” like the mature adult listener. We believe that such conclusions are unwarranted given the results of the present study, which demonstrates comparable context effects and relational processing for nonspeech stimuli with adults, and the recent findings of Jusczyk et al. (1983), which demonstrate comparable context effects for discrimination of nonspeech signals by young infants. The present findings demonstrate that context effects in labeling due to stimulus duration and the internal structural organization of a complex stimulus pattern are not due to the interpretation or coding of the stimuli as speech signals per se or to a distinctive phonetic mode of response that is unique to processing speech signals by human listeners.

Acknowledgments

The research reported here was supported by NIMH Research Grant MH-24027, NINCDS Research Grant NS-12179, and NICHHD Research Grant HD-11915 to Indiana University. We thank Paul Luce for his assistance in running subjects, Tom Jonas for his help in programming, and Jerry C. Forshee for his expertise with the instrumentation and computer facilities in our laboratory. We also thank Peter W. Jusczyk for helpful comments and suggestions on an earlier version of the manuscript.

References

Aslin, RN.; Pisoni, DB. Some developmental processes in speech perception. In: Yeni-Komshian, G.; Kavanagh, JF.; Ferguson, CA., editors. Child phonology: Perception and production. New York: Academic Press; 1980.

Aslin RN, Pisoni DB, Hennessy BL, Perey AJ. Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. Child Development. 1981; 52:1135–1145. [PubMed: 7318516]

Eimas PD, Miller JL. Contextual effects in infant speech perception. Science. 1980; 209:1140–1141. [PubMed: 7403875]

Grunke ME, Pisoni DB. Some experiments on perceptual learning of mirror-image acoustic patterns. Perception & Psychophysics. 1982; 31:210–218. [PubMed: 7088662]

Jusczyk, PW. Auditory versus phonetic coding of speech signals during infancy. In: Mehler, J.; Walker, ECT.; Garrett, MF., editors. Perspectives on mental representation. Hillsdale, N.J: Erlbaum; 1982.

Jusczyk PW, Pisoni DB, Reed MA, Fernald A, Myers M. Infants’ discrimination of the duration of rapid spectrum changes in nonspeech signals. Science. 1983; 222:175–177. [PubMed: 6623067]

Jusczyk PW, Pisoni DB, Walley AC, Murray J. Discrimination of relative onset time of two-component tones by infants. Journal of the Acoustical Society of America. 1980; 67:262–270. [PubMed: 7354194]

Klatt DH. Speech perception: A model of acoustic-phonetic analysis and lexical access. Journal of Phonetics. 1979; 7:279–312.

Klatt DH. Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America. 1980; 67:971–995.

Lane HL. The motor theory of speech perception: A critical review. Psychological Review. 1965; 72:275–309. [PubMed: 14348425]

Liberman AM. On finding that speech is special. American Psychologist. 1982; 37:148–167.

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychological Review. 1967; 74:431–461. [PubMed: 4170865]

Liberman, AM.; Studdert-Kennedy, M. Phonetic perception. In: Held, R.; Leibowitz, H.; Teuber, HL., editors. Handbook of sensory physiology: Perception. New York: Springer-Verlag; 1978.

(a). Miller JL. Contextual effects in the discrimination of stop consonant and semivowel. Perception & Psychophysics. 1980; 28:93–95. [PubMed: 7413419]

(b). Miller, JL. The effect of speaking rate on segmental distinctions: Acoustic variation and perceptual compensation. In: Eimas, PD.; Miller, JL., editors. Perspectives on the study of speech. Hillsdale, N.J: Erlbaum; 1980.

Miller JL. Some effects of speaking rate on phonetic perception. Phonetica. 1981; 38:159–180. [PubMed: 7267718]

Miller JL, Eimas PD. Studies on the categorization of speech by infants. Cognition. 1983; 13:135–165. [PubMed: 6682742]

Miller JL, Grosjean F. How the components of speaking rate influence perception of phonetic segments. Journal of Experimental Psychology: Human Perception & Performance. 1981; 7:208–215. [PubMed: 6452497]

Miller JL, Liberman AM. Some effects of later-occurring information on the perception of stop consonant and semivowel. Perception & Psychophysics. 1979; 25:457–465. [PubMed: 492910]

Pisoni DB. Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops. Journal of the Acoustical Society of America. 1977; 61:1352–1361. [PubMed: 881488]

Pisoni, DB. Perception of speech vs. nonspeech: Evidence for different modes of processing. Proceedings of the Ninth International Congress of Phonetic Sciences,; Copenhagen. August 1979; Copenhagen: Institute of Phonetics, University of Copenhagen; 1979.

Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. Science. 1981; 212:947–950. [PubMed: 7233191]

Schwab, EC. Unpublished doctoral thesis. State University of New York; Buffalo: Aug. 1981 Auditory and phonetic processing for tone analogs of speech.

Summerfield Q. Articulatory rate and perceptual constancy in phonetic perception. Journal of Experimental Psychology: Human Perception and Performance. 1981; 7:1074–1095. [PubMed: 6457109]

Woodworth, RS. Experimental psychology. New York: Holt; 1938.

REFERENCE NOTE

1. Kewley-Port, D. Research on Speech Perception Progress Report No. 3. Bloomington: Department of Psychology, Indiana University; 1976. A complex-tone generating program.

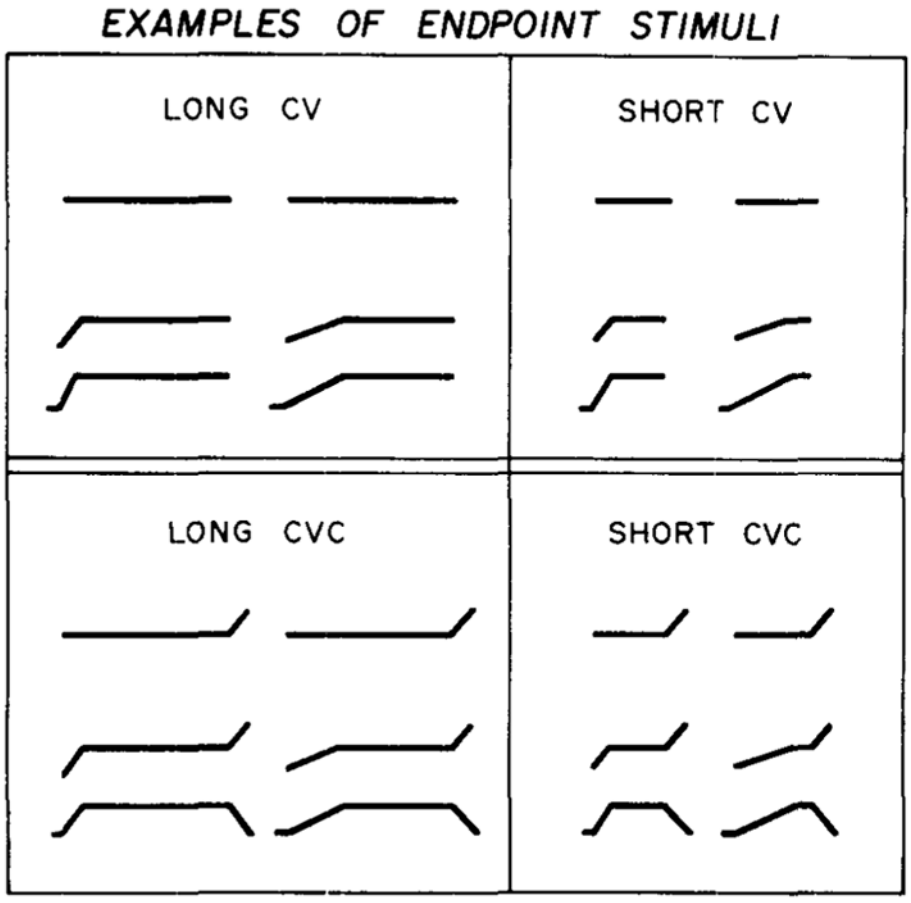

Figure 1.
Schematic representations of the formant motions of the endpoint stimuli corresponding to [b] and [w]. The top panel shows the long and short CV syllables; the bottom panel shows the long and short CVC syllables which had final transitions added to the steady-state vowel.

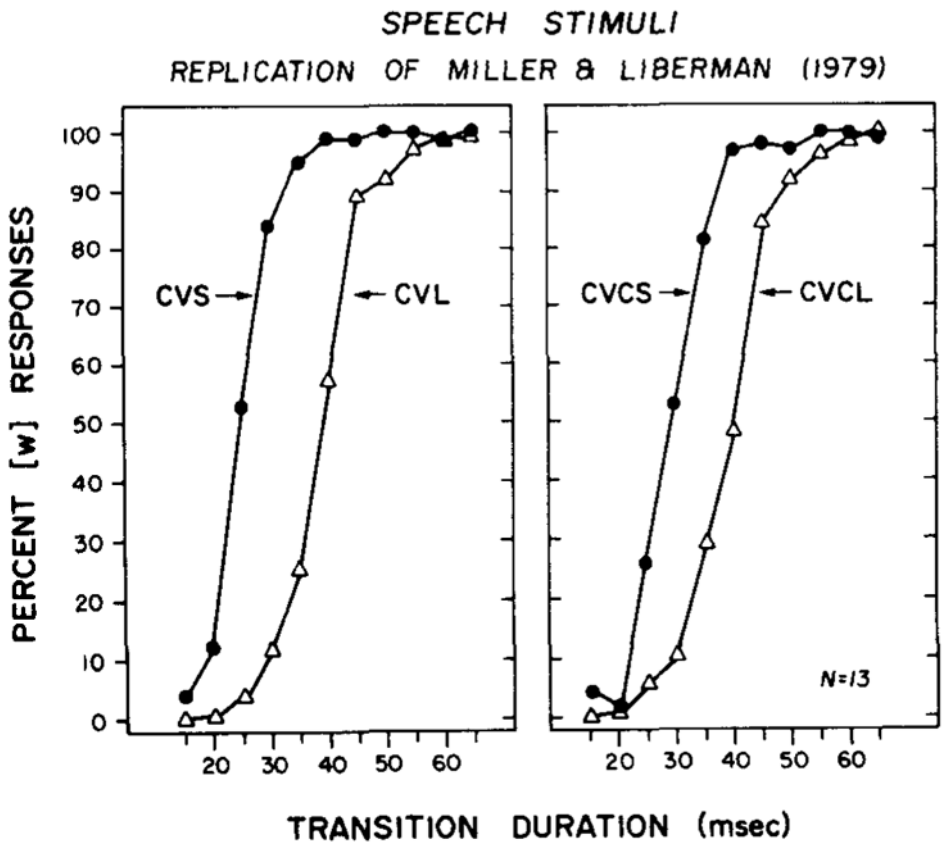

Figure 2.

Labeling data for the four sets of synthetic speech stimuli differing in the duration of their formant transitions for [b] and [w]. Percent [w] responses are shown as a function of transition duration. The data in the left panel are for the CV syllables; the data in the right panel are for the CVC syllables. Filled circles represent the short stimulus in each condition, and open triangles represent the corresponding long stimulus.

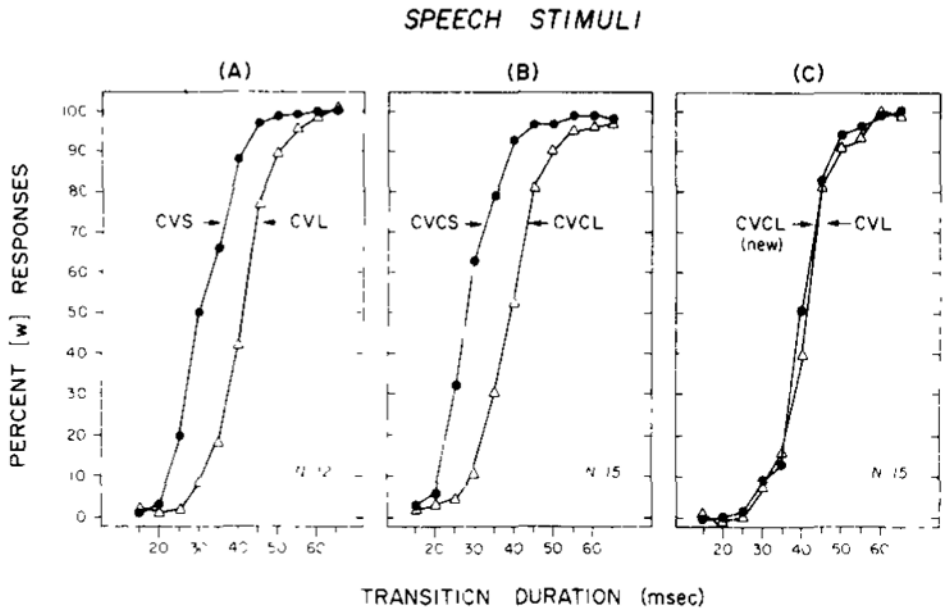

Figure 3.

Labeling data for synthetic CV and CVC syllables. Panels A and B display the data for CV and CVC syllables that differ in overall duration; the filled circles are short stimuli, and the open triangles are long stimuli. Panel C displays data for CV and CVC syllables that were matched for overall duration.

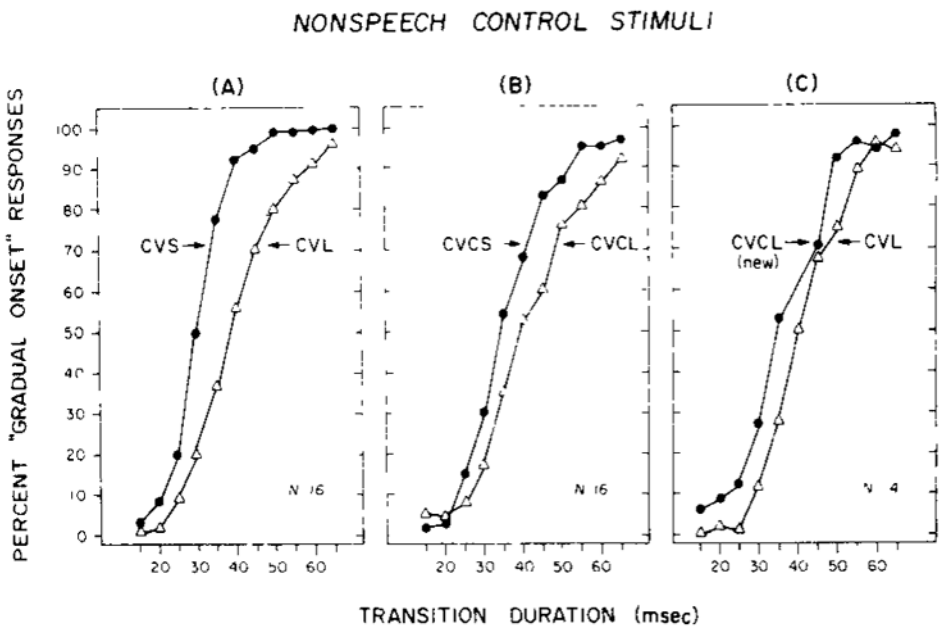

Figure 4.

Labeling data for nonspeech control stimuli that were modeled after the CV and CVC syllables used in Experiment 1. The three panels display the percent of "gradual onset" responses as a function of transition duration. Panels A and B show the data for stimulus comparisons that differ in overall duration; Panel C shows the data for nonspeech stimuli that were matched in terms of overall duration and differed only in terms of the internal organization of the stimulus pattern. The nonspeech data shown here are exactly parallel to the speech data shown in the previous figure.

Table 1

Means and Standard Deviations in Milliseconds for Subjects' Labeling Functions in Experiment 1 Using Speech Stimuli

	CVS - CVL				CVCS - CVCL			
	CVS		CVL		CVCS		CVCL	
Ss	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	16.85	12.82	41.72	7.46	24.07	11.49	38.71	7.60
2	27.36	9.38	40.77	7.33	34.91	7.95	41.30	7.41
3	23.65	13.83	48.22	12.17	31.96	11.86	46.07	9.51
4*	28.86	8.93	40.56	7.35	35.49	7.31	45.02	7.86
5	28.85	8.91	37.54	7.85	34.66	7.52	38.71	7.57
6	28.38	9.50	39.26	7.06	31.66	8.35	39.58	7.93
7	25.36	10.56	37.03	7.70	34.77	7.75	39.13	8.26
8	32.23	8.13	40.00	6.82	31.90	8.33	38.30	6.94
9	22.15	12.16	41.62	7.50	29.65	8.68	39.08	7.88
10	31.85	8.27	41.58	7.20	31.77	8.74	42.70	7.25
11	9.90	16.63	35.75	7.75	28.03	10.02	37.38	8.03
12	15.39	13.59	29.83	9.14	17.30	13.35	29.18	9.38
13	24.99	9.96	37.28	7.07	28.33	9.03	38.40	7.42
$\bar{x}$ = 24.30		$\bar{x}$ = 39.32		$\bar{x}$ = 29.58		$\bar{x}$ = 39.71		

Table 2

Means and Standard Deviations in Milliseconds for Subjects' Labeling Functions in Experiment 2 Using Nonspeech Stimuli

	GROUP A					GROUP B					GROUP C				
	CVS - CVL					CVCS - CVCL					CVCL (new) - CVL				
	CVS		CVL			CVCS		CVCL			CVCL (new)		CVL		
Ss	Mean	SD	Mean	SD	Ss	Mean	SD	Mean	SD	Ss	Mean	SD	Mean	SD	
1	27.36	9.75	45.41	9.38	1	36.91	8.30	41.70	9.58	1	36.18	7.66	35.08	8.07	
2	31.67	9.08	49.27	12.97	2	38.23	9.45	41.19	9.82	2	36.11	10.39	40.61	9.01	
3	29.63	9.41	43.60	9.29	3	36.09	8.36	37.27	9.77	3	42.20	9.35	38.87	8.12	
4	24.87	10.24	29.93	9.38	4	35.21	8.90	44.39	11.91	4	38.26	8.53	41.27	9.02	
5	20.21	12.80	45.47	11.64	5	35.65	7.72	54.45	13.17	5	36.31	8.91	43.70	8.19	
6	35.12	8.09	36.74	9.17	6	39.11	17.75	33.18	15.38	6	33.82	12.48	40.93	9.65	
7	33.56	8.26	31.48	8.62	7	33.64	8.20	42.99	9.04	7	31.48	20.47	60.23	15.86	
8	34.44	8.12	32.68	9.93	8	37.26	8.46	39.14	9.18	8	34.87	20.85	44.81	11.46	
9	30.49	9.60	44.33	10.87	9	44.53	9.04	48.96	14.10	9	38.51	8.78	44.21	9.20	
10	32.30	11.81	36.19	8.35	10	29.49	9.78	50.77	13.02	10	33.71	8.78	42.96	10.71	
11	34.50	8.30	38.89	8.20	11	48.45	10.57	27.67	11.58	11	34.67	8.61	39.48	8.31	
12	31.50	11.96	50.47	13.87	12	31.00	13.15	44.05	20.97	12	36.07	17.26	45.43	11.91	
13	35.62	7.90	45.23	9.95	13	33.69	14.04	41.29	11.06	13	36.23	9.05	39.92	8.05	
14	32.05	8.14	37.74	7.37	14	41.98	10.28	42.54	9.07	14	33.43	13.00	44.00	8.13	
15	35.17	14.12	50.25	18.91	15	38.39	9.69	33.66	13.68						
16	31.22	9.10	40.19	9.12	16	34.08	10.09	57.58	16.20						
$\bar{x}$ = 31.21			$\bar{x}$ = 41.24		$\bar{x}$ = 37.37			$\bar{x}$ = 42.32			$\bar{x}$ = 35.75		$\bar{x}$ = 44.50		