Perception & Psychophysics
2005, 67 (6), 939-950

Perceptual effects of preceding nonspeech rate on temporal properties of speech categories

TRAVIS WADE and LORI L. HOLT
Carnegie Mellon University, Pittsburgh, Pennsylvania

The rate of context speech can influence phonetic perception. This study investigated the bounds of rate dependence by observing the influence of nonspeech precursor rate on speech categorization. Three experiments tested the effects of pure-tone precursor presentation rate on the perception of a [ba]–[wa] series defined by duration-varying formant transitions that shared critical temporal and spectral characteristics with the tones. Results showed small but consistent shifts in the stop–continuant boundary distinguishing [ba] and [wa] syllables as a function of the rate of precursor tones, across various manipulations in the amplitude of the tones. The effect of the tone precursors extended to the entire graded structure of the [w] category, as estimated by category goodness judgments. These results suggest a role for durational contrast in rate-dependent speech categorization.

Since speech unfolds over time, rate is a potential source of variability in the realization of phonological contrasts whose primary cues are temporal. Indeed, it is readily observable that time cues to contrasts involving vowel identity (e.g., Gay, 1978; Port, 1981), gemination (e.g., Pickett & Decker, 1960), consonant voicing (e.g., Miller, Green, & Reeves, 1986; Summerfield, 1975), and manner of articulation (Mack & Blumstein, 1983; for a review, see Miller, 1981, 1987; Miller & Baer, 1983) vary substantially in their realization across speaking rates. The result is often overlapping classes of sounds that are not readily separable by time-invariant criteria. However, human perceivers appear to overcome this variability by interpreting the relevant acoustic cues in what may be described as a rate-dependent manner. For example, listeners' perceptual categorization boundaries for a series of stimuli varying acoustically in the duration of an initial formant transition and perceptually from [ba] to [wa] steadily increase along the formant-transition duration dimension (such that there are more [ba] responses) when rate, as conveyed by overall syllable duration, decreases (Miller & Liberman, 1979). This shift parallels the shift observed in speech production; formant transition durations are typically longer in slower speech. Thus, rate-dependent perception appears to compensate for patterns in the natural language environment.

There is some indication that this rate dependence is diminished, and perhaps is less necessary, when targeted contrasts possess more complex, natural sets of cues. For example, Shinn and colleagues (Shinn & Blumstein, 1984; Shinn, Blumstein, & Jongman, 1985) observe that the effect of syllable length on perception of the [b] versus [w] distinction gradually disappears as [b] and [w] tokens are synthesized in a less stylized manner and are differentiated by additional, nontemporal cues such as formant onset frequencies and formant-frequency trajectories. However, Miller and Wayland (1993) show that in the presence of multitalker babble noise, rate-dependent category boundary shifts hold up even without these additional cues, suggesting that rate-dependent categorization is useful in natural speaking situations.

Parallel findings of rate dependence in the perception of other temporal contrasts are abundant (e.g., Ainsworth, 1974; Fujisaki, Nakamura, & Imoto, 1975; Summerfield, 1981). Furthermore, it has been demonstrated that speaking rate affects not only the locations of perceptual category boundaries, but also the internal structure of the categories. When trained participants rate consonants as voiceless stops based on voice onset time (VOT; Miller & Volaitis, 1989b) or glides based on formant-transition durations (Miller, O'Rourke, & Volaitis, 1997), entire category goodness rating curves, including both ends of best-exemplar ranges, shift with changes in speaking rate. As speaking rate decreases, stimuli with increasingly longer formant-transition durations are judged to be the best-sounding [w] category members. This is in marked contrast to patterns observed for other, higher level factors affecting phonetic categorization; the effects of lexical status, for example, seem to be limited to ambiguous boundary regions and do not affect overall category structure (Allen & Miller, 2001).

The standard interpretation of the pervasive perceptual effects of rate is that they serve to accommodate recovery of intended phonetic units; that is, perception is mediated by knowledge of the rate-dependent nature of tem-

This work was supported by a James S. McDonnell Foundation award for Bridging Mind, Brain, and Behavior to L.L.H., NIH Grant 5 RO1 DC04674-02 to L.L.H., and a fellowship from the NIH Postdoctoral Training Grant on Individual Differences in Cognition. We thank Christi Adams and Ashley Episcopo for help in conducting the experiments. Correspondence should be addressed to T. Wade, PositScience, 114 Sansome Street, 5th Floor, San Francisco, CA 94104 (e-mail: travis.wade@positscience.com).

poral contrasts, or recovery of their articulatory sources (e.g., Fowler, 1980; Fowler, 1990; Miller & Baer, 1983; Summerfield, 1981). One observation that might give pause to such a causal view of the link between production and perception is that listeners generally fail to compensate optimally for the rate-dependent patterns that actually occur in speech. For example, continuous increases in formant-transition lengths of English onset [w] glides (Miller & Baer, 1983) and in VOT duration of syllable-initial [p] consonants (Miller et al., 1986) have been observed for speech produced across very fast to very slow speaking rates. Pairing these speech production measurements with similar observations of rate-dependent [b] productions, it is possible to calculate optimal rate-dependent shifts in perceptual boundaries for consonant voicing and manner of articulation. However, listener categorization reflects these patterns of speech production only qualitatively and, particularly for the [b]–[p] boundary, perceptual shifts are more limited and negatively accelerated at longer VOT values (Miller et al., 1986) than is predicted from production measurements. Pind (1995) has observed a similar pattern for VOT in Icelandic consonants.

Miller and colleagues offer two possible explanations for such inconsistencies: (1) limits on the flexibility of perceptual boundaries and (2) limits on the effect of later occurring information (syllable duration) on categorization (Miller et al., 1986). An alternative explanation is that rate-dependent categorization patterns do not actually reflect detection of, and compensation for, the rate of articulation at all, but rather stem fortuitously from perceivers' sensitivity to the rate of auditory events more generally. It is well known that sensorineural systems are more sensitive to change than to absolute levels in perceived signals, enabling them to process stimuli over wider physical ranges than would otherwise be possible (e.g., Purves & Lotto, 2003). Such contrastive perceptual patterns are in line with the directionality of rate-dependent speech categorization. Faster speech contexts result in the perception of subsequent temporal information as being relatively slower than the same information preceded by slower context stimuli. Contrast may have additional benefits in speech perception, since due to vocal-tract mechanics and coarticulation the absolute physical properties of speech events are highly influenced by adjacent events. It has been suggested, in fact, that much of the observed context dependence in speech perception results from a tendency to perceive acoustic events in a relative manner (Holt, 2005; Holt & Kluender, 2000; Kluender, Coady, & Kiefte, 2003; Lotto & Kluender, 1998; Lotto, Kluender, & Holt, 1997). Along these lines, Diehl and Walsh (1989) assert that apparent speaking rate effects may result from durational contrast whereby temporal cues, such as the length of an acoustic segment, are perceived relative to nearby segments rather than absolutely. Thus, for example, listeners may accept longer formant transitions for [b] at slower speaking rates not because they reveal something about typical articulation, but simply because they appear shorter than longer surrounding segments, a pattern previously documented for nonspeech tone durations (Goldstone, Boardman, & Lhamon, 1959; Goldstone, Lhamon, & Boardman, 1957; Walker & Irion, 1979; Walker, Irion, & Gordon, 1981). Oller, Eilers, Miskiel, Burns, and Urbano (1991) speculated further on the precise workings of a durational contrast effect. To account for the nonlinear relation of boundary shifts and apparent rate, they introduced a "comparable range model" whereby contrast effects only occur between two sounds (e.g., a transition and following steady state) with absolute durations falling within a similar range, perhaps a 1:1 to 1:2 ratio. Oller et al. also introduced a possible mechanism ("pivot point error") to account for the effect, postulating that sluggish perceptual recognition of a change in formant trajectory results in perceptual overestimation of the duration of transitions preceding short vowels.

Some evidence that speaking rate effects are not unique to speech was offered by Pisoni, Carrell, and Gans (1983), who observed boundary shifts analogous to those found for speech (Miller & Liberman, 1979) when listeners characterized nonspeech sine wave approximations of [ba]–[wa] formant transitions with different durations as either "abrupt" or "gradual." Diehl and Walsh (1989), addressing the concern that these sounds may have been too speechlike, achieved the same effect with single frequency-modulated sine wave stimuli. Whether stimuli tracked a normal bilabial $F1$ frequency trajectory or a reverse pattern that is impossible in normal speech, participant classification of transition abruptness was robustly dependent on overall stimulus length.

The significance of these and similar findings has been challenged, perhaps most notably by Fowler (1990; see also Fowler, 1992), on the grounds that perception of nonspeech analogs cannot be directly compared with speech perception, since speech has a clear, identifiable environmental source, whereas nonspeech analogs (pure tones, for example) do not. Fowler bolsters this argument with comparison of perception of a set of natural recordings of sounds produced by real mechanical events and sine wave analogs of these sounds. Listeners showed contrastive perceptual patterns in the classification of sine wave analogs, but perception of the natural events the sine waves modeled was contrastive in some cases, but not in others. Thus, contrary to the expectations of a durational contrast account, in some cases, perception of the event and its sine wave analog diverged, although their acoustics shared common temporal characteristics. Fowler argues that this is because, whereas the events possessed a real acoustic source that could be apprehended from the waveform, the sine wave analogs did not. By analogy, Fowler suggested that studies investigating perception of nonspeech stimuli are not helpful in understanding speech perception because speech, like mechanical events, originates from an environmental source, whereas simple nonspeech analogs do not. Perception of the two classes of stimuli, by this view, arises from different origins.

Perhaps informative to these theoretical positions are additional data from studies with infant and nonhuman participants. Infants demonstrate rate-dependent perception of the [b]–[w] contrast, responding to more sounds as [w] before shorter vowels in a sucking-habituation study (Eimas & Miller, 1980). So do budgerigars, which have the ability to mimic human sounds by using a suprasyringeal cavity (Dent, Brittan-Powell, Dooling, & Pierce, 1997) and Japanese macaques, which do not (Sinnott, Brown, & Borneman, 1998). These studies indicate that significant knowledge of rate-dependent speech patterns may not be necessary for rate-dependent speech categorization. Nevertheless, these results do little to differentiate whether rate-dependent speech perception arises from general perceptual mechanisms like durational contrast or from recovery of the environmental source (i.e., articulatory gestures), as suggested by Fowler (1990). Either a sufficiently developed auditory system, or alternatively, a sufficiently general ability to directly recover the natural (articulatory) sound source could potentially account for infant and nonhuman patterns of behavior.

Thus, despite differing theoretical accounts, the present data by and large leave open the question of how rate-dependent speech categorization arises. Nevertheless, Fowler's (1990) criticism does reveal theoretical limitations of experimental designs in which the nature of stimuli (speech vs. nonspeech) is manipulated and compared across experiments or conditions. In determining the nature of rate-dependent speech perception, it is informative to test not only whether context rate affects speech and nonspeech perception in the same way (Diehl & Walsh, 1989; Pisoni et al., 1983), but also whether rate effects observed for *speech* may be elicited by manipulating the rate of nonspeech contexts that are not readily attributable to an articulatory source. The present study was designed to test this possibility.

The studies of rate dependence in speech perception discussed so far have concentrated primarily on a single source of rate-relaying context—namely, the length of the vowel immediately following a target segment. Indeed, the vast majority of studies to date have manipulated the duration of target-adjacent segments, since nearby sounds appear to play a critical role in determining a context rate. Newman and Sawusch (1996), for example, observed that the rate of nonadjacent speech following a target segment affected its perception only if it fell within a short (around 300 msec) window after the target. This apparent primacy of local segments in driving rate effects is consistent with intrinsic models of timing (e.g., Fowler, 1980; Summerfield, 1981) that assume that perception mainly considers a target segment's immediate articulatory context when accounting for rate.

This stimulus paradigm is unsuitable for the aims of the present study. Substitution of target-adjacent segments with nonspeech acoustic events is problematic, since these segments often overlap acoustically with the target, carrying critical information about its identity. However, it is also known that rate information further removed from a given sound can affect its phonetic categorization (e.g., Kidd, 1989; Summerfield, 1981) and internal category structure (Wayland, Miller, & Volaitis, 1994). The present study, therefore, adopts methods from the latter studies, using sentence-length nonspeech sequences as precursors to target speech segments. A series of experiments was designed to measure the perception of formant-transition durations as cues to the [b]–[w] distinction and [w] category goodness in the presence of these precursors, to determine whether nonspeech as well as speech contexts can affect speech categorization of stimuli defined by a temporal acoustic cue.

Gordon (1988) takes an important step in showing that the speech rate information cued by more temporally distant events may not be entirely speech specific in nature, observing that the rate of a preceding carrier sentence can influence a consonant voice distinction even when the sentence is severely degraded (by low-pass filtering at 375 Hz or imposing its amplitude envelope on a sine wave, but not on white noise). Gordon stops short of suggesting that the effects are contrastive or general in nature, however, addressing only which aspects of the speech signal might be responsible for syllable-extrinsic rate detection.

Of the studies that have tested directly for nonspeech effects on temporal speech contrasts, nonspeech rates notable failure to influence speech categorization is reported by Summerfield (1981). In this study, listeners heard a portion of a familiar melody followed by a syllable. Listeners' judgment of the syllable's initial consonant ([b] vs. [p]) was not affected by the rate of the melody, although the length of the following vowel did have an effect. Within a durational contrast account of rate effects, there is at least one major reason to predict this null result (see also Gordon,1988, for more general reasons). Whereas the precise workings of durational contrast are as yet unclear, it seems that some degree of spectral continuity (Walker & Irion, 1979) and temporal similarity (Oller et al., 1991) are necessary between contrast-providing context and target stimuli. In the case of phonetic contrasts such as stop–continuant ([b]–[w]) and voiced–voiceless ([b]–[p]), temporal cues for which rate-dependent perception has been investigated include voicing offset and formant transitions. Both of these acoustic events involve temporal durations in the range of tens of milliseconds, and both (particularly the stop–continuant distinction) involve critical spectral movement in the $F1$–$F2$ range. Summerfield's (1981) melodic precursors, on the other hand, were composed of tones with durations of hundreds of milliseconds, on the order of the syllable duration rather than the duration of the syllable–initial VOT that distinguished the stimuli. These tones, described as "machine-like buzzes," did contain spectral energy in speech-formant ranges, with two resonances held constant at 1.0 and 3.0 kHz and a third oscillating at 100 Hz between 1.5 and 2.5 kHz. However, the melodic line, for which rate was varied, consisted of notes with frequencies in the range of the fundamental frequency ($f0$) of the following speech, not its formants.

Thus, it is possible that issues of acoustic continuity, rather than speech-specific processing, were responsible for the lack of rate dependence.

It seems reasonable, therefore, to resurrect the melodic precursor paradigm in examining possible nonspeech effects on categorization of speech along a temporal dimension, provided that the precursor stimuli possess acoustic characteristics well matched to the acoustic properties of the targeted speech contrast. The present experiments examine the extent to which sequences of precursor tones sharing spectral and temporal properties with following formant transitions can affect the stop–continuant perceptual boundary between initial [b] and [w] consonants in phonetic categorization. In addition, goodness judgments of [w] segments were elicited in these same contexts to observe the nature of any effects with respect to overall category structure.

EXPERIMENT 1

The purpose of Experiment 1 was to determine whether sequences of pure tones, comparable in duration and frequency to following formant transitions, affect speech categorization along a series of stimuli varying perceptually from [ba] to [wa]. Also of interest was whether tones in the range of a particular formant ($F1$ vs. $F2$, for which transition rate varies across speech stimuli) would produce superior effects.

Method

Participants. Thirty-four listeners participated, with half assigned to one condition and half to another in a random manner. The participants were undergraduate college students. All were native English speakers with no known or obvious speaking or hearing disorders. They received psychology course credit for their participation.

Stimuli. The stimuli were synthetic syllables from a [ba]–[wa] continuum preceded by pure tone sequences. The stimuli sampling the [ba]–[wa] continuum were modeled after intermediate-length stimuli used in the acoustically simplest (least "natural") conditions of previous studies by Shinn and colleagues (Shinn & Blumstein, 1984; Shinn et al., 1985). Syllables were generated at 11.025 kHz using the HLsyn (Sensimetrics Corporation) implementation of the Klatt (1980) speech synthesizer's cascade branch. All syllables were 171 msec in total duration. The first and second formant resonances, respectively, were held constant at 234 and 616 Hz for 40 msec before rising linearly to 769 and 1232 Hz. Continuum members varied only in length of $F1$ and $F2$ transitions; this duration was varied from 15 to 65 msec in eleven 5-msec increments. The fundamental and third and fourth formant frequencies were held constant at 115, 2862, and 3500 Hz, respectively, for the entire duration of a syllable. Amplitude of voicing (AV) rose from 0 to 60 dB during the syllables' initial 5 msec and decreased to zero over the final 5 msec. Formant bandwidths for $F1$–$F4$ were 200, 90, 150, and 250 Hz, respectively. All remaining synthesis parameters were held constant at default values across continuum members. Following synthesis, the stimuli's spectral and temporal dimensions (MATLAB; The MathWorks, Inc.) were inspected and found to be in close accord with synthesizer parameters. In particular, formant-transition duration was observed to vary uniformly from 15 to 65 msec across the continuum. Finally, continuum members were equalized for RMS amplitude.

The precursor stimuli were designed to approximate the length of a short carrier sentence and to be spectrally and temporally matched with the acoustic characteristics of the initial formant transitions of the target syllables while remaining acoustically simple and unambiguously nonspeech in nature. Each precursor was composed of a 1.2-sec sequence of sine wave tones with frequencies sampled randomly from either the entire $F1$–$F2$ range (Condition 1; 234–1232 Hz) or only the $F2$ range (Condition 2; 769–1232 Hz) of the following syllable. Each stimulus in the experiment possessed a unique random sampling of tone frequencies, so the frequency ranges of each condition were well represented within and across precursor stimuli. Sequences consisted entirely of either short or long tones. Fast sequences were made up of 30 short tones with interonset intervals (IOIs) of 40 msec (a duration 15 msec shorter than the 55-msec onset-plus-transition duration of the most [ba]-like target syllable). Slow sequences were made up of 10 long tones with IOIs of 120 msec (15 msec longer than the 105-msec duration of the most [wa]-like target). To ensure that precursors were perceived as sequences of tones rather than a continuous stream, a short silent gap was added to separate the tones in time. The duration of this gap was 10-msec for both fast and slow sequences and provided an audible segmentation while also allowing the tones' periodicity to be clearly perceived. A "tonal event" in fast sequences, then, consisted of a 30-msec tone followed by a 10-msec silent interval; slow-sequence tonal events were made up of a 110-msec tone and a 10-msec silent interval. All tones had 5-msec onset and offset amplitude ramps. Target syllables were appended to the sequences following the final 10-msec silent interval.

Twenty fast and 20 slow sequences (each with a different sampling of tone frequencies) were randomly generated and paired with each [ba]–[wa] continuum member, resulting in 440 different stimuli. RMS amplitude of each tone sequence was normalized to that of the following syllable. Representative waveforms of fast and slow stimulus types for Experiment 1 are shown in the top panel of Figure 1.

Procedure. The participants heard stimuli from either Condition 1 or Condition 2 in a single session. Acoustic presentation was under the control of Tucker-Davis Technologies System II hardware; stimuli were converted from digital to analog, low-pass filtered at 4.8 kHz, amplified and presented diotically over linear headphones (Beyer DT-150) at approximately 70 dB SPL(A) to the participants, who were seated in sound-attenuated booths. They were instructed to listen to the entire sequence and indicate whether the syllable most resembled "ba" or "wa" by using response buttons.

Results

Responses for both conditions are shown in the top panel of Figure 2. A repeated measures ANOVA of probit-defined boundary locations (Finney, 1971) showed a significant shift of the [b]–[w] boundary in the direction of shorter transition durations (more [wa] responses) for syllables preceded by fast sequences than by slow sequences, in both Condition 1 [$F(1,16) = 12.81$, $p = .003$] and Condition 2 [$F(1,16) = 13.55$, $p = .002$]. This is the same direction as the speaking rate effects commonly observed in speech production (Miller & Baer, 1983) and perception (Miller & Liberman, 1979). A mixed design ANOVA comparing the categorization patterns produced by the different range of tone frequencies in Conditions 1 and 2 revealed a significant effect of tone duration [$F(1,32) = 25.83$, $p < .001$], but no effect of tone frequency range [$F(1,32) = .341$, $p = .564$], and no range × duration interaction [$F(1,32) = .32$, $p = .575$]. Thus, a similar small-but-reliable effect of precursor tone rate was observed, whether tone frequencies sampled the $F1$–$F2$ range or only the $F1$ range. This sug-

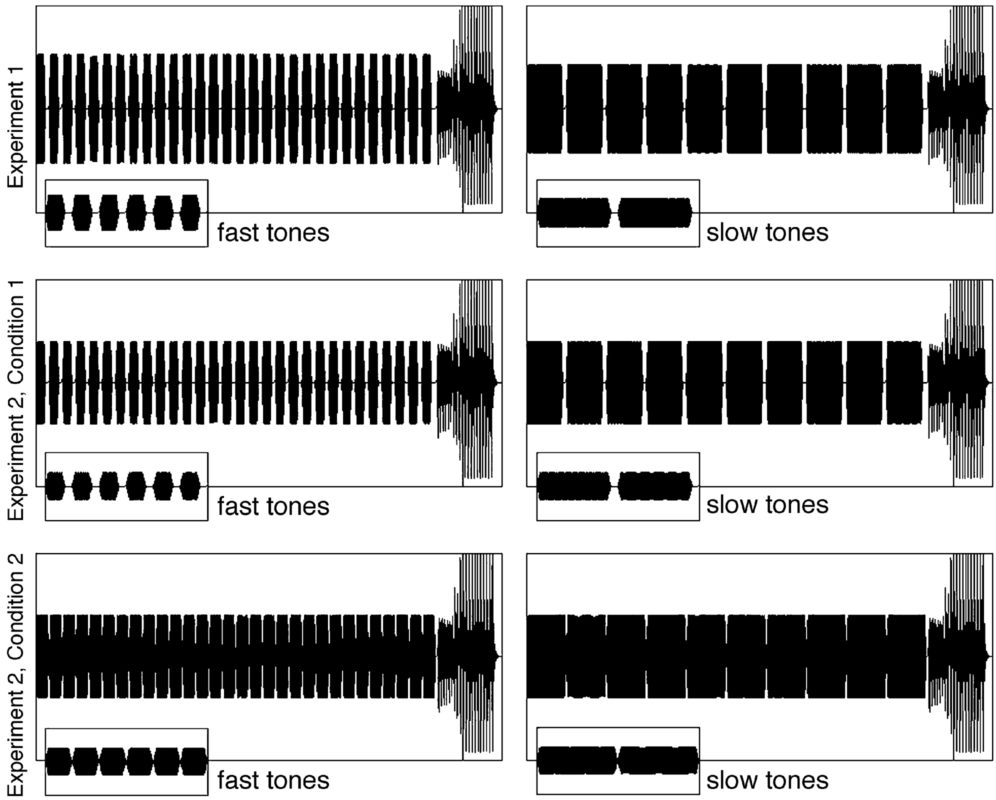

Figure 1. Representative waveforms showing amplitude and duration characteristics of precursor sequences used in Experiments 1 and 2. The insets in the bottom left corner of each graph show an expanded view of the first 240 msec of the waveform.

gests that if Summerfield's (1981) failure to observe such an effect was due to lack of spectral continuity, that continuity may be obtained by aligning nonspeech precursors' frequency with that of either of the following formants ($F1$ or $F2$), with transition lengths contributing to the temporal distinction.

The nonspeech tone precursors possessed no information about articulatory gestures, so these findings present immediate problems for an account of rate effects that depends on the recovery of articulatory events. The results are consistent with a durational contrast account. Nevertheless, the fast and slow tone sequences of Experiment 1 did not differ *only* in duration. Since identical intertone silent intervals and on–off ramps were used for both tone types and RMS amplitude was normalized across the entire precursor sequence, it can be seen in Figure 1 that individual fast (short) tones were somewhat greater in maximum amplitude than individual slow tones, compensating for the larger numbers of ramps and silent intervals. Thus, although fast and slow precursor sequences carried the same amount of total power, the maximum intensity level of individual short-tone waveforms was approximately 1.22 times that of individual long tones, a 0.8-dB difference in sound level. As a result, fast sequences possessed slightly more power than did slow sequences across the 30-msec period preceding the last silent interval before a target syllable. It is unlikely that a short-term peripheral process such as forward masking, operating on such a fast time scale, could result in the categorization differences observed in Experiment 1 (e.g., Moore & Glasberg, 1983). However, it is necessary to remove any confounds potentially produced by differences in maximum or overall amplitude across precursors in determining whether differences in [b]–[w] perception resulted from nonspeech precursor rate.

EXPERIMENT 2

Experiment 2 was designed to address this issue, examining whether the rate of the precursor tones indeed caused the observed context effects in Experiment 1 rather than the precursors' amplitude characteristics. In one condition, the absolute maximum amplitude of precursor sequences, rather than the overall RMS amplitude, was held constant across tone types. In another condition, the amplitude envelope of tones was also held constant, creating amplitude envelopes that were virtually flat with near-continuous tones in both conditions.

Method

Participants. The participants were 24 undergraduate college students. All were native English speakers with no known or obvious speaking or hearing disorders. They received psychology course credit for their participation.

Stimuli. As in Experiment 1, the stimuli were synthetic syllables from a [ba]–[wa] continuum preceded by pure tone sequences. The syllables were identical to those used in Experiment 1; tone sequences were identical in frequency composition and intertone interval to those in Experiment 1's Condition 2, differing only in tone amplitude and intertone silent interval duration. In two conditions, steady-state portions of individual tones, rather than entire precursor sequences including on–off ramps and silent intervals, were adjusted to the RMS amplitude of the following syllable, effectively normalizing the maximum amplitude of sequences across precursor rates. In Condition 1, this was accomplished by elongating individual tones to completely eliminate silent intervals and adding 5-msec on–off ramps only after individual tone amplitudes were adjusted, resulting in a steady stream of connected tone events. Thus, overall sequences of fast (40-msec) tones possessed slightly less energy overall but had the same maximum amplitude as sequences of slow (120-msec) tones. This was more extremely the case in Condition 2, which had tone-ramp-silence composition identical to that of Experiment 1's Condition 2 but with tone sequences normalized for maximum, rather than overall, RMS amplitude. Representative waveforms of stimuli for these conditions are shown in the second and third rows of Figure 1. Number, composition, and randomization of stimulus types within tests were identical to Experiment 1.

Procedure. Fifteen randomly selected participants heard stimuli from Condition 1 in a single session; a subset of 7 of the same participants and a remaining 9 participants heard stimuli from Condition 2. The procedures and apparatus were identical to those in Experiment 1. The participants were instructed to listen to an entire sequence and indicate whether the syllable most resembled "ba" or "wa" by using response buttons.

Results

Data from 2 participants (1 from Condition 1, 1 from Condition 2) were discarded for failing to reflect a reliable stop–continuant distinction based on formant transition (probit-defined boundaries were outside the limits of the continuum). Responses for the remaining tests are shown in the middle panel of Figure 2. As in Experiment 1, a repeated measures ANOVA of probit-defined boundaries revealed a significant shift in the [b]–[w] boundary in the direction of shorter transition durations for syllables preceded by fast sequences (more [wa] responses) than by slow sequences, in both Condition 1 [$F(1,13) = 9.8$, $p = .008$] and Condition 2 [$F(1,14) = 15.5$, $p = .001$]. Taken with the results of Experiment 1, this indicates that the *rate* at which tones occur in a precursor sequence effects the perception of a following consonant, regardless of small differences in RMS amplitude, maximum amplitude, or amplitude envelope across sequence types.

EXPERIMENT 3

A final experiment was designed to determine whether nonspeech precursors affect the entire category structure of a speech sound, as speech precursors do (Miller et al., 1997; Wayland et al., 1994), or whether nonspeech effects are limited to the ambiguous stop–glide boundary region. Specifically, goodness ratings for the category [w] were obtained as formant transitions similar to those de-

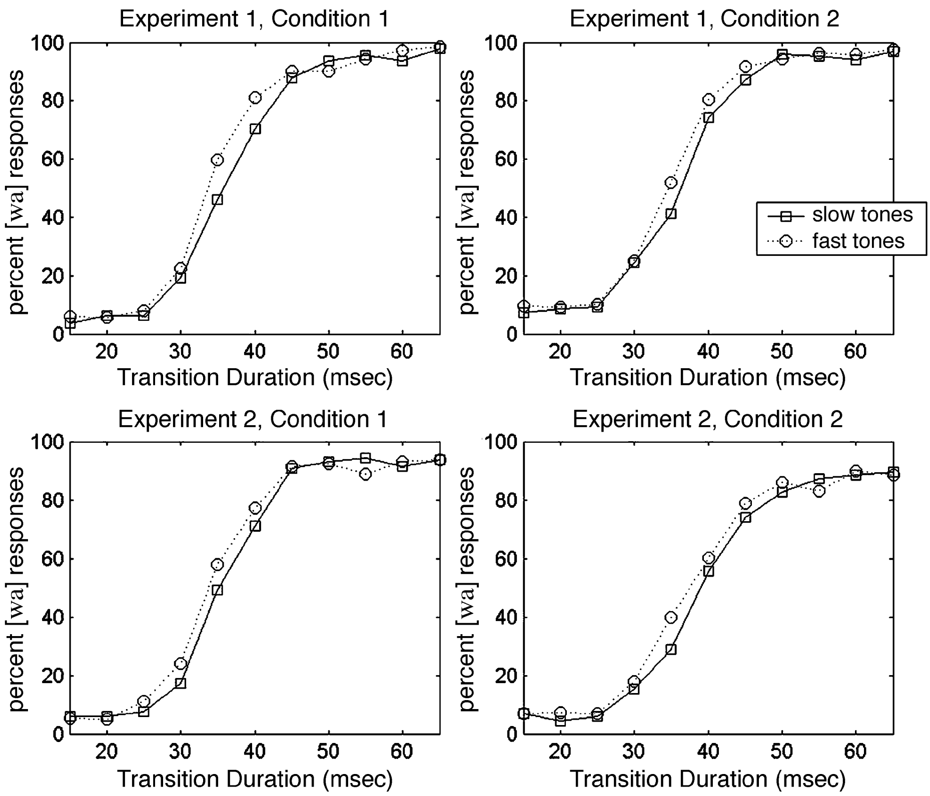

Figure 2. Categorization responses from Experiments 1 and 2.

scribed in Experiments 1 and 2 increased in length from very short ([b]-like) to long ([w]-like) to exaggeratedly long ([*w], by Miller's notation). In designing this study, one concern was the recent suggestion (Utman, 1998) that shifts in category goodness curves as a function of apparent context rate do not properly reflect rate-dependent processing and, instead, are a product of the experimental design used by Miller and colleagues. Utman used natural voiced and voiceless consonants produced at various speaking rates as stimuli to elicit category goodness ratings for voiceless consonants at these different rates. The expected dependence on speaking rate was observed only when participants were provided explicit instructions and prior exposure to tokens labeled *voiced*, *voiceless*, or *exaggerated, breathy versions* (this training provided even more familiarization than did previous experiments by Miller et al., 1997), and not when such training was absent. To address this question, the present experiment was designed to take an intermediate approach to pretest training; participants were given no prior exposure to the [ba]–[wa]–[*wa] continuum members they would encounter during the experiment but were verbally informed of the types of productions they might expect to hear.

Method

Participants. The participants were 30 undergraduate college students. All were native English speakers with no known or obvious speaking or hearing disorders. They received psychology course credit for participation.

Stimuli. The stimuli were synthetic syllables from a [ba]–[wa]–[*wa] (exaggerated [wa]) continuum preceded by tone sequences. The syllables were modeled after those used in Experiments 1 and 2, except that it was necessary to increase syllable length to 200 msec to create transition durations yielding a sufficiently exaggerated [*wa] percept. Fundamental and formant values were identical to those of the previous two experiments; $F1$ and $F2$ transition durations were varied across the continuum from 15 msec to 160 msec in thirty 5-msec steps. Precursor tone sequences were generated in the same manner as those used in Experiment 2, Condition 2. Eight fast and eight slow precursor sequences were created for each continuum member, for a total of 480 different stimuli.

Procedure. The participants heard the stimuli in sound-attenuated booths over headphones at approximately 70 dB SPL. They were instructed to listen to an entire sequence and then to decide how well the syllable resembled the syllable "wa" as it occurs in English words. The participants were not given any explicit training or exposure to the speech stimuli before testing, but they were told that some syllables should sound very much like "wa," whereas others might sound more like "oo-a," "ba," "ma," or some sound that is not used in English at all. For each sound, they were instructed to use a mouse to click somewhere along a sliding scale that appeared on the screen with the labels *not WA*, *ok WA*, and *good WA* positioned to its left, center, and right. The stimuli were presented in random order using ALVIN, a software system recently developed by Hillenbrand and Gayvert (2005). The response scale was quantized to 1,000 equally spaced points; the participants were instructed to make full use of the scale and could move a sliding cursor as desired before clicking a button labeled "OK" to register the selected rating and proceed to the next stimulus. Response patterns suggested that this two-step task resulted in at least two types of questionable responses. The participants would occasionally either (1) accidentally click the "OK" button before adjusting the *WA* rating scale, registering the same rating to two consecutive stimuli, or (2) take too long in adjusting the scale, disrupting the pace of stimulus presentation. Although these problems did not affect results qualitatively, it was determined that excluding those responses corresponding to each participant's shortest and longest 5% (24 trials) of reaction times from consideration provided a more accurate representation of participant ratings.

Results

Overall response patterns are shown in the left panel of Figure 3. Following Wayland et al. (1994), the stimulus to which a participant assigned the highest mean *WA* rating for a given condition was taken as his or her best [w] exemplar for the condition. These peak locations are shown as the isolated data points labeled "Fast peak" and "Slow peak" in Figure 3. A repeated measures ANOVA revealed that the formant transition corresponding to the best [w] exemplar depended on precursor rate [$F(1,29) = 8.01, p = .008$], in the expected direction. The best stimulus corresponding to the best [w] exemplar was 13.2 msec longer, on average, with slow precursors than with fast precursors. This strongly indicates that perception of the transition durations across the continuum, and not just near the perceptually ambiguous boundary, was influenced by the precursor tones such that they were perceived as being longer following faster precursors. Thus, nonspeech precursors affected the internal perceptual structure of the [w] category in the same way context speaking rate does, causing not only category boundaries but also the category best-exemplar to shift in a contrastive direction. This effect does not appear to be due merely to specific task characteristics, because (following the caution of Utman, 1998) explicit training on typical [b], [w], and [*w] tokens was not provided. Rather, it suggests that differences between Utman's results and previous findings may be due to other factors, including the naturalness of tokens (cf. Shinn et al., 1985) and the distribution of VOT.

Lack of explicit training in the present study seems to have resulted in considerable cross-participant variability in responses to the extended transition-length continuum. As is shown in Figure 3, participants were generally reluctant to rate exaggerated [*wa] tokens as low as those they did [ba]-like tokens. This trend varied substantially from listener to listener, resulting in the large standard errors at longer transitions seen in Figure 3. Some participants demonstrated a precipitous decline in *WA* rating at very long durations, whereas others reached a plateau at some intermediate value or increased across the entire continuum. As a result, the variability in Figure 3 prohibits clear examination of patterns in category structure other than peak locations.

Therefore, to facilitate comparison of category structures across precursor types, an additional analysis following the methods of Wayland et al. (1994) considered the responses of only those participants who demonstrated the expected decline in *WA* rating for the exaggeratedly long transitions. The criterion for demonstrating this decline was whether a participant assigned an average rating for the two longest transition lengths that was

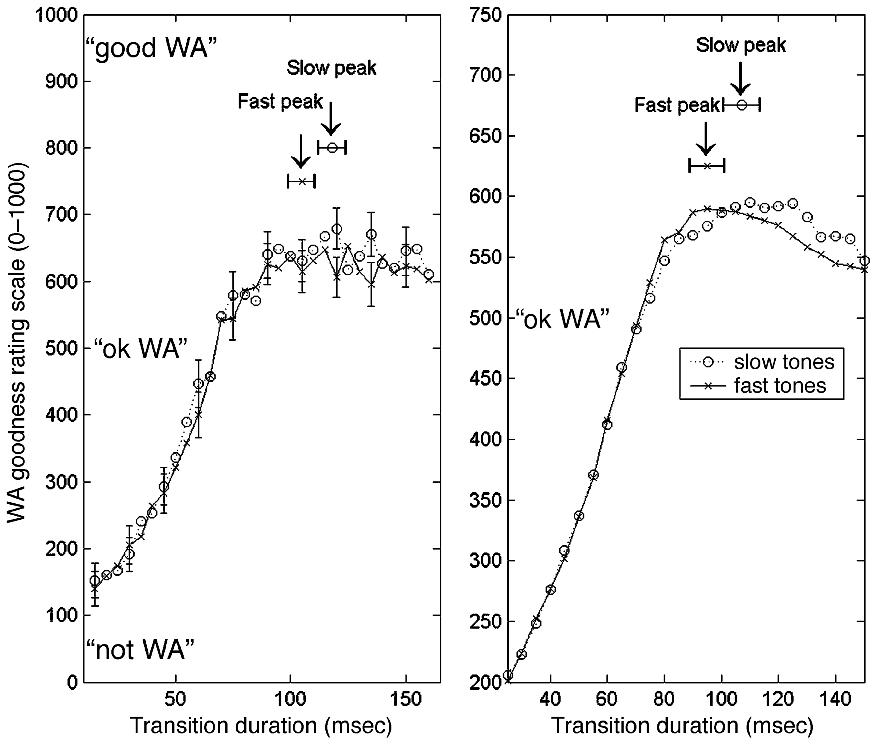

Figure 3. On the left, average [w] goodness ratings with standard error of the mean are plotted as a function of syllable transition duration. Listeners' average peak [w] ratings with fast versus slow nonspeech precursors are shown as the isolated symbols. For these symbols, the *y*-axis is arbitrary, but placement along the *x*-axis reveals a difference in the best-rated [w] exemplar along the transition duration dimension as a function of nonspeech precursor rate. The right panel presents smoothed ratings and peak locations for 20 participants whose ratings decreased to 90% of peak level at the longest transition duration values.

less than 90% of the highest rating they assigned to any continuum member (i.e., the peak *WA* rating) in the relevant precursor condition. Twenty of the 30 participants met this criterion. The right panel of Figure 3 shows the response curves for these participants, with absolute rating values smoothed over 5 consecutive duration points to mitigate additional variability due to the thousand-point scale and the relatively small number of responses to each stimulus. Here, the effects of nonspeech precursors can be seen more clearly to resemble those of the surrounding precursor rate, with a systematic rise and fall in category goodness occurring at roughly 10–15 msec longer transition durations for slow precursors than fast precursors. It can also be seen that participants' average absolute peak *WA* locations across conditions correspond well to the peaks in the overall rating functions, verifying that this is indeed an accurate measure for quantifying differences in category structure as a function of precursor rate.

In summary, nonspeech precursors affected the perceptual [w] category in the same way that context speaking rate does (Miller et al., 1997), in that category best-exemplar locations shifted in a contrastive direction on the basis of precursor rate. Another measure often used by Miller and colleagues to represent category structure has been the size and orientation of a best-exemplar range surrounding this peak location, typically the estimated range (Miller & Volaitis, 1989a) of stimuli on either side of the absolute peak for which listeners assign an average rating at least 90% of that assigned to the peak stimulus. Although variability due to the task (described above) precluded estimation of similar ranges for participants in the present study, the smoothed data in Figure 3 (right panel) suggest that this entire range of ratings was in fact shifted on the basis of nonspeech precursor rate.

GENERAL DISCUSSION

The results reported here demonstrate that simple sequences of pure tones, selected to provide continuity with formant transitions in following speech, are sufficient to bring about rate effects on speech categorization that are remarkably similar to previously observed effects of speaking rate context. For a speech contrast differentiated entirely by the length of initial formant transitions, the perceived category boundary between the stop [b] and the continuant [w] shifted reliably, depending on the rate of tones occurring in preceding sequences, in the same direction as it has been shown to shift for speaking rate (Miller & Liberman, 1979). The boundary consistently shifted in the direction of shorter transition durations (more [wa] responses) for syllables preceded by fast sequences than by slow sequences, whether tones occurred in the entire $F1$–$F2$ range or only the $F2$ frequency range (Experiment 1), and regardless of their overall amplitude or amplitude envelope (Experiment 2). Perhaps most convincingly, these effects penetrated the [w] category to affect its entire graded structure (Experiment 3). The best [w] exemplar, and probably the best-exemplar range along the transition-duration dimension, was shifted in a manner analogous to that previously observed for rate-varying speech context (Miller et al., 1997; Wayland et al., 1994). Taken together, these findings suggest that the apparent dependence on speaking rate in human perception of temporal speech cues need not rely entirely on a mechanism that is in any sense speech specific. Since it is unlikely that perception of pure tone sequences is mediated by knowledge of speech categories or their articulations or their direct recovery, a more general mechanism such as durational contrast must have been responsible for the results observed here.

Although it is not feasible to predict *optimal* rate-dependent [b]–[w] perceptual category boundaries on the basis of nonspeech precursor rate like those derived from measurements of speech production (Miller & Baer, 1983), the difference between the size of the category boundary shift observed in the present experiments (1–2 msec for Experiments 1 and 2 and 10–15 msec for Experiment 3) and the actual acoustic difference in tone lengths across conditions (80 msec between tones of fast and slow precursors) is striking. Figure 4 compares perceptual shifts observed in the present experiments with those of previous studies using speech precursor sequences (Summerfield, 1981; Wayland et al., 1994).

No reliable differences in effect sizes were observed between Experiments 1 and 2; the effects were, however, all generally smaller than those observed previously for speech precursors. As is shown in the left panel of Figure 4, Summerfield (1981) observed a much larger shift in the [b]–[p] boundary when the rate of speech immediately preceding a target segment was manipulated. However, Summerfield did observe smaller effects when the rate-providing speech was displaced in time from the target. When the duration of the word "Why" instead of the word "you" in the precursor phrase "Why are you…" was manipulated, the observed rate effects on the target were more similar to those observed for nonspeech precursors in the present experiments. Thus, one reason for the small category-boundary shifts in the present study may have been the long temporal window across which nonspeech rate information was presented.

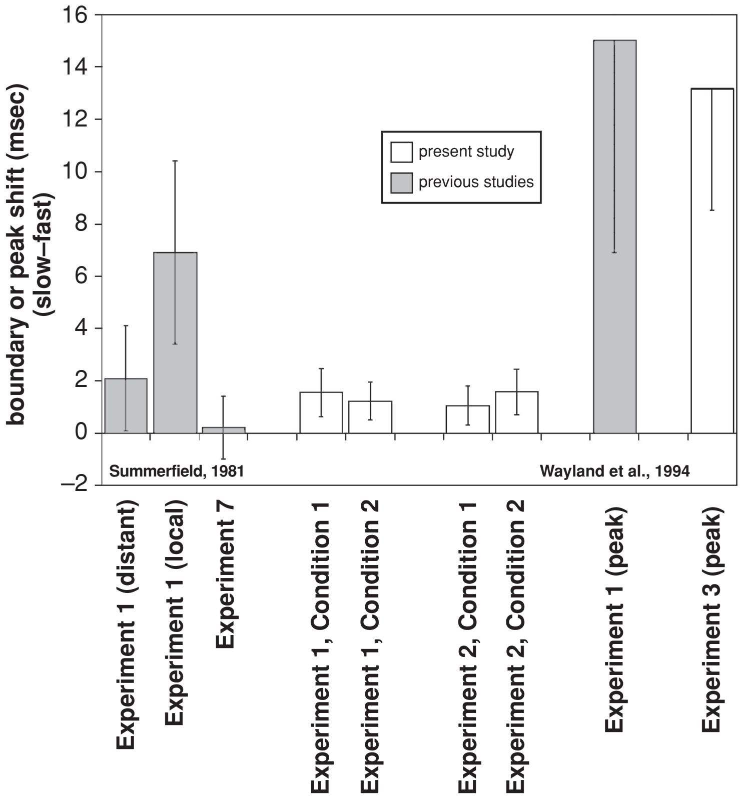

Figure 4. Comparison of the effect sizes (in milliseconds of boundary shift across conditions) produced by nonspeech precursor rate manipulations of the present study and speech precursor rate of previous studies (Summerfield, 1981; Wayland, Miller, & Volaitis, 1994). Error bars for Summerfield are estimates from visual data.

Another cause for the small effect sizes may have been insufficient acoustic continuity between nonspeech precursors and syllables, despite control of the rate and frequency characteristics of the precursors. It is likely that periodic, spectral, and spatial continuity are all important in achieving integration (and, presumably, providing for contrast) across sequential acoustic events (e.g., Bregman, 1990). The present pure tone precursor stimuli offered no periodic and only partial (single frequency band) spectral continuity with following syllables. In addition, the effectiveness of pure tones as precursors in general might well be questioned from a neuroethological standpoint, on the grounds that they insufficiently resemble the complex harmonic structures the auditory system has been developed to perceive (for a review, see Eggermont, 2001; Lewicki, 2002). The present stimuli were chosen solely for their unambiguously nonspeech nature; it will be important to determine whether manipulations to the continuity of precursors or their naturalness change the extent of their perceptual influence on speech categorization.

Task differences may also be important in assessing the effects of the rate of context stimuli, whether speech or nonspeech. The effect size observed in Experiment 3, for which category goodness ratings were measured, was considerably greater than that observed in Experiments 1 and 2, for which categorization boundaries were estimated. This difference is akin to that observed for speech categories. In a similar category-goodness task, Wayland et al. (1994) observed effects of sentence rate on the best [p] exemplar ratings larger than the effect sizes observed by Summerfield (1981) for [b]–[p] category boundary shifts. The target stimuli used may also have influenced the size and nature of the observed effects. Acoustically simple synthetic syllables differing only in transition duration were used to encourage [b]–[w] identification along a purely temporal dimension. Had additional cues to the two segments been present, either in more complex synthetic syllables or in natural speech, it is possible that nonspeech rate effects would have diminished as speech effects seem to do (Shinn & Blumstein, 1984; Shinn et al., 1985). It is even possible that the unnatural quality of the stimuli made them more continuous acoustically with the simple nonspeech precursors, further facilitating the observed effects. However, since the experiments demonstrate that nonspeech rate is taken into account under these simplistic conditions, and since speaking rate effects emerge in noisier listening conditions even for very realistic stimuli (Miller & Wayland, 1993), it seems that effects of the type observed could play a role in natural speaking situations. Effects of nonspeech precursors on the perception of stimuli derived from natural speech have been demonstrated previously (e.g., Holt, 2005). Further experimentation will be needed to explore these possibilities.

Finally, it remains possible that the general contrastive effect demonstrated in the present study represents just one of several factors contributing to the rate-dependent nature of speech perception. Another factor may well be speech- or language-specific knowledge. For individual speakers, dialects, languages and for human speech sounds in general, there exists a multitude of highly predictable, and therefore presumably learnable, patterns (e.g., Miller, 1981) whereby the temporal properties of some speech segments covary with those of nearby segments. Contrastive effects generally work in the direction of perceptually mitigating rate-dependent variability. They exaggerate the differences between a given sound and its acoustic context, making it perceptually more distinct from those sounds with which it is coarticulated and therefore more like a prototypical example of a given category. A given formant transition, for example, will sound shorter and therefore more [b]-like, the longer the surrounding segments. However, since this effect bears no functional relation to context dependence in production, it is unlikely to result in perfect perceptual compensation for every contrast in every language. It seems likely, then, that speakers compensate for any remaining variability by learning the rate-dependent covariance patterns specific to various contrasts. Additional research, perhaps involving cross-linguistic comparison of nonspeech precursor effects, will be needed to determine how linguistic knowledge and contrast effects work together to accommodate perception. Beddor, Harnsberger, and Lindemann (2002), for example, observe language-specific patterns in apparent perceptual compensation for vowel coarticulation, citing differences between English and Shona speakers' category boundary shifts depending on the position of context-providing vowels. Comparison of such differences with analogous cross-linguistic observation of perceptual dependence on nonspeech contexts will be helpful in identifying the mechanisms responsible for context-dependent speech perception.

A final issue is the distinction between rate and duration as the concepts are presented in this and similar studies. It appears that transition duration, and not rate, provides the primary temporal cue to the stop–continuant contrast (Schwab, Sawusch, & Nusbaum, 1981), so it seems reasonable that preceding temporal information would affect the distinction by means of durational contrast (e.g., Goldstone et al., 1959; Goldstone et al., 1957; Walker & Irion, 1979; Walker et al., 1981). However, the precursor sequences in the experiments described here varied in both the duration and presentation rate of tones, creating a possible confound in this respect. The present study, therefore, does not rule out the possibility that a general compensation for precursor rate (such as a nonspeech-specific extrinsic timing mechanism), and not durational contrast, was responsible for the effects observed.

In conclusion, the present results demonstrate that the temporal properties of nonspeech precursor sequences influence the perception of temporal speech contrasts in a manner parallel to previously observed effects of speaking rate. Across experimental manipulations, the durations of pure tones were found to affect perception of formant transition durations in following syllables, as evidenced by shifts in both the stop–continuant, [b]–[w], category boundary and the overall graded structure of the [w] category. This is taken to suggest that a general auditory effect, such as durational contrast, may play a role in the apparent rate-dependent nature of speech perception.

REFERENCES

Ainsworth, W. A. (1974). The influence of precursive sequences on the perception of synthesized vowels. *Language & Speech*, **17**, 103-109.

Allen, J. S., & Miller, J. L. (2001). Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception & Psychophysics*, **63**, 798-810.

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, **30**, 591-627.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., & Pierce, A. (1997). Perception of synthetic /ba/–/wa/ speech continuum by budgerigars (*Melopsittacus undulatus*). *Journal of the Acoustical Society of America*, **102**, 1891-1897.

Diehl, R. L., & Walsh, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America*, **85**, 2154-2164.

Eggermont, J. J. (2001). Between sound and perception: Reviewing the search for a neural code. *Hearing Research*, **157**, 1-42.

Eimas, P. D., & Miller, J. L. (1980). Discrimination of information for manner of articulation by young infants. *Infant Behavior & Development*, **3**, 367-375.

Finney, D. J. (1971). *Probit analysis*. Cambridge: Cambridge University Press.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, **8**, 113-133.

Fowler, C. A. (1990). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, **88**, 1236-1249.

Fowler, C. A. (1992). Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *Journal of Phonetics*, **20**, 143-165.

Fujisaki, H., Nakamura, K., & Imoto, T. (1975). Auditory perception of duration of speech and non-speech stimuli. In G. Fant & M. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 197-219). London: Academic Press.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, **63**, 223-230.

Goldstone, S., Boardman, W. K., & Lhamon, W. T. (1959). Intersensory comparisons of temporal judgments. *Journal of Experimental Psychology*, **57**, 243-248.

Goldstone, S., Lhamon, W. T., & Boardman, W. K. (1957). The time sense: Anchor effects and apparent duration. *Journal of Psychology*, **44**, 145-153.

Gordon, P. C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception & Psychophysics*, **43**, 137-146.

Hillenbrand, J. M., & Gayvert, R. T. (2005). Open source software for experiment design and control. *Journal of Speech, Language, & Hearing Research*, **48**, 45-60.

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, **16**, 305-312.

Holt, L. L., & Kluender, K. R. (2000). General auditory processes contribute to perceptual accommodation of coarticulation. *Phonetica*, **57**, 170-180.

Kidd, G. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 736-748.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, **67**, 971-995.

Kluender, K. R., Coady, J. A., & Kiefte, M. (2003). Sensitivity to change in perception of speech. *Speech Communication*, **41**, 59-69.

Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature Neuroscience*, **5**, 356-363.

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, **60**, 602-619.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, **102**, 1134-1140.

Mack, M., & Blumstein, S. (1983). Further evidence of acoustic invariance in speech production: The stop–glide contrast. *Journal of the Acoustical Society of America*, **73**, 1739-1750.

Miller, J. L. (1981). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 39-74). Hillsdale, NJ: Erlbaum.

Miller, J. L. (1987). Rate-dependent processing in speech perception. In A. Ellis (Ed.), *Progress in the psychology of language* (pp. 119-157). Hillsdale, NJ: Erlbaum.

Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of [b] and [w]. *Journal of the Acoustical Society of America*, **73**, 1751-1755.

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, **43**, 106-115.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, **25**, 457-465.

Miller, J. L., O'Rourke, T. B., & Volaitis, L. E. (1997). Internal structure of phonetic categories: Effects of speaking rate. *Phonetica*, **54**, 121-137.

Miller, J. L., & Volaitis, L. E. (1989a). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, **46**, 505-512.

Miller, J. L., & Volaitis, L. E. (1989b). Internal structure of phonetic categories: Effects of speaking rate. *Phonetica*, **54**, 121-137.

Miller, J. L., & Wayland, S. C. (1993). Limits on the limitations of context-conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics*, **54**, 205-210.

Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, **74**, 750-753.

Newman, R. S., & Sawusch, J. R. (1996). Perceptual normalization for speaking rate: Effects of temporal distance. *Perception & Psychophysics*, **58**, 540-560.

Oller, D. K., Eilers, R. E., Miskiel, E., Burns, R., & Urbano, R. (1991). The stop/glide boundary shift: Modelling perceptual data. *Phonetica*, **48**, 32-56.

Pickett, J., & Decker, L. (1960). Time factors in perception of a double consonant. *Language & Speech*, **32**, 693-703.

Pind, J. (1995). Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues. *Perception & Psychophysics*, **57**, 291-304.

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.

Port, R. (1981). Linguistic timing factors in combination. *Journal of the Acoustical Society of America*, **69**, 262-274.

Purves, D., & Lotto, R. B. (2002). *Why we see what we do: An empirical theory of vision*. Sunderland, MA: Sinauer.

Schwab, E. C., Sawusch, J. R., & Nusbaum, H. C. (1981). The role of second formant transitions in the stop-semivowel distinction. *Perception & 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.

Shinn, P. C., Blumstein, S. E., & Jongman, A. (1985). Limitations of context conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics*, **38**, 397-407.

Sinnott, J. M., Brown, C. H., & Borneman, M. A. (1998). Effects of syllable duration on stop–glide identification in syllable-initial and syllable-final position by humans and monkeys. *Perception & Psychophysics*, **60**, 1032-1043.

Summerfield, Q. (1975). Aerodynamics versus mechanics in the control of voicing onset in consonant–vowel syllables. In *Speech perception* (No. 4). Belfast: Queen's University, Department of Psychology.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 1074-1095.

Utman, J. (1998). Effects of local speaking rate context on the perception of voice-onset time in initial stop consonants. *Journal of the Acoustical Society of America*, **103**, 1640-1653.

Walker, J. T., & Irion, A. L. (1979). Two new contingent aftereffects: Perceived auditory duration contingent on pitch and on temporal order. *Perception & Psychophysics*, **26**, 241-244.

Walker, J. T., Irion, A. L., & Gordon, D. G. (1981). Simple and contingent aftereffects of perceived duration in vision and audition. *Perception & Psychophysics*, **29**, 475-486.

Wayland, S. C., Miller, J. L., & Volaitis, L. E. (1994). The influence of sentential speaking rate and the internal structure of phonetic categories. *Journal of the Acoustical Society of America*, **95**, 2694-2701.

(Manuscript received March 19, 2004;
revision accepted for publication October 13, 2004.)