

# Effects of contrast between onsets of speech and other complex spectra

Jeffrey A. Coady<sup>a)</sup>

*Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706-1611*

Keith R. Kluender<sup>b)</sup>

*Department of Psychology and Department of Physiology, University of Wisconsin, Madison, Wisconsin 53706-1611*

William S. Rhode<sup>c)</sup>

*Department of Physiology, University of Wisconsin, Madison, Wisconsin 53706*

(Received 25 June 2003; accepted for publication 10 July 2003)

Previous studies using speech and nonspeech analogs have shown that auditory mechanisms which serve to enhance spectral contrast contribute to perception of coarticulated speech for which spectral properties assimilate over time. In order to better understand the nature of contrastive auditory processes, a series of CV syllables varying acoustically in  $F_2$ -onset frequency and perceptually from /ba/ to /da/ was identified following a variety of spectra including three-peak renditions of [e] and [o], one-peak simulations of only  $F_2$ , and spectral complements of these spectra for which peaks are replaced with troughs. Results for three-versus one-peak (or trough) precursor spectra were practically indistinguishable, suggesting that effects were spectrally local and not dependent upon perception of precursors as speech. Effects of complementary (trough) spectra had complementary effects on perception of following stops; however, effects for spectral complements were particularly dependent upon the interval between precursor and CV onsets. Results from these studies cannot be explained by simple masking or adaptation of suppression. Instead, they provide evidence for the existence of processes that selectively enhance contrast between onset spectra of neighboring sounds, and these processes are relevant for perception of connected speech. © 2003 Acoustical Society of America. [DOI: 10.1121/1.1608955]

PACS numbers: 43.71.Es, 43.71.Pc, 43.66.Dc [CWT]

## I. INTRODUCTION

All sensorineural systems are particularly responsive to change; perception of stimulus energy is always relative to properties of temporally or spatially proximal energy. The broad descriptor for many instances of responsivity to change is "contrast." Demonstrations of perceptual contrast are ubiquitous and have a long history of being recognized as central to understanding perception (e.g., Locke, 1706/1974). For vision, contrast has been demonstrated for perception of lightness (Koffka, 1935; Wallach, 1948), line orientation (Gibson, 1933, 1937; Gibson and Radner, 1937), size, position, and curvature (Kohler and Wallach, 1944), spatial frequency (Blakemore and Sutton, 1969), depth (Ames, 1935; Bergman and Gibson, 1959; Kohler and Emery, 1947), and color (Cathcart and Dawson, 1928–1929). Contrast has been found for tempo of behavior (Cathcart and Dawson, 1927–1928) and lifting of weights (Guilford and Park, 1931; Sherif, Taub, and Hovland, 1958). For hearing, contrast effects for pitch (Cathcart and Dawson, 1928–1929; Christman, 1954) and spatial location of sounds (Flügel, 1920–1921) have been demonstrated.

Not surprisingly, contrastive processes have been introduced as putative explanations for a variety of perceptual

effects found with speech sounds. For example, Diehl and colleagues (Diehl, Kluender, and Walsh, 1990) have posited durational contrast as an explanation for effects of vowel length preceding medial stops and following initial stops and glides. For the former case, listeners are more likely to report hearing short-closure [+voice] stops [b] following longer vowels, and to report long-closure [–voice] stops [p] following shorter vowels (Denes, 1955; Port and Dalby, 1982; Raphael, 1972). For stops with short formant transitions such as [b] versus longer-transition glides such as [w], longer following vowels encourage perception of the shorter [b], while [w] is more likely to be perceived preceding shorter vowels (Miller and Liberman, 1979). In each of these cases, analogous patterns of results were found for nonspeech stimuli that maintained durational characteristics equivalent to those for speech syllables tested (Diehl and Walsh, 1989; Kluender and Diehl, 1988; Pisoni, Carrell, and Gans, 1983).

Perceptual contrast between adjacent spectral composition has been suggested to explain some aspects of perception of coarticulated speech (e.g., Holt and Kluender, 2000). Coarticulation, the spatial and temporal overlap of adjacent articulatory activities, is reflected in the acoustic signal by severe context dependence. Acoustic information specifying one speech sound varies substantially depending on surrounding sounds. This context always follows the same pattern; adjacent sounds assimilate toward spectral characteristics of one another. Owing to mass and inertia of articulators (as well as planning), articulatory movements are compro-

<sup>a)</sup>Electronic mail: jacoady@facstaff.wisc.edu

<sup>b)</sup>Electronic mail: krkluend@facstaff.wisc.edu

<sup>c)</sup>Electronic mail: rhode@physiology.wisc.edu

mises between where articulators have been and where they are going. Because the acoustic signal directly reflects these articulatory facts, the spectrum assimilates in the same fashion that speech articulation assimilates in coarticulation.

Lindblom (1963) and Öhman (1965) provided some of the best early descriptions of how context systematically influences speech production. Lindblom reported that the frequency of the second formant ( $F_2$ ) was higher in the productions of [did] and [dud] than for the vowels [i] and [u] in isolation, and that  $F_2$  for each vowel was lower in the productions of [bib] and [bub]. In both contexts,  $F_2$ s approached the  $F_2$  of flanking consonants, which are higher for [d] than for [b]. In a subsequent study, Lindblom and Studdert-Kennedy (1967) demonstrated how perception of coarticulated vowels is complementary to these facts of articulation. For CVC syllables with  $F_2$  of the vowel varying acoustically from higher to lower and perceptually from [i] to [u], listeners reported hearing [i] (higher  $F_2$ ) more often in the [wVw] (low  $F_2$ ) context, and [u] (lower  $F_2$ ) more often in the [yVy] (high  $F_2$ ) context. Lindblom and Studdert-Kennedy (1967) wrote: "...mechanisms of perceptual analysis whose operations contribute to enhancing contrast in the above-mentioned sense are precisely the type of mechanisms that seem well suited to their purpose given the fact that the slurred and sluggish manner in which human speech sound stimuli are often generated tends to reduce rather than sharpen contrast." (p. 842).

Findings from more recent studies more clearly demonstrate the role of general auditory processes for providing this contrast. Using CVCs and hybrid speech–non-speech analogs, Holt, Lotto, and Kluender (2000) demonstrated that replacing syllable-initial and syllable-final consonantal information with single frequency-modulated sine waves (following  $F_2$  trajectories) or with constant-frequency sine waves (set at onset and offset frequencies of  $F_2$ ) resulted in the same pattern of perceived vowel quality as was found for full-spectrum CVCs.

For VCCVs, Mann and Repp (1981) entertained the hypothesis that contrast might account in part for their finding that listeners were more likely to report hearing higher-frequency [s] following [r] (low  $F_3$ ) and lower-frequency [ʃ] following [l] (high  $F_3$ ) in syllables varying from [arʃa] to [arsa] and from [alʃa] to [alsa]. Mann (1980) had earlier shown that preceding [r] and [l] had differential effects of perception of following stops [d] and [g], such that listeners are more likely to report hearing higher- $F_3$  [d] following low- $F_3$  [r] and hearing lower- $F_3$  [g] following high- $F_3$  [l]. Lotto and Kluender (1998) later demonstrated that the same pattern of identification for a series of stops varying from [d] to [g] was obtained when preceding [ar] and [al] were replaced either by single frequency-modulated sine waves following the trajectory of  $F_3$  for [r] and [l], or by single constant-frequency sine waves set to the offset frequency of  $F_3$  for [r] and [l].

Data from the foregoing studies are consistent with the universal principle that sensorineural systems respond most vigorously to stimulus change. Results are consistent with an account by which spectral contrast between adjacent speech or nonspeech spectra predicts changes in perception as a

function of neighboring spectral composition. Alternatively, based upon results from a McGurk-type experiment investigating bimodal perception of disyllables, Fowler, Brown, and Mann (2000) argued that their audio-visual demonstration of contrast effects rendered the auditory spectral contrast account unviable, or at least incomplete. However, the auditory process of spectral contrast is not purported to be an exclusive explanation of speech perception, auditory or audio-visual, and other processes are expected to be involved. More recent evidence (Holt, Stephens, and Lotto, unpublished; Stephens and Holt, 2002) serves to question the original Fowler *et al.* (2000) results and conclusions. Holt and colleagues could only replicate Fowler *et al.*'s original findings with the original stimulus materials, but not with similar materials. Upon examining the original video materials, Holt and colleagues found a critical confound in the video coinciding with test syllables. Different visual properties of the test CVs alone were sufficient to produce Fowler *et al.*'s original results, even without contextual precursors.

There are a large number of experimental precedents for spectral contrast. Summerfield *et al.* (1984) established the existence of an "after effect" in vowel perception. When a uniform harmonic spectrum was preceded by a spectrum that was complementary to a particular vowel with troughs replacing peaks and vice versa, listeners reported hearing a vowel during presentation of the uniform spectrum. The vowel percept (for the uniform spectrum) was appropriate for a spectrum with peaks at frequencies where there were troughs in the preceding spectrum. Summerfield *et al.* (1984) noted that perceiving vowel sounds in uniform spectra (following appropriate complementary spectral patterns) has a well-known precedent in psychoacoustics. This oft-reported finding is that, if just one member of a set of harmonics of equal amplitude is omitted from a harmonic series and is reintroduced, then it stands out perceptually against the background of pre-existing harmonics (Cardozo, 1967; Green, McKey, and Licklider, 1959; Houtgast, 1972; Schouten, 1940; Viemeister, 1980).

Summerfield and colleagues (Summerfield *et al.*, 1984; Summerfield, Sidwell, and Nelson, 1987) suggested that either simple adaptation or adaptation of suppression could account for their findings of enhanced perception of energy in spectral regions where previously there had been relatively little energy. By one simple adaptation account, neurons adapt (become less sensitive), and the enhanced prominence of the added harmonic(s) is due to the fact that neurons tuned to its frequency were not adapted prior to its onset. In contrast, some researchers (e.g., Houtgast, 1974; Moore and Glasberg, 1983) have suggested that adaptation serves mostly to enhance onsets selectively. Suppression, then, is hypothesized to be a process through which differences in level of adjacent spectral regions in complex spectra (e.g., formants in speech signals) are preserved and/or enhanced. Viemeister and Bacon (1982) showed that not only was an "enhanced" target tone more detectable; the tone also served as a more effective masker of a following tone. They suggested that suppression must be included in an adaptation scenario to place it in closer accord to this finding.

Different frequency components of a signal serve to sup-

press one another, and Viemeister and Bacon (1982) suggest that what is relevant is not so much adaptation to absolute amplitude of specific frequency components, but rather individual spectral channels are less able to suppress neighboring channels. This explanation is consistent with studies of two-tone suppression which has been cast as an instance of lateral inhibition in hearing (Houtgast, 1972). Investigators have argued that suppression helps to provide sharp tuning (e.g., Festen and Plomp, 1981; Wightman, McGee, and Kramer, 1977). With respect to speech perception, Houtgast (1974) has argued that this process serves to sharpen the neural projection of a vowel spectrum in a fashion that effectively provides formant extraction. One way to conceptualize these processes is that they serve to provide simultaneous spectral contrast, enhancing prominences versus spectral regions of lesser energy.

There exist several neurophysiological observations that bear upon enhancement effects. In particular, a number of neurophysiological studies of auditory-nerve (AN) recordings (e.g., Smith, 1979; Smith, Brachman, and Frisina, 1985; Smith and Zwislocki, 1971) provide evidence for peripheral adaptation. Delgutte and colleagues (Delgutte, 1980, 1986, 1996; Delgutte *et al.*, 1996; Delgutte and Kiang, 1984) have established the case for a much broader role of peripheral adaptation for perception of speech. They note that peaks in AN discharge rate correspond to spectro-temporal regions that are rich in phonetic information, and that adaptation increases the resolution with which onsets are represented. This role of adaptation for encoding onset information is consistent with earlier observations noted above. Delgutte and colleagues (1996) note neurophysiological evidence that “adaptation enhances spectral contrast between successive speech segments” (p. 3.) This enhancement arises because a fiber adapted by stimulus components close to its CF is relatively less responsive to subsequent energy at that frequency, while stimulus components not present immediately prior are encoded by fibers that are unadapted—essentially the same process offered by psychoacousticians but now grounded in physiology. Delgutte also notes that adaptation takes place on many time scales. In general, adaptation extends over longer intervals with increasing level in the auditory system. Some of the effects described above (particularly those that are temporally extended) may be less likely to have very peripheral origin.

Inspired by Summerfield and his colleagues’ vowel after-effect studies (1984, 1987), the present series of experiments investigates the extent to which the same or similar processes may be responsible for the findings with very simple sine-wave flanking stimuli (e.g., Holt *et al.*, 2000; Lotto and Kluender, 1998) and those earlier studies which used rich spectra that were complementary to those for vowel sounds. Although there is an extended history of using sine waves and FM glides as nonspeech proxies for formants, such sounds have only limited resemblance to speech formants. While it is true that spectrograms illustrate formants as bands of energy and formant transitions as bands of energy traversing frequency, such pictorial descriptions can be misleading. For example, if fundamental frequency ( $f_0$ ) is constant, individual harmonics of the fundamental do not

change frequency; only relative amplitudes of harmonics change. Individual frequency components of the speech spectrum change frequency no more than  $f_0$  changes. In an early report concerning effects of simple nonspeech context (FM and constant-frequency sine waves), Lotto and Kluender (1998) used the term “frequency contrast” to describe their effects. This may be a misnomer because the frequency of spectral components did not change in their speech stimuli. Describing effects as frequency contrast implied effects due to frequency differences rather than as differences in relative amplitude or spectral envelope.

The present experiments bridge this gap between simple sine waves as weak renditions of speech and the empirical precedents of Summerfield and his colleagues by investigating whether “after-image” effects translate into the same sorts of changes in phonetic perception brought about by simpler nonspeech precursors. They examine the perception of VCV sequences for which characteristics of the initial vowel have dramatic effects on the spectral characteristics of the following consonant in production (Öhman, 1965). And, like the findings reported above for perception of vowels in CVCs, perception of spectral information for the consonant in VCVs is complementary to the facts of production (Holt, 1999).

Experiments reported below examine the effects of vowels and spectral complements of vowels on perception of a subsequent consonant–vowel series. Instead of simple sine-wave nonspeech analogs of vowel precursors, spectral complement precursors are complex harmonic spectra that here are hypothesized to affect consonant perception in a manner complementary to vowel precursors. First, for typical speech VCVs, spectral prominences (formants) in vowel precursors should influence perception of following stops by enhancing the perceptual prominence of spectral energy away from prominences (at contrasting frequencies.) For example, greater energy in the higher-frequency region of  $F_2$  in [e] should encourage perception of a lower-frequency  $F_2$  stop [b] following the vowel, and greater energy in the lower-frequency region of  $F_2$  in [o] should encourage perception of the higher-frequency  $F_2$  stop [d]. Following the same assumptions, spectral complements of [e] and [o] should provide complementary patterns of perception. The absence of spectral energy in the spectral complement precursors should enhance perception of those frequencies not represented in the precursors.

## II. EXPERIMENT 1

### A. Method

#### 1. Subjects

Eighteen native-English speaking undergraduates at the University of Wisconsin-Madison participated in return for course credit in Introductory Psychology. All subjects reported normal hearing.

#### 2. Stimuli

Half of the stimuli consisted of VCV disyllables with digitally created renditions of vowels [e] or [o] followed by each member of a six-step series of CV syllables varying

perceptually from [ba] to [da]. The other stimuli consisted of the same six-step CV series, except these were preceded by spectral complements [ $\sim$ e] and [ $\sim$ o]. To create precursors, a 500-ms harmonic spectrum with a fundamental frequency of 120 Hz and approximately 6-dB/octave roll-off was created using a MATLAB<sup>®</sup> implementation of the Klatt (1980) speech synthesizer (Kieffe, Kluender, and Rhode, 2002). The source output was extracted before going through formant filters. Spectra were filtered using FIR filtering in MATLAB<sup>®</sup>. A 2000-order filter with a Blackman window was used to ensure maximal stopband attenuation. To create the vowel-like precursors, the harmonic spectrum was passed through a passband filter, leaving harmonics corresponding to formants. For the [e] precursor,  $F1$  was represented by harmonics at 360 and 480 Hz,  $F2$  by harmonics at 1800, 1920, and 2040 Hz, and  $F3$  by harmonics at 2520 and 2640 Hz. The [o] precursor was the same as [e] except harmonics for  $F2$  were at 720, 840, and 960 Hz. To create the spectral complement precursors [ $\sim$ e] and [ $\sim$ o], the harmonic spectrum was passed through stopband filters at the same values, leaving no energy in regions occupied by harmonics in the vowel-like precursors. Owing to the length of the impulse response for such steep filters, the initial portions of the original 500-ms waveform are not filtered as accurately as the latter portions. Consequently, the final 100 ms of each precursor was excised at zero crossings at the beginning and end of pitch pulses.

Duration of precursors was varied because it may influence the magnitude of effects upon perception of following CV. In particular, effects of some processes of adaptation may be expected to be larger with increasing duration of precursor. There was a second unexpected benefit to using variable-duration precursors. Pilot testing revealed that some listeners may actively attend away from nonspeech precursors to the extent that precursors seem superfluous to their task of identifying the following CV. In some related earlier work (e.g., Holt, 1999; Holt *et al.*, 2000) listeners were asked to explicitly label precursors in some fashion, so listeners did not have the opportunity to neglect precursors. The use of precursors of variable durations made it more difficult for listeners to neglect precursors because the CV to be labeled could occur at different intervals following the onset of the precursor. While study of effects of attention may be of interest as a separate question, in this case, variable precursor durations can simply be taken as more representative of normal continuous processing of connected speech for which information can continually be gleaned from the signal. Precursors were therefore 100-, 200-, and 300 ms in duration. The 200- and 300-ms precursors were created by digitally iterating the 100-ms samples. Each of the 12 precursors ([e], [o], [ $\sim$ e], and [ $\sim$ o] at 3 durations) had 5-ms linear ramps at onset and offset.

A six-step [ba–da] CV series varying only in  $F2$ -onset frequency was synthesized at a 10-kHz sampling rate with 12-bit resolution on the cascade branch of the Klatt speech synthesizer (1980). Syllables were 250 ms in total duration with a constant  $f_0$  of 120 Hz. Nominal  $F2$  onset frequency varied in 120-Hz steps from 1080 to 1680 Hz. Over the first 40 ms,  $F2$  ramped linearly from onset frequency to 1200 Hz,

and then held constant. Frequency of  $F1$  rose linearly from 400 to 700 Hz over 40 ms and was constant for the remainder of the syllables. Frequencies of  $F3$ ,  $F4$ , and  $F5$  were set at constant values of 2580, 3500, and 4500 Hz, respectively. This series was used as the test series in all of the present experiments. Precursors and CV syllables were rms matched in amplitude, and each of the 12 precursors ([e], [o], [ $\sim$ e], and [ $\sim$ o] at 3 durations) was concatenated with each of the six CV syllables separated by 10-ms intervening silence. Figure 1 provides schematic spectrograms for each precursor with the CV series.

### 3. Procedure

Stimulus presentation and response collection were under the control of an 80486-25 microcomputer. Following D/A conversion (Ariel DSP-16), stimuli were low-pass filtered (4.8 kHz cutoff frequency, Frequency Devices, #677), amplified (Stewart HDA4), and presented to subjects via headphones (Beyer DT-100) at a level of 75 dB SPL.

Listeners participated in a 2AFC identification task. One to three subjects were tested concurrently in individual sound-attenuated chambers during a single experimental session. In a completely mixed design ([eCa], [oCa], [ $\sim$ eCa], and [ $\sim$ oCa]), participants identified CVs as [ba] or [da] by pressing either of two buttons on a handheld electronic response box with buttons labeled “BA” and “DA.” Over two blocks of 360 presentations, participants responded to each stimulus ten times. The entire session lasted approximately 40 min. After the session, participants were informally asked to identify the precursor sounds. Vowel precursors were easily identified, while spectral complement precursors were more difficult to label. Those listeners who attempted to identify them reported hearing a buzz.

### B. Results and discussion

To ensure that listeners performed competently, data from listeners who respond with at least 90-percent consistency on unambiguous endpoint stimuli were analyzed. Only two listeners failed to meet this criterion. Figure 2 displays mean consonant identification for the vowel and vowel complement precursors in experiment 1. Data for vowel precursors at each duration were analyzed via paired t-tests.<sup>1</sup> Consistent with previous results (e.g., Holt, 1999), there was a significant effect of vowel context on consonant identification for all precursor durations: 100 ms,  $t(15)=3.48$ ,  $p<0.01$ ; 200 ms,  $t(15)=2.68$ ,  $p<0.05$ ; 300 ms,  $t(15)=4.92$ ,  $p<0.001$ . Results of t-tests performed on probit data revealed the same pattern of results.<sup>2</sup> Listeners were much more likely to report hearing [da] in the [oCa] context than in the [eCa] context.

Data for spectral complement precursors were analyzed in the same fashion. The 100-ms spectral complement precursors [ $\sim$ e] and [ $\sim$ o] had a significant effect on consonant identification,  $t(15)=3.42$ ,  $p<0.001$ . However, 200- and 300-ms spectral complement precursors did not,  $t(15)=-0.07$ ,  $p=0.94$ , and  $t(15)=-0.38$ ,  $p=0.66$ , respectively.

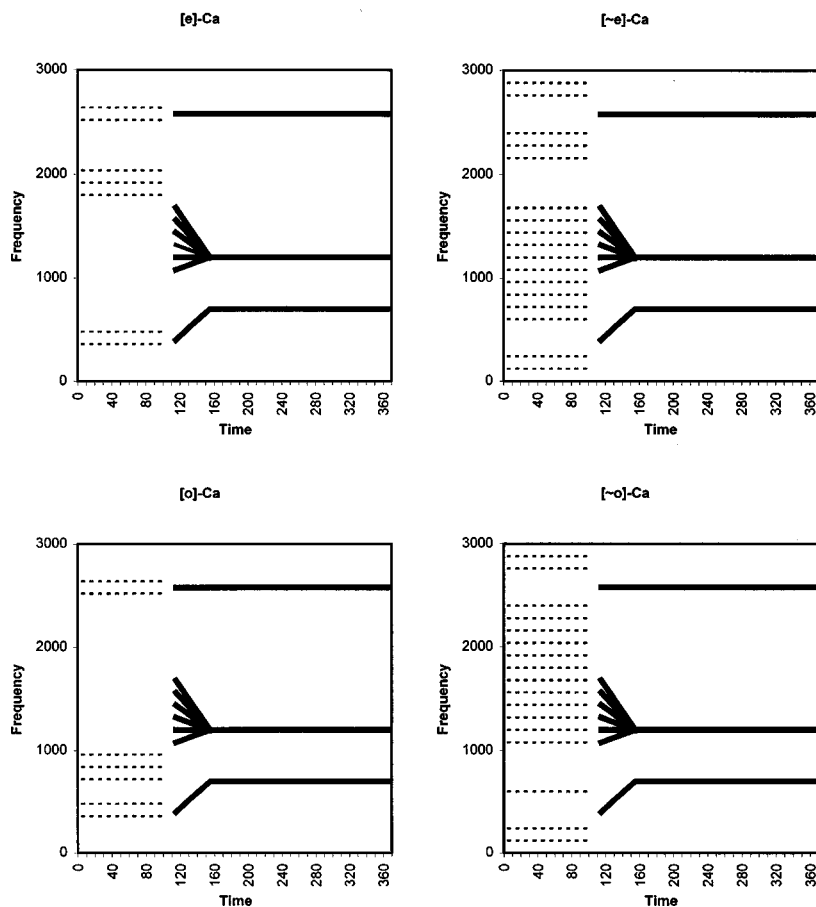


FIG. 1. Schematic renditions of stimuli for experiment 1. All harmonics of precursors are illustrated, and are represented by thin dotted lines. Formant center frequencies of the [ba]–[da] series are represented by thick solid lines (all harmonics present in test stimuli). The left column includes [e] (top) and [o] precursors preceding the [ba–da] series, and the right column includes spectral complement [~e] and [~o] precursors.

Only for the briefest precursors, listeners were more likely to report hearing [da] after spectral complement [~e] than after [~o].

While complementary effects for spectrally complementary precursors were predicted, the presence of this effect for only the briefest precursor durations was not fully expected. The lack of effect for longer precursors is not expected if one assumes either adaptation or adaptation of suppression (e.g., Viemeister and Bacon, 1982) as an explanation for this effect. However, adaptation of suppression was but one of several instances of adaptation in the auditory system. In experiment 2, the importance of precursor duration will be investigated *vis à vis* other instances of adaptation and their

putative effects on encoding of stimulus onsets.

There are other reasons one might expect that nonspeech precursors should provide less effect on perception of following stops. Listeners have a tremendous amount of experience hearing coarticulated VCVs. Repp (1982) argues that “listeners make continuous use of their tacit knowledge of speech patterns” (p. 81) in perceiving coarticulated speech. Coarticulation yields multiple covariances in the signal that are orderly in as much as they reflect dependable regularities in physical constraints upon articulators. Perceptual learning would be expected to seize upon such dependable relationships between stimulus qualities, and one would expect results from experienced listeners to reflect perceptual experi-

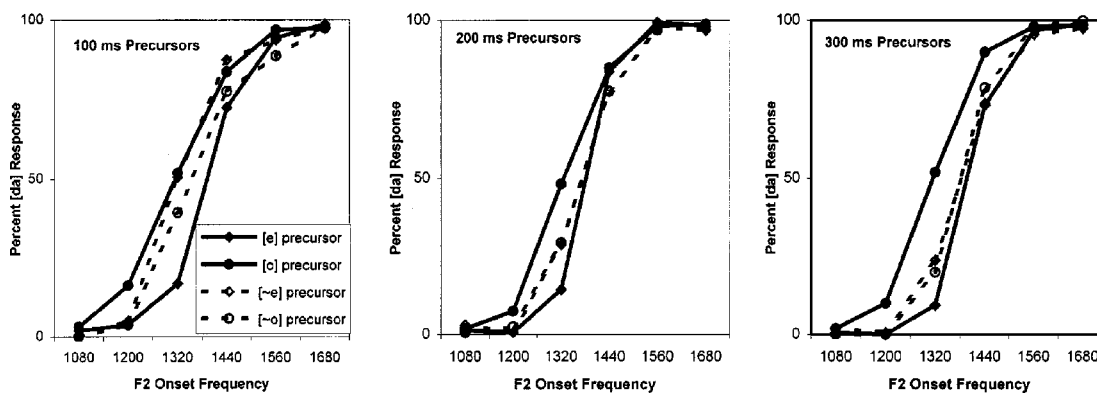


FIG. 2. Data from identification (/b/,/d/) task in experiment 1 for each precursor duration. Percent /da/ responses as a function of second formant onset frequency are plotted by precursor context. Solid lines denote the functions for [e] and [o] precursors. Data for spectral complements [~e] and [~o] are shown with dashed lines.

ence with speech in addition to less domain specific auditory processes (Holt and Kluender, 2000; Lindblom and Studdert-Kennedy, 1967).

### C. Additional conditions

A series of follow-up experiments was conducted to examine whether spectral properties of the precursors might explain why longer spectral complement precursors failed to elicit the predicted effects. For the sake of brevity, and because the results across conditions are very consistent, the follow-up experiments are summarized below.

#### 1. Single peak or trough

First, the spectral contrast effects were replicated with precursors with just a single spectral prominence or trough. A spectral contrast account suggests that effects in the previous experiment result, in whole or in part, from the presence or absence of energy in the frequency range of  $F2$ . While precursors in experiment 1 differ only in the frequency of the second spectral peak or trough, they are nevertheless acoustically rich stimuli. In the case of renditions of [e] and [o], precursors were clearly perceived as /e/ and /o/. And, listeners came to the laboratory with ample experience about acoustic properties of consonants following these two vowel sounds. Single-band (pass and stop) precursors that share much less acoustic resemblance to speech sounds were used in the interest of attenuating effects that are due to experience with coarticulated speech sounds. In these conditions, results matched those for the three-formant vowel and spectral complement precursors. Effects were significant for single-passband precursors at all durations, while effects for single-stopband precursors were only significant at the shortest duration.

#### 2. Equalized cochlear distance

Next, cochlear area was considered as a potentially confounding variable. Because the cochlea is organized in a roughly logarithmic fashion above 1 kHz, a three-harmonic spectral prominence or trough centered at a lower frequency (typical of  $F2$  for [o]) will map onto a greater cochlear area than one centered at a higher frequency (typical of  $F2$  for [e]). Therefore, bandwidths were adjusted using equal rectangular bandwidth (ERB) estimates of auditory filter width (Moore and Glasberg, 1983). The ERB scale also corresponds well to estimates of human cochlear distance (Greenwood, 1990). In this instance, three harmonics at a lower frequency (typical of  $F2$  for [o]) map onto an equivalent cochlear area as five harmonics at higher frequencies (typical of  $F2$  for [e]). Therefore, three-formant vowel and spectral complement precursors, along with single spectral prominence and trough precursors were created with and without ERB-rate normalization. Because of the large number of precursors, 300-ms precursors were omitted from this condition. For the vowel and single-spectral prominence precursors, effects were significant for 100- and 200-ms precursors, replicating experiment 1. For the spectral complement precursors, results generally matched those for experiment 1. When comparing ERB-rate normalized three-formant and single-formant spectral complement precursors, spectral contrast ef-

fects were significant for shorter, 100-ms precursors, but not for longer 200-ms ones. However, when comparing three-formant spectral complement precursors with and without ERB-rate normalization, spectral contrast effects were significant for both 100- and 200-ms precursors. Because the same non-ERB-normalized stimuli were used in experiment 1 and this follow-up, it is not clear what accounts for this difference. It may be due to subject variance or to some unknown effect of removing 300-ms precursors from the task. Data from additional experiments, however, indicate that the likely explanation is that the temporal extent of these effects begins to wane at 200 ms, but is not yet extinguished.

#### 3. Random phase harmonics

Finally, we consider that listeners may somehow “fill in” spectral troughs over the longer time course of 300-ms, and sometimes 200-ms, precursors. For natural harmonic spectra, harmonics are in phase. Duifhuis (1970) and Wightman (1973) observed that, when listeners are presented with harmonic complexes from which specific harmonics have been removed, missing harmonics could potentially be derived from the temporal waveform. Horst, Javel, and Farley (1990) found that, when a center component of a harmonic complex was left out, this missing component was restored in the pattern of discharges in AN at higher stimulus levels. In recent studies of basilar-membrane mechanics, Rhode and Recio (2001) presented equal-interval seven-tone complexes. When the center component was deleted and only six tones were presented, the center component was restored in the basilar-membrane response as a result of distortion-product generation in the nonlinear cochlea.

In all experiments reported here thus far, precursors were filtered versions of a harmonic spectrum created via a MATLAB<sup>®</sup> implementation of the Klatt80 synthesizer (Kiefte *et al.*, 2002; Klatt, 1980), and all harmonics had the same phase. It is possible that listeners in the experiments above perceptually restored missing harmonics for longer precursor stimuli, but not for the shortest precursors. Consequently, spectral troughs may have been only a temporary characteristic of these stimuli as represented in the auditory spectrum. Randomizing phase of harmonics makes it impossible to recover missing harmonics. Spectral complement precursors synthesized by Summerfield and colleagues (1984, 1987) had harmonics in random phase. Their effects maintained even though precursors were 1 s length. To control for this possibility, new precursors were created such that their harmonics were in random phase. Results for vowel precursors match previous results, with significant spectral contrast effects for precursors of all durations. For spectral complement precursors, spectral contrast effects were significant for the shorter 100- and 200-ms precursors, but not for the longer 300-ms precursors.

Thus far, the results for three- (vowel) and one-passband precursors are consistent with a general description of spectral contrast in speech perception. Spectral energy at higher frequencies makes subsequent spectral energy at lower frequencies more effective, and vice versa. This is true at each of the three precursor durations tested. These effects are amenable to explanation via processes of adaptation or adap-

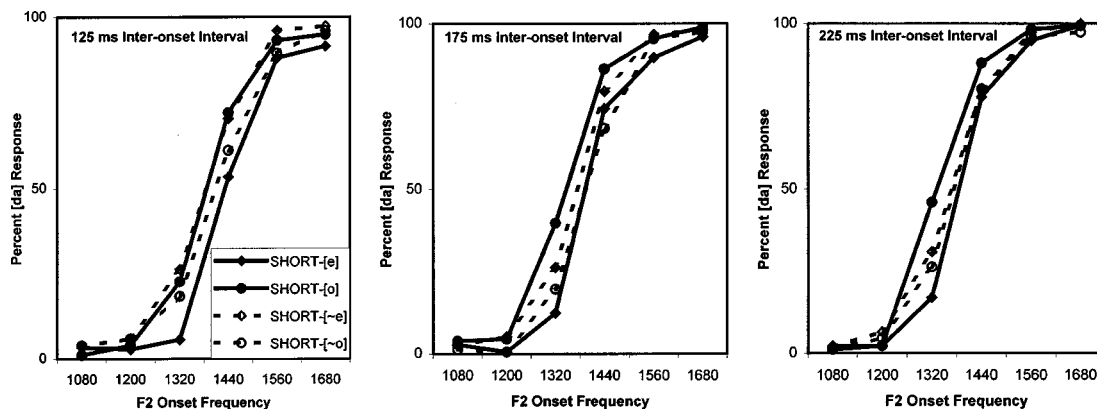


FIG. 3. Data from identification (/b,/d/) task in experiment 2 for very short (25-ms) precursors and three interonset intervals. Percent /da/ responses as a function of second formant onset frequency are plotted by precursor type. Solid lines indicate data for three-peak SHORT-[e] and SHORT-[o] precursors, and data from spectral complement SHORT-[~e] and SHORT-[~o] precursors are shown with dashed lines.

tation of suppression in a relatively straightforward way. Results for three- and single-stopband (complementary) precursors appear more difficult to interpret. Effects of troughs in the spectrum are consistent across conditions when duration is 100 ms. These effects are less reliable for 200-ms precursors, and are functionally absent for 300-ms precursors.

One hypothesis for the effect of spectral troughs at short but not long precursor durations is that effects are due to some change in the processing of onset properties of stimuli. While, up to this point, longer precursors may be thought of as providing greater adaptation, what has been neglected is the fact that longer precursors result in onsets of precursors being displaced further in time from onsets of CVs. It is widely appreciated that onset properties of speech sounds are very important for listeners (e.g., Blumstein and Stevens, 1981). As noted above, Delgutte (1996) argues that adaptation in AN increases the spectral resolution with which onsets are represented. From psychoacoustics, others (e.g., Houtgast, 1974; Moore and Glasberg, 1983) have suggested that adaptation serves mostly to enhance onsets selectively. Experiment 2 is designed to test the importance of contrast between successive (precursor and CV) onsets. To the extent that precursor onset properties give rise to contrast effects with the following CV, precursor duration *per se* does not matter. Experiment 2, therefore, uses only very short (25 ms) precursors presented at variable intervals preceding CVs.

### III. EXPERIMENT 2

#### A. Method

##### 1. Subjects

Two groups of native-English speaking undergraduates at the University of Wisconsin–Madison participated in return for course credit in Introductory Psychology. All reported normal hearing. The first group of 22 listeners heard only the CV series with SHORT-[e] and SHORT-[o] precursors. A second group of 23 listeners heard only the CV series with spectral complement SHORT-[~e] and SHORT-[~o] precursors.

#### 2. Stimuli

Precursors were created from the same three-passband ERB-normalized vowels [e] and [o] and their spectral complements [~e] and [~o] used previously. Based on the 120-Hz fundamental frequency, each pitch pulse is 8.33 ms in duration. Precursors that were 25 ms in duration were created from three pitch pulses. Three interonset intervals of 125, 175, and 225 were created by inserting 100, 150, or 200 ms of silence between precursor offset and CV onset.

#### 3. Procedure

Stimulus presentation and response collection were identical to experiment 1. Each group of listeners heard only one set of stimuli, either with spectral peak or spectral trough precursors. Each stimulus was presented 8 times in random order for a total of 288 presentations (2 precursors×3 intervals×6 CV syllables×4 presentations×2 blocks). Experimental apparatus and procedure were identical to those for experiment 1. The experimental session lasted approximately 20 min.

### B. Results and discussion

Figure 3 displays mean consonant identification curves for experiment 2. Data for spectral peak and spectral trough precursors for each interval were examined in paired t-tests. Results for SHORT-[e] and SHORT-[o] precursors correspond well with those from all experiments reported above with significant effects for all onset–onset durations: 125 ms,  $t(21)=4.03$ ,  $p<0.001$ ; 175 ms,  $t(21)=5.57$ ,  $p<0.0001$ ; and 225 ms,  $t(21)=4.63$ ,  $p<0.0001$ . As in all previous conditions, listeners heard [d] more often in the [o] context than in the [e] context, even with these very short precursors. In spectral complement conditions, four listeners failed to reach 90-percent consistency on unambiguous endpoint stimuli, and results from the remaining 18 listeners were analyzed. Onset properties of these 25-ms SHORT-[~e] and SHORT-[~o] precursors significantly alter perception of subsequent CVs as they did in previous experiments when precursor onsets were separated by 125 ms,  $t(18)=2.80$ ,  $p<0.05$ , and by 175 ms,  $t(18)=3.40$ ,  $p<0.01$ . Consistent with the inconsistent effects for 200-ms precursors in the previous experi-

ments, the difference between precursors was not significant when 225 ms intervened between onsets,  $t(18) = 1.75$ ,  $p = 0.10$ .

Even very short 25-ms precursors affect perception of subsequent CVs in a manner consistent with contrast between onset spectra. Across all conditions, effects of vowel precursors have been constant across variations in precursor duration, bandwidth in the region of  $F2$ , presence or absence of lower and higher spectral energy (in the region of  $F1$  and  $F3$ ), phase of harmonics, and duration of intervening silence. In all cases, no matter the manipulation, listeners are more likely to report hearing higher-frequency [da] following lower-frequency energy corresponding to  $F2$  in [o].

Effects for spectral complement precursors, however, appear more circumscribed. As is the case for one- and three-peak passband stimuli, patterns of results for one- and three-trough stopband precursors are strikingly similar. However, unlike effects for spectral peaks, the strength of the effects of spectral complements appears to depend critically upon the duration of the interval between onset of precursor and onset of CV. Aside from this difference, the patterns of findings correspond well to prediction on the basis of spectral contrast between precursor and CV onsets. Listeners are more likely to hear higher-frequency [da] following [~e] for which there is an absence of energy in the higher-frequency region corresponding to  $F2$  in [e]. This effect of spectral troughs wanes as the time between the onset of the precursor and the onset of the test syllable increases. Interestingly, it does not seem to be influenced by the presence or absence of intervening acoustic energy. Identification of subsequent consonants was influenced by spectral complement precursors when precursors were 100 ms in duration followed by 10 ms of silence, and when they were 25 ms in duration followed by 100-ms silence.

For a number of reasons, the fact that onset properties have such a great effect for both vowel and spectral complement stimuli in experiment 2 and earlier experiments is not entirely surprising. Perhaps most obviously, it is the perception of CV onsets that is being affected by precursors. The [ba–da] series is defined by variation in onset frequency of  $F2$ . Transitions were 40 ms in duration, and all members of the series become increasingly similar over the first 40 ms until all six stimuli share common spectral properties. More generally, it is well known that onsets of sounds are physiologically significant, not only in AN neurons, but also in neurons in cochlear nucleus (e.g., Rhode, 1991; Winter and Palmer, 1995) and at successive stages of the auditory system through to auditory cortex (Heil, 1997a, 1997b; Phillips and Hall, 1990).

The spectral complement precursors used in the current series of experiments provide a strong test of spectral contrast accounts for some phenomena of speech perception. They are acoustically rich signals, created from harmonic spectra modeled after the human voice source. They do not mimic speech, but rather, are complementary to it. However, listeners seldom hear such sounds outside of the speech laboratory. Nevertheless, one may expect that processes underlying effects demonstrated above should be well suited to the perception of running speech. In fluent speech, onsets are

temporally displaced, and acoustic energy typically exists between onsets. In order to test the extent which the present findings generalize to more complex speech signals that are more representative of connected speech, experiment 3 was conducted to test the extent to which the onset spectra of a preceding syllable ([ya] and [wa]) affect perception of the initial stop in following syllables ([ba–da]). Glide-vowel syllables [ya] and [wa] were chosen because onset spectra could be crafted to mimic spectral properties of precursors used above.

## IV. EXPERIMENT 3

### A. Method

#### 1. Subjects

Forty-one native-English speaking undergraduates at the University of Wisconsin-Madison participated in return for course credit in Introductory Psychology. All reported normal hearing.

#### 2. Stimuli

Precursor syllables were synthesized at a 10-kHz sampling rate with 12-bit resolution on the cascade branch of the Klatt speech synthesizer (1980). Syllables were 100 ms in duration with a constant fundamental frequency of 120 Hz. Across 50 ms,  $F1$  increased from 400 to 700 Hz before remaining at 700 Hz until the end of the syllable. Onset of  $F2$  for [ya] was 1920 Hz, and that for [wa] was 840 Hz. These values match those for the center frequencies of the vowels [e] and [o], respectively, used in the previous experiments.  $F2$  changed linearly from onset frequency to 1200 Hz over 50 ms and then held constant. Values for  $F3$ ,  $F4$ , and  $F5$  were held constant at 2580, 3500, and 4500 Hz, respectively, throughout the syllables. Owing to the relatively brief (50-ms) vowel portion of the [ya] and [wa] syllables, these were clearly heard as glides, not stops. Perception of these syllables as glides is consistent with earlier findings by Miller and Liberman (1979) demonstrating perception of stops and semivowels at different durations of transitions and following vowels. Precursor glide-vowel syllables and following stop-glide syllables were rms matched in amplitude before being concatenated with 40-ms intervening silence. Thus, the interval between syllable onsets is 140 ms. Figure 4 provides formant tracks for both precursors with the CV series.

#### 3. Procedure

Stimulus presentation and response collection were identical to experiment 1. Each stimulus was presented 10 times in random order for a total of 120 presentations (2 precursors [ya], [wa]) $\times$ 6 CV syllables $\times$ 5 presentations $\times$ 2 blocks). Experimental apparatus and procedure were identical to those for experiment 1. The experimental session lasted approximately 15 min.

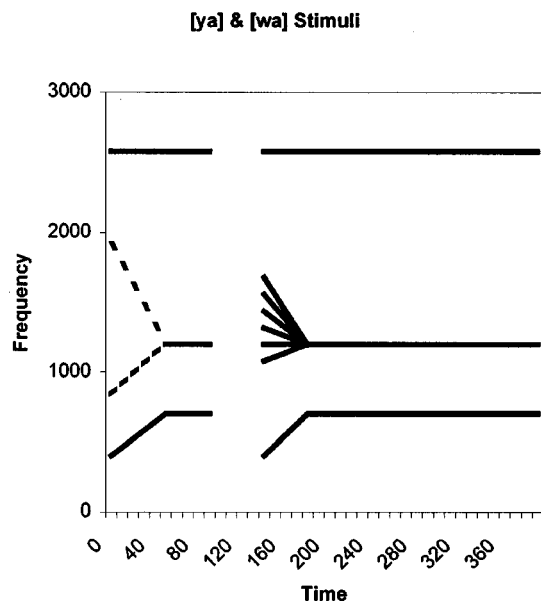


FIG. 4. Schematic depictions of center frequencies of the first three formants of stimuli for experiment 3. Onsets of glide-vowel syllable precursors (left) have identical center frequencies as [e] and [o] precursors used in previous experiments.

## B. Results and discussion

Figure 5 displays the mean consonant identification curves for experiment 3. Data from the 20 listeners who met the endpoint criterion were submitted to a one-sample t-test.<sup>3</sup> Listeners reported hearing [da] more often after [wa] than after [ya],  $t(19) = 2.99$ ,  $p < 0.01$ . Spectral characteristics of the onsets of the precursor stimuli provided sufficient spectral contrast to affect the perception of subsequent speech, even following considerable intervening acoustic information.

## V. DISCUSSION

Significance of spectral contrast for perception of coarticulated speech previously had been demonstrated for a variety of phonetic environments (e.g., CVVC: Lotto and Kluender, 1998; CVC: Holt, Lotto, and Kluender, 2000; VCV: Holt, 1999). In all previous work, very simple sine-wave flanking stimuli were used. Here, investigation of spectral contrast was extended to flanking stimuli with richer spectral composition. Single bands of spectral energy, approximating a single speech formant ( $F_2$ ), proved adequate to shift perception of a following stop to virtually the same extent as was found for replicas of vowels [e] and [o].

The most novel aspect of the present effort concerned the use of complementary precursor spectra for which spectral peaks were replaced with spectral troughs. Inasmuch as they were harmonic complexes, these complementary spectra were comparable to speech in complexity. Yet, they rendered no speech percept. As would be predicted by a spectral contrast account, complementary spectral precursors affected perception of following stops in a manner complementary to that for speech stimuli upon which they were modeled. These spectral complements were similarly effective in modifying perception of the following stop with either one or three

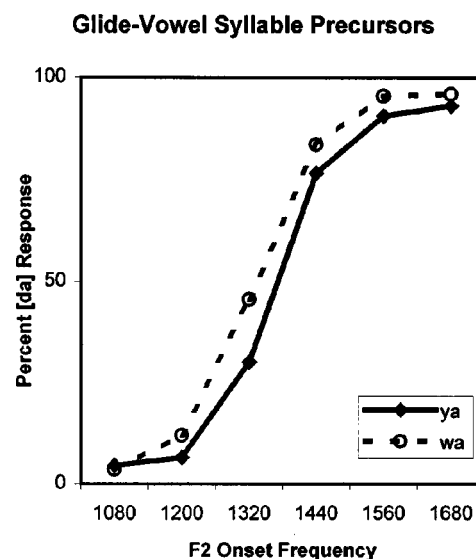


FIG. 5. Data from identification (b/d) task following [ya] (filled diamonds) and [wa] (open circles) syllables in experiment 3.

stopbands, as was the case for one- and three-passband stimuli. From this, it can be inferred that the principal source of effects on perception of following stops rests in characteristics of the preceding spectrum in the region of  $F_2$ —the principal acoustic determinant of perception of labial versus alveolar stops in the test series.

Results for complementary spectra differed in two important ways. First, contrast between spectral properties of the onsets of precursors versus CV syllables was most important. Second, due to the importance of onsets, the temporal distance between successive onsets proved critical. Effects of contrast between onsets appears to have a temporal extent that exceeds 100 ms and is less than 300 ms. Across experiments, it can be inferred that effects of precursor onsets are diminishing and become less reliable at around 200-ms interonset interval. The importance of interonset interval, versus precursor duration *per se*, appears less consistent with adaptation of suppression (Viemeister and Bacon, 1982) and more consistent with the perspective that adaptation serves mostly to enhance characteristics of onsets (e.g., Delgutte, 1996; Houtgast, 1974; Moore and Glasberg, 1983). It is important to note, however, that adaptation and adaptation of suppression are not mutually exclusive models. All that can be said is that adaptation provides the best explanation of the present data for spectral complements, and adaptation of suppression likely remains the best explanation for the many instances of auditory enhancement. This probably includes Summerfield and colleagues' demonstrations of vowel aftereffects and some of the effects of flanking speech (bandpass) and nonspeech (sine-wave) energy on perception of adjacent speech sounds.

Results from experiment 2 demonstrate that contrast between onsets is perceptually significant for both spectral peaks and spectral troughs, and results from experiment 3 provide evidence that these effects are sufficiently robust to play a role in connected speech for which there exists ample energy intervening between onsets.

Contrast effects are ubiquitous in perception, occurring

in every modality. This universality finds its cause in the fact that sensorineural systems respond predominantly to change. Within a modality, one may expect to find effects of contrast for every dimension within that modality. This certainly appears to be the case for vision, for which contrast plays a role in perception of lightness, line orientation, size, position, curvature, spatial frequency, depth, and color. One may similarly expect processes of contrast to be present for every auditory dimension including contrast between spectral properties of onsets.

## ACKNOWLEDGMENTS

This work was supported by the National Institute of Deafness and Communicative Disorders, Grant No. R01 DC 04072. The authors thank Chris Darwin, Christopher Turner, and an anonymous reviewer for insightful suggestions on an earlier version of this manuscript.

<sup>1</sup>Data for number of [da] responses were entered into a 2 ([e], [o]) $\times$ 2 (passband, stopband) $\times$ 3 (duration) $\times$ 6 (stimulus steps) within-subjects ANOVA. Consistent with the large effect apparent in Fig. 2, statistical analyses reveal a significant interaction between passband/stopband and duration.

<sup>2</sup>There is no reason to believe that effects of spectral contrast are restricted to regions near the identification crossover (boundary). Consequently, total response is the most appropriate measure of effect size (see also Samuel, 1986). This is especially true for the series of CV stimuli used here because all six members of the series are within three steps from identification crossover points. Nevertheless, crossover points were estimated using probit analysis (Finney, 1971) for every experiment reported here, and in only four cases did probit values result in statistical patterns different from those for mean response data. In all but one case (200-ms spectral complements in the second follow-up experiment,  $p=0.106$ , two-tail), probit differences were significant in a one-tail comparison but fell short of significance for two-tail.

<sup>3</sup>It is not clear why there was such a high attrition rate in this study. Syllables [ya] and [wa] are the most realistic stimuli used across all experiments reported here. Apparently listeners found this task very difficult, perhaps because of the similarity between precursor and test syllables. Some listeners fell just short of the 90-percent endpoint criterion. Others responded based on the precursor syllable identities. Still others seemed to respond randomly for unknown reasons. However, one can be confident that the remaining listeners heard the syllables as intended, because those who did not were removed based on the endpoint criterion. This study was replicated, with 13 of 28 listeners failing to meet the endpoint criterion. The results of the remaining 15 listeners replicate the original finding,  $t(14)=2.58$ ,  $p<0.05$ .

Ames, A. (1935). "Aneiseikonia—A factor in the functioning of vision," *Am. J. Ophthalmol.* **28**, 248–262.  
 Bergman, R., and Gibson, J. J. (1959). "The negative aftereffect of a surface slanted in the third dimension," *Am. J. Psychol.* **72**, 364–374.  
 Blakemore, C., and Sutton, P. (1969). "Size adaptation: A new aftereffect," *Science* **166**, 245–247.  
 Blumstein, S. E., and Stevens, K. N. (1981). "Phonetic features and acoustic invariance in speech," *Cognition* **10**, 25–32.  
 Cardozo, B. L. (1967). "Ohm's Law and masking," in *IPO Annual Progress Report, Institute for Perception Research* (Eindhoven, The Netherlands), **2**, pp. 59–64.  
 Cathcart, E. P., and Dawson, S. (1928–1929). "Persistence (2)," *Br. J. Psychol.* **19**, 343–356.  
 Christman, R. J. (1954). "Shifts in pitch as a function of prolonged stimulation with pure tones," *Am. J. Psychol.* **67**, 484–491.  
 Delgutte, B. (1980). "Representation of speech-like sounds in the discharge patterns of auditory nerve fibers," *J. Acoust. Soc. Am.* **68**, 843–857.  
 Delgutte, B. (1986). "Analysis of French stop consonants with a model of the peripheral auditory system," in *Invariance and Variability of Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 131–177.

Delgutte, B. (1996). "Auditory neural processing of speech," in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver (Blackwell, Oxford), pp. 507–538.  
 Delgutte, B., Hammond, B. M., Kalluri, S., Litvak, L. M., & Cariani, P. A. (1996). "Neural encoding of temporal envelope and temporal interactions in speech," in *Auditory Basis of Speech Perception*, edited by W. Ainsworth and S. Greenberg (European Speech Communication Association), pp. 1–9.  
 Delgutte, B., and Kiang, N. Y. S. (1984). "Speech coding in the auditory nerve IV. Sounds with consonant-like dynamic characteristics," *J. Acoust. Soc. Am.* **75**, 897–907.  
 Denes, P. (1955). "Effect of duration on perception of voicing," *J. Acoust. Soc. Am.* **27**, 761–764.  
 Diehl, R. L., Kluender, K. R., and Walsh, M. A. (1990). "Some auditory bases of speech perception and production," in *Advances in Speech, Hearing, and Language Processing*, edited by W. A. Ainsworth (JAI, London), pp. 243–267.  
 Diehl, R. L., and Walsh, M. A. (1989). "An auditory basis for the stimulus-length effect in the perception of stops and glides," *J. Acoust. Soc. Am.* **85**, 2154–2164.  
 Duifhuis, H. (1970). "Audibility of high harmonics in a periodic pulse," *J. Acoust. Soc. Am.* **48**, 888–893.  
 Festen, J. M., and Plomp, R. (1981). "Relations between auditory functions in normal hearing," *J. Acoust. Soc. Am.* **70**, 356–369.  
 Finney, D. J. (1971). *Probit Analysis* (Cambridge University Press, New York).  
 Flügel, J. C. (1920–1921). "On local fatigue in the auditory system," *Br. J. Psychol.* **11**, 105–134.  
 Fowler, C. A., Brown, J. M., and Mann, V. A. (2000). "Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans," *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 877–888.  
 Gibson, J. J. (1933). "Adaptation, after-effect and contrast in the perception of curved lines," *J. Exp. Psychol.* **16**, 1–31.  
 Gibson, J. J. (1937). "Adaptation with negative after-effect," *Psychol. Rev.* **44**, 222–244.  
 Gibson, J. J., and Radner, M. (1937). "Adaptation, after-effect, and contrast in the perception of tilted lines. I. Quantitative studies," *J. Exp. Psychol.* **20**, 453–467.  
 Green, D. M., McKey, M. J., and Licklider, J. C. R. (1959). "Detection of a pulsed sinusoid in noise as a function of frequency," *J. Acoust. Soc. Am.* **31**, 1146–1152.  
 Greenwood, D. D. (1990). "A cochlear frequency-position function for several species—29 years later," *J. Acoust. Soc. Am.* **87**, 2592–2605.  
 Guilford, J. P., and Park, D. G. (1931). "The effect of interpolated weights upon comparative judgments," *Am. J. Psychol.* **43**, 589–599.  
 Heil, P. (1997a). "Auditory cortical onset responses revisited. I. First-spike timing," *J. Neurophysiol.* **77**, 2616–2641.  
 Heil, P. (1997b). "Auditory cortical onset responses revisited. II. Response strength," *J. Neurophysiol.* **77**, 2642–2660.  
 Holt, L. L. (1999). "Auditory constraints on speech perception: An examination of spectral contrast," unpublished Ph.D. dissertation, University of Wisconsin–Madison.  
 Holt, L. L., and Kluender, K. R. (2000). "General auditory processes contribute to perceptual accommodation of coarticulation," *Phonetica* **57**, 170–180.  
 Holt, L. L., Lotto, A. J., and Kluender, K. R. (2000). "Neighboring spectral content influences vowel identification," *J. Acoust. Soc. Am.* **108**, 710–722.  
 Holt, L. L., Stephens, J. D., and Lotto, A. J. (unpublished).  
 Horst, J. W., Javel, E., and Farley, G. R. (1990). "Coding of fine spectral structure in the auditory nerve. II. Level-dependent nonlinear responses," *J. Acoust. Soc. Am.* **88**, 2656–2681.  
 Houtgast, T. (1972). "Psychophysical evidence for lateral inhibition in hearing," *J. Acoust. Soc. Am.* **51**, 1885–1894.  
 Houtgast, T. (1974). "Auditory analysis of vowel-like sounds," *Acustica* **31**, 320–324.  
 Kieffe, M. J., Kluender, K. R., and Rhode, W. S. (2002). "Synthetic speech stimuli spectrally normalized for nonhuman cochlear dimensions," *ARLO* **3**, 41–46.  
 Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," *J. Acoust. Soc. Am.* **67**(3), 971–995.  
 Kluender, K. R., and Diehl, R. L. (1988). "Vowel-length differences before voiced and voiceless consonants: An auditory explanation," *J. Phonetics* **16**, 153–169.

- Koffka, K. (1935). *Principles of Gestalt Psychology* (Harcourt, Brace & World, New York).
- Kohler, W., and Emery, D. A. (1947). "Figural aftereffects in the third dimension of visual space," *Am. J. Psychol.* **60**, 159–201.
- Kohler, W., and Wallach, H. (1944). "Figural aftereffects: An investigation of visual processes," *Proc. Am. Philos. Soc.* **88**, 269–357.
- Lindblom, B. E. F. (1963). "Spectrographic study of vowel reduction," *J. Acoust. Soc. Am.* **35**, 1773–1781.
- Lindblom, B. E. F., and Studdert-Kennedy, M. (1967). "On the role of formant transitions in vowel recognition," *J. Acoust. Soc. Am.* **42**, 830–843.
- Locke, J. (1706/1974). *An Essay Concerning Human Understanding*, 5th ed., edited by A. D. Woolzley (New American Library, New York). Original work published in 1706.
- Lotto, A. J., and Kluender, K. R. (1998). "General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification," *Percept. Psychophys.* **60**, 602–619.
- Mann, V. A. (1980). "Influence of preceding liquid in stop-consonant perception," *Percept. Psychophys.* **28**, 407–412.
- Mann, V. A., and Repp, B. H. (1981). "Influence of preceding fricative on stop consonant perception," *J. Acoust. Soc. Am.* **69**, 548–558.
- Miller, J. L., and Liberman, A. M. (1979). "Some effects of later-occurring information on the perception of stop consonant and semivowel," *Percept. Psychophys.* **25**, 457–465.
- Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulas for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.
- Öhman, S. E. G. (1965). "Coarticulation in VCV utterances: Spectrographic measurements," *J. Acoust. Soc. Am.* **39**(1), 151–168.
- Phillips, D. P., and Hall, S. E. (1990). "Response timing constraints on the cortical representation of sound time structure," *J. Acoust. Soc. Am.* **88**, 1403–1411.
- Pisoni, D. B., Carrell, T. D., and Gans, S. J. (1983). "Perception of the duration of rapid spectrum changes in speech and nonspeech signals," *Percept. Psychophys.* **34**, 314–322.
- Port, R. F., and Dalby, J. (1982). "Consonant/vowel ratio as a cue for voicing in English," *Percept. Psychophys.* **32**, 141–152.
- Raphael, L. F. (1972). "Preceding vowel duration as a cue to the perception of the voicing characteristics of word-initial consonants in English," *J. Acoust. Soc. Am.* **51**, 1296–1303.
- Repp, B. (1982). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," *Psychol. Bull.* **92**, 81–110.
- Rhode, W. S. (1991). "Physiological–morphological properties of the cochlear nucleus," in *Neurobiology of Hearing: The Central Auditory System*, edited by R. A. Altschuler (Raven Press, New York).
- Rhode, W. S., and Recio, A. (2001). "Basilar membrane response to multi-component stimuli in chinchilla," *J. Acoust. Soc. Am.* **110**, 981–994.
- Samuel, A. G. (1986). "Red herring detectors and speech perception: In defense of selective adaptation," *Cognit. Psychol.* **18**, 452–499.
- Schouten, J. F. (1940). "The residue, a new component in subjective analysis," *Proc. K. Ned. Akad. Wet.* **43**, 356–365.
- Sherif, M., Taub, D., and Hovland, C. I. (1958). "Assimilation and contrast effects of anchoring stimuli on judgments," *J. Exp. Psychol.* **55**, 150–155.
- Smith, R. L. (1979). "Adaptation, saturations, and physiological masking in single auditory-nerve fibers," *J. Acoust. Soc. Am.* **65**, 166–178.
- Smith, R. L., Brachman, M. L., and Frisina, R. D. (1985). "Sensitivity of auditory-nerve fibers to changes in intensity: A dichotomy between decrements and increments," *J. Acoust. Soc. Am.* **78**, 1310–1316.
- Smith, R. L., and Zwislocki, J. J. (1971). "Responses of some neurons of the cochlear nucleus to tone-intensity increments," *J. Acoust. Soc. Am.* **50**, 1520–1525.
- Stephens, J. D., and Holt, L. L. (2002). "Are context effects in speech perception modulated by visual information?" 43rd Annual Meeting of the Psychonomic Society, Kansas City, MO.
- Summerfield, Q., Haggard, M. P., Foster, J., and Gray, S. (1984). "Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect," *Percept. Psychophys.* **35**, 203–213.
- Summerfield, Q., Sidwell, A., and Nelson, T. (1987). "Auditory enhancement of changes in spectral amplitude," *J. Acoust. Soc. Am.* **81**, 700–707.
- Viemeister, N. F. (1980). "Adaptation of masking," in *Psychophysical, Physiological, and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (University Press, Delft), pp. 190–197.
- Viemeister, N. F., and Bacon, S. P. (1982). "Forward masking by enhanced components in harmonic complexes," *J. Acoust. Soc. Am.* **71**, 1502–1507.
- Wallach, H. (1948). "Brightness constancy and the nature of achromatic colors," *J. Exp. Psychol.* **38**, 310–324.
- Wightman, F. L. (1973). "Pitch and stimulus fine structure," *J. Acoust. Soc. Am.* **54**, 397–406.
- Wightman, F., McKee, T., and Kramer, M. (1977). "Factors influencing frequency selectivity in normal and hearing-impaired listeners," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 295–310.
- Winter, I. M., and Palmer, A. R. (1995). "Level dependence of cochlear nucleus onset unit responses and facilitation by second tones or broadband noise," *J. Neurophysiol.* **73**, 141–159.