Spectral tilt change in stop consonant perception

Joshua M. Alexander[a] and Keith R. Kluender
Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706

(Received 22 January 2007; revised 26 September 2007; accepted 5 November 2007)

There exists no clear understanding of the importance of spectral tilt for perception of stop consonants. It is hypothesized that spectral tilt may be particularly salient when formant patterns are ambiguous or degraded. Here, it is demonstrated that relative change in spectral tilt over time, not absolute tilt, significantly influences perception of /b/ vs /d/. Experiments consisted of burstless synthesized stimuli that varied in spectral tilt and onset frequency of the second formant. In Experiment 1, tilt of the consonant at voice onset was varied. In Experiment 2, tilt of the vowel steady state was varied. Results of these experiments were complementary and revealed a significant contribution of relative spectral tilt change only when formant information was ambiguous. Experiments 3 and 4 replicated Experiments 1 and 2 in an /aba/-/ada/ context. The additional tilt contrast provided by the initial vowel modestly enhanced effects. In Experiment 5, there was no effect for absolute tilt when consonant and vowel tilts were identical. Consistent with earlier studies demonstrating contrast between successive local spectral features, perceptual effects of gross spectral characteristics are likewise relative. These findings have implications for perception in nonlaboratory environments and for listeners with hearing impairment. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2817617]

PACS number(s): 43.71.Es [JES] Pages: 386–396

I. INTRODUCTION

Experience with multiple acoustic differences between speech sounds gives rise to trading relations between attributes or cues for identification (e.g., Repp, 1982), thereby making speech perception robust (Kluender and Alexander, in press). When one attribute becomes less informative or ambiguous, perceptual constancy can be maintained by information provided by other attributes. For example, voice onset time (VOT) and fundamental frequency ($f0$) often covary in the production of stop consonants, with shorter VOTs and lower $f0$'s corresponding to voiced consonants. Whether because of learned covariation (Holt *et al.*, 2001) or general auditory processes (Kingston and Diehl, 1994), increases in VOTs can be offset to some extent by decreases in $f0$ in order to maintain the perception of voicing (Whalen *et al.*, 1993). In other words, perception of voicing is not an absolute function of VOT or $f0$, but is dependent on these and other attributes of the signal. The literature is replete with examples of these trading relations (e.g., Repp 1982), which are by-products of the perceptual system's natural ability to extract relationships among multiple sources of information.

Spectrally global sources of information, such as gross spectral tilt (i.e., the balance of low- and high-frequency energy), may contribute to perception of place of articulation in stop consonants. However, the majority of research concerning gross spectral properties of stop consonants varying in place of articulation has been conducted to support or to critique claims of acoustic invariance (e.g., Stevens and Blumstein, 1978, 1981; Blumstein and Stevens, 1979, 1980; Blumstein *et al.*, 1982; Walley and Carrell, 1983; Kewley-Port, 1983; Lahiri *et al.*, 1984; Kewley-Port and Luce, 1984; Dorman and Loizou, 1996). It is now generally accepted that multiple sources of information influence perception and that no one source is absolute or invariant across all acoustic contexts for all listeners. It is likely that the importance of a particular acoustic attribute of speech is at least related to (1) its reliability (mean differences and variance) in distinguishing phonemes across speech contexts (2) listeners' sensitivity (ability to discriminate differences) to the attribute across phonemes, and (3) the presence of other acoustic attributes. Much work has focused on the first of these (see the following), but much less work has been done on the last two. For example, little is understood about how changes in the relative sensitivity of different acoustic attributes accompanying sensorineural hearing loss influence their importance in speech perception. This is the focus of other work in this lab. The focus of the current work on normal-hearing listeners is to examine how a spectrally global attribute, gross spectral tilt, conspires with spectrally local information, formant peaks, to inform perception of place of articulation in stop consonants.

Much of the research on gross spectral tilt in speech emanated from Stevens and Blumstein's work (e.g., 1979, 1980; Stevens and Blumstein, 1978, 1981) which advanced a strong argument that an invariant acoustic marker for place of articulation in stop consonants can be found by integrating the spectral energy of the first 25 ms or so following the onset of the release burst. For voiced consonants with absent or short-duration bursts, this integration window includes not only burst energy but also the first one or two pitch pulses associated with the onset of voicing. Stevens and Blumstein argued that perception of place of articulation is primarily determined by the shape of the stimulus onset spectrum, which is governed by the size and shape of resonator cavities created by different constriction points in the oral tract. In a

[a]Current address: Boys Town National Research Hospital, Omaha, NE 68131. Electronic mail: alexanderj@boystown.org

preemphasized signal (+6 dB/oct.), a labial place of articulation is characterized by a "diffuse-falling spectrum" (i.e., negative spectral tilt), an alveolar place of articulation by a "diffuse-rising spectrum" (i.e., positive spectral tilt), and a velar place of articulation by a "prominent midfrequency spectral peak." Kewley-Port (1983), who advocated for classification templates based on kinematic spectral features through the first 40 ms following stimulus onset, also adopted this classification convention for place of articulation, in addition to the timing of voice onset and the presence of midfrequency peaks. In contrast to early work (e.g., Liberman *et al.*, 1952; Halle *et al.*, 1957) that was event based and focused on spectral shape information exclusively in the burst, the spectral features in both the Stevens and Blumstein framework and the Kewley-Port framework extend to the vocalic portion of the syllable because they are time based and not limited to the burst.

Refinement of the above-mentioned approach as a template for invariance was motivated in part by the failure to accurately classify stops in Malayalam and French (Lahiri *et al.*, 1984) and by conflicting-cue experiments that seemed to indicate that onset frequencies of formant peaks influenced perception of stop consonants much more than absolute spectral tilt (Blumstein *et al.*, 1982; Walley and Carrell, 1983). Both Blumstein (Lahiri *et al.*, 1984) and Kewley-Port (Kewley-Port and Luce, 1984) altered their original templates based on static spectral shapes to include a kinematic component based on the relative change in tilt. That is, a positive change in tilt was used as characteristic marker for a labial place of articulation and a negative change in tilt for an alveolar place of articulation.

Lahiri *et al.* (1984) devised a perceptual experiment to support their classification data. Synthesizing "prototypical" consonant-vowel (CV) syllables as produced by a French speaker, they altered either the tilt of the burst or the tilt of the voicing onset of [b] and [d] in five different vowel contexts ([i], [e], [a], [o], [u]). For the [b] exemplars, relative tilt was altered to more closely resemble the pattern for [d]. That is, the tilt of the burst was increased to a diffuse-rising spectrum or the tilt of the voicing onset was decreased while the tilt of the burst was held constant so the change in tilt was negative. The opposite was done for the [d] exemplars. Lahiri *et al.* found that most of the stimuli were classified as the phoneme cued by relative tilt, and tilt manipulation of the burst was slightly more effective at altering perception than tilt manipulation of voicing onsets.

Several limitations of the Lahiri *et al.* (1984) perceptual experiment motivate the need for a clearer demonstration of the use of relative spectral tilt in speech perception. First, relative spectral tilt was considered only as a categorical variable using an arbitrary metric based on the ratio of the energy difference in the spectral envelope between burst and voicing onset at 3500 Hz and the difference at 1500 Hz. At the time, one was limited to this crude metric of relative tilt given the difficulty in generating the different tilts via formant amplitude manipulation in the parallel branch of the Klatt synthesizer (Klatt, 1980).[1] As noted by the authors, "because there was overlap in the skirts of the filters, a change in the amplitude of one formant peak often resulted in a change of spectral shape for other formant peaks" (p. 401). Finally, as noted by Dorman and Loizou (1996), the exemplar stimuli of Lahiri *et al.* were perceptually impoverished; 12 of the 30 subjects could not reliably identify the exemplar stimuli above 70% in three of the five vowel contexts.

Dorman and Loizou (1996) took a slightly different approach to the perceptual experiment of Lahiri *et al.* (1984). First, they used naturally produced speech rather than synthesized speech, which resulted in near perfect identification of the exemplar stimuli. Next, using the fast-Fourier transform (FFT) magnitude spectra of the stimuli, they altered the tilt of the burst or of the voicing onset to match the metric defined by Lahiri *et al.* (1984). Unlike Lahiri *et al.*, Dorman and Loizou found that altering relative tilt change did not substantially influence perception of /b/ and /d/ in any of the vowel contexts, except for the syllable /bi/.

While Dorman and Loizou (1996) demonstrated that relative spectral tilt is not an invariant cue for place of articulation, this does not imply that tilt plays no role in perception. As noted earlier, the importance of a particular acoustic attribute (e.g., spectral tilt) depends on a number of factors, including the quality of information provided by other acoustic attributes (e.g., formant frequency). From this perspective, the importance of Dorman and Loizou's findings with clearly spoken natural speech is a demonstration that the effects of relative tilt are overwhelmed when most other potential acoustic information is available to listeners. However, there are multiple circumstances when formant information is compromised. Listeners with hearing loss have diminished frequency resolution that obscures spectral peaks, and the presence of background noise likely has a similar effect for normal-hearing listeners. Outside the laboratory, fluent speech is often hypoarticulated such that ambiguous information specifying place might be the rule more than the exception. From this perspective, the greater importance of spectral tilt in Lahiri *et al.* (1984) can be explained by the fact that their stimuli had compromised information for perception of place.

In the following series of experiments, we evaluate the interaction between spectral tilt and formant frequency in perception of voiced stop consonants. While controlling for phonetic context and all other sources of information, each attribute independently varied along a series from [ba] to [da], including several intermediate or ambiguous levels.

II. STIMULI

A. Rationale

The goals in creating our stimuli were to control for all extraneous variables except those under study and to make the acoustic manipulations systematically. For this reason, we first synthesized a burstless /ba/ to /da/ series that varied acoustically in the onset frequency of the second formant (F2) transition. We then used digital filters to manipulate the tilt trajectory of the formant transitions. We chose a labial to alveolar series because the spectral shape manipulation is easily defined from negative to positive. In our case, because spectral shape is described without preemphasis, this series is

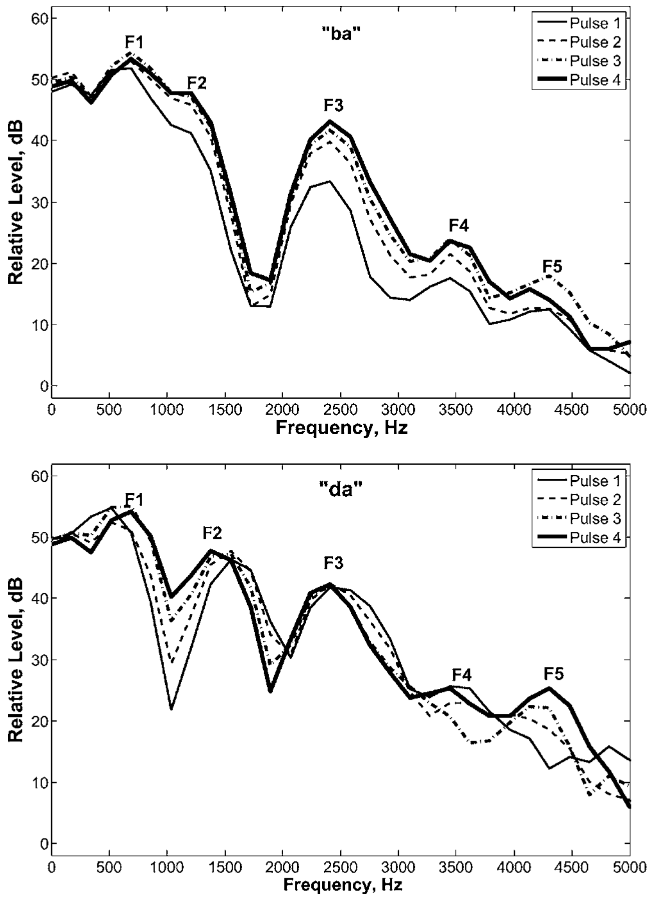

FIG. 1. The short-term spectra for the first four pitch pulses following the onset of voicing from a male talker with fundamental frequency of about 105 Hz are displayed. The top and bottom panels show the short-term spectral history of a production of the syllable /ba/ and the syllable /da/, respectively. Each pitch pulse (different lines) was about 9.5 ms in duration ($1/f0$), and was analyzed without preemphasis using a 256-point FFT with a 50% Hamming window overlap.

described as going from steep negative tilt (labial) to shallow tilt (alveolar). The low, back vowel /a/ was chosen because it has a relatively neutral spectral tilt and because phonetically trained and untrained listeners identified it as /a/ despite wide manipulations of spectral tilt. This contrasts with perception of high vowels, for example, which Kiefte and Kluender (2005) demonstrated to be quite sensitive to changes in spectral tilt. There is precedence for synthesizing burstless stops (e.g., Fruchter and Sussman, 1997) as it simplifies conclusions concerning the parameter under study. Furthermore, because stop bursts in vocalic positions other than the utterance initial position are often either low in energy or absent, synthesizing stops without the bursts extends the generality of our findings to typical running speech for which bursts are a less frequent event particularly for voiced consonants.

As noted earlier, the use of tilt as a cue for stop consonant perception is not limited to bursts, but also extends to the following vocalic segment. Changes in spectral tilt related to resonance properties (versus source properties) are to be expected as vocal tract shape changes. Figure 1 shows the short-term spectra for the first four pitch pulses following the onset of voicing from an adult male production (fundamental frequency of about 105 Hz) of the syllables [ba] (top panel) and [da] (bottom panel) extracted from a sample of running nonsense speech recorded with a sampling rate of 44 100 Hz. Here and elsewhere in this report, each pitch pulse was analyzed without preemphasis using a 256-point FFT with a 50% Hamming window overlap. The top panel shows that for this production of [ba], the short-term spectra of the initial pitch pulse ($t=0-9.5$ ms) following the onset of voicing (thin solid line) is steeply negative and quickly transitions to the shallower vowel tilt by the second or third pitch pulse (about 19–28.5 ms). In contrast, the short-term spectra of the initial pitch pulse for the production of [da] in the bottom panel is relatively shallow and changes very little over the time course of the formant transitions.[2]

B. Synthesis of F2 series

A burstless CV series varying perceptually from /ba/ to /da/ and acoustically in eight steps of F2-onset frequency (1000–1700 Hz, respectively) was synthesized at a sampling rate of 22 050 Hz with 16 bits of resolution and with a 5 ms update rate using the parallel branch of the Klatt synthesizer (Klatt and Klatt, 1990). The CVs were 250 ms total duration, including 30 ms formant transitions with a linear amplitude rise of 6 dB. Fundamental frequency was 100 Hz and decreased to 90 Hz during the final 50 ms. Synthesis parameters are provided in Table I. Because F4 was the highest

TABLE I. Klatt synthesis parameters for the eight-step series varying in the onset frequency of the second formant (F2). The amplitudes of F1 and F2 (A1V and A2V, respectively) were varied as a function of formant frequency, in order to maintain a relatively constant spectral tilt of −3 dB/oct. All other parameters were kept constant throughout the duration of the 250 ms stimuli, except $f0$ which started at 100 Hz and decreased to 90 Hz during the final 50 ms.

F1	A1V	BW1	F2	A2V	BW2	F3	A3V	BW3	F4	A4V	BW4
$t=0$ (constant onset)											
300	50	80	1000	70	90	2400	72	150	3600	80	350
			1100	69							
			1200	68							
			1300	67							
			1400	66							
			1500	65							
			1600	64							
			1700	63							
$t=30$ ms (vowel steady state)											
800	55	80	1200	67	90	2400	72	150	3600	80	350

formant frequency synthesized, its nominal amplitude (A4V) was set to maximum (80 dB) and nominal values in the synthesizer for the remaining formant amplitudes (A1V–A3V) were manipulated so that each of the CVs had a reasonably constant spectral tilt of −3 dB/oct. throughout its duration. Spectral tilt was referenced to the energy in F4 because it did not change frequency during the duration of the stimulus. This is expressed in the following:

$$(\mathrm{Amp}_{F2})_{dB} = -3 \times \log_2(F2/F4) + (\mathrm{Amp}_{F4})_{dB}, \quad (1)$$

where $(\mathrm{Amp}_{F2})_{dB}$ is the desired amplitude of F2 in dB, −3 is the designated tilt, $\log_2(F2/F4)$ is the number of octaves between the fourth and second formant, and $(\mathrm{Amp}_{F4})_{dB}$ is the measured amplitude of F4 in dB. F1 and F3 were substituted for F2 in Eq. (1) to derive the desired amplitudes of the first and third formants. Formant frequency and amplitude values were linearly interpolated by the synthesizer between $t=0$ ms (consonant onset) and $t=30$ ms (beginning of vowel steady state). It is important to note that the amplitudes of the formants were empirically measured from spectra of the individual pitch pulses (about every 10 ms) and were used to set the nominal amplitude values in the synthesizer so that spectral tilt for all four formants was reasonably constant [cf. Eq. (1)].

III. EXPERIMENT 1: EFFECT OF CONSONANT ONSET TILT ON CV PERCEPTION

A. Rationale

We predict that spectral tilt will have the greatest influence for stimuli in which F2-onset frequency is ambiguous between [ba] and [da]. When formant cues to place of articulation are ambiguous, steeper tilts at onset (more negative) should encourage perception of /ba/ and shallower tilts at onset (more positive) should encourage perception of /da/. To test this hypothesis, stimuli varying in F2-onset frequency (see earlier text) were filtered to generate a series of five spectral tilts that varied at consonant onset from −12 to 0 dB/oct. These end point tilts are equivalent to Blumstein and Stevens's templates for labial and alveolar stop consonants, respectively (e.g., Blumstein and Stevens, 1979). For all stimuli, spectral tilt transitioned along with F1 and F2 frequency from an initial onset value at $t=0$ ms to a common value in the vowel portion of the syllables at $t=30$ ms.

B. Methods

1. Listeners

Listeners for all experiments were undergraduate students from the University of Wisconsin–Madison and participated as part of course credit. No listener participated in more than one experiment. All reported that they were native speakers of American English and had normal hearing. One to three listeners ran in the experiment concurrently. Each individual was seated in an isolated single-walled sound chamber and had a unique presentation order of the stimuli. Listeners were recruited for each experiment until data were collected on at least 20 participants. Twenty-three listeners (3 male, 20 female) participated in Experiment 1.

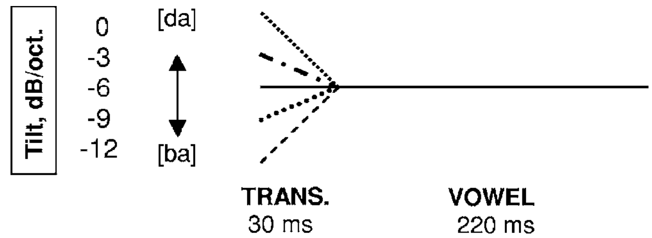

FIG. 2. Schematic representing the change in tilt for the stimuli in Experiment 1 in which different consonant onset tilts converged to a −6 dB/oct. vowel tilt. Consonant onset tilts steeper than the vowel tilt are expected to result in more labial responses and consonant onset tilts shallower than the vowel tilt are expected to result in more alveolar responses.

2. Stimuli

Using the eight-step series varying in F2-onset frequency, parametric manipulations of spectral tilt were made on time slices cut at the zero crossings corresponding to the midpoint and the end of each pitch pulse (about every 5 ms). Stimuli were filtered between 212 and 4800 Hz, using 90-order finite impulse response (FIR) filters created in MATLAB, to have one of five different spectral tilts at consonant onset ranging from −12 to 0 dB/oct. (Because stimuli already had a constant −3 dB/oct. tilt, the actual slopes of the filters varied from −9 to +3 dB/oct.) During formant transitions ($t \leqslant 30$ ms), spectral tilt converged linearly to the vowel steady state, which was filtered as a whole segment to a tilt of −6 dB/oct. (see Fig. 2). Because initial portions of wave forms are not filtered accurately when the length of the impulse response approaches that of the wave form, each input wave form was concatenated several times and then convolved with the FIR filters. From the medial portion of this filtered wave form, the output wave form was extracted at zero crossings corresponding to the original input wave form and then scaled to the RMS amplitude of the original time slice. The CVs were unsampled to 48 828 Hz with 24 bits of resolution and low-pass filtered with an 86-order FIR filter with a passband at 4800 Hz and a stopband of −90 dB of at 6400 Hz. The CVs were then scaled to a constant rms amplitude.

Figure 3 displays the short-term spectra of the first four pitch pulses from sample stimuli with F2-onset frequencies of 1400 Hz. Each pitch pulse is approximately 10 ms in duration. In each panel, the thin solid line is the short-term spectra of the first pitch pulse ($t \approx 0-10$ ms) and represents the consonant onset and the thick solid line is the short-term spectra of the fourth pitch pulse ($t \approx 30-40$ ms) and represents the beginning of the steady state portion of the following vowel. The short-term spectra of the second and third pulses show a continuous change in tilt and formant frequency between the first and fourth pitch pulse. Following Blumstein and Stevens's templates (e.g., Blumstein and Stevens, 1979), stimuli with steeply negative consonant onset tilts like the one represented in the top panel are expected to lead to more /ba/ responses and stimuli with relatively shallow onset tilts as in the bottom panel are expected to lead to more /da/ responses.

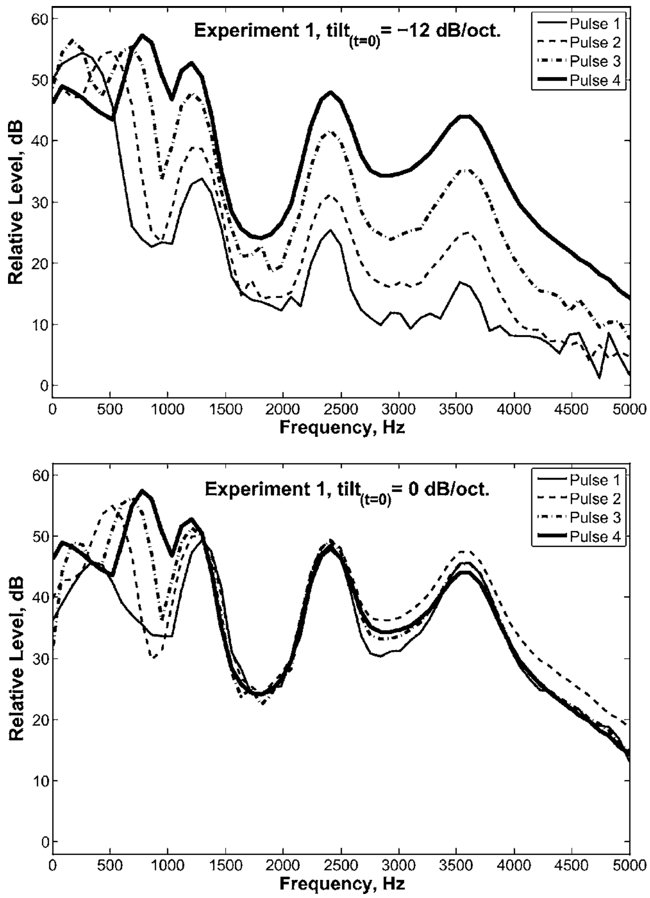

FIG. 3. Short-term spectra for the first four pitch pulses (about 10 ms each) for the tilt end point stimuli in Experiment 1. This example represents the stimuli with an F2-onset frequency of 1400 Hz. It is hypothesized that the stimuli in the top panel will lead to more labial responses because the tilt is steeply negative for the initial pitch pulse (thin solid line) and becomes shallower over the duration of the formant transition until it reaches a steady state at the vowel onset (thick solid line). The stimuli represented in the bottom panel are hypothesized to lead to more alveolar responses because the tilt is flat for the initial pitch pulse and becomes steeper over the duration of the formant transition until it reaches a steady state at the vowel onset.

3. Procedure

Participants listened to each of the 40 CV tokens (eight F2-onset frequencies by five spectral tilt trajectories) once per trial block in randomized order. Following two warm-up blocks (80 trials), data were collected on eight subsequent blocks (320 trials). Stimuli were presented diotically to participants through Beyerdynamic DT150 headphones at an average level of 73 dBA. In a two-alternative, forced choice task participants indicated their responses by pressing one of two buttons labeled "BA" and "DA."

C. Results

For every listener, at each tilt manipulation, the probability of responding /da/ as a function of F2-onset frequency was fit to a logistic function using the psignifit toolbox for MATLAB[3] which implements the maximum-likelihood method described by Wichmann and Hill (2001). The upper and lower asymptotes were free to vary in order to achieve the best fit. Frequency of F2 corresponding to the $p=0.5$ point on the ordinate, where /ba/ and /da/ responses are equally likely, hereafter the boundary, was obtained from the fits. The higher the boundary frequency, the greater the likelihood of a /ba/ response over the range of F2-onset frequencies. Figure 4 displays identification functions of the mean data for each consonant onset tilt; circles, asterisks, squares, crosses, and triangles represent the mean data for consonant onset tilts of −12, −9, −6, −3, and 0 dB/oct., respectively, and the different lines represent fitted functions for the mean data (see the legend).

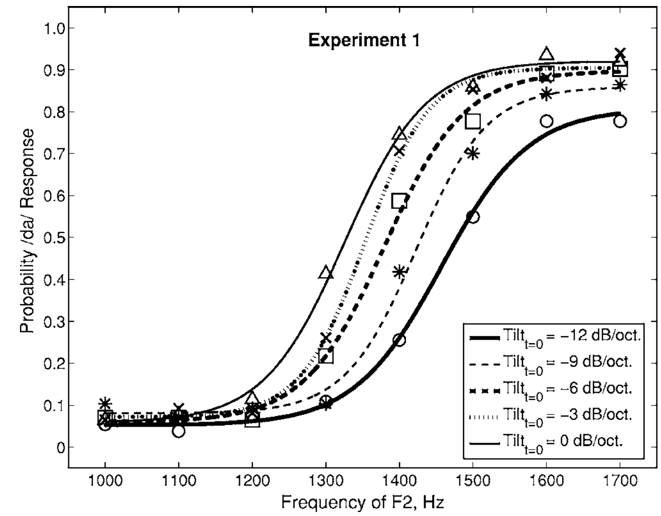

FIG. 4. Mean data for Experiment 1 in which the probability of responding /da/ as a function of F2-onset frequency is plotted separately for each consonant onset tilt. Circles, asterisks, squares, crosses, and triangles represent the mean data for consonant onset tilts of −12, −9, −6, −3, and 0 dB/oct., respectively. Maximum-likelihood fits of the identification functions are displayed for the mean data at each consonant onset tilt as different lines (see the legend).

Listeners' identifications were primarily influenced by onset frequencies of F2 as indicated by the floor/ceiling identifications of the formant frequencies near the series end points. However, tilt of consonant onset also influenced listeners' identifications, especially at intermediate values of F2 (e.g., 1400 Hz) where formant information for place of articulation was ambiguous. Listeners were more likely to perceive /da/ for consonant onsets with shallower tilt compared to consonant onsets with steeper tilt. The influence of consonant onset tilt can be quantified by shifts in the identification functions along the F2 axis—the boundary frequencies. Figure 5 displays the mean frequency and 95% confidence intervals (CIs) of boundaries at each consonant onset tilt. As can be seen, there was a systematic decrease in boundary frequency (increase in bias for /da/ responses) for progressively shallower consonant onset tilts (e.g., toward 0 dB/oct.). A within-subjects analysis of variance (ANOVA) confirmed that the effect of consonant onset tilt was statistically significant [$F(4,88)=16.1$, $p<0.0001$]. Tukey HSD post hoc tests revealed that the mean boundary frequency for the −12 dB/oct. consonant onset tilt was significantly greater than the mean boundary frequencies of the other consonant onset tilts except the −3 dB/oct. consonant onset tilt. Also, the mean boundary frequency for the 0 dB/oct. consonant onset tilt was significantly less than the mean boundary frequencies of the other consonant onset tilts except the −3 dB/oct. consonant onset tilt.[4] In addition, mean boundary frequencies for the −9 and −3 dB/oct. consonant onset tilts

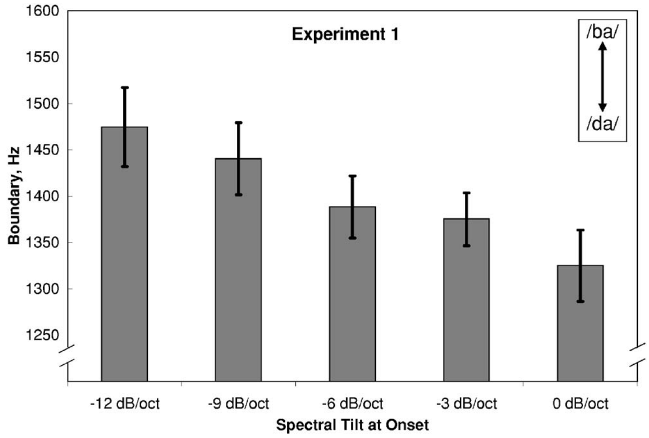

FIG. 5. Mean boundary frequencies of the identification functions at each consonant onset tilt from Experiment 1. Lower-frequency boundaries are associated with a bias toward /da/ responses. Error bars represent the 95% confidence intervals.

were significantly different from one another. This pattern of results indicates that shallower consonant onset tilts progressively increase the bias for alveolar responses and vice versa for the steeper consonant onset tilts creating a bias for labial responses.

IV. EXPERIMENT 2: EFFECT OF FOLLOWING VOWEL TILT ON CV PERCEPTION

A. Rationale

The results of Experiment 1 clearly indicate that despite an absence of bursts, the gross shape of the spectrum during the voiced portions of consonant onset influences perception of /ba/ and /da/. Following the templates of Blumstein and Stevens (e.g., Blumstein and Stevens, 1979), which were based on preemphasized spectra (+6 dB/oct.), negative spectral tilts at onset in our experiment (less than −6 dB/oct.) were associated with more labial (/ba/) responses and positive spectral tilts at onset (greater than −6 dB/oct.) were associated with more alveolar (/da/) responses. However, as noted by others (Lahiri *et al.* 1984; Kewley-Port and Luce, 1984), the important feature of the onset spectrum is its shape relative to the following vowel. Spectral tilt that becomes shallower (more positive) from consonant onset to vowel steady state (e.g., relative negative consonant onset tilt) should encourage perception of /ba/ and spectral tilt that becomes steeper (more negative) from consonant onset to vowel steady state (e.g., relative positive consonant onset tilt) should encourage perception of /da/. If relative tilt change is the perceptual cue used by listeners, then the pattern of results in Experiment 1 should be maintained if consonant onset tilt is held constant and vowel tilt varied. Experiment 2 tests this hypothesis.

B. Methods

Twenty-one college students who reported normal hearing (6 male, 15 female) identified as /ba/ or /da/ a series of 40 CVs that varied along both F2-onset frequency in eight steps and along spectral tilt of the following vowel in five steps ranging from −12 to 0 dB/oct. (see Fig. 6). Stimuli were filtered using the same techniques as described for Experiment 1. Participants listened to each of the CV tokens once per trial block in randomized order. Following two warm-up blocks (80 trials), data were collected on eight subsequent blocks (320 trials). Stimuli were presented diotically at an average level of 72 dBA.

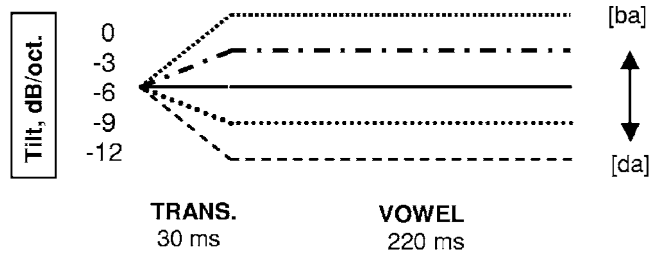

FIG. 6. Schematic representing the change in tilt for the stimuli in Experiment 2 in which a −6 dB/oct. consonant onset tilt diverged to different vowel tilts. Vowel tilts shallower (more positive) than the consonant onset tilt are expected to result in more labial responses and vowel tilts steeper (more negative) than the consonant onset tilt are expected to result in more alveolar responses.

Figure 7 displays the short-term spectra of the first four pitch pulses from sample stimuli in Experiment 2 with F2-

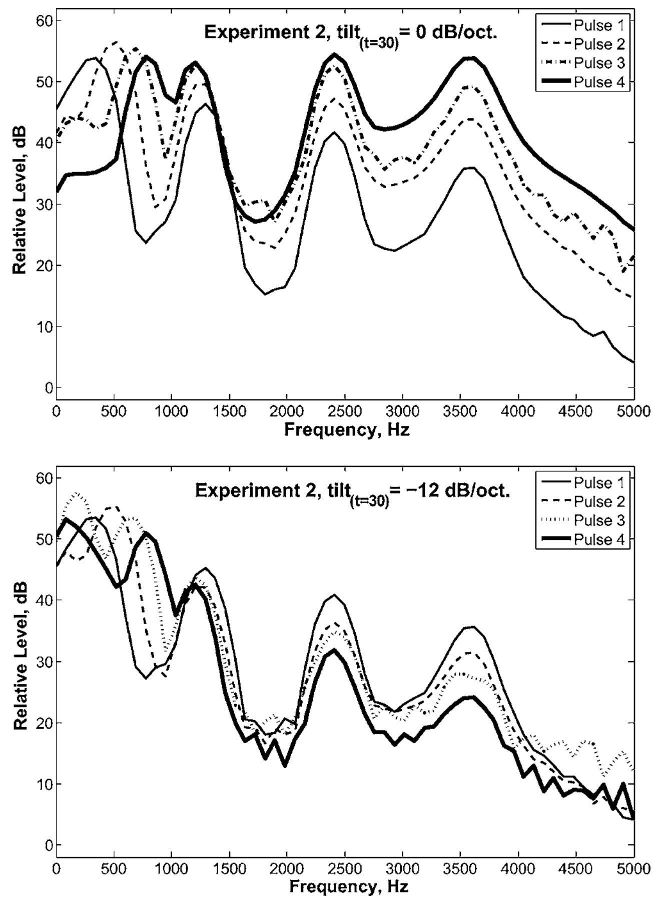

FIG. 7. Short-term spectra for the first four pitch pulses for the tilt endpoint stimuli in Experiment 2. As with Fig. 3, this example represents the stimuli with an F2-onset frequency of 1400 Hz. It is hypothesized that the stimuli in the top panel will lead to more labial responses and that the stimuli in the bottom panel will lead to more alveolar responses despite having identical stimulus onset spectra because the change in tilt is different. For the top panel, tilt becomes shallower until it reaches a flat spectrum that is sustained during the duration of the vowel, whereas, for the bottom panel, tilt becomes steeper until it reaches a steeply negative spectrum for the vowel.

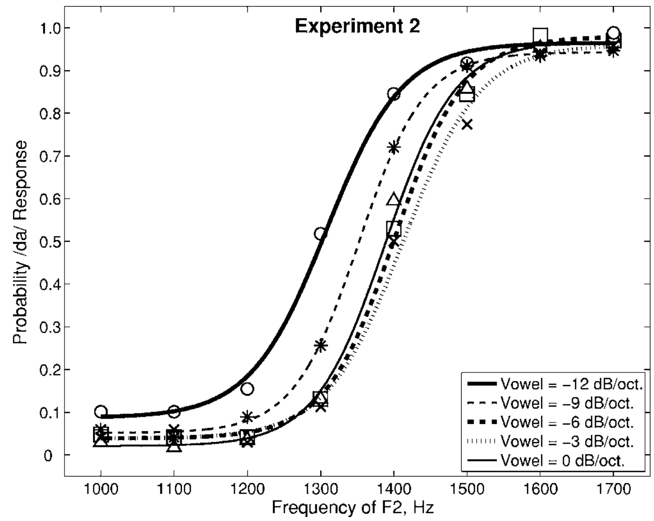

FIG. 8. Mean data for Experiment 2 in which the probability of responding /da/ as a function of F2-onset frequency is plotted separately for each vowel tilt. Circles, asterisks, squares, crosses, and triangles represent the mean data for vowel tilts of −12, −9, −6, −3, and 0 dB/oct., respectively. Maximum-likelihood fits of the identification functions are displayed for the mean data at each vowel tilt as different lines (see the legend).

onset frequencies of 1400 Hz (cf. Fig. 3). Notice that the initial pitch pulses (thin solid lines) in both the top and bottom panels have the same tilt but diverge to two different steady state vowel tilts by the fourth pulse (thick solid line). We hypothesize that manipulation of vowel tilt should lead to a complementary pattern of results as Experiment 1. The top panel shows a relative flattening of spectral tilt (−6 to 0 dB/oct.) from consonant onset (t=0 ms) to vowel steady state (t=30 ms). This pattern of change is predicted to increase the perception of a labial stop consonant. In contrast, the bottom panel is predicted to increase the perception of an alveolar stop consonant because spectral tilt becomes steeper over the course of the consonant transition.

C. Results

Figure 8 displays the mean data and identification functions for each vowel tilt. Circles, asterisks, squares, crosses, and triangles represent the mean data for vowel tilts of −12, −9, −6, −3, and 0 dB/oct., respectively, and the different lines represent the fitted functions for the mean data (see the legend). As with Experiment 1, listeners' perceptions were primarily influenced by onset frequency of F2. However, they were also influenced by tilt of the following vowel, especially at intermediate values of F2. Listeners were more likely to perceive /da/ for steeper vowel tilts. Figure 9 displays the mean frequency and 95% CIs of boundaries at each vowel tilt. With the exception of the boundary for the 0 dB/oct. vowel tilt, there was a systematic increase in the mean boundary frequency (increase in bias for /ba/ responses) with shallower vowel tilts. A within-subjects ANOVA confirmed that the effect of vowel tilt was statistically significant [$F(4,80)=21.7, p<0.0001$]. Tukey HSD post hoc tests revealed that the mean boundary frequency for the −12 dB/oct. vowel tilt was significantly less than the mean boundary frequencies of the other vowel tilts. In addition, the mean boundary frequency for the −9 dB/oct. vowel

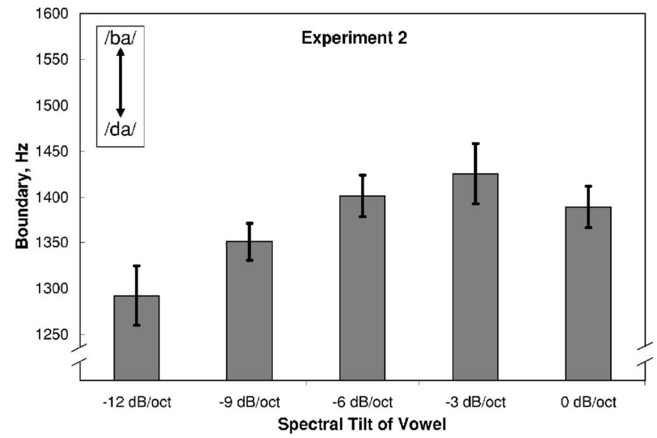

FIG. 9. Mean boundary frequencies of the identification functions at each vowel tilt from Experiment 2. Error bars represent the 95% confidence intervals.

tilt was significantly less than the mean boundary frequencies of the −6 and −3 dB/oct. vowel tilts. None of the differences in mean boundary frequencies for the vowel tilts between −6 and 0 dB/oct. were statistically significant.

D. Discussion of Experiments 1 and 2

The results of Experiments 1 and 2 demonstrate that for burstless, voiced stops, the spectral tilt of the consonant formant transitions is perceived relative to the tilt of the following vowel. Perception of consonant onset tilt is contrastive to the tilt of the following vowel. Relatively *shallow* vowel tilts encourage perception of /ba/ (*steeper* consonant onset tilt) and relatively *steep* vowel tilts encourage perception of /da/ (*shallower* consonant onset tilt). Experiments 1 and 2 also indicate that the influence of relative tilt depends on the ambiguity of formant cues. When F2-onset frequency is near the series end points and most appropriate for /ba/ (1000 Hz) or most appropriate for /da/ (1700 Hz), relative tilt has little or no influence on perception. This result is consistent with results from Dorman and Loizou (1996), who used naturally produced stimuli with very high identification rates. However, when F2-onset frequency is intermediate between [ba] and [da], relative tilt has a substantial effect on stop consonant identification. This result is consistent with Lahiri *et al.* (1984), who used perceptually impoverished stimuli with regard to place of articulation. In summary, perception of the place of articulation in CVs (specifically, /ba/ and /da/) when formant information is ambiguous can be largely influenced by the relative change in spectral tilt.

V. EXPERIMENTS 3 AND 4: EFFECT OF RELATIVE TILT CHANGE ON VCV PERCEPTION

A. Rationale

Experiments 1 and 2 demonstrate the relative and contrastive influence of spectral tilt change as a perceptual cue to stop consonant identification in a CV context without bursts. In contrast to an utterance initial position, stop consonants in medial position are much more common in natural speech and much less likely to have a significant burst release. The next set of experiments were designed to test the influence

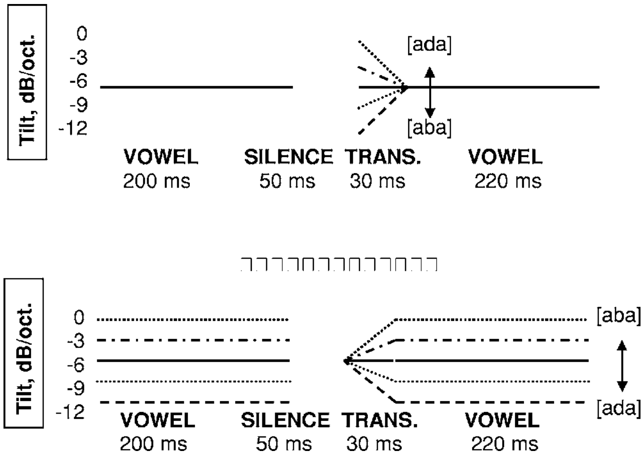

FIG. 10. Schematic for the stimuli in Experiment 3 (top panel) and Experiment 4 (bottom panel). The addition of a preceding vowel should enhance the perceptual cue associated with the spectral tilt change during the consonant transition.

spectral tilt change on the perception of burstless stop consonants in medial position. Specifically, Experiments 3 and 4 tested if effects of tilt are enhanced in a VCV context in which preceding vowels share the same critical acoustic features as the following vowels.

B. Methods

Experiments 3 and 4 replicated Experiments 1 and 2, respectively, using 25 (10 male, 15 female) and 22 (6 males, 16 females) college students who reported normal hearing, respectively. Stimuli were created by appending a 200 ms [a] followed by 50 ms of silence to the stimuli in Experiments 1 and 2. Formant frequencies and spectral tilt of the preceding [a] were matched to the following [a] to enhance the perception of tilt and formant frequency change associated with the consonant onset. As shown in Fig. 10, spectral tilt of the preceding [a] in Experiment 3 was always −6 dB/oct. (top panel) but varied in Experiment 4 from trial to trial in the same way the tilt of the following vowel varied (bottom panel). In a diotic presentation, listeners identified the series of 40 VCVs as /aba/ or /ada/ eight times in separate blocks (320 trials) following two warm-up blocks (80 trials).

C. Results and discussion

Figure 11 displays the mean data and fitted identification functions for Experiment 3, and Fig. 12 displays the mean boundary frequencies and 95% CIs. Compared to the data points in Fig. 4 (CVs), the data points in Fig. 11 (VCVs) show a greater divergence as a function of consonant onset tilt at intermediate F2 frequencies, especially at 1300 Hz. The effect of consonant onset tilt was highly significant [$F(4,96)=64.9$, $p<0.0001$] in a within-subjects ANOVA. Tukey HSD post hoc tests revealed that each paired comparison of the mean boundary frequencies for the different consonant onset tilts was statistically significant except for the mean boundary frequencies of the −3 and 0 dB/oct. consonant onset tilts. Analyses of the mean boundary frequencies across Experiments 1 and 3 reveal that providing an additional comparison for consonant onset tilt in the form of a

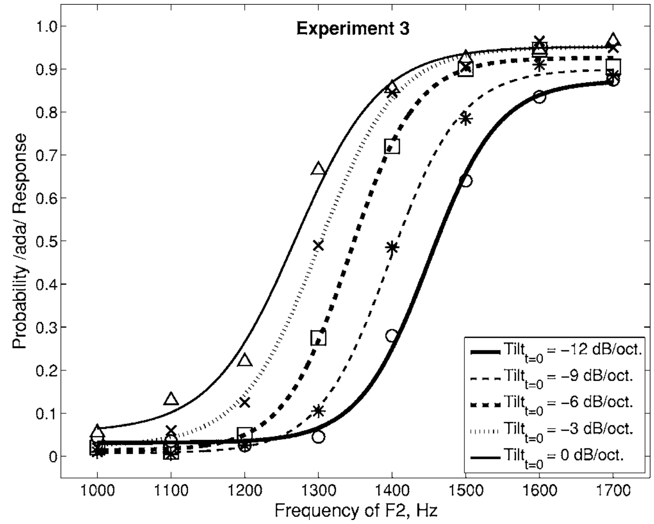

FIG. 11. Mean data and maximum-likelihood fits for Experiment 3 plotted in the same way as Fig. 4 in which the probability of responding /ada/ as a function of F2-onset frequency is plotted separately for each consonant onset tilt.

preceding vowel resulted in significantly negative shifts in the mean boundary frequencies (increase in bias for alveolar responses) for the stimuli with consonant onset tilts of −3 dB/oct. [$t(46)=4.14$, $p=0.0001$] and 0 dB/oct. [$t(46)=2.16$, $p<0.05$].

Figure 13 displays mean data and fitted identification functions for Experiment 4, and Fig. 14 displays the mean boundary frequencies and 95% CIs. Compared to data for Experiment 2, addition of the preceding vowel contrast in Experiment 4 resulted in a significant positive shift in mean boundary frequency (increase in bias for labial responses) for the stimuli with 0 dB/oct. vowel tilts [$t(41)=-2.17$, $p<0.05$]. This is evidenced in Fig. 14, which now shows an orderly increase in labial responses with progressively shallower vowel tilts. The overall effect of vowel tilt in Experiment 4 was significant [$F(4,84)=14.4$, $p<0.0001$] according to a within-subjects ANOVA. Tukey HSD post hoc tests revealed that the mean boundary frequency for the −12 dB/oct. vowel tilt was significantly less than the mean

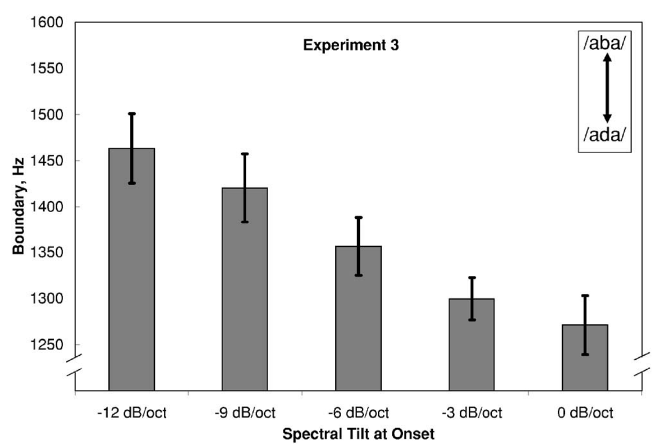

FIG. 12. Mean boundary frequencies of the identification functions at each consonant onset tilt in Experiment 3. Lower-frequency boundaries are associated with a bias toward /ada/ responses. Error bars represent the 95% confidence intervals.

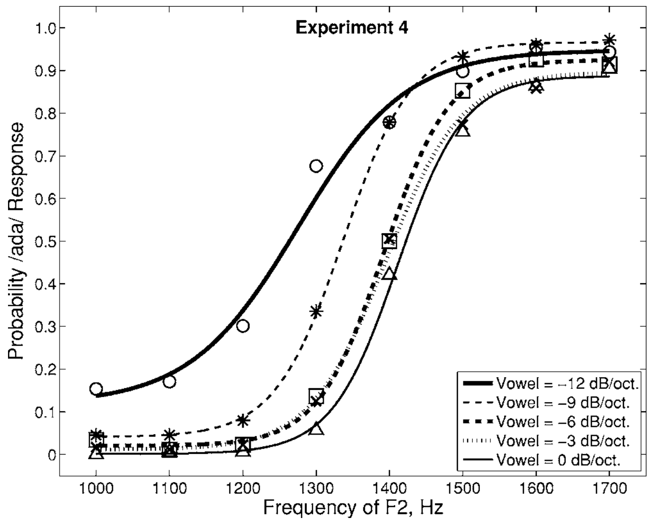

FIG. 13. Mean data and maximum-likelihood fits for Experiment 4 plotted in the same way as Fig. 8 in which the probability of responding /ada/ as a function of F2-onset frequency is plotted separately for each vowel tilt.

boundary frequencies of the stimuli with vowel tilts between −6 and 0 dB/oct. Additionally, the mean boundary frequency for the −9 dB/oct. vowel tilt was significantly less than the mean boundary frequencies of the −3 and 0 dB/oct. vowel tilts.

Results of Experiments 3 and 4 demonstrate the effects of relative spectral tilt change on the perception of medial position stop consonants. As is the case with following vowel tilt, the influence of preceding vowel tilt on the perception of labial versus alveolar voiced stops is contrastive.

VI. EXPERIMENT 5: EFFECTS OF ABSOLUTE TILT

A. Rationale

The above-mentioned experiments establish that relative spectral tilt change from consonant to vowel influences voiced stop consonant perception. The purpose of this final experiment is to test whether the influence of spectral tilt in stop consonant perception depends on a change in tilt or

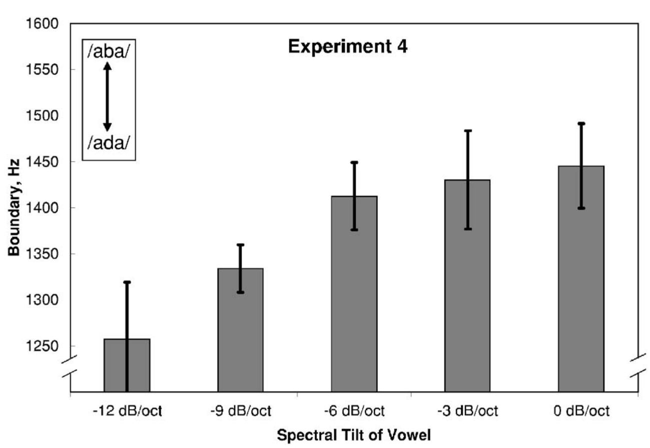

FIG. 14. Mean boundary frequencies of the identification functions at each vowel tilt in Experiment 4. Error bars represent the 95% confidence intervals.

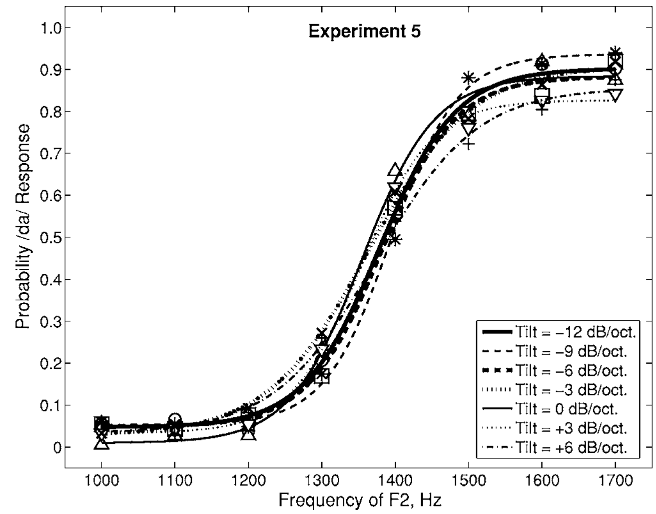

FIG. 15. Mean data for Experiment 5 in which the probability of responding /da/ as a function of F2-onset frequency is plotted separately for each absolute spectral tilt. Circles, asterisks, squares, crosses, triangles, inverted triangles, and plus signs represent the mean data for CVs with absolute tilts of −12, −9, −6, −3, 0, +3, and +6 dB/oct. respectively. Maximum-likelihood fits of the identification functions are displayed for the mean data at each tilt as different lines (see the legend).

whether static differences in consonant onset tilt are enough to influence stop consonant perception (cf. Blumstein and Stevens, 1979).

B. Methods

Twenty-three college students who reported normal hearing (9 male, 14 female) identified a series of 56 CVs that varied from [ba] to [da] along both F2-onset frequency (eight steps) and absolute tilt (constant tilt throughout duration) in seven steps (−12 to +6 dB/oct.). Participants listened to every CV token once per trial block in randomized order. Following two warm-up blocks (112 trials), data were collected on eight subsequent blocks (448 trials). Stimuli were presented diotically at an average level of 73 dBA.

C. Results

Figure 15 displays the mean data and fitted identification

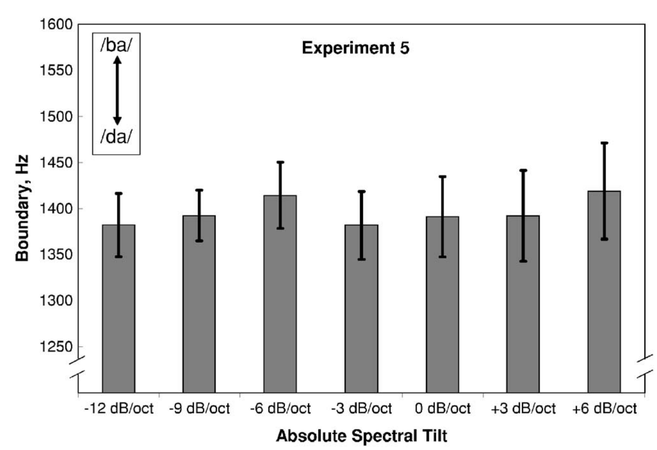

FIG. 16. Mean boundary frequencies of the identification functions at each absolute tilt from Experiment 5. Error bars represent the 95% confidence intervals.

functions for Experiment 5 and Fig. 16 displays the mean boundary frequencies and 95% CIs. From the data, it is clear that absolute tilt, over a wide range, had virtually no effect on listeners' perception $[F(5,132)=1.0]$.[5] From this, we can conclude that consonant onset tilt alone, absent relative change, is not used by listeners to classify stop consonants. Matching the tilt of the following vowel to the tilt of the formant transitions, effectively nullifies the effect of consonant onset tilt on perception.

VII. GENERAL DISCUSSION

Results of the experiments in this report establish that the influence of spectral shape on stop consonant perception is dependent on the availability of other cues to place of articulation, especially formant peaks. When formant information was at the series' end points and most appropriate for [ba] or [da], spectral tilt had little to no influence on identification (cf. Dorman and Loizou, 1996). When formant information was ambiguous and intermediate between [ba] and [da], spectral tilt had a substantial influence on identification (cf. Lahiri *et al.*, 1984). A similar pattern of results was observed by Abramson and Lisker (1985) for the influence of VOT and f_0 on the perception of voicing, with the influence of f_0 maintaining only at ambiguous VOT values. Moreover, the critical feature of spectral tilt is the tilt of the consonant onset relative to the tilt of the following vowel. A negative (steeper) consonant onset tilt relative to vowel tilt encourages perception of a labial place of articulation and a positive (shallower) consonant onset tilt relative to vowel tilt encourages perception of an alveolar place of articulation. Manipulations of consonant onset tilt (Experiments 1 and 3) and vowel tilt (Experiments 2 and 4) are complementary, and either is sufficient to influence perception.

There are several reasons to believe that our results are conservative with respect to natural speech and to the wide inventory of speech sounds and acoustic contexts. First, consider fluent conversational speech in the presence of competing sounds as encountered by a normal-hearing listener. We have already shown that adding preceding context can further enhance the influence of tilt in burstless, voiced stop consonants. Furthermore, articulation is less precise (i.e., formant frequencies for different phonemes are less extreme) in connected speech. Addition of ambient sound sources further undermines resolution of spectral peaks. Second, consider listeners with sensorineural hearing loss (SNHL), for whom spectral detail (e.g., formant peaks) is often compromised. Because only a gross characterization of the spectrum is necessary to encode spectral tilt, it likely takes on greater importance in speech perception by hearing-impaired listeners. In ongoing research, the current experiments have been replicated by listeners with SNHL. For these listeners, spectral tilt can and often does dominate perception of place of articulation even for frequencies near the F2 end points. Finally, our depiction of tilt effects for normal-hearing listeners may be conservative inasmuch as the appearance of mitigated effects near F2-onset end points could be the result of ceiling and floor effects.

There is evidence that our choice of stimuli may also have worked against stronger effects of spectral tilt. First, one can expect that had we extended our manipulations to include bursts, our findings would be an exaggerated version of the present findings because of the additional spectral tilt information provided by the burst. Furthermore, in a series of deleted-cue and conflicting-cue experiments, Smits *et al.* (1996) found that the relative effectiveness of bursts (spectrally gross information, including shape, level, and duration) and formant transitions (spectrally detailed information) in the perception of place of articulation in stop consonants depended on consonant voicing and vowel context. Specifically, the relative influence of spectrally global information in the bursts was more dominant for voiceless stops compared to voiced stops and more dominant for front vowel contexts (e.g., /i/) compared to nonfront vowel contexts (e.g., /a/). Interestingly, the only effect for Dorman and Loizou's (1996) spectral tilt manipulation occurred in the /i/ context, and the effects of Lahiri *et al.* (1984) were strongest for the /i/ context and weakest for the /a/ context. Smits (1996) argued that these findings have less to do with the identity of the vowel and more to do with the reliability (mean differences and variance) of the acoustic information. That is, an acoustic analysis of voiceless stop consonants across the three places of articulation revealed that in the /a/ context, the formant frequencies were more different from one another while bursts were more similar to one another. The situation was opposite for the /i/ context in which the formant frequencies were more similar to one another and the bursts were more different from one another.

Results of these experiments also establish that change in spectral tilt, not absolute tilt, is perceptually effective. When tilt varied over a wide range from trial to trial, but did not change within a stimulus, tilt had no effect on perception (Experiment 5). Further, perception of relative spectral tilt was enhanced when a preceding vowel tilt also contrasted with the consonant onset tilt (Experiments 3 and 4). The idea that acoustic features are encoded relative to one another across time is not new. For example, locus equations, which are linear regression equations that compare the F2 frequency at voicing onset to the F2 frequency of the following vowel steady state, were used by Lindblom (1963) to describe the contextual relationship of F2. Sussman and colleagues (e.g., Sussman *et al.*, 1991) have further developed the concept of locus equations in efforts to define an invariant cue for place of articulation. Our findings with respect to spectrally global information (i.e., tilt) are similar to Sussman's locus equations for spectrally local information in that they both are relational in nature, however, we cannot make any claims of invariance given the restricted set of stimuli employed here. We demonstrated that listeners can and do use the change in spectral tilt as information to classify labial and alveolar voiced stop consonants in an /a/ context, especially when other information specifying place of articulation is absent or ambiguous. However, it could be that, if our general relations were expanded to resemble specific locus equations, then separate equations would need to be calculated for different speakers, vowel contexts, and for voicing/unvoicing just as has been done for locus equations. Such determinations for any model of consonant identification are likely to be influenced by the relative reliability of the acoustic information across contexts and the presence of other acoustic attributes.

Finally, the importance of relative tilt, or tilt contrast, is consistent with multiple demonstrations that perception of coarticulated speech, signaled by formant changes, is facilitated by spectral contrast between local spectral prominences (formants) (e.g., Coady *et al.*, 2003; Holt and Kluender, 2000; Holt *et al.*, 2000; Lotto and Kluender, 1998; Lotto *et al.*, 1997; Kluender *et al.*, 2003). The fact that the perceptual efficacy of spectral tilt is dependent on change is expected given that sensorineural systems, in general, respond predominantly to change relative to what is predictable or does not change (Kluender *et al.*, 2003; Kluender and Alexander, in press; Kluender and Kiefte, 2006).

ACKNOWLEDGMENTS

The authors would like to thank two anonymous reviewers whose comments on an earlier draft proved extremely helpful. The authors also thank Amanda Baum, Rebecca Edds, Tricia Nechodom, and Rebecca ("Hallie") Strauss for their efforts during the data collection process. This research was supported by a grant to J.M.A. from The National Organization for Hearing Research Foundation ("*The 2006 Graymer Foundation Grant in Auditory Science*") and a grant to K.R.K. from the National Institutes of Health, NIDCD (R01DC04072). Portions of this manuscript were written while receiving support from NIDCD T32 DC000013 (BTNRH).

[1]In Klatt's synthesizers (1980, Klatt and Klatt, 1990), spectral tilt is increased or decreased by adjusting a single-pole filter such that the broad high-frequency skirt of the filter alters the overall shape of the spectrum.
[2]The point of this speech production example is not to make a claim of acoustic invariance (cf. Stevens and Blumstein, 1978; 1981; Blumstein and Stevens, 1979; 1980). The important point is that when spectra tilt information is present it can influence perception.
[3]Software version 2.5.41. See http://bootstrap-software.org/psignifit.
[4]Unless otherwise stated, statistical significance is assumed to be $p<0.05$ on a two-tailed test.
[5]One listener who did show an effect for absolute tilt responded /aba/ for all stimuli with a +6 dB/oct. tilt, hence, no boundary frequency could be computed. The boundary frequency for +3 dB/oct., 1717 Hz, was substituted for the missing data point. Unexplainably, increases in absolute tilt resulted in more /aba/ responses for this listener.

Abramson, A. S., and Lisker, L. (**1985**). "Relative power of cues: F_0 shift versus voice timing," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. Fromkin (Academic, Orlando, FL), pp. 25–33.

Blumstein, S. E., Isaacs, E., and Mertus, J. (**1982**). "The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants," J. Acoust. Soc. Am. **72**, 43–50.

Blumstein, S. E., and Stevens, K. N. (**1979**). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," J. Acoust. Soc. Am. **66**, 1001–1017.

Blumstein, S. E., and Stevens, K. N. (**1980**). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," J. Acoust. Soc. Am. **67**, 648–662.

Coady, J. A., Kluender, K. R., and Rhode, W. S. (**2003**). "Effects of contrast between onsets of speech and other complex spectra," J. Acoust. Soc. Am. **114**, 2225–2235.

Dorman, M. F., and Loizou, P. C. (**1996**). "Relative spectral change and formant transitions as cues to labial and alveolar place of articulation," J. Acoust. Soc. Am. **100**, 3825–3830.

Fruchter, D., and Sussman, H. M. (**1997**). "The perceptual relevance of locus equations," J. Acoust. Soc. Am. **102**, 2997–3008.

Halle, M., Hughes, G. W., and Radley, J.-P. A. (**1957**). "Acoustic properties of stop consonants," J. Acoust. Soc. Am. **29**, 107–116.

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., Lotto, A. J., and Kluender, K. R. (**2001**). "Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariance or auditory enhancement?," J. Acoust. Soc. Am. **109**, 764–774.

Kewley-Port, D. (**1983**). "Time-varying features as correlates of place of articulation in stop consonants," J. Acoust. Soc. Am. **73**, 322–335.

Kewley-Port, D., and Luce, P. A. (**1984**). "Time-varying features of initial stop consonants in auditory running spectra: A first report," Percept. Psychophys. **35**, 353–360.

Kiefte, M., and Kluender, K. R. (**2005**). "The relative importance of spectral tilt in monophthongs and diphthongs," J. Acoust. Soc. Am. **117**, 1395–1404.

Kingston, J., and Diehl, R. L. (**1994**). "Phonetic knowledge," Language **70**, 419–454.

Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.

Kluender, K. R., and Alexander, J. M., (in press). "Perception of speech sounds," in *Handbook of the Senses: Audition*, edited by P. Dallos and D. Oertel (Elsevier, London).

Kluender, K. R., Coady, J. A., and Kiefte, M. (**2003**). "Sensitivity to change in perception of speech," Speech Commun. **41**, 59–69.

Kluender, K. R., and Kiefte, M. (**2006**). "Speech perception within a biologically-realistic information-theoretic framework," in *Handbook of Psycholinguistics*, edited by M. A. Gernsbacher and M. Traxler (Elsevier, London), pp. 153–199.

Lahiri, A., Gewirth, L., and Blumstein, S. E. (**1984**). "A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study," J. Acoust. Soc. Am. **76**, 391–404.

Liberman, A. M., Delattre, P. C., and Cooper, F. S. (**1952**). "The role of selected stimulus variables in the perception of unvoiced stop consonants," Am. J. Psychol. **65**, 497–516.

Lindblom, B. (**1963**). "On vowel reduction," Rep. No. 29, The Royal Institute of Technology, Speech Transmission Laboratory, Stockholm, Sweden.

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.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1997**). "Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica)," J. Acoust. Soc. Am. **102**, 1134–1140.

Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode," Psychol. Bull. **92**, 81–110.

Smits, R. (**1996**). "Context-dependent relevance of burst and transitions for perceived place in stops: It's in production, not perception," Proceedings of the Fourth International Conference on Spoken Language, Philadelphia, PA, Vol. **4** (IEEE, Piscataway, NJ), pp. 2470–2473.

Smits, R., ten Bosch, L., and Collier, R. (**1996**). "Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment," J. Acoust. Soc. Am. **100**, 3852–3864.

Stevens, K. N., and Blumstein, S. E. (**1978**). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. **64**, 1358–1368.

Stevens, K. N., and Blumstein, S. E. (**1981**). "The search for invariant acoustic correlates of phonetic features," *Perspectives in the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ), pp. 1–38.

Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (**1991**). "An investigation of locus equations as a source of relational invariance for stop place categorization," J. Acoust. Soc. Am. **90**, 1309–1325.

Walley, A. C., and Carrell, T. D. (**1983**). "Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants," J. Acoust. Soc. Am. **73**, 1011–1022.

Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (**1993**). "*F*0 gives voicing information even with unambiguous voice onset times," J. Acoust. Soc. Am. **93**, 2152–2159.

Wichmann, F. A., and Hill, N. J. (**2001**). "The psychometric function. I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1293–1313.