Spectral motion contrast as a speech context effect

Ningyuan Wang[a] and Andrew J. Oxenham
Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 30 January 2014; revised 11 July 2014; accepted 21 July 2014)

Spectral contrast effects may help "normalize" the incoming sound and produce perceptual constancy in the face of the variable acoustics produced by different rooms, talkers, and backgrounds. Recent studies have concentrated on the after-effects produced by the long-term average power spectrum. The present study examined contrast effects based on spectral motion, analogous to visual-motion after-effects. In experiment 1, the existence of spectral-motion after-effects with word-length inducers was established by demonstrating that the identification of the direction of a target spectral glide was influenced by the spectral motion of a preceding inducer glide. In experiment 2, the target glide was replaced with a synthetic sine-wave speech sound, including a formant transition. The speech category boundary was shifted by the presence and direction of the inducer glide. Finally, in experiment 3, stimuli based on synthetic sine-wave speech sounds were used as both context and target stimuli to show that the spectral-motion after-effects could occur even with inducers with relatively short speech-like durations and small frequency excursions. The results suggest that spectral motion may play a complementary role to the long-term average power spectrum in inducing speech context effects. © *2014 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4892771]

PACS number(s): 43.66.Mk, 43.71.An, 43.71.Rt [VB] Pages: 1237–1245

I. INTRODUCTION

Perceptual systems encode stimuli in a way that is highly dependent on contextual information. Speech is no exception to this general rule, and our perception of individual speech sounds can depend strongly on the context in which they are presented. In a pioneering study, Ladefoged and Broadbent (1957) tested 60 subjects in a word identification task. They observed that altering the first two formants within a context sentence ("Please say what this word is") dramatically changed subjects' identification of the following tests words. For example, a test word was perceived as "bit" by 53 subjects out of 60 when the unfiltered sentence was presented as the context, whereas the same word was perceived as "bet" by 54 of the subjects after the first formant (F1) of the preceding sentence was lowered somewhat. In a later example, Mann (1980) found that ambiguous syllables along a /ga/-/da/ continuum were generally perceived as /ga/ when preceded by the syllable /al/ and were perceived as /da/ when preceded by the syllable /ar/.

Since these early studies, it has been debated whether such context effects are specific to speech, or whether they reflect more general auditory processes. Soon after Mann's study, Fowler (1981) suggested that this "compensation for coarticulation" must reflect speech processes, since subjects' strategy for perceiving vowels was tightly coupled to their strategy for producing them. However, other researchers have since argued that such context effects may reflect more general auditory processes (Diehl *et al.*, 2004). For instance, Lotto and Kluender (1998) observed a smaller but significant effect even when using sine-wave tones or glides corresponding to F3 of /al/ and /ar/ as the precursor, demonstrating that it was not necessary for the precursor to be perceived as speech for context effects to occur. In addition, Lotto *et al.* (1997) found similar context effects in a behavioral study of Japanese quails, suggesting that knowledge of speech was also not necessary. Both these and other studies (e.g., Holt, 2006), have suggested that the average power spectrum of the preceding sound plays a dominant role in determining context effects, and that the effects are contrastive. Summerfield *et al.* (1984) found that listeners were able to identify a flat-spectrum harmonic tone complex as a vowel, if it followed a sound with a similar spectrum, but with components at frequencies corresponding to the first three formants of the vowel omitted. Wang *et al.* (2012) also observed similar effects with cochlear-implant users. Such contrastive effects are common in other sensory modalities (Gibson, 1933), and may reflect the tendency of perceptual systems to normalize or "whiten" the incoming stimuli to improve coding efficiency (e.g., Barlow, 1961; Dean *et al.*, 2008).

Aside from average power spectrum, other stimulus properties may also induce after-effects that may be relevant to speech perception. For instance, both speech and non-speech contexts affect the perception of the fundamental-frequency (F0) contour of lexical tones in a contrastive way: following a context with a higher mean F0, the target syllable is more likely to be identified as a lexical tone starting from a lower F0 and vice versa (Huang and Holt, 2012).

In addition to spectral contrast effects, temporal contrast effects also occur in speech perception (e.g., Diehl and Walsh, 1989; Wade and Holt, 2005). For instance, Wade and Holt (2005) measured the influence of the presentation rate of a preceding sequence of pure tones on the perception of

[a]Author to whom correspondence should be addressed. Electronic mail: wang2087@umn.edu

stimuli generated from a continuum between /ba/ and /wa/, as defined by the duration of formant transitions. They observed that a rapid presentation rate of the preceding pure tones resulted in more /wa/ responses, corresponding to the perception of a longer formant transition, while a slower presentation rate resulted in more /ba/ responses, corresponding to the perception of a shorter formant transition. Thus, contrastive after-effects have been shown in speech in both spectral and temporal domains.

Dynamic spectral changes may also play a role in inducing context effects. In a demonstration with some similarities to the visual-motion after-effect (Gibson, 1933), often referred to as the "waterfall effect," Shu *et al.* (1993) found that preceding glides in the center frequency of narrowband noise induced the perception of spectral motion in the opposite direction, such that a downward sweep, repeated over 2–3 min, caused listeners to hear a stationary noise band as increasing in frequency, and vice versa. Beyond that initial report on the spectral-motion after-effect, little is known concerning the underlying mechanisms, or its relevance to everyday auditory perception. One earlier study (Holt *et al.*, 2000) reported that preceding contexts that included formant transitions had a larger effect on synthesized vowel identification than conditions with only a steady-state spectral context, suggesting that spectral motion may also play a role in speech context effects.

The present study investigates spectral-motion after-effects and their influence on the perception of non-speech and synthesized-speech sounds. The first experiment confirms the presence of spectral-motion after-effects with stimulus durations closer to those approximating speech sounds. The second experiment reports after-effects of spectral motion on perceptual judgments of speech sounds. Finally, the third experiment examines possible trade-offs between average spectrum and spectral motion, using precursors that were designed to more closely resemble speech sounds.

II. EXPERIMENT 1: AUDITORY SPECTRAL-MOTION AFTER-EFFECTS WITH WORD-LENGTH INDUCERS

A. Methods

1. Subjects

Eight (2 males, 6 females) native speakers of American English participated in this experiment and were compensated for their time. Their ages ranged from 18 to 28 years (mean age 23.6 years). They had normal hearing, as defined by audiometric thresholds below 20 dB hearing level (HL) at octave frequencies between 0.25 and 8 kHz.

2. Stimuli

Each trial consisted of a single 500-ms precursor tone, followed by a single 50-ms target tone. The precursor and target were separated by a 50-ms silent gap. All the stimuli were gated on and off with 20-ms raised-cosine ramps. As illustrated in Fig. 1, the precursor was centered in the high (2200 Hz), middle (2000 Hz), or low (1800 Hz) frequency region, and was a rising or falling linear frequency glide, or remained at the same frequency. The combination of three

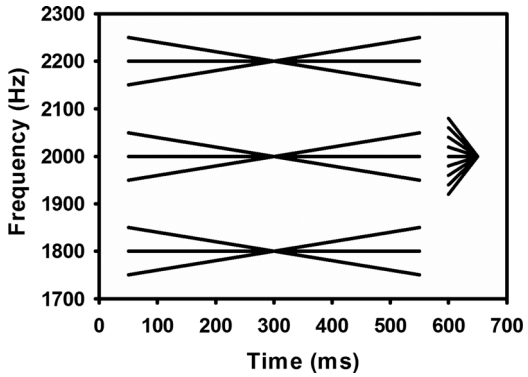

FIG. 1. Schematic diagram of the stimuli used in experiment 1. The precursor, or inducer, was a rising, falling, or steady 500-ms glide that was centered at one of three frequencies. The test stimulus, or target, was a 50-ms tone, selected from one of the rising, falling, or steady lines shown at the right of the figure.

frequency regions and three temporal patterns resulted in a total of nine precursor conditions. The nominal beginning and end frequencies of the precursors are listed in Table I. The nominal beginning frequency of target stimulus was selected from the range between 1920 Hz and 2080 Hz in steps of 20 Hz, and the nominal end frequency was always 2000 Hz. The overall frequency content of both precursor and target was roved together by ±10% across trials, so that the frequency relationship between the precursor and the target remained constant. The rove was designed to discourage listeners from using potential cues based on absolute frequency.

The stimuli were generated digitally and played out diotically from a LynxStudio L22 24-bit soundcard at a sampling rate of 22.5 kHz via Sennheiser HD650 headphones to subjects seated in a double-walled sound-attenuating chamber. The equivalent diffuse-field presentation level for all the sounds was 65 dB sound pressure level (SPL).

3. Procedure

Subjects were asked to judge whether the target tone was "rising" or "falling" and to respond via virtual buttons on the computer display. Prior to the actual experiment, all subjects underwent a training session, during which they were presented with just the target and no precursor. Eight target conditions were tested, including all the target conditions tested in the actual experiment, with the exception of the "flat" target. Each of the conditions was presented 10 times within a block of trials. Feedback was provided during training. In order to progress to the actual experiment, subjects had to achieve at least 80% correct responses on average within 3 blocks in discriminating rising from falling glides. Two of the initial 10 subjects failed to reach this criterion, so only the remaining 8 were tested further. In the actual experiment, all 9 target conditions were tested 10 times each within each block in random order, for a total block length of 90 trials with a single precursor condition. The 10 precursor conditions (9 precursors and 1 no-precursor reference condition) were presented in separate blocks and were repeated 5 times, each in random order, for

TABLE I. Onset and offset frequencies of each precursor condition.

Conditions	No precursor	High-rising	High-flat	High-falling	Middle-rising	Middle-flat	Middle-falling	Low-rising	Low-flat	Low-falling
Onset (Hz)	N/A	2150	2200	2250	1950	2000	2050	1750	1800	1850
Offset (Hz)	N/A	2250	2200	2150	2050	2000	1950	1850	1800	1750

a total of 50 blocks. Thus, each of the 90 conditions (9 target by 10 precursor conditions) was repeated 50 times, and the proportion of "rising" and "falling" responses was calculated for each subject and condition from these 50 responses. No feedback was provided in the test sessions. All subjects provided informed written consent prior to participating, and the experimental protocols were approved by the Institutional Review Board of the University of Minnesota.

B. Results

The mean results are shown in Fig. 2. The left, middle, and right panels show the results using the precursor in the low, middle, and high spectral region, respectively. For comparison, the results from the condition with no precursor are shown as circles in all three panels. Considering first the condition with the precursor in the middle spectral region (Fig. 2, middle panel), it seems that on average the rising precursor led to more "falling" responses, and the falling precursor led to more "rising" responses, relative to the "flat" precursor condition. In other words, the results from the precursor in the middle region are consistent with predictions based on a contrastive spectral-motion after-effect. Similar differences between the falling and rising precursor can be observed in the lower and higher spectral regions (Fig. 2, left and right panels, respectively), although the relationship between those responses and the responses to the flat or no precursor are not so clear cut.

To quantify the effects of the precursor, we used probit analysis to fit each of the curves shown in Fig. 2 for each subject individually. Then we calculated the point at which each curve crossed the 50% point (i.e., the point at which a "falling" response was as likely as a "rising" response), which is termed the "category response boundary." The mean category response boundaries, averaged across subjects, are shown in Fig. 3. A boundary value of 2000 Hz implies that a flat target was perceived veridically; higher boundary values imply that flat targets were more likely to be reported as rising, whereas lower boundary values imply that flat targets were more likely to be reported as falling. The category response boundaries were subjected to a two-way within-subjects analysis of variance (ANOVA), with precursor glide direction (up, down, or flat) and spectral region (low, medium, or high) as the two factors. Significant main effects were observed for both glide direction [$F(2,14) = 5.6$; $p = 0.016$] and frequency region [$F(2,14) = 12.6$; $p = 0.001$], and for their interaction [$F(4,28) = 4.05$; $p = 0.01$]. The main effect of glide direction reflects the trend visible in Fig. 3 that the rising precursor tended to lead to lower boundary values than the falling precursor. *Post hoc* contrast analysis showed that the response boundary in the rising condition was significantly different from that in the falling condition ($p = 0.049$). However, no significant difference was observed between the response boundary in the flat condition and that in either the rising or falling condition. The main effect of spectral region reflects the trend for decreasing boundary value from low to high precursor spectral region. The interaction presumably reflects the impression that the effect of spectral motion seems greater in the middle spectral region than in the low or high region.

C. Discussion

The results from this experiment, showing a rising precursor leading to more "falling" responses, and vice versa, is consistent with the original report of a contrastive spectral-motion after-effect (Shu *et al.*, 1993), and extends the original finding by showing that a relatively short, word-length,

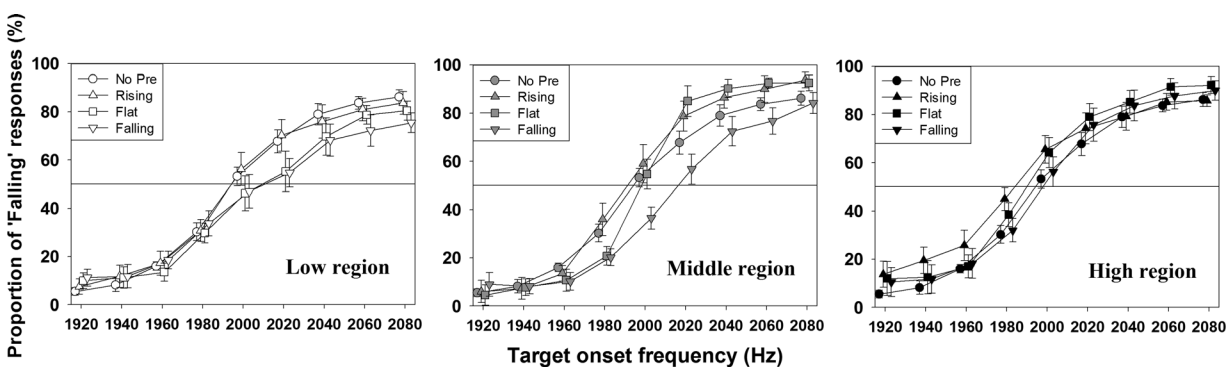

FIG. 2. Psychometric functions showing the average proportion of "falling" responses in percent as a function of the onset frequency of the target glide. The left, center, and right panels show results from the precursors in the low, medium, and high spectral regions, respectively. Upward- and downward-pointing triangles denote conditions with rising and falling precursors, respectively. Squares denote conditions with the constant-frequency (flat) precursors. The same data from the condition with no precursor (circles) are shown in each panel for ease of comparison. Error bars represent 1 standard error (s.e.) of the mean across subjects. The horizontal lines mark the category response boundary of 50% of "falling" responses. Symbols in the different conditions are offset slightly in the horizontal direction for clarity.

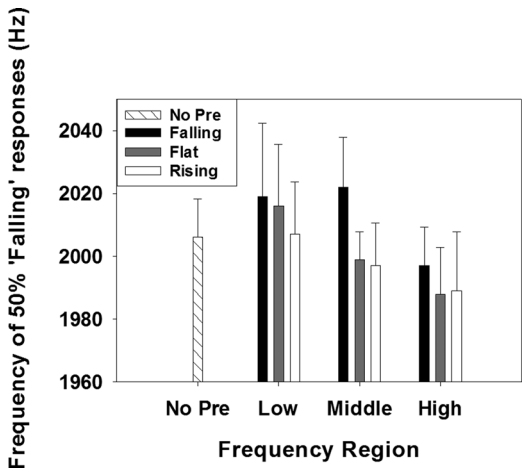

FIG. 3. Mean category response boundary frequencies for each condition. The different bar shadings represent the different precursor motion conditions, as shown in the legend. The results from the three spectral regions are shown in separate groups, as listed along the horizontal axis. Error bars represent 1 s.e. of the mean.

precursor of 500 ms is sufficient to produce a measurable effect. Relatively short spectral motion on this time scale could come from pitch glides in speech, particularly in tone languages, where it has already been shown that F0 contrast effects can be measured (Huang and Holt, 2009).

The effect of spectral region produced an interesting trend, which might be described as "continuity": if the precursor was in the high spectral region, then the target was more likely to be reported as "falling," i.e., moving from the region of the precursor to the center, whereas if the precursor was in the low spectral region, the target was more likely to be reported as "rising." This is the opposite of what would be expected based on spectral contrast, where a high precursor would be expected to lower the perceived beginning of the precursor. One potential reason for why our results are not consistent with expectations based on spectral contrast was that the target consisted of just a short glide, whereas earlier studies have used speech-like sounds that began with a short glide, simulating a formant transition, and ended with a longer steady-state portion. The lack of a steady-state portion at the end of the glide may have reduced the extent to which spectral contrast differentially affected the beginning and end of the target sound.

We have assumed that the differences produced by the rising and falling precursors, particularly in the middle spectral region, are due to their spectral-motion properties. It is clear that the *average* spectrum of the precursor in the middle region cannot explain the effects, as the average frequency of the rising, falling and flat precursors are the same. Nevertheless, it is possible that the results reflect primarily the end frequency of the precursor, rather than spectral motion *per se*. This interpretation is rendered less likely by the fact that the end frequency does not provide a good predictor of all the results. Progressing from the low spectral region to the high, there is a 100-Hz difference between the end frequency of the falling and rising precursor within each spectral region, and between the rising precursor of one spectral region and the falling precursor of the next (going from left to right in Fig. 3, ignoring the flat precursor conditions). Therefore, if the end frequency of each precursor predicted the results, the category response boundary should monotonically (and perhaps linearly) decrease with increasing end frequency. Although this pattern holds within each of the three spectral regions, it does not hold across spectral regions; for instance, going from low-rising to middle-falling leads to an increase in category response boundary, rather than the expected decrease predicted by the end frequency of the precursor. However, the results are somewhat variable, leaving potential room for doubt. In the next experiment we used sine-wave speech targets where the perceived glide direction of a synthetic formant changed the identity of the speech sound. Based on earlier studies, we expected long-term spectral contrast effects to predict the opposite pattern of results from spectral-motion after-effects, thereby making it easier to distinguish between the two.

III. EXPRIMENT 2: SPECTRAL-MOTION AFTER-EFFECTS WITH SYNTHETIC SINE-WAVE SPEECH TARGETS

A. Methods

1. Subjects

Eight (3 males, 5 females) native speakers of American English participated in this experiment and were compensated for their time. Their ages ranged from 18 to 61 years (mean age at 29.3 years). They had normal hearing, as defined by audiometric thresholds below 20 dB HL at octave frequencies between 0.25 and 8 kHz. Three of them had also participated in experiment 1.

2. Stimuli

A synthetic syllable identification task (/ba/-/da/), similar to that of Holt and Lotto (2002), was used, as shown in Fig. 4. Target syllables of 250 ms duration were synthesized with sine waves representing the first three formants. The frequency of the tone representing F1 began at 450 Hz and was swept linearly to 700 Hz over the first 50 ms, where it remained for the final 200 ms. The tone representing F3 remained steady at 2600 Hz. The onset frequency of F2 varied from 800 to 1600 Hz in steps of 100 Hz. During its first 50 ms, the frequency of the F2 tone was swept linearly to 1200 Hz, where it remained for the final 200 ms.

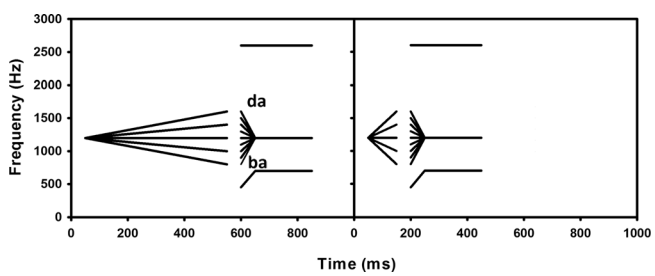

FIG. 4. Schematic diagram of the stimuli used in experiment 2. A 500-ms (left) or 100-ms (right) precursor was followed by a synthetic target syllable that listeners were asked to categorize as either /da/ or /ba/.

Two precursor durations were tested. The first was 500 ms, which was shown to produce a spectral-motion after-effect in experiment 1. The second was 100 ms, which is of a more relevant duration for formant transitions in speech. The precursor frequency always began at 1200 Hz, and was swept linearly over its entire duration to an end frequency that varied parametrically between 800 and 1600 Hz in steps of 200 Hz. The gap between the precursor and target was always 50 ms.

3. Procedure

Initially, subjects took part in a training session in which no precursor was presented. Only the two endpoints of the /ba/-/da/ continuum were presented, with beginning F2 frequencies of 800 and 1600 Hz for /ba/ and /da/, respectively. Subjects were required to achieve at least 80% correct identification in the training phase in order to proceed to the test phase. Within each of the training blocks, there were 20 repetitions for each of the two target conditions, resulting in 40 trials per block. All eight subjects passed the training phase. In the test phase, there were 5 repetitions of each of the 9 targets (4 falling, 4 rising, 1 flat) per block (45 trials). Each precursor was tested in a separate block and each of these blocks was presented 8 times for a total of 80 blocks, and 3600 total trials per subject (40 per subject and condition). All conditions were presented in random order, selected independently for each subject. Feedback was provided for the training blocks, but not during the test blocks.

B. Results

The results of experiment 2 are presented in Figs. 5 and 6, using the same format as those of experiment 1. Figure 5 shows the average identification curves in terms of proportion of /ba/ responses as a function of the F2 onset frequency with the preceding glides for both long (left panel) and short (right panel) precursor conditions. As with the results from experiment 1, a probit analysis was performed using the psychometric functions from the individual listeners to derive a 50% category response boundary for each listener and condition. A one-way within-subjects ANOVA was conducted, with the frequency at the category response boundary as the dependent variable, and precursor glide slopes (difference between start and end points of 400, 200, 0, −200, −400 Hz) as the factor for both long and short precursor conditions separately (Fig. 6). No significant main effect of precursor was observed for 500-ms precursor conditions [$F(4,28) = 0.388$; $p = 0.815$]. However, a significant main effect was found for the 100-ms precursor conditions [$F(4,28) = 2.85$; $p = 0.042$]. In pairwise comparison contrast tests, the short + 400 condition differed significantly from the short −200 ($p = 0.013$) and short −400 ($p = 0.04$) conditions. A further contrast analysis revealed a significant linear trend [$F(1,7) = 6.03$; $p = 0.044$], confirming that there was a systematic trend for increasing boundary value with decreasing slope value of the precursor.

C. Discussion

The main finding from experiment 2 is the existence of a spectral-motion after-effect using a synthesized sine-wave speech sound as a target. The effect found with the shorter precursor is in the same direction as the spectral-motion after-effect found in experiment 1: a falling precursor glide led to a greater proportion of responses corresponding to the rising target, which in this case corresponds to the syllable /ba/. Thus, as in experiment 1, the results are consistent with a contrastive after-effect of spectral motion.

It is not clear why the shorter, but not the longer, precursor resulted in a measurable after-effect. The frequency excursion of the longer, 500-ms precursor, relative to that of the target, was similar to what was used in experiment 1, also with a 500-ms precursor. However, there are also multiple differences between the two experiments. First, the nature of the task was different, with glide direction identification in experiment 1, compared with consonant identification in experiment 2. Second, there were large differences in the stimuli, including a much wider range of frequency excursion for both the precursor and the target (+/−4 % of the end frequency in experiment 1, compared to +/− 33% of the end frequency in experiment 2), a lack of overall frequency roving in experiment 2, and the addition of F1 and F3 in experiment 2. Third, the long-term (average) spectrum of the three precursors was the same in experiment 1, whereas in experiment 2 the rising precursor was higher, and the falling precursor was lower, in average spectrum than the steady (flat) precursor. The difference in average spectrum might have counteracted part of the effects of spectral motion, through a spectral contrast effect, whereby a higher average spectrum would be expected to lead to more /ba/

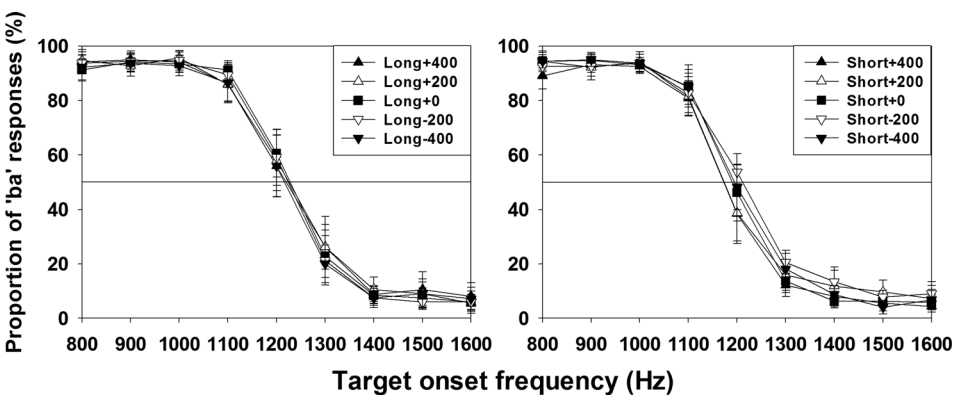

FIG. 5. Proportion of trials identified as /ba/ as a function of the F2 onset frequency. Left and right panels show results from long and short precursor conditions, respectively. Error bars represent 1 s.e. of the mean across subjects. The horizontal lines mark 50% of /ba/ responses. Numbers in the legend represent the frequency difference between the beginning and end of the precursor glide, with negative numbers indicating a falling glide and positive numbers indicating a rising glide.

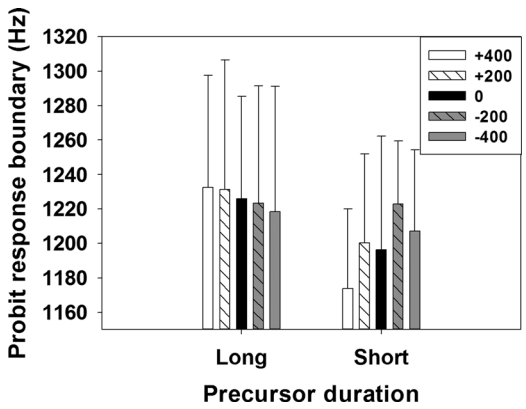

FIG. 6. Mean 50% boundary response frequencies from experiment 2. Numbers in the legend represent the frequency difference between the onset and offset frequency of the precursor, with negative numbers indicating a falling glide and positive numbers indicating a rising glide. Error bars represent 1 s.e. of the mean across subjects.

responses. Thus, the effects of long-term spectrum may have reduced the (opposite) effect produced by spectral motion.

Overall, the results suggest that spectral-motion after-effects can affect speech category boundaries. The outcome cannot be easily explained in terms of long- or short-term average spectrum of the precursor, and instead seems to reflect genuine spectral-motion after-effects. It is possible that similar phenomena could arise with speech sounds, not just artificial glides, as context or precursors. In the final experiment, materials based on synthetic sine-wave speech sounds were used as both context and target stimuli, to investigate whether relatively small formant transitions could themselves produce spectral-motion after-effects in speech, beyond long-term spectral contrasts.

IV. EXPERIMENT 3: SPECTRAL-MOTION AFTER-EFFECTS WITH SYNTHESIZED SPEECH CONTEXT AND TARGET

A. Methods

1. Subjects

Eight (2 males, 6 females) native speakers of American English participated in this experiment and were compensated for their time. Their ages ranged from 18 to 61 years (mean age at 28.1 years). They had normal hearing, as defined by audiometric thresholds below 20 dB HL at octave frequencies between 0.25 and 8 kHz. Five of them had also participated in experiment 2.

2. Stimuli

The same nine target stimuli were used as in experiment 2. A total of 7 different precursors were used. All precursors

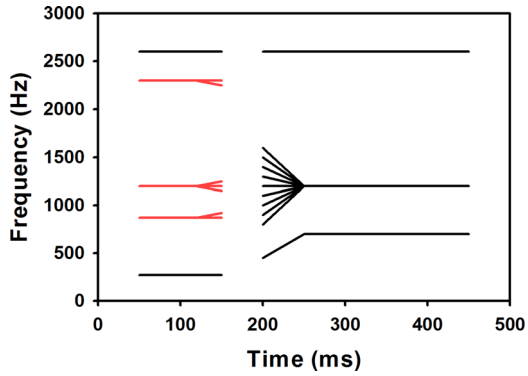

FIG. 7. (Color online) Schematic diagram of stimuli used in experiment 3.

consisted of three 100-ms tones, resembling formant frequencies. The lowest and highest tones remained constant at 870 and 2300 Hz, respectively. The middle tone began at one of three frequencies, as shown in Table II and illustrated in Fig. 7. The final 30 ms of the middle precursor tone was either constant, or was a linear rising or falling sweep. Two of the precursors were designed to resemble the speech sounds /i/ and /u/, as indicated in Table II. The others were variations of these speech sounds that were designed to test the relative importance of the average spectrum and the sweep direction. The precursor and the target were separated by a 50-ms silent gap.

3. Procedure

The subjects first completed the same training session that was used in experiment 2. They were again required to achieve at least 80% correct in the /ba/-/da/ identification task before progressing to the test phase. Again, all eight subjects passed the training phase. In the test phase, each block tested a single precursor and each of the 9 targets was presented 5 times in random order. Each of the 7 precursor conditions was tested in 8 blocks for a total of 56 blocks, and 40 (8 × 5) repetitions of each condition per subject. As in the previous experiments, feedback was provided only in the training phase, and not in the test phase.

B. Results

The results from experiment 3 are shown in Fig. 8. Again, probit analysis was used to derive the 50% point of the psychometric functions for each subject in each condition. These category response boundaries, averaged across subjects, are shown in Fig. 9 and were used as the dependent variable in two separate ANOVAs. Consider first the conditions where the precursor F2 was in the highest or lowest spectral region (Fig. 8, left panel). A two-way repeated-measures ANOVA with main factors of spectral location (high or low)

TABLE II. Stimulus conditions from experiment 3. The upper and lower rows show the onset and offset frequencies of F2 for each precursor condition separately.

Conditions	High-flat	High-falling (/i/)	Middle-rising	Middle-flat	Middle-falling	Low-rising (/u/)	Low-falling
Onset (Hz)	2300	2300	1200	1200	1200	870	870
Offset (Hz)	2300	2250	1250	1200	1150	920	870

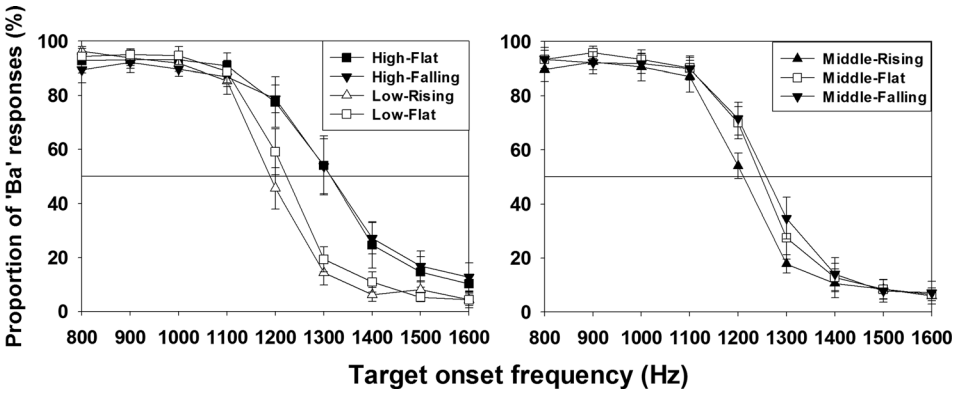

FIG. 8. Proportion of trials identified as /ba/, as a function of the target F2 onset frequency. The left panel shows responses in conditions where the precursor F2 was in the high or low spectral region. The right panel shows responses in conditions where the precursor F2 was in the middle spectral region. Error bars represent 1 s.e. across subjects. The horizontal lines mark 50% of /ba/ responses.

and spectral motion (moving or steady) showed a significant main effect of spectral location [$F(1,7) = 9.11$, $p = 0.02$] but no effect of spectral motion [$F(1,7) = 1.53$, $p = 0.26$], and no interaction [$F(1,7) = 0.036$, $p = 0.86$]. Thus, in cases where the spectral motion in the precursor was remote from the target, no effect of spectral motion was found. Instead, an effect of spectral contrast was observed, with the high-frequency precursor resulting in the target formant being perceived as beginning from a lower frequency, and the low-frequency precursor resulting in the target formant being perceived as beginning from a higher frequency.

Consider next the three conditions with the precursor F2 in the central spectral region (Fig. 8, right panel). A one-way repeated-measures ANOVA found a significant main effect of spectral motion [$F(2,14) = 6.9$, $p = 0.008$]. Pairwise comparisons revealed that the precursor with the rising formant transition produced a significantly different boundary than the precursor with the falling formant transition ($p = 0.025$). Similarly, a contrast analysis revealed a significant linear trend, confirming that the boundary value increased with decreasing slope value [$F(1,7) = 13.19$, $p = 0.008$]. Thus, for the precursor at the target frequency, a small but contrastive spectral-motion after-effect was observed.

C. Discussion

When the precursor was in the same spectral region as the target, a small but significant spectral-motion after-effect was observed, even though the precursor motion was as short as that of the target itself, and the frequency excursions of the precursor were actually smaller than that of the smallest frequency excursion of the target. The lack of an effect of spectral motion with the low- and high-region precursors suggests that, as in experiment 1, the strongest effects of spectral motion are observed when the precursor and target fall in the same spectral region. Again, as in the previous two experiments, the results cannot be easily explained in terms of the average long- or short-term spectrum of the precursor; in fact, as in experiment 2, any averaging of the precursor would lead to predictions in the opposite direction of those observed in the results, with the higher precursor predicted to produce more rising responses. Thus, as in the previous experiments, the outcome is more easily explained in terms of sensitivity to spectral motion *per se*.

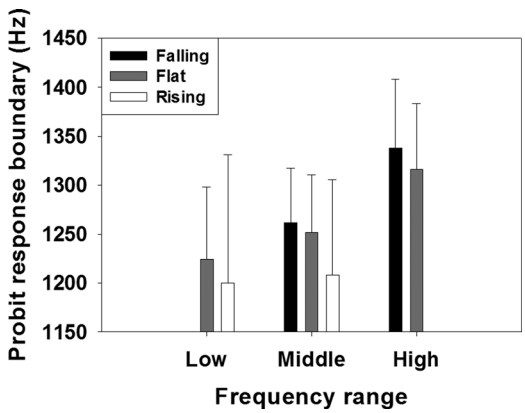

FIG. 9. Mean category response boundaries, at which /ba/ and /da/ responses are equally likely. The *x* axis indicates the frequency region of the precursor F2. Different shaded bars represent the different spectral motion of the patterns of formant transition. Error bars represent 1 s.e. of the mean across subjects.

Comparing the results from the low- and high-region precursors, the effect is similar to that observed in earlier studies (Holt and Lotto, 2002; Holt, 2006) in that the higher precursor led to the report of more /ba/ responses, corresponding to a lower perceived starting frequency of the target glide, and the lower precursor led to the report of more /da/ responses, corresponding to a higher perceived starting frequency of the target glide. Note that this outcome, although in line with earlier studies, is not consistent with the results from experiment 1, where a "continuity" effect was observed. As mentioned in the discussion of experiment 1, one explanation for this apparent discrepancy relates to the nature of the target stimulus: in earlier studies (and in the current experiment), the target remained at a constant frequency after the initial glide, whereas in experiment 1 the target consisted only of a brief glide. It may be necessary to have a longer target (or stable end frequency) for the precursor to have a differential effect on the beginning and end of the target. Another difference between the experiments is that the target glide in experiment 3 led to the perception of one of two speech categories, /ba/ and /da/, for which subjects have established long-term category representations. Further experiments will be required to test the possibility that the nature of the target can affect the direction of the context effects.

The duration and size of the precursor's spectral motion were chosen in this experiment to be representative of the motion found in speech. Therefore, the fact that spectral-motion after-effects were found suggests that they may also

play some role in more natural situations involving speech perception.

V. GENERAL DISCUSSION

The three experiments presented here all provide evidence that spectral motion on frequency and time scales that are relevant to speech can produce contrastive after-effects. The after-effects are relatively small in absolute terms, but can be induced with surprisingly small frequency excursions and short precursor durations. The after-effects also appear to be spectrally local, in that precursor glides that are remote in frequency from the target do not produce significant after-effects. The finding that the effects are spectrally local can be compared to the conclusions drawn from an earlier study of temporal contrast effects. As mentioned earlier, Wade and Holt (2005) found that a sequence of pure tones presented at a rapid rate resulted in more /wa/ responses, whereas a sequence presented at a slower rate resulted in more /ba/ responses to the target. The results were therefore consistent with a temporal contrast effect, in which a faster precursor rate led listeners to judge the following transitions as slower. When comparing their effect with a null result reported in earlier study by Summerfield (1981), Wade and Holt (2005) suggested that one important difference might have been the lack of spectral and temporal continuity between the relevant precursor dimensions and the target.

It is important to establish whether the spectral-motion after-effect is in fact mediated by spectral motion, rather than the long- or short-term spectrum of the precursor. In this respect the results from all three experiments converge to suggest that it is the spectral motion *per se*, rather than the spectrum that determines the effect, as outlined below.

In the first experiment, the average spectra of all the precursors were the same. If one assumes that only the final part of the precursor contributes to the aftereffect, then the direction of the after-effect was the same as that predicted by just the spectrum, as illustrated by the effect of the precursors in the higher and lower spectral regions. However, as discussed in experiment 1, considering just the end points of the precursor frequency cannot explain why the rising precursor in the low spectral region produced a lower category response boundary than the falling precursor in the middle spectral region. Instead, the overall pattern of results are more consistent with an explanation based on spectral motion within the local spectral region of the target.

In the second experiment, all the precursors began at the same frequency, so that the precursor with the upward spectral motion also had a higher long-term (and short-term) spectrum than the precursor with the downward spectral motion. In this case, spectral-motion contrast predicts that an upward precursor glide should lead to more perceived downward target glides and a downward precursor glide should lead to more perceived upward target glides, consistent with the obtained data, whereas an explanation based on simple spectral contrast predicts the opposite. Therefore, in this case it is clear that an explanation based on spectral-motion contrast provides a better account of the data.

In the third experiment, spectral-motion contrast and simple spectral contrast make opposite predictions. Presenting the precursor in different spectral regions resulted in outcomes consistent with spectral contrast, whereas the precursor in the spectral region of the target produced results consistent with the predictions of spectral-motion contrast. Thus, a parsimonious explanation of all three experiments is that the spectral motion of the precursor can induce after-effects beyond those predicted by the long-term (or short-term) spectrum.

Having established the existence of a spectral-motion after-effect that may be relevant for speech perception, a next step is to determine the underlying mechanisms. Just as spectral-contrast context effects could be explained in terms of neural adaptation or forward suppression of frequency-selective cortical and/or sub-cortical neurons, spectral-motion after-effects could be explained in terms of adaptation or forward suppression of neurons that are tuned to the direction of spectral motion. Such neurons have been identified in other mammals (Weinberger and Mckenna, 1988; McKenna *et al*., 1989; Brosch and Schreiner, 2000). In addition, there are other psychophysical results involving tone detection and discrimination experiments that have led researchers to propose the presence of "pitch-shift detectors" (Demany and Ramos, 2005; Demany *et al*., 2009), which could also be invoked to explain the results of the present experiment. Further studies could explore in more detail the parametric effects of precursor duration and rate of frequency change to better define the nature of these hypothetical frequency glide detectors.

VI. SUMMARY

This study explored the potential role of spectral motion in inducing context effects in non-speech and synthesized-speech stimuli. Experiment 1 confirmed the existence of a contrastive spectral-motion after-effect in judging the motion of a target tone glide, and extended previous findings by showing that significant after-effects could be produced using a relatively short (500-ms) inducer.

Experiment 2 found that the glide direction of a shorter inducer, of only 100 ms, could influence phonemic judgments along a /ba/-/da/ continuum in a way that was also consistent with spectral-motion contrast, although the longer 500-ms precursor had no significant influence on the phonemic judgments.

In experiment 3 a precursor was constructed with three tones to resemble the formant structure of two vowels, /i/ and /u/, along with more artificial variants. Consistent with previous studies, the long-term spectrum of the precursor affected judgments of artificial stimuli constructed along the /ba/-/da/ continuum. In addition, when the spectral motion of the precursor was in the same spectral region as the formant transition of the target, a contrastive spectral-motion after-effect was observed.

Overall, the results demonstrate that spectral motion can induce changes in the responses to both non-speech and speech-like stimuli, and suggest that spectral-motion

after-effects may play a role in more natural situations involving speech perception.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. R01 DC012262. N.W. was supported by Advanced Bionics.

Barlow, H. B. (**1961**). "Possible principles underlying the transformations of sensory message," in *Sensory Communication*, edited by W. Rosenblith (MIT Press, Cambridge, MA), pp. 217–234.

Brosch, M., and Schreiner, C. E. (**2000**). "Sequence sensitivity of neurons in cat primary auditory cortex," Cereb. Cortex **10**, 1155–1167.

Dean, I., Robinson, B. L., Harper, N. S., and McAlpine, D. (**2008**). "Rapid neural adaptation to sound level statistics," J. Neurosci. **28**, 6430–6438.

Demany, L., Pressnitzer, D., and Semal, C. (**2009**). "Tuning properties of the auditory frequency-shift detectors," J. Acoust. Soc. Am. **126**, 1342–1348.

Demany, L., and Ramos, C. (**2005**). "On the binding of successive sounds: perceiving shifts in nonperceived pitches," J. Acoust. Soc. Am. **117**, 833–841.

Diehl, R. L., Lotto, A. J., and Holt, L. L. (**2004**). "Speech perception," Ann. Rev. Psychol. **55**, 149–179.

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.

Fowler, C. A. (**1981**). "Production and perception of coarticulation among stressed and unstressed vowels," J. Speech Hear. **24**, 127–139.

Gibson, J. J. (**1933**). "Adaptation, after-effect and contrast in the perception of curved lines," J. Exp. Psychol. **16**, 1–31.

Holt, L. L. (**2006**). "The mean matters: Effects of statistically defined non-speech spectral distributions on speech categorization," J. Acoust. Soc. Am. **120**, 2801–2817.

Holt, L. L., and Lotto, A. J. (**2002**). "Behavioral examinations of the level of auditory processing of speech context effects," Hear. Res. **167**, 156–169.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**, 710–722.

Huang, J., and Holt, L. L. (**2009**). "General perceptual contributions to lexical tone normalization," J. Acoust. Soc. Am. **125**, 3983–3994.

Huang, J., and Holt, L. L. (**2012**). "Listening for the norm: Adaptive coding in speech categorization," Front. Psychol. **3**, 10.

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

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.

Mann, V. A. (**1980**). "Influence of preceding liquid on stop-consonant perception," Percept. Psychophys. **28**, 407–412.

McKenna, T. M., Weinberger, N. M., and Diamond, D. M. (**1989**). "Responses of single auditory cortical neurons to tone sequences," Brain Res. **481**, 142–153.

Shu, Z. J., Swindale, N. V., and Cynader, M. S. (**1993**). "Spectral motion produces an auditory after-effect," Nature **364**, 721–723.

Summerfield, Q. (**1981**). "Articulatory rate and perceptual constancy in phonetic perception," J. Exp. Psychol. Hum. Percept. Perform. **7**, 1074–1095.

Summerfield, Q., Haggard, M., Foster, J., and Gray, S. (**1984**). "Perceiving vowels from uniform spectra: Phonetic exploration of an auditory after-effect," Percept. Psychophys. **35**, 203–213.

Wade, T., and Holt, L. L. (**2005**). "Perceptual effects of preceding non-speech rate on temporal properties of speech categories," Percept. Psychophys. **67**, 939–950.

Wang, N., Kreft, H., and Oxenham, A. J. (**2012**). "Vowel enhancement effects in cochlear-implant users," J. Acoust. Soc. Am. **131**, EL421–EL426.

Weinberger, N. M., and Mckenna, T. M. (**1988**). "Sensitivity of single neurons in auditory-cortex to contour—Toward a neurophysiology of music perception," Mus. Percept. **5**, 355–389.