Journal of Experimental Psychology: Human Perception and Performance

Asymmetric Discrimination of Nonspeech Tonal Analogues of Vowels

Matthew Masapollo, T. Christina Zhao, Lauren Franklin, and James L. Morgan

Online First Publication, December 20, 2018. http://dx.doi.org/10.1037/xhp0000603

American Psychological Association

Journal of Experimental Psychology:
Human Perception and Performance

0096-1523/18/$12.00

http://dx.doi.org/10.1037/xhp0000603

Asymmetric Discrimination of Nonspeech Tonal Analogues of Vowels

Matthew Masapollo
Brown University and Boston University

T. Christina Zhao
University of Washington

Lauren Franklin and James L. Morgan
Brown University

Directional asymmetries reveal a universal bias in vowel perception favoring extreme vocalic articulations, which lead to acoustic vowel signals with dynamic formant trajectories and well-defined spectral prominences because of the convergence of adjacent formants. The present experiments investigated whether this bias reflects speech-specific processes or general properties of spectral processing in the auditory system. Toward this end, we examined whether analogous asymmetries in perception arise with nonspeech tonal analogues that approximate some of the dynamic and static spectral characteristics of naturally produced /u/ vowels executed with more versus less extreme lip gestures. We found a qualitatively similar but weaker directional effect with 2-component tones varying in both the dynamic changes and proximity of their spectral energies. In subsequent experiments, we pinned down the phenomenon using tones that varied in 1 or both of these 2 acoustic characteristics. We found comparable asymmetries with tones that differed exclusively in their spectral dynamics, and no asymmetries with tones that differed exclusively in their spectral proximity or both spectral features. We interpret these findings as evidence that dynamic spectral changes are a critical cue for eliciting asymmetries in nonspeech tone perception, but that the potential contribution of general auditory processes to asymmetries in vowel perception is limited.

Public Significance Statement

The present research investigated the extent to which directional asymmetries in vowel perception may reflect general auditory processes by examining discrimination of nonspeech tones that approximate certain spectro-temporal properties of vowel sounds, but are not explicitly recognized as speech. Specifically, we examined the relative contribution of 2 key acoustic properties hypothesized to differ between vowel signals generating asymmetries in phonetic discrimination tasks: the dynamics and proximity of spectral energies. The findings demonstrate that qualitatively similar but weaker asymmetries emerge only with tones varying in their spectral dynamics. Although these findings suggest that asymmetries in nonspeech tone perception may reflect a sensitivity to dynamic changes in spectral energies, the potential role of general auditory processes on asymmetries in vowel perception is limited.

Keywords: speech perception, Natural Referent Vowel framework, focal vowels, auditory perception, auditory cognitive science

It is well-known that we experience a biased sense of the world, where some stimuli are more perceptually prominent or salient than others. Such differences in salience are often experimentally demonstrated as directional asymmetries in discrimination tasks,

Matthew Masapollo, Department of Cognitive, Linguistic and Psychological Sciences, Brown University, and Department of Speech, Language, and Hearing Sciences, Boston University; T. Christina Zhao, Institute for Learning and Brain Sciences, University of Washington; Lauren Franklin and James L. Morgan, Department of Cognitive, Linguistic and Psychological Sciences, Brown University.

The research reported here was supported by National Institutes of Health (NIH) Grant R01 HD068501 (James L. Morgan, principal investigator) and the Ready Mind Project at the University of Washington's Institute for Learning and Brain Sciences. Matthew Masapollo was also supported by NIH Grant R01 DC002852 (Frank H. Guenther, principal investigator) during the preparation and revision of the article. We are grateful to Ellen Macaruso, Leah Mann, and Lori Rolfe at Brown University for assistance with subject recruitment and data-collection. This work benefited from helpful discussions with, or comments from, Linda Polka, Navin Viswanathan, Frank H. Guenther, Stephen Politzer-Ahles, and Elaine Kearney.

Correspondence concerning this article should be addressed to Matthew Masapollo, Department of Speech, Language, and Hearing Sciences, Boston University, 677 Beacon Street, Boston, MA 02215. E-mail: mmasapol@bu.edu

where A is confused more frequently with B than B is with A (Medin & Barsalou, 1987; Miller & Nicely, 1955; Rosch, 1975). Asymmetrical perceptual relations are widespread in human perception and cognition, having been well-documented in a wide range of stimulus domains including with speech, face, color, and music stimuli (see Polka & Bohn, 2003, for discussion). In the speech realm, among other phenomena, listeners (both adult and infant) tend to perform better at discriminating many vowel contrasts presented in one order compared with the same change presented in the reverse order (Polka & Bohn, 2003, 2011). These findings have been reviewed and discussed extensively by Polka and Bohn (Bohn & Polka, 2014; Polka & Bohn, 2003, 2011; Polka, Bohn, & Weiss, 2015), and recently compiled in several meta-analyses (Polka, Ruan, & Masapollo, 2018; Tsuji & Cristia, 2017). While the existence of such directional effects has been clearly established, attention has turned toward increasing our understanding of the *stimulus properties* and *perceptual processes* that underlie them. The issue we address in the present research concerns the nature of the information in speech that contributes to driving these asymmetrical discrimination patterns.

Polka and Bohn (2003), in their first review of asymmetries, noted that in general such effects could be predicted by considering the relative positions of the contrasting vowels within traditional acoustic vowel space (as defined by *F*1–*F*2). Specifically, listeners tend to perform better at discriminating a change from a relatively less peripheral vowel (e.g., /e/) to a relatively more peripheral vowel (e.g., /i/), compared with the same change presented in the reverse direction. This perceptual pattern is summarized in Figure 1A, which shows many of the vowel contrasts that have been examined in infant vowel discrimination studies with arrows indicating the direction of change that was reported to be easier to discriminate (see Polka & Bohn, 2003, 2011, for the list of studies these results are based on). In early infancy, these asymmetries have been reported to occur during the discrimination of both native and nonnative (foreign language) vowel contrasts, indicating that they do not derive from specific linguistic experience. In adulthood, analogous asymmetry effects emerge most clearly during the discrimination of within-category or nonnative vowel contrasts (Dufour, Brunellière, & Nguyen, 2013; Polka & Bohn, 2011; Pons, Albareda-Castellot, & Sebastian-Gallés, 2012; Tyler, Best, Faber, & Levitt, 2014). Thus, the evidence indicates that listeners are universally sensitive to the relative position of vowels within acoustic space, although long-term linguistic experience also clearly plays a significant role in modulating perception.

On the basis of these and subsequent findings, Polka and Bohn offered a theoretical framework termed the Natural Referent Vowel (NRV) framework (Polka & Bohn, 2011), which focused on explicating the processes underlying asymmetries, as well as the nature of the information that those processes operate on. Another focus was on how factors such as phonological working memory capacity, attention, and particular task demands interact to influence asymmetries. This framework has been used to guide a number of recent studies of vowel perception with both adults and infants (see, e.g., Kriengwatana & Escudero, 2017; Masapollo, Franklin, Morgan & Polka, 2018; Masapollo, Polka, & Ménard,

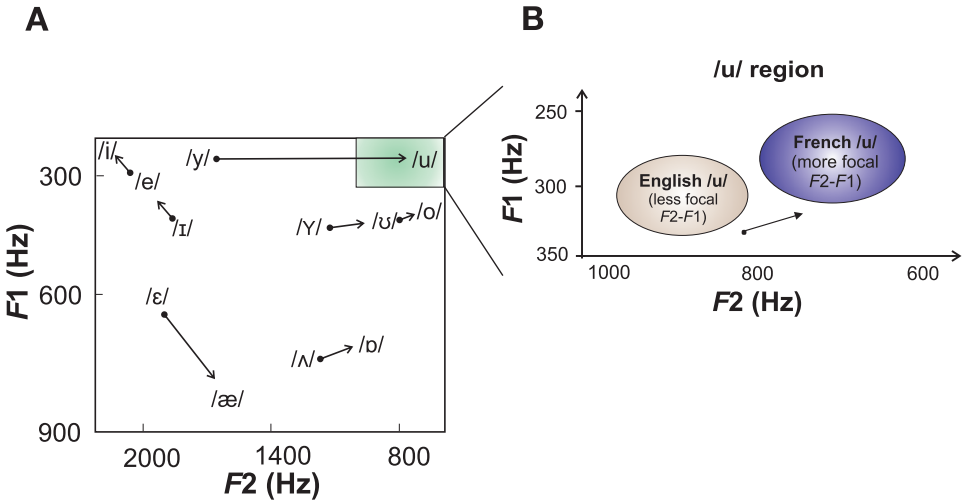

Figure 1. (A) Schematic illustration of acoustic vowel space (defined by the first two formant frequencies [*F*1 and *F*2]; adapted from Polka & Bohn, 2011). Vowel contrasts reported to show directional asymmetries in studies of infant vowel perception are plotted (see Polka & Bohn, 2003 [Table 1, p. 225], for a list of studies these results are based on). Arrows indicate the direction of vowel change that is easier to discriminate. The green rectangle delimits the portion of acoustic space that corresponds to the acoustic realization of the /u/ ("oo") vowel category across languages. (B) Magnified view of the /u/ portion of acoustic space. The precise location in the acoustic space of the /u/ category in English and French is shown; the beige ellipse delimits the region corresponding to prototypic English /u/, and the blue ellipse delimits the region corresponding to prototypic French /u/. As the plot shows, French /u/ is more acoustically peripheral and more focal (between *F*1 and *F*2) than English /u/. The arrow points in the direction that has been found to be easier to discriminate by both English- and French-speaking adults (see text for explanation). See the online article for the color version of this figure.

2017; Masapollo, Polka, et al., 2018; Masapollo, Polka, Molnar, & Ménard, 2017; Pons et al., 2012; Tyler et al., 2014), which have informed our understanding of the nature of the interplay between initial discrimination abilities and biases and linguistic experience in vowel perception.

In its current form, NRV proposes that perceptual asymmetries are phonetically grounded in the way that the human articulatory system shapes the physical speech signal. During vowel production, movements of the articulators, particularly those of the tongue, change the overall configuration of the vocal tract, which in turn, shifts the resonances of vowels—that is, the formant frequencies—in systematic and predictable ways (see Stevens, 1999, for a thorough review). According to NRV, it is hypothesized that asymmetries arise because listeners are biased toward vowels produced with extreme articulatory configurations, which lead to salient acoustic vowel signals characterized by the convergence of two or more adjacent formants (Polka & Bohn, 2011; see also Schwartz, Abry, Boë, Ménard, & Vallée, 2005). Peripheral vowels are produced when the tongue body is in its most extreme posture and displacement (either front or back, high, or low) from a "neutral" (i.e., schwa-like) vocal tract configuration. As well some peripheral vowels (e.g., /u/ and /y/) are produced with an extreme lip posture (i.e., the lips are compressed and/or protruded). These extreme articulatory configurations result in acoustic speech signals in which formants merge close together in frequency. For example, *F*2, *F*3, and *F*4 converge during the production of /i/ (that is the highest front vowel), *F*2 and *F*3 converge during the production of /y/ (the highest front rounded vowel), and *F*1 and *F*2 converge during the production of /a/ (that is the lowest back vowel) as well as /u/ (that is the highest back vowel).

There is acoustic and perceptual evidence that vowels produced with a high degree of formant convergence are auditorily more salient and perceptible to listeners. When two neighboring formants move close together in frequency there is a mutual reinforcement of their acoustic energy, such that the amplitude of one or both formants is enhanced. As a result, acoustic energy becomes focused into a narrow spectral region (see Kent & Read, 2002; Stevens, 1999, for a discussion). This focalization of spectral energy is hypothesized to give rise to vowels with well-defined spectral prominences that are easier for perceivers to detect, encode, and retain in memory (see, e.g., Masapollo, Polka, & Ménard, 2017; Masapollo, Polka, Molnar, et al., 2017; Schwartz & Escudier, 1989). The peripheral vowels (/i/, /y/, /a/, and /u/) have been referred to as "focal" vowels in the speech literature because they exhibit maximal degrees of formant convergence (Schwartz et al., 2005; Schwartz, Boë, Vallée, & Abry, 1997). However, focalization is not all-or-nothing. Rather, it is a graded property that gives rise to salience differences across vowel space (see Polka & Bohn, 2011, for discussion).

Two important acoustic cues are potentially available to the listener to aid in identifying how focal a given vowel stimulus is: (a) The proximity between formants (i.e., spectral proximity) and (b) steeper formant slopes (that may coincide with two or more energy bands moving toward each other at a higher velocity; i.e., spectral dynamics). There is currently evidence that listeners are sensitive to the proximity of formants, even when discriminating subtle acoustic differences in vowel quality. To take one example, Schwartz and Escudier (1989) tested European French-speaking listeners on their ability to discriminate among variants of /e/ synthesized with fixed *F*1, *F*2, and *F*4, but different *F*3 contours. More precisely, the *F*3 path was either fixed at an equal psychophysical distance between *F*2 and *F*4, or converged very close in frequency to either *F*2 or to *F*4. Thus, the variants systematically differed in their degree of formant proximity. Discrimination was assessed using an AX discrimination task. The results showed an asymmetry such that perceivers performed better at discriminating the changes from the tokens with less formant convergence to those with more formant convergence, compared with the reverse.

In another example, Masapollo and colleagues (Masapollo, Polka, Molnar, et al., 2017) tested Canadian-English and Canadian-French speakers on a range of synthetic vowels that fell within the /u/ category. The stimuli were created by systematically varying the proximity between *F*1 and *F*2 in equal psychophysical steps. The results of a phonetic identification and goodness rating task showed that although all of the members of the speech series were consistently identified as /u/ by subjects in both language groups, the best French /u/ exemplars had a higher degree of formant convergence than did the best English /u/ exemplars. These differences in category goodness are consistent with findings from cross-language vowel production studies showing that *F*1 and *F*2 converge more in French /u/ than English /u/ (Escudero & Polka, 2003; MacLeod, Stoel-Gammon, & Wassink, 2009; Noiray, Cathiard, Ménard, & Abry, 2011). In subsequent AX discrimination tests, subjects from both language groups performed better at discriminating changes from instances of the less-focal/English-prototypic /u/ to instances of the more-focal/French-prototypic /u/, compared with the reverse direction (shown in Figure 1B). Moreover, the magnitude of the asymmetry did not interact with native language. Masapollo, Polka, and Ménard (2017) replicated and extended these results using natural /u/ vowel stimuli (in dynamic CV syllables), confirming that this is a robust effect that is not limited to highly controlled artificial stimuli. Taken together, these findings bolster the claim that perceivers display a universal bias favoring vowels with a high degree of formant convergence and that this bias operates independently of language-specific prototype categorization processes (cf. Kuhl, 1991).

An extremely important point—that is often missed in discussions of directional asymmetries in the speech literature—is that the NRV framework proposes that effects of formant proximity on vowel discrimination reflect a phonetic bias that emerges when listeners are perceiving speech, rather than a low-level sensitivity to raw acoustic energy (Polka & Bohn, 2011; Polka et al., 2018). By this account, the foregoing findings do not derive from basic psychoacoustic processes. Indeed, recent results have provided evidence that is compatible with this view (Masapollo, Franklin, et al., 2018; Masapollo, Polka, et al., 2018; Polka & Bohn, 2011). For example, Masapollo, Franklin, et al. (2018) found that asymmetries in adult vowel perception were diminished in discrimination tasks that reduced demands on phonological working memory. Specifically, these authors found that asymmetries emerged when listeners discriminated Masapollo, Polka, and Ménard (2017) English /u/ and French /u/ tokens with a relatively long interstimulus interval (ISI; 1,500 ms) but not with relatively short ISIs (500 and 1,000 ms), suggesting that the effect of formant convergence is sensitive to processing load. If asymmetries derive from low-level sensory processes, then they should have also been present with

the shorter ISIs, because task performance would have reflected basic auditory sensitivity.

Nevertheless, while the previously noted findings are consistent with the NRV account that formant proximity contributes to driving perceptual asymmetries, at least under certain task demands, it is important to point out that the focalization of acoustic energy is also highly correlated with movements of acoustic energy (i.e., spectral dynamics[1]) during the production of vowels. The execution of more peripheral vowels entails more tongue (and, in some cases, lip) movement from a starting, neutral (schwa) position, a difference that must be reflected in spectral dynamics either by steeper formant transition slopes or greater duration of spectral change (see, e.g., Dromey, Jang, & Hollis, 2013; Lee, Shaiman, & Weismer, 2016; Mefferd, 2016; Mefferd & Green, 2010). Thus, dynamic spectral change patterns could also play a role in driving asymmetries. Indeed, substantial research on the perception of coarticulated vowels indicates that listeners represent vowels not only in terms of their canonical "target" formant patterns, but in terms of their *formant trajectory* patterns (see, e.g., Hillenbrand, 2013; Strange, 1989). Thus, an alternative, but not mutually exclusive, account of asymmetries is one in which such effects derive from articulatory kinematics, which are acoustically specified via frequency modulation in the formants of the acoustic signal (i.e., formant movements). On this account, vowels with highly focal spectral configurations may be auditorily salient, at least in part, because they necessarily involve more dynamic spectral changes during their execution. Moreover, such a sensitivity, assuming it exists, could be a consequence of general aspects of spectral processing in the human auditory system, rather than processes specific to speech. To this date, no known study has specifically examined the role of spectral dynamics in driving perceptual asymmetries.

To address this gap in the existing literature, the present investigation examined the relative contributions of dynamic and static spectral cues to directional asymmetries, by exploring whether analogous directional effects emerge using nonspeech tonal analogues that capture these key spectro-temporal properties of vowels, without being explicitly recognized as speech. While tones do differ from vowel formants in a number of key respects, particularly in their bandwidth and harmonic structure, researchers have previously used nonspeech tonal stimuli that approximate critical acoustic properties of speech stimuli in a variety of experiments and paradigms to investigate whether the mechanisms and processes involved in perceiving speech can be explained from a general auditory cognitive science perspective (e.g., Hillenbrand, Clark, & Baer, 2011; Holt, 2005; Holt, Lotto, & Kluender, 2000; Lotto & Kluender, 1998; Remez, Rubin, Pisoni, & Carrell, 1981; Viswanathan, Fowler, & Magnuson, 2009; Viswanathan, Magnuson, & Fowler, 2014). Such studies have yielded mixed results with some reporting similarities between the perception of speech and nonspeech and others reporting dissimilarities (see Fowler, 1990, for discussion). The present findings have the potential to offer insights into whether general aspects of spectral processing, that may not be specific to speech, contribute to directional asymmetries in vowel perception.

In the present experiments, both types of spectral convergence cues were manipulated, alone or in combination, to address three questions: (a) Can we observe comparable asymmetric perceptual responses with two-component tones that retain both types of spectral cues? (b) Can we observe directional asymmetries with two-component tones that solely differ in their spectral proximity? (c) Can we observe asymmetries with single-component tones that solely differ in their degree of spectral modulation? We report the results of five experiments. Experiment 1 was designed to first examine whether we could observe asymmetries using nonspeech tonal analogues that approximate some of the key spectro-temporal properties of Masapollo, Polka, and Ménard (2017) natural English /u/ and French /u/ vowel stimuli; namely, the center frequencies of the vowels' *F*1 and *F*2 trajectories. These authors reported the results of articulatory and acoustic-phonetic analyses confirming that their English /u/ and French /u/ tokens differed in both their degree and rate of lip compression and protrusion, which led to acoustic differences in their formant proximity (*F*1 and *F*2) and spectral change patterns. Specifically, the French /u/ tokens were more acoustically peripheral, and focal between *F*1 and *F*2 than the English /u/ tokens throughout their vocalic trajectories. In addition, the slopes of the *F*2 contours leading into the vowel nucleus were steeper for the French /u/ tokens compared with the English /u/ tokens. Experiments 2–5 then separately examined the effects of spectral proximity and dynamics on directional asymmetries, and in doing so, focused on better explicating the nature of the stimulus properties that may be contributing to asymmetries. Specifically, we tested whether listeners show asymmetries while discriminating tones that differ exclusively in either the relative distance between their frequency components, or in their magnitude of frequency modulation.

Experiment 1: Replicating Directional Asymmetries Using Nonspeech Tonal-Analogues

The goal of Experiment 1 was to examine whether the asymmetric discrimination of English /u/ and French /u/ vowels observed by Masapollo, Polka, and Ménard (2017; see Figure 1B) is also found with nonspeech tonal-analogues designed to approximate some of their spectro-temporal properties without sounding like vowels. We required a reliable directional asymmetry in discrimination performance to proceed to our subsequent experiments, which were designed to explore the competing roles of spectral proximity and frequency modulation on asymmetries.

Materials and Method

All experiments complied with the principles of research involving human subjects as stipulated by Brown University.

Subjects. An a priori power analysis for a paired *t* test was conducted in R (R Core Team, 2013) to determine a sufficient sample size using an α of 0.05, a power of 0.80, a large effect size ($d = .71$), and two tails. The effect size was observed in earlier research (Masapollo, Polka, & Ménard, 2017), which is nearly identical in design to the present experiments. Based on the aforementioned assumptions, the minimum desired sample size is 16.

[1] The term "spectral dynamics" typically refers in a broader sense to changes in the spectral properties of an acoustic event over time. We use this term here in the context of speech to refer to cases where spectral energies (i.e., formants) converge towards each other in frequency. However, other authors in different contexts may use this term in a less restricted sense to also refer to parallel or divergent spectral movements.

We recruited 20 students from Brown University to participate in the experiment (mean age = 19.9 years [*SD* = 1.7]; 8 men). The subjects here and in all future experiments were native, monolingual American-English speakers who reported normal hearing and no history of a hearing, speech, language or other neurological disorder. Subjects' language profiles were assessed using the Language Experience and Proficiency Questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007). In addition, none of the subjects reported more than 8 years of formal musical or vocal training (i.e., such experience might have enhanced subjects' perceptual sensitivity for discriminating the present tonal stimuli). All subjects received course credit or pay for their participation in this half-hour experiment.

Stimuli. The stimuli were nonspeech tones constructed to be similar to the center of the *F*1 and *F*2 formant paths of the natural English /u/ and French /u/ vowel stimuli used by Masapollo, Polka, and Ménard (2017, Experiment 1). All vowels were recorded by a simultaneous English-French bilingual speaker in a dynamic CV (i.e., /gu/) context, as opposed to in isolation, because in a later experiment the authors examined whether asymmetries emerged during bimodal audio-visual vowel discrimination, and audio-visual dubbing was performed by aligning the initial consonantal release burst with the video frame in which the consonantal release was first visible. Because the production of the initial consonantal portion of the recorded syllables differed along several acoustic-phonetic dimensions (i.e., stop closure duration, voice-onset-time, and amplitude of prevoicing) in English and French, the researchers cross-spliced the stop portion from a clear, intelligible English /gu/ with the vocalic portion from each of the acoustic English /u/ and French /u/ tokens. In doing so, this ensured that each acoustic token of /gu/ had the same acoustic specification of the stop consonant and, therefore, any differences observed in perception would be attributable to the vocalic portion of the signal.

Example spectrograms of each vowel type (in /gV/ contexts) and its corresponding tonal analogue are shown in Figure 2. As the figure illustrates, the nonspeech stimuli combined a low-frequency tone (characterizing the center of the *F*1 path) and a high-frequency tone (characterizing the center of the *F*2 path). The low-tone was fixed at 300 Hz for all the stimuli since *F*1 values were fairly steady around that frequency across the vocalic trajectories for both the English /u/ and French /u/ tokens. The high-tone had a constant onset frequency of 1,800 Hz, which decreased to a lower offset frequency that varied from stimulus to stimulus (as described below). The slope of the high-tone was derived by linearly interpolating between the onset and offset of *F*2 in the naturally spoken /u/ tokens. The high-tone was attenuated by 12 dB in relation to the low-tone, reflecting the intensity difference between the center of *F*1 and *F*2.

For the stimuli approximating the less-focal (less-dynamic, less proximal)/English /u/ tokens, the high-tone started at 1,800 Hz and decreased to 1,300, 1,200, or 1,100 Hz; whereas for the stimuli mimicking the more-focal (more-dynamic, more proximal)/French /u/ tokens, the high-tone also started at 1,800 Hz, but decreased more steeply in frequency to 700, 800, or 900 Hz. These differ-

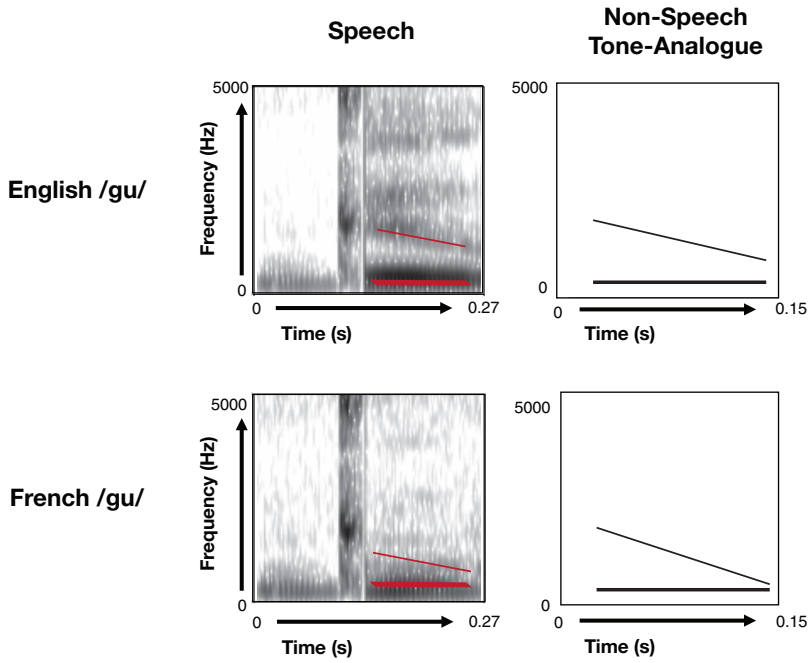

Figure 2. Example spectrograms of naturally spoken English /gu/ (upper-left) and French /gu/ syllables (lower-left; from Masapollo, Polka, & Ménard, 2017) and nonspeech tone analogues of the vowels in each syllable (upper- and lower-right, respectively). The nonspeech analogues were composed of a low- and high-frequency tone characterizing the center frequencies of *F*1 and *F*2, respectively, of the vocalic portion of the acoustic signal (highlighted in red); the frequency values of the tones were determined by interpolating the onset and offset frequencies of *F*1 and *F*2 using linear interpolation. See the online article for the color version of this figure.

ences in the offsets of the high-tone for the two stimulus types led to differences in both the dynamics and proximity of their spectral energies. A schematic representation of this stimulus structure is shown in Figure 3A. All of the tones had symmetric onset and offset ramps of 10 ms, and were 150 ms in duration, close to the duration of Masapollo et al.'s vowel stimuli. Stimuli were synthesized using the Audacity software (Version 2.0.3, Audacity Team).

It is important to note that while these tones retained some of the frequency characteristics of Masapollo et al.'s vowels, there were important acoustic differences between these nonspeech tones and the natural speech upon which they were modeled. More important, the spectral peaks tracking the center frequencies of the vowel formants had a much narrower bandwidth than the formants in the natural vowels (as shown in Figure 2). Thus, these tones only model certain aspects of the acoustics of vowels.

To ensure that the stimuli were not perceived as speech or "speech-like," all of the subjects tested were informally interviewed regarding the sounds after the completion of the experiment. Specifically, subjects were asked in an open-ended manner to describe their general impression of the sounds and whether the sounds resembled any environmental events. Critically, none of the subjects tested interpreted the stimuli as speech. This raises an additional issue: whereas the vowels used by Masapollo et al. exemplified phonetic categories with which their subjects were more or less familiar, it was not a priori clear whether participants in the present experiments would categorize the tonal stimuli in a speech-like manner (we return to this issue below).

Procedure and design. Subjects completed a categorical same/different (AX) discrimination task. On each trial, subjects heard a pair of tones, separated by an ISI of 1,500 ms, and then judged whether they were the "same" or "different." For each same trial, different tones of the same stimulus type were paired (i.e., two different tonal-analogues from the more-dynamic, more-proximal group were paired or two different tonal-analogues from the less-dynamic, and less-proximal group were paired). For each different trial, tokens from the two different tone types were paired (i.e., a tonal-analogue from the more-dynamic and proximal group was paired with a tonal-analogue from the less-dynamic and proximal group). Thus, subjects had to indicate whether pairs of physically different stimuli were members of the same tone set or members of the two different tone sets. A long ISI was used to ensure that the task placed sufficient demands on attention and auditory working memory (e.g., Masapollo, Franklin, et al., 2018; Polka & Bohn, 2011; Strange, 2011; Werker & Logan, 1985). As previously noted, recent work indicates that asymmetries are not present in experimental conditions that use relatively short ISIs (500 or 1,000 ms; Masapollo, Franklin, et al., 2018; Polka & Bohn, 2011). As well, this was the same ISI used in Masapollo, Polka, and Ménard (2017). Subjects initiated a trial by pressing a response key, and then pressed one of two labeled buttons to indicate whether the second stimulus was the same or different from the first. Their response for each trial was recorded.

Subjects were tested individually in a sound-proof laboratory room. The experiment was programmed using the SuperLab 5.0 software package (Cedrus Corporation, San Pedro, CA), which controlled the presentation of the stimuli, and collected subjects' responses. The stimuli were presented through loudspeakers (NAD Electronics, Pickering, Ontario, Canada) at a comfortable listening level (60 dB SPL). Subjects were seated about 45 cm from a 22-in flat screen monitor. The loudspeakers were located below the screen on either side.

Before the start of the experiment, subjects were informed that they would be presented with pairs of tones, and that the pairs would contain either two different instances of the same type of tone (same pairs) or instances of two different types of tones (different pairs). Subjects were then instructed to attempt to differentiate between these two different types of tone pairings.

Before the test trials, subjects completed a short practice session (6 trials: 3 same, 3 different) to confirm that they understood the instructions. After the practice session, subjects heard every possible type of pairing of the six stimuli, five times, in both presentation orders. The test trials were organized into five blocks. Each block had 30 trials, which consisted of each possible pairing (i.e., 18 different-type trials and 12 same-type trials). This resulted in a total of 150 test trials (90 different-type, 60 same-type). Note that in this task, there were no trials with a stimulus token being paired with itself. Because these stimulus pairs did not consist of acoustically identical pairings, subjects had to generalize across small acoustic differences to perceptually group the stimuli. Subjects took a short break after completing each block. No feedback was provided on either the practice or test trials. Finally, there was no reference to speech processing in the description of the study or in any of the task instructions. Before obtaining informed consent, subjects here (and in all future experiments) were told that the purpose of the study was to examine the nature of human auditory perception. After the test session, the experimenter then informed the subjects that the tonal stimuli were nonspeech analogues of human vowel sounds. All of the subjects were surprised to learn this and reported that they did not interpret the sounds as speech (or speech-like) during the discrimination task.

Results

To ensure that differences in discrimination performance did not reflect an inherent bias to respond same or different, we used a signal detection analysis (see Grier, 1971). Each subject's performance on the different pairs was converted to an *A'* score. *A'* is a nonparametric unbiased index of performance that ranges from .50 (chance) to 1.0 (perfect discrimination). The following formula (from Grier, 1971) was used: $A' = 0.5 + (H - FA)(1 + H - FA)/[4H(1 - FA)]$, where H = proportion of hits (i.e., the proportion of trials in which subjects correctly responded to a category difference between two vowel stimuli) and FA = proportion of false alarms (i.e., the proportion of trials in which subjects incorrectly responded to a category difference between two vowel stimuli). The false alarm rate was the combined error rate observed on same trials involving each vowel within the stimulus pair.

The first question we addressed in our analyses was whether subjects perceived the tonal stimuli as falling into categories as we had intended. *A'* scores were computed relative to our categorization of the stimuli. If subjects did not share these categories, there is no reason that their *A'* scores would differ from chance. However, an analysis of the subjects' overall mean *A'* scores over all trial types showed that they was significantly greater than would be expected by chance ($M = .83$, $SD = .08$, $t(19) = 17.506$, $p < .01$, $d = 3.92$). Furthermore, the overall mean *A'* scores in the present experiment were not significantly different from those reported in Masapollo, Polka, and Ménard (2017, English-

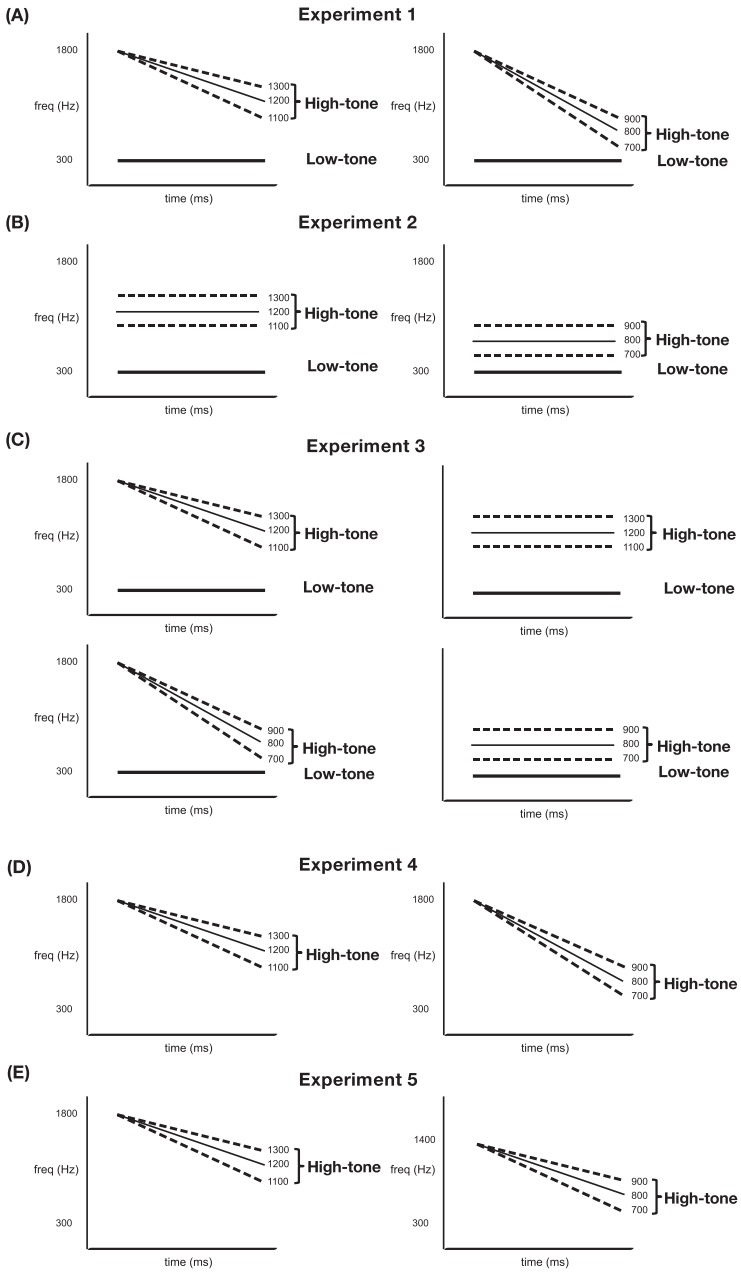

Figure 3. Stimulus structure for Experiments 1–5. (A) Schematic spectrograms of the nonspeech tone-analogues of the English /u/ (left) and French /u/ tokens (right) used in Experiment 1. All stimuli consisted of a low-tone (characterizing the center frequency of the *F*1 path) and a high-tone (characterizing the center frequency of the *F*2 path; see text for explanation). Critically, the low- and high-tones merged closer in frequency in the French /u/ tone-analogues than in the English /u/ tone-analogues. (B) Schematic spectrograms of the low- and high-tones used in Experiment 2. Note that while the tones were closer in proximity in the French /u/ analogues, the high-tone showed no change in frequency across time. (C) Schematic spectrograms of the low- and high-tones used in Experiment 3. The stimuli shown on the left are more spectrally distal, but more dynamic, whereas the stimuli shown on the right are spectrally proximal, but less dynamic. Note that the more-dynamic/less-proximal tones have a higher spectral average than the less-dynamic/more-proximal tones, but both sets of tones are matched in their spectral offsets. (D) Schematic spectrograms of the stimuli used in Experiment 4. Each stimulus consisted of only a high-tone characterizing the center frequency of the *F*2 path of each vowel type. Here, the two stimulus types vary in both the slope and offset frequency of the high-tone. (E) Schematic spectrograms of the stimuli used in Experiment 5. Each stimulus consisted of only a high-tone, but with a fixed slope and varying (onset and) offset frequencies.

speaking subjects only; $M = .83$, $SD = .06$) vowel discrimination tests, $t(33) = .058$, $p = .954$, $d = .01$. Thus, subjects were treating these artificial tonal categories in a similar way to the speech categories that they were designed to emulate, at least at some level of auditory processing.

Our analyses next focused on comparing subjects' discrimination depending on the order of stimulus presentation—from a less-dynamic/proximal tonal-analogue to a more-dynamic/proximal tonal-analogue, compared with the reverse order. For each subject, mean *A'* scores were computed for each order of stimulus presentation (see Figure 4). These scores were then compared using a paired samples *t* test. There was a significant effect ($t(19) = 2.551$, $p = .020$, $d = .37$), such that subjects performed better at discriminating the changes from the less-dynamic/less-proximal tones to the more-dynamic/more-proximal tones ($M = .84$, $SD = .10$), compared with the reverse ($M = .81$, $SD = .09$). While subjects showed a qualitatively similar effect to that previously observed using speech, the effect size was much weaker (Masapollo, Polka, & Ménard, 2017, Experiment 1, $d = .71$).

Discussion

Experiment 1 was conducted to assess whether listeners would show directional asymmetries, analogous to those shown with vowels, while discriminating nonspeech tones that approximate some of the spectral properties of Masapollo et al.'s less-focal/English /u/ and more-focal/French /u/ stimuli. Despite the mechanical quality of the tones, subjects nevertheless performed significantly better at discriminating a change from a tone whose spectral energies were further apart in frequency and less dynamic to a tone whose spectral energies were closer in frequency and more dynamic, compared with the reverse. That said, the effect size was much weaker than that previously reported by Masapollo, Polka, and Ménard (2017) using natural vowels. This is perhaps unsurprising given that our artificial tones are not nearly as rich in acoustic cues as naturally produced vowels. However, it is also possible that the differences in physical properties between formants and tones, as noted previously, underlie variation in the magnitude of the directional effect between speech and nonspeech. In other words, a

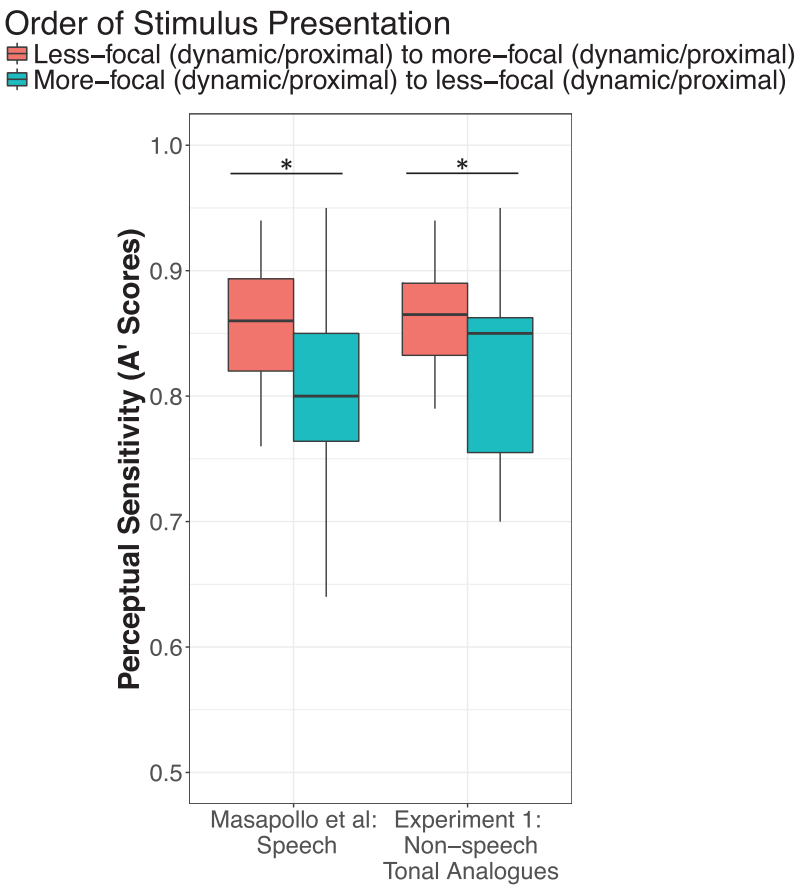

Figure 4. Boxplots of *A'* scores for Masapollo et al. (2017; Experiment 1, English-speakers only) and Experiment 1. These scores are grouped according to the order of stimulus presentation (vowels [left]: less-focal/English /u/ to more-focal/French /u/ vs. more-focal/French /u/ to less-focal/English /u/; tones [right]: less-dynamic/proximal tones to more-dynamic/distal tones vs. more-dynamic/distal tones to less-dynamic/proximal tones). * $p < .05$. See the online article for the color version of this figure.

strong directional asymmetry may only occur for stimuli containing spectral energies with larger associated bandwidths that reinforce each other when they merge in frequency. Yet, another possibility is that the perceptual processes underlying asymmetries in vowel perception are broadly tuned to certain "second-order" (Remez et al., 1981) signal properties that are shared by the speech and nonspeech stimuli, a point we will return to in the general discussion. Given these findings, we proceeded to Experiment 2.

Experiment 2: Effects of Spectral Proximity

The purpose of Experiment 2 was to test whether asymmetries, comparable with those observed in Experiment 1, would emerge while listeners attempted to discriminate tones that differed exclusively in the proximity between their low- and high-tones. To do so, we modified the stimuli in Experiment 1 such that both the high-tone (approximating the center of the *F*2 path) and the low-tone (mimicking the center of the *F*1 path) were level (see Figure 3A and 3B). In this way, the stimuli differed in the proximity between their spectral peaks, but there was no frequency modulation. If the asymmetries documented in Experiment 1 derive predominantly from differences in the spectral proximity between tones, then they should also emerge in Experiment 2. Alternatively, if the directional effect observed in Experiment 1 is predominantly attributable to frequency modulation, then this manipulation should yield no asymmetry. If, however, spectral proximity and spectral dynamics both contribute to directional asymmetries in nonspeech tone discrimination, then this might give rise to a weaker asymmetry relative to Experiment 1.

Materials and Method

Subjects. Twenty students from Brown University served as participants (mean age = 19.6 years [*SD* = 1.2]; 7 men).

Stimuli. The stimuli for this experiment (shown in Figure 3B) were synthesized using the same procedure described above. However, in this case, the high-tone was fixed at a given value, based on the high-tone offsets specified in Experiment 1 (shown in Figure 3A). For the stimuli approximating the less proximal tokens, the high-tone was fixed at 1,300, 1,200, or 1,100 Hz. For the stimuli mimicking the more proximal tokens, the high-tone was fixed at 900 Hz, 800 Hz, or 700 Hz. All other aspects of the stimuli remained the same; in particular, the low-tone was fixed at 300 Hz. Thus, although the two stimulus types differed in the proximity between their low- and high-tones, they lacked the dynamic spectral cues present in the nonspeech stimuli used in Experiment 1.

Procedure and design. The experimental protocol was identical to that of Experiment 1.

Results

Overall mean *A'* scores were again significantly greater than chance ($M = .76$, $t(19) = 10.311$, $p < .001$, $d = 2.36$). The critical question in Experiment 2 was whether subjects would show asymmetries, comparable with those observed in Experiment 1, when the frequencies of both the low- and high-tones were fixed (at their respective offset frequencies in Experiment 1). As in Experiment 1, mean *A'* scores were computed for each subject for each order of stimulus presentation (see Figure 5). A paired samples *t* test revealed no significant differences between the two stimulus orders ($t(19) = -1.291$, $p = .212$, $d = .27$).

In a second analysis, we directly compared the results of Experiment 1 and 2. *A'* scores were submitted to a two-way mixed analysis of variance (ANOVA) with experiment (Experiment 1 vs. 2) as a between-subjects factor, and order of stimulus presentation (less to more focal vs. more to less focal) as a within-subjects factor. There was no significant main effect of order of stimulus presentation ($F(1, 38) = .036$, $p = .851$, $\eta_p^2 = .001$). There was, however, a main effect of experiment ($F(1, 38) = 5.046$, $p = .031$, $\eta_p^2 = .117$), such that subjects showed greater overall sensitivity in Experiment 1 ($M = .83$, $SD = .08$) than in Experiment 2 ($M = .76$, $SD = .11$). Critically, there was a significant interaction ($F(1, 38) = 5.813$, $p = .021$, $\eta_p^2 = .133$). Discrimination was asymmetric in Experiment 1, but not in Experiment 2.

Discussion

The results of Experiment 2 revealed a discrimination pattern inconsistent with that predicted by the spectral proximity account. We found no asymmetries when subjects attempted to discriminate steady-state (i.e., fixed) tones that differed exclusively in the proximity of their spectral prominences. In addition, overall discrimination performance was higher in Experiment 1 (i.e., when the high-tone modulated in frequency) than in Experiment 2 (i.e., when the high-tone was fixed). Taken together, these findings support our "spectral dynamics" account that tones with a greater degree of frequency modulation are perceptually more salient than fixed tones with more acoustic energy focused into a narrow spectral region.

Caution, however, should be taken in interpreting the importance of this null result for two reasons. First, the directional asymmetry observed with the nonspeech tones used in Experiment 1 showed a weaker effect size than that observed in Masapollo, Polka, and Ménard (2017) using natural vowels. Thus, asymmetries, analogous to those observed with speech, appear to be subtle when using more impoverished, artificial stimulus materials. It is possible, then, that there may be a very small directional effect while discriminating the tones used in Experiment 2, but that we simply cannot observe it with the present sample size.

Second, despite the fact that the relative distance between the low- and high-tones systematically differed across the two types of stimuli, it is possible that these artificial frequency components may acoustically interact in a fundamentally different way when they converge, compared with when formants converge (a point we will return to in the general discussion). However, the present experiment suggests that differences in the relative distance between two spectral peaks alone is not sufficient to elicit an asymmetry with nonspeech tones.

Experiment 3: Dissociating Spectral Proximity and Spectral Dynamics

From the results, so far, we can conclude that dynamic spectral cues are playing some role in eliciting directional asymmetries during the discrimination of nonspeech tones. However, in an

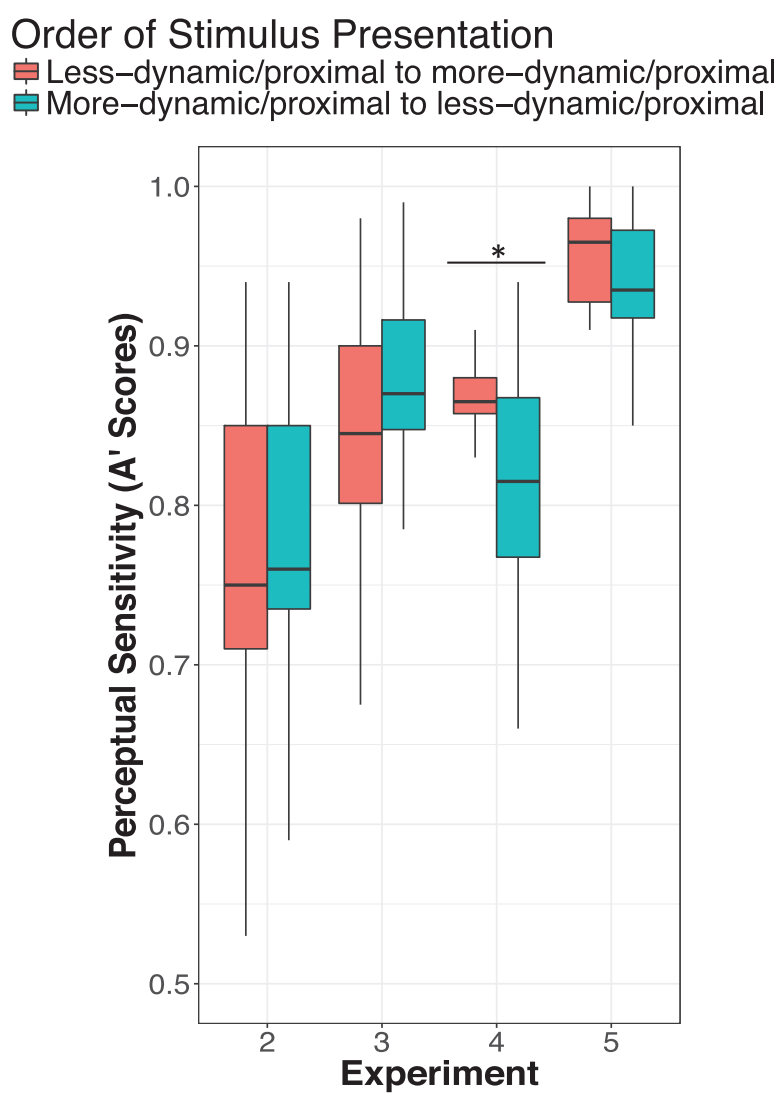

Figure 5. Boxplot of *A'* scores for Experiments 2–5. The means are grouped according to the order of stimulus presentation (less-dynamic/proximal tones to more-dynamic/distal tones vs. more-dynamic/distal tones to less-dynamic/proximal tones). * $p < .05$. See the online article for the color version of this figure.

attempt to strengthen this argument, a third experiment was run to directly pit our spectral proximity and spectral dynamics accounts against each other.[2] To this end, Experiment 3 examined whether we could elicit an asymmetry effect during the discrimination of tones that are dynamic, but more spectrally distal (Figure 3C, left column) versus tones that are not dynamic (i.e., fixed), but more spectrally proximal (see Figure 3C, right column). If dynamic spectral cues drive asymmetries, then we should find a directional effect, analogous to Experiment 1, such that listeners perform better at discriminating a change from the less-dynamic/more-proximal tones to the more-dynamic/less-proximal tones. If, however, the spectral proximity cues drive asymmetries, then listeners should show the reverse asymmetry, that is, perform better at discriminating a change from more-dynamic/less-proximal tones to the less-dynamic/more-proximal tones.

Materials and Method

Subjects. Twenty students from Brown University served as participants (mean age = 19.8 years, [*SD* = 1.18]; 6 men).

Stimuli. The stimuli for this experiment (shown in Figure 3C) were identical to those used in Experiments 1 and 2 (shown in Figure 3A and B, respectively). The stimuli from Experiment 1 served as the more-dynamic/less-proximal tones, whereas the stimuli from Experiment 2 served as the less-dynamic/more-proximal tones.

Procedure and design. The experimental design was nearly identical to that of Experiment 1, except that subjects did not

[2] We thank Navin Viswanathan for this suggestion.

discriminate every possible pairing of the 12 stimuli. Rather, a subset of the more-dynamic/less-proximal tones from Experiment 1 were paired with a subset of the less-dynamic/more-proximal tones from Experiment 2 for discrimination. As schematized by the two different rows in Figure 3C, only tones that were matched in the spectral offsets of both their high- and low-tones were contrasted, that is, we examined discrimination of the dynamic versus steady-state English /u/ tonal analogues, and discrimination of the dynamic versus steady-state French /u/ tonal analogues. Subjects heard each of these pairings, five times, in both presentation orders. The test trials were organized into five blocks. Each block had 36 trials (12 different-type trials and 24 same-type trials). This resulted in a total of 180 test trials (60 different-type, 120 same-type). As in the previous experiments, there were no physical same-same pairs. Because these stimulus pairs did not consist of acoustically identical pairings, subjects had to generalize across small acoustic differences to perceptually group the stimuli. All other aspects of the experimental protocol remained the same.

Results

We examined subjects' mean *A'* scores for each order of stimulus presentation (see Figure 5). Contrary to the predictions of both the spectral proximity and dynamics accounts, there was no significant difference ($t(19) = -1.601$, $p = .126$, $d = .19$), such that subjects performed equally well at discriminating the changes from the less-dynamic/more-proximal to the more-dynamic/less-proximal tones ($M = .84$, $SD = .07$) compared with the reverse ($M = .86$, $SD = .07$).

Discussion

Having found evidence (across Experiments 1 and 2) that directional asymmetries also emerge during the discrimination of tonal analogues of vowels, and that such asymmetry effects may arise from differences in the dynamics, as opposed to the proximity, of spectral energies, we sought to further strengthen our spectral dynamics account in Experiment 3. Specifically, we examined whether we could elicit an asymmetry effect during the discrimination of tones that are dynamic, but more spectrally distal versus tones that are not dynamic (i.e., fixed), but more spectrally proximal. In contrast to the predictions of both the spectral proximity and dynamics accounts, we found no evidence of a directional asymmetry. However, the near-ceiling discrimination of these stimuli suggests it may have been relatively easy for listeners to distinguish between a fixed and frequency modulating tone. Thus, it is possible that there may indeed be differences in the perceptual salience between static and dynamic tones, but we were not able to observe it here because discrimination performance did not deviate enough from ceiling.

Experiment 4: Effects of Degree of Spectral Dynamics

We designed Experiment 4 to examine further the role that dynamic spectral cues might be playing in the asymmetry observed in Experiment 1. We reasoned that if asymmetries in nonspeech tone perception reflect a general bias favoring acoustic signals with more dynamic frequency modulation, then we should find a directional effect, comparable with that found in Experiment 1, using only the dynamic high-tones, even when the low-tones are absent. If, however, the convergence of two (or more) spectral peaks provide the necessary conditions to induce this bias in tone perception, then we should fail to find an asymmetry.

Materials and Method

Subjects. Twenty students from Brown University served as participants (mean age = 20.6 years [$SD = 2.5$]; 6 men).

Stimuli. The stimuli for this experiment were identical to those used in Experiment 1 (Figure 3A), except that the low-tone was removed (see Figure 3C). In this way, the tones retained the dynamic spectral change patterns as the stimuli in Experiment 1, but not the spectral proximity cues.

Procedure and design. The experimental protocol was identical to that of Experiment 1.

Results

As in the previous experiments, we examined subjects' mean *A'* scores for each order of stimulus presentation (see Figure 5). There was a significant difference ($t(19) = 2.515$, $p = .021$, $d = .67$), such that subjects were better at discriminating the change from the less-dynamic tones to the more-dynamic tones ($M = .86$, $SD = .05$), compared with the reverse ($M = .81$, $SD = .09$).

In a second analysis, we directly compared the results of Experiment 1 and 4. *A'* scores were submitted to a two-way mixed ANOVA with experiment (Experiment 1 vs. 4) as a between-subjects factor, and order of stimulus presentation (less-to-more dynamic vs. more-to-less dynamic) as a within-subjects factor. Here, there was only a significant main effect of order of stimulus presentation ($F(1, 38) = 12.255$, $p = .001$, $\eta_p^2 = .244$), such that subjects were better at discriminating the change from the less-dynamic tones to the more-dynamic tones ($M = .85$, $SD = .07$), compared with the reverse ($M = .82$, $SD = .08$). However, there was no main effect of experiment ($F(1, 38) = .065$, $p = .801$, $\eta_p^2 = .002$), or interaction ($F(1, 38) = .505$, $p = .482$, $\eta_p^2 = .013$).

Discussion

The findings of Experiment 4 provide support for the view that the asymmetry observed in Experiment 1 derives from differences in frequency modulation of the high-tone, rather than the proximity between the low- and high-tones. Notably, the directional effect showed a larger effect size here ($d = .67$) than in Experiment 1 ($d = .37$), even though the stimuli were unlike speech, further supporting the view that asymmetries in nonspeech tone discrimination simply reflect differences in spectral dynamics rather than in spectral proximity. However, the asymmetries observed here and in Experiment 1 could also arise because of the mere presence of spectral dynamics, rather than because of the differences in the degree of frequency modulation, because all of the tones were fixed in Experiment 2. If this is the case, then asymmetries might still emerge with dynamic single-component tones that are matched in their spectral slopes but different in onset/offset frequencies (see Figure 3D for illustration). If, however, the asymmetries are a consequence of differences in the degree of frequency modulation, then they should fail to emerge when tones are matched in their slopes. We tested this hypothesis in Experiment 5.

Experiment 5: Controlling for Effects Related to Presence of Dynamic Cues

The tones used in Experiment 4 shared the same 1,800 Hz onset frequency but differed in their offset frequencies, so that the slope of frequency changes (i.e., spectral dynamics) differed. In these stimuli, offset frequency and slope were confounded. To disentangle this, the tones from Experiment 4 were edited by altering the onset frequency of the "more-dynamic" stimulus set, so that the slopes of all tones matched across conditions. In this case, one group had a higher average frequency and one group had a lower average frequency. If offset frequency in the presence of dynamics contributes to directional asymmetry, we would expect to observe some degree of an asymmetry in this experiment. Otherwise, the differences in the degree of dynamics would be the key contribution to the asymmetry observed in Experiment 4.

Materials and Method

Subjects. Twenty students from Brown University served as participants (mean age = 19.4 years [*SD* = 1.1]; 5 men).

Stimuli. The stimuli in this experiment were similar to those in Experiment 4, except that the slopes of all of the tones were matched to those of the high-tones in the less-focal stimuli in Experiment 1 by lowering the onset frequency of one group (see Figure 3D). Therefore, the two stimulus groups were equally dynamic, but one group had higher average frequency and the other had lower average frequency.

Procedure and design. The experimental protocol was identical to that of Experiment 1.

Results

We analyzed subjects' *A'* scores in the same way as in the previous experiments, comparing the differences for each order of stimulus presentation. As shown in Figure 5, no significant difference emerged ($t(19) = .677$, $p = .507$, $d = .11$) for dynamic tones matched in their rate of spectral change. That is, discrimination was comparable when the dynamic tones with a higher spectral average changed to the dynamic tones with a lower spectral average ($M = .94$, $SD = .05$) compared with the reverse order ($M = .94$, $SD = .03$). However, it should also be noted that discrimination of these tones was close to ceiling ($M = .94$, $SD = .04$), and task performance must deviate substantially from ceiling to measure a directional asymmetry.

In a second analysis, we directly compared the results of Experiment 2, 4, and 5. *A'* scores were submitted to a two-way mixed ANOVA with experiment (2 vs. 4 vs. 5) as a between-subjects factor, and order of stimulus presentation as a within-subjects factor. There was a significant main effect of experiment ($F(1, 57) = 27.269$, $p < .001$, $\eta_p^2 = .489$), indicating that absolute discrimination varied across experiments. The effect of order of stimulus presentation did not reach significance ($F(1, 57) = .825$, $p = .368$, $\eta_p^2 = .014$). There was, however, a significant interaction ($F(2, 57) = 4.751$, $p = .021$, $\eta_p^2 = .143$), indicating that discrimination was asymmetric in Experiment 4, but not in Experiments 2 and 5.

To further examine the differences in overall task performance across experiments, we conducted pairwise postdoc least significant difference (LSD) *t* tests. The results indicated that mean *A'* scores were higher in Experiments 4 ($M = .84$, $SD = .06$) and 5 ($M = .95$, $SD = .05$) compared with Experiment 2 ($M = .76$, $SD = .11$; Experiment 2 vs. 4: $t(38) = -2.668$, $p = .011$, $d = .28$; Experiment 2 vs. 5: $t(38) = -6.802$, $p < .001$, $d = .54$). In addition, mean *A'* scores were higher in Experiment 5 than in Experiment 4 ($t(38) = -6.189$, $p = <.001$, $d = 2.07$).

Discussion

The results of Experiment 5, in relation to the previous experiments' results, suggest an effect of degree of frequency modulation on asymmetries in nonspeech tone perception. Two aspects of the results are noteworthy. First, we found no evidence of an asymmetry when tones were matched in their rate of spectral change. Recall that in Experiment 4, we found an asymmetry such that subjects performed better at discriminating a change from the less-dynamic high tones to the more-dynamic high tones compared with the reverse. The results of Experiment 5 suggest that the mere presence of spectral dynamic cues did not contribute to that asymmetry. Thus, the directional effect observed in Experiment 1 may derive from the differences in spectral dynamics of the high-tones, rather than from differences in the proximity between the low- and high-tones or the offset frequencies of the high-tones. Although these results should be interpreted with caution, given the near-ceiling performance in Experiment 5, they are consistent with our logic about the effects of dynamic spectral cues on directional asymmetries in tone discrimination.

A second finding compatible with the view that asymmetries in tone perception are driven by differences in frequency modulation was obtained in the across-experiment analysis. Overall task performance was significantly greater in Experiments 4 and 5 compared with Experiment 2. Again, if spectrally dynamic acoustic signals are perceptually more salient, then we might expect to observe such an improvement in absolute discrimination across experiments. The finding that discrimination was higher still in Experiment 5 compared with Experiment 4 could have arisen because their respective tones differed in both their onset and offset frequencies, and swept through partially nonoverlapping frequency ranges (see Figure 3E).

General Discussion

In the present research, we investigated the nature of the perceptual processes and stimulus properties that might contribute to directional asymmetries in vowel discrimination. According to the NRV framework (Polka & Bohn, 2011), asymmetries reflect a speech-specific bias favoring "focal" vowels (i.e., vowels with adjacent formants close in frequency). By this account, the convergence of formants gives rise to well-defined spectral peaks that increase the auditory salience and perceptibility of a given vowel stimulus, and listeners are highly attuned to those convergent spectral patterns. An alternative account is that asymmetries reflect a more general auditory processing bias favoring the dynamic spectral change patterns inherent to focal vowels, rather than to the proximity and interaction of their spectral energies. On this view, focal vowels are perceptually more salient, at least in part, because extreme vocalic articulations result in more dramatic movements of acoustic energy. To begin to systematically evaluate the role

that each of these two spectral features might contribute to asymmetries, we assessed discrimination of nonspeech tones that approximate certain dynamic and static spectral properties of vowels. Moreover, the use of nonspeech tones also allowed us to probe the specificity of directional effects in vowel perception.

First, in Experiment 1, we tested whether directional asymmetries previously reported with natural and formant-synthesized vowels (e.g., Masapollo, Polka, & Ménard, 2017; Masapollo, Polka, Molnar, et al., 2017; Schwartz & Escudier, 1989) would be observed using nonspeech tonal analogues, or, instead, would be limited to speech stimuli. The nonspeech tones consisted of two-component tones with spectral dynamics and proximity properties that were similar to the *F*1 (low-tone) and *F*2 (high-tone) formant paths of Masapollo, Polka, and Ménard (2017) less-focal/English /u/ and more-focal/French /u/ stimuli. The results revealed a qualitatively similar but much weaker directional effect: subjects showed increased discrimination sensitivity for stimulus pairs contrasting a less-dynamic/proximal tonal-analogue with a more-dynamic/proximal tonal-analogue than the other direction. This directional effect may have been weaker because the tones had less robust acoustic cues than speech.

However, apart from the differences in physical properties between tones and formants, it is also possibility that the directional effect for tones (in Experiment 1) might have been weaker than that for vowels because they were not explicitly recognized as speech. On this view, speech-specific processes may underlie asymmetries in vowel perception, and such processes may be engaged to a greater extent when subjects interpret the stimuli as speech. Consistent with this view, Masapollo and colleagues (Masapollo, Polka, et al., 2018) reported that adults show directional asymmetries when discriminating English /u/ and French /u/ visemes and schematic nonspeech visual analogues of them (i.e., point light speech), but that the effect size was stronger for the nonspeech conditions when subjects were informed that the visual displays were simulating the configuration and motion of a talking mouth.

In Experiments 2–5, we then tested whether the asymmetries in tone discrimination documented in Experiment 1 result from a bias favoring dynamic changes in spectral energies and/or the proximity of spectral energies. Consistent with the spectral dynamics account, we found that asymmetries emerged when tones only manifested differences in dynamic spectral change (Experiment 4). In contrast, no asymmetries emerged in the discrimination of flat tones that nonetheless differed in their degree of spectral proximity (Experiment 2) or single-component tones varying in frequency but matched in their degree of frequency modulation (Experiment 5). That said, the spectral dynamics account failed to provide a rigorous explanation of why no asymmetries emerged during the discrimination of tones that were spectrally distal, but dynamic versus tones that were spectrally proximal, but flat (Experiment 3). We speculate that this null effect may be because listeners found it relatively easy to discriminate a flat tone from a frequency modulating tone, which in turn, might have masked any potential directional effect. Furthermore, as noted earlier, it is possible that the null effect observed in Experiment 2 might reflect an issue of statistical power. That is, there may be a very small directional effect while discriminating the flat tones, but that we simply cannot observe it with the present sample size. Nevertheless, we interpret the evidence in its entirety as consistent with the view that the degree of frequency modulation plays an important role in eliciting asymmetries in nonspeech tone perception.

Although the present findings suggest some broad generality of the stimulus properties and perceptual processes underlying asymmetries in speech and nonspeech discrimination, that conclusion can be challenged because manipulating the proximity of spectral energies in tones does not fully capture the acoustic consequences of formant convergence in speech signals. Tones are not vocal resonances and do not interact with each other in the same manner as do formants. In speech, when two formants get close in frequency, they become acoustically amplified, while in our nonspeech stimuli intensity was controlled. Furthermore, several studies have shown asymmetries using isolated steady-state vowel stimuli (e.g., Masapollo, Polka, Molnar, et al., 2017; Repp, Healy, & Crowder, 1979; Swoboda, Kass, Morse, & Leavitt, 1978). Such experiments may be interpreted as evidence that asymmetries are not driven by dynamic onset and offset formant transitions because there was little to no spectral change throughout the course of the vocalic trajectories.

However, such findings with steady-state vowels may not provide definitive evidence against a spectral dynamics account. That asymmetries can be elicited with steady-state vowels does not preclude a perceptual mechanism in which articulatory dynamics are inferred from formant proximity information alone without any overt differences in spectral change patterns.[3] According to one prominent speech perception theory, Analysis-by-Synthesis (e.g., Poeppel & Monahan, 2011; Skipper, van Wassenhove, Nusbaum, & Small, 2007), listeners (implicitly) mentally simulate what actions they would have to produce with their own vocal apparatus to generate the perceived speech signal. Data from numerous functional brain-imaging studies suggest that during passive speech perception, listeners use information in the speech signal to generate an internal forward model to synthesize and mimic the intended gesture of the speaker, and that feedback from the motor system (in the form of an efference copy) influences perception (e.g., Kuhl, Ramírez, Bosseler, Lotus Lin, & Imada, 2014; Skipper, Nusbaum, & Small, 2005; Skipper et al., 2007). From this conceptual perspective, during the internal generation of an incoming vowel signal, listeners might use implicit articulatory knowledge to infer that a larger displacement of the vocal-tract (from a neutral posture) is required to produce a relatively more focal vowel signal compared with a relatively less focal vowel. Accordingly, a more focal vowel—even in a steady-state situation without any overt temporal formant dynamics—would be internally synthesized and perceived as a more dynamic vocal-tract event. In this way, the existing data may be interpreted as being consistent with the hypothesis that asymmetries derive from perceived spectral dynamics, rather than formant proximity per se.

If the perceived dynamics of a given stimulus are, in fact, the critical factor driving asymmetries in discrimination, then this may again depend upon whether the stimulus is explicitly recognized as speech. Aside from not having been generated by a natural source in the environment that is apparent to the perceiver, the nonspeech tones have no definite causal source. Most studies in which the perception of speech and nonspeech analogues have been compared typically suffer from this confounding (but see Brancazio,

[3] We thank Linda Polka for pointing this out to us.

Best, & Fowler, 2006; Fowler & Rosenblum, 1990; see also Fowler, 1990, for discussion). In this light, vowels with relatively more focal spectral configurations will be perceived as more dynamic events, even without any physical differences in formant movement, because the perceiver implicitly knows from their own motor competence that such an acoustic signal had to be generated by a more extreme articulatory configuration. In contrast, for tones, recovery of such information cannot be inferred from tonal proximity alone because the source itself is disembodied. Consequently, overt spectro-temporal differences in the acoustic signature of the stimulus may be needed for the perceiver to infer differences in stimulus dynamics; thus, accounting for difference in observed results between Experiments 1 and 2.

An alternative interpretation derives from ecological psychological approaches to the study of speech perception, such the Direct Realist theory (e.g., Fowler, 1990). The Direct Realist theory shares with the Analysis-by-Synthesis theory the prediction that differences should emerge during the perception of speech and nonspeech sounds. However, unlike the Analysis-by-Synthesis theory, the Direct Realist account predicts that such differences would emerge independent of implicit motor knowledge. Rather, such differences are thought to emerge because listeners cannot recover information about the distal sound-producing source from proximal stimulation patterns if the source itself is disembodied (cf. Diehl, Walsh, & Kluender, 1991). Thus, by this account, speech signals are unique and distinct from other sounds in that they carry information about the distal vocal tract movements that gave rise to them, and listeners attune to spectral dynamics to perceive those dynamic speech movement patterns (see Viswanathan et al., 2014, for supporting evidence).

One potential avenue for testing the spectral dynamics account further would be to examine whether asymmetries are present during the discrimination of nonspeech tones with time-varying characteristics that are atypical of natural speech (e.g., reversed speech). If asymmetries reflect a general auditory processing bias favoring dynamic spectral cues, then one might observe analogous effects with such stimuli regardless of whether they could actually be generated by a human vocal tract.

An important challenge for the spectral dynamics account is to further explicate how spectral dynamics of formants might vary depending on local phonetic context. Even though the production of a more-focal /u/ articulation might lead to more spectral change than a less-focal /o/ articulation when executed from a neutral schwa position (because the tongue has to move farther from schwa for /u/ than /o/), vowels are rarely produced in isolation during the typical communicative speech. Rather, vowels are almost always coarticulated with flanking consonants and vowels, and the extent of tongue movement required to produce a given vowel will be systematically influenced by the articulatory configuration and movements of those surrounding segments. For example, during the production of the second vowel in the /ubu/ context, the tongue would have to move less than in the /əbu/ context because the position of the tongue body for the first vowel will carry over to some degree into the second vowel, even across the intervening consonant. One possibility is that listeners might learn from their own experience producing and perceiving speech that even though /u/ might not require much movement in certain coarticulatory contexts, in general it usually does require more and, therefore, perceive it as a more dynamic vocal-tract event, independent of context. In that case, one might predict to observe effects of spectral dynamics regardless of preceding context.

It is important to note, however, that there are also other types of evidence in the existing literature suggesting that the possible role of general auditory processes in vowel perception asymmetries is limited (Masapollo, Franklin, et al., 2018; Masapollo, Polka, et al., 2018; Polka & Bohn, 2011; Polka et al., 2015). Perhaps the strongest evidence comes from recent studies, as noted above, demonstrating that information from sources outside of audition can modulate asymmetries (Masapollo, Polka, & Ménard, 2017; Masapollo, Polka, et al., 2018). Specifically, Masapollo and colleagues examined whether visual articulatory cues provided by the speaker's face also play a role in eliciting directional asymmetries. Their logic was as follows: If purely auditory processes apply, then perceivers should show asymmetries when vowels are not heard, but perceived visually. Furthermore, co-occurring visual speech information should not be capable of modulating asymmetries during bimodal (audio-visual) vowel perception. These authors reported analogous asymmetries when subjects heard or lip-read English /u/ and French /u/ vowels. In addition, they found asymmetries, comparable with those found for unimodal vowels, for bimodal vowels when the audio and visual channels were phonetically congruent. In contrast, when the audio and visual channels were phonetically incongruent (as in the "McGurk illusion"), such asymmetries were disrupted. Collectively, these results suggest that the perceptual processes underlying asymmetries are sensitive to information available across sensory modalities.

Regardless of whether the spectral dynamics account ultimately turns out to be correct, other studies suggest that the convergence of formants play an important role in other aspects of vowel perception (e.g., "speaker normalization"). For example, there is considerable evidence that formant convergence may provide a mechanism for stabilizing a given part of a vowel spectrum across various sources of acoustic-phonetic variability, such as those associated with changes in talker identity. More specifically, perceptual experiments with adults reveal that when two adjacent vowel formants fall within a critical psychophysical distance of 3–3.5 Bark, the auditory system effectively averages the two spectral prominences, resulting in a percept that is intermediate in frequency (see, Beddor & Hawkins, 1990, for discussion). This perceptual phenomenon (referred to as "the center of gravity effect") was first suggested by research showing that it was easier to synthesize one-formant back vowels (where $F1$ and $F2$ frequencies are close) than one-formant front vowels (where $F1$ and $F2$ frequencies are widely spaced; Delattre, Liberman, Cooper, & Gerstman, 1952). In subsequent experiments, Chistovich and colleagues found that when listeners were asked to select the one-formant (F') vowel that best matches a two-formant ($F1$, $F2$) reference vowel, they choose F' between $F1$ and $F2$ if and only if $F1$ and $F2$ fell within 3–3.5 Bark of each other (Chistovich, 1985; Chistovich & Lublinskaya, 1979; Chistovich, Sheikin, & Lublinskaya, 1979; see also, Beddor & Hawkins, 1990; Fox, Jacewicz, & Chang, 2011). The convergence of two or more vowel formants, then, may help listeners achieve perceptual invariance because variability in the acoustics appears to have a relatively small impact on perception (see, e.g., Schwartz et al., 1997; Stevens, 1999; Syrdal & Gopal, 1986). The cross-linguistic regularities in vowel distributions suggest that many languages exploit this nonlinear acoustic-perceptual relation by selecting vowels found at the

extremes of phonetic space, which are not only acoustically disperse from one another, but also intrinsically focal (Polka & Bohn, 2011; Schwartz et al., 1997; Stevens, 1999).

In summary, the present findings establish that directional asymmetries in auditory perception are not limited to speech stimuli, but also occur with nonspeech tonal-analogues that approximate some of the spectro-temporal properties of natural vowels (i.e., spectral proximity and frequency modulation). Critically, however, asymmetric perceptual responses with nonspeech tones are much weaker than those found with speech, and they can only be elicited when information about frequency modulation is preserved. These results suggest limitations on the possible role of general auditory processes in vowel perception asymmetries. Collectively, these findings provide critical data in support of the NRV framework (Polka & Bohn, 2011), which posits that asymmetries in vowel perception reflect speech-specific processes that are sensitive to the way that articulatory gestures shape the acoustic structure of speech. These data will help to motivate further experimentation aimed at further explicating the nature of the perceptual processes underlying asymmetries in vowel perception, as well as the nature of the information that those processes operate on.

References

Beddor, P. S., & Hawkins, S. (1990). The influence of spectral prominence on perceived vowel quality. *The Journal of the Acoustical Society of America, 87,* 2684–2704. http://dx.doi.org/10.1121/1.399060

Bohn, O.-S., & Polka, L. (2014). Fast phonetic learning in very young infants: What it shows, and what it doesn't show. *Frontiers in Psychology, 5,* 511. http://dx.doi.org/10.3389/fpsyg.2014.00511

Brancazio, L., Best, C. T., & Fowler, C. A. (2006). Visual influences on perception of speech and nonspeech vocal-tract events. *Language and Speech, 49,* 21–53. http://dx.doi.org/10.1177/00238309060490010301

Chistovich, L. A. (1985). Central auditory processing of peripheral vowel spectra. *The Journal of the Acoustical Society of America, 77,* 789–805. http://dx.doi.org/10.1121/1.392049

Chistovich, L. A., & Lublinskaya, V. V. (1979). The "center of gravity" effect in vowel spectra and critical distance between formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research, 1,* 185–195. http://dx.doi.org/10.1016/0378-5955(79)90012-1

Chistovich, L. A., Sheikin, R. L., & Lublinskaya, V. V. (1979). Centres of gravity and spectral peaks as the determinants of vowel quality. In B. Lindblom & S. Ohman (Eds.), *Frontiers of speech communication research* (pp. 143–157). New York, NY: Academic.

Delattre, P., Liberman, A., Cooper, F., & Gerstman, L. (1952). An experimental study of the acoustic determinants of vowel colour: Observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word, 8,* 195–210. http://dx.doi.org/10.1080/00437956.1952.11659431

Diehl, R. L., Walsh, M. A., & Kluender, K. R. (1991). On the interpretability of speech/nonspeech comparisons: A reply to Fowler. *The Journal of the Acoustical Society of America, 89,* 2905–2909. http://dx.doi.org/10.1121/1.400728

Dromey, C., Jang, G.-O., & Hollis, K. (2013). Assessing correlations between lingual movements and formants. *Speech Communication, 55,* 315–328. http://dx.doi.org/10.1016/j.specom.2012.09.001

Dufour, S., Brunellière, A., & Nguyen, N. (2013). To what extent do we hear phonemic contrasts in a non-native regional variety? Tracking the dynamics of perceptual processing with EEG. *Journal of Psycholinguistic Research, 42,* 161–173. http://dx.doi.org/10.1007/s10936-012-9212-8

Escudero, P., & Polka, L. (2003). A cross-language study of vowel categorization and vowel acoustics. In M. J. Sole, D. Recansens, & J. Romero (Eds.), *Proceedings of the International Congress of Phonetic Sciences* (pp. 861–864). Barcelona, Spain: Causal Productions.

Fowler, C. A. (1990). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *The Journal of the Acoustical Society of America, 88,* 1236–1249. http://dx.doi.org/10.1121/1.399701

Fowler, C. A., & Rosenblum, L. D. (1990). Duplex perception: A comparison of monosyllables and slamming doors. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 742–754. http://dx.doi.org/10.1037/0096-1523.16.4.742

Fox, R. A., Jacewicz, E., & Chang, C. Y. (2011). Auditory spectral integration in the perception of static vowels. *Journal of Speech, Language, and Hearing Research, 54,* 1667–1681. http://dx.doi.org/10.1044/1092-4388(2011/09-0279)

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75,* 424–429. http://dx.doi.org/10.1037/h0031246

Hillenbrand, J. M. (2013). Static and dynamic approaches to vowel perception. In G. S. Morrison & P. F. Assman (Eds.), *Vowel inherent spectral change* (pp. 9–30). Heidelberg, Germany: Springer-Verlag. http://dx.doi.org/10.1007/978-3-642-14209-3_2

Hillenbrand, J. M., Clark, M. J., & Baer, C. A. (2011). Perception of sinewave vowels. *The Journal of the Acoustical Society of America, 129,* 3991–4000. http://dx.doi.org/10.1121/1.3573980

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science, 16,* 305–312. http://dx.doi.org/10.1111/j.0956-7976.2005.01532.x

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *The Journal of the Acoustical Society of America, 108,* 710–722. http://dx.doi.org/10.1121/1.429604

Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech*. Stamford, CT: Singular/Thomson Learning.

Kriengwatana, B. P., & Escudero, P. (2017). Directional asymmetries in vowel perception of adult nonnative listeners do not change over time with language experience. *Journal of Speech, Language, and Hearing Research, 60,* 1088–1093. http://dx.doi.org/10.1044/2016_JSLHR-H-16-0050

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50,* 93–107. http://dx.doi.org/10.3758/BF03212211

Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J. F., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 11238–11245. http://dx.doi.org/10.1073/pnas.1410963111

Lee, J., Shaiman, S., & Weismer, G. (2016). Relationship between tongue positions and formant frequencies in female speakers. *The Journal of the Acoustical Society of America, 139,* 426–440. http://dx.doi.org/10.1121/1.4939894

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics, 60,* 602–619. http://dx.doi.org/10.3758/BF03206049

MacLeod, A., Stoel-Gammon, C., & Wassink, A. B. (2009). Production of high vowels in Canadian English and Canadian French: A comparison of early bilingual and monolingual speakers. *Journal of Phonetics, 37,* 374–387. http://dx.doi.org/10.1016/j.wocn.2009.07.001

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50,* 940–967. http://dx.doi.org/10.1044/1092-4388(2007/067)

Masapollo, M., Franklin, L., Morgan, J. L., & Polka, L. (2018). *Asymmetries in unimodal auditory and visual vowel perception derive from phonetic encoding strategies*. Manuscript submitted for publication.

Masapollo, M., Polka, L., & Ménard, L. (2017). A universal bias in adult vowel perception – By ear or by eye. *Cognition, 166,* 358–370. http://dx.doi.org/10.1016/j.cognition.2017.06.001

Masapollo, M., Polka, L., Ménard, L., Franklin, L., Tiede, M., & Morgan, J. (2018). Asymmetries in unimodal visual vowel perception: The roles of oral-facial kinematics, orientation, and configuration. *Journal of Experimental Psychology: Human Perception and Performance, 44,* 1103–1118. http://dx.doi.org/10.1037/xhp0000518

Masapollo, M., Polka, L., Molnar, M., & Ménard, L. (2017). Directional asymmetries reveal a universal bias in adult vowel perception. *The Journal of the Acoustical Society of America, 141,* 2857–2869. http://dx.doi.org/10.1121/1.4981006

Medin, D. L., & Barsalou, L. W. (1987). Categorical processes and categorical perception. In S. Harnad (Ed.), *Categorical perception*. United Kingdom: Cambridge University Press.

Mefferd, A. S. (2016). Associations between tongue movement pattern consistency and formant movement pattern consistency in response to speech behavioral modifications. *The Journal of the Acoustical Society of America, 140,* 3728–3737. http://dx.doi.org/10.1121/1.4967446

Mefferd, A. S., & Green, J. R. (2010). Articulatory-to-acoustic relations in response to speaking rate and loudness manipulations. *Journal of Speech, Language, and Hearing Research, 53,* 1206–1219. http://dx.doi.org/10.1044/1092-4388(2010/09-0083)

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America, 27,* 338–352. http://dx.doi.org/10.1121/1.1907526

Noiray, A., Cathiard, M.-A., Ménard, L., & Abry, C. (2011). Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in French and English speakers. *The Journal of the Acoustical Society of America, 129,* 340–349. http://dx.doi.org/10.1121/1.3518452

Poeppel, D., & Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and Cognitive Processes, 26,* 935–951. http://dx.doi.org/10.1080/01690965.2010.493301

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication, 41,* 221–231. http://dx.doi.org/10.1016/S0167-6393(02)00105-X

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics, 39,* 467–478. http://dx.doi.org/10.1016/j.wocn.2010.08.007

Polka, L., Bohn, O.-S., & Weiss, D. J. (2015). Commentary: Revisiting vocal perception in non-human animals: A review of vowel discrimination, speaker voice recognition, and speaker normalization. *Frontiers in Psychology, 6,* 941. http://dx.doi.org/10.3389/fpsyg.2015.00941

Polka, L., Ruan, Y.-F., & Masapollo, M. (2018). *Understanding vowel perception biases—It's time to take a meta-analytic approach*. Manuscript submitted for publication.

Pons, F., Albareda-Castellot, B., & Sebastián-Gallés, N. (2012). The interplay between input and initial biases: Asymmetries in vowel perception during the first year of life. *Child Development, 83,* 965–976. http://dx.doi.org/10.1111/j.1467-8624.2012.01740.x

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212,* 947–949. http://dx.doi.org/10.1126/science.7233191

Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 129–145. http://dx.doi.org/10.1037/0096-1523.5.1.129

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 7,* 532–547.

Schwartz, J.-L., Abry, C., Boë, L.-J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion-Focalization Theory. *Speech Communication, 45,* 425–434. http://dx.doi.org/10.1016/j.specom.2004.12.001

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997). The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics, 25,* 255–286. http://dx.doi.org/10.1006/jpho.1997.0043

Schwartz, J.-L., & Escudier, P. (1989). A strong evidence for the existence of a large-scale integrated spectral representation in vowel perception. *Speech Communication, 8,* 235–259. http://dx.doi.org/10.1016/0167-6393(89)90004-6

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: Motor cortical activation during speech perception. *NeuroImage, 25,* 76–89. http://dx.doi.org/10.1016/j.neuroimage.2004.11.006

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex, 17,* 2387–2399. http://dx.doi.org/10.1093/cercor/bhl147

Stevens, K. N. (1999). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Strange, W. (1989). Evolving theories of vowel perception. *The Journal of the Acoustical Society of America, 85,* 2081–2087. http://dx.doi.org/10.1121/1.397860

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics, 39,* 456–466. http://dx.doi.org/10.1016/j.wocn.2010.09.001

Swoboda, P. J., Kass, J., Morse, P. A., & Leavitt, L. A. (1978). Memory factors in vowel discrimination of normal and at-risk infants. *Child Development, 49,* 332–339. http://dx.doi.org/10.2307/1128695

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *The Journal of the Acoustical Society of America, 79,* 1086–1100. http://dx.doi.org/10.1121/1.393381

Tsuji, S., & Cristia, A. (2017). Which acoustic and phonological factors shape infants' vowel discrimination? Exploiting natural variation in InPhonDB. *Proceedings of Interspeech, 2017,* 2108–2112. http://dx.doi.org/10.21437/Interspeech.2017-1468

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica, 71,* 4–21. http://dx.doi.org/10.1159/000356237

Viswanathan, N., Fowler, C. A., & Magnuson, J. S. (2009). A critical examination of the spectral contrast account of compensation for coarticulation. *Psychonomic Bulletin & Review, 16,* 74–79. http://dx.doi.org/10.3758/PBR.16.1.74

Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2014). Information for coarticulation: Static signal properties or formant dynamics? *Journal of Experimental Psychology: Human Perception and Performance, 40,* 1228–1236. http://dx.doi.org/10.1037/a0036214

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37,* 35–44. http://dx.doi.org/10.3758/BF03207136

Received December 15, 2017
Revision received September 14, 2018
Accepted September 19, 2018 ■