JSLHR

Research Article

The Effect of Pitch Auditory Feedback Perturbations on the Production of Anticipatory Phrasal Prominence and Boundary

Allison Hilger,[a] Jennifer Cole,[b] Jason H. Kim,[a] Rosemary A. Lester-Smith,[c] and Charles Larson[a]

Purpose: In this study, we investigated how the direction and timing of a perturbation in voice pitch auditory feedback during phrasal production modulated the magnitude and latency of the pitch-shift reflex as well as the scaling of acoustic production of anticipatory intonation targets for phrasal prominence and boundary.
Method: Brief pitch auditory feedback perturbations (±200 cents for 200-ms duration) were applied during the production of a target phrase on the first or the second word of the phrase. To replicate previous work, we first measured the magnitude and latency of the pitch-shift reflex as a function of the direction and timing of the perturbation within the phrase. As a novel approach, we also measured the adjustment in the production of the phrase-final prominent word as a function of perturbation direction and timing by extracting the acoustic correlates of pitch, loudness, and duration.
Results: The pitch-shift reflex was greater in magnitude after perturbations on the first word of the phrase, replicating the results from Mandarin speakers in an American English–speaking population. Additionally, the production of the phrase-final prominent word was acoustically enhanced (lengthened vowel duration and increased intensity and fundamental frequency) after perturbations earlier in the phrase, but more so after perturbations on the first word in the phrase.
Conclusion: The timing of the pitch perturbation within the phrase modulated both the magnitude of the pitch-shift reflex and the production of the prominent word, supporting our hypothesis that speakers use auditory feedback to correct for immediate production errors and to scale anticipatory intonation targets during phrasal production.

Auditory feedback is important for a speaker's control of fundamental frequency (f0) in the production of phrasal prosody (Chen et al., 2007; Liu et al., 2010, 2009; Natke & Kalveram, 2001; R. Patel et al., 2011, 2015; Xu et al., 2004). According to the DIVA (Directions Into Velocities of Articulators; Guenther, 2016) model, the auditory feedback control system compares the actual and expected speech output and, if necessary, sends error signals for correction in ongoing speech (Guenther, 2016).

Auditory feedback control of f0 has been investigated by perturbing the pitch auditory feedback of the vocal signal through headphones while the speaker is talking or producing a sustained vowel sound (Burnett et al., 1998; Chen et al., 2007). When individuals are presented with briefly perturbed pitch auditory feedback through headphones while speaking, they produce a reflexive, corrective response usually in the opposite direction of the perturbation, which has been termed the *pitch-shift reflex* (Bauer & Larson, 2003; Bauer et al., 2006; Burnett et al., 1998; Fairbanks & Guttman, 1958; Hain et al., 2000; Larson et al., 2000; Liu & Larson, 2007; Munhall et al., 2009; Sivasankar et al., 2005). When the reflexive response occurs in the opposite direction of the perturbation, it is categorized as an opposing response and is thought to reflect corrective measures of the feedback control system (Burnett et al., 1998). A less common following

[a]Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL
[b]Department of Linguistics, Northwestern University, Evanston, IL
[c]Department of Physical Medicine & Rehabilitation, Northwestern University, Chicago, IL

Rosemary A. Lester-Smith is now at the Department of Communication Sciences & Disorders, Moody College of Communication, The University of Texas at Austin.

Correspondence to Charles Larson: clarson@northwestern.edu

Editor-in-Chief: Bharath Chandrasekaran
Editor: Chao-Yang Lee

Received July 2, 2019
Revision received November 25, 2019
Accepted April 15, 2020
https://doi.org/10.1044/2020_JSLHR-19-00043

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

response occurs when the reflexive response is in the same direction of the perturbation (Behroozmand et al., 2012). While it is still unclear why the following response sometimes occurs, a possible explanation is that speakers follow the direction of the perturbation when they perceive the perturbation as an external referent rather than an internal mismatch (Hain et al., 2000). Regardless of response direction, the pitch-shift reflex is a reflection of the role of the auditory feedback control system, that is, to detect the difference in the expected and actual acoustic output and to send an error signal for correction.

Research findings from auditory feedback perturbation studies in phrasal production have revealed that vocal control mechanisms are more sensitive to productions for linguistic intent, such as phrase production compared to simple sustained-vowel production (Chen et al., 2007; Liu & Larson, 2007; Natke & Kalveram, 2001; Xu et al., 2004). The pitch-shift reflex is greater in magnitude in phrase production than sustained-vowel production for American English and Mandarin phrases (Chen et al., 2007; Xu et al., 2004). Additionally, vocal control mechanisms may be more sensitive to perturbations at time points within the phrase that are critical for planning upcoming changes in intonation (Liu et al., 2009). Liu et al. (2009) reported greater response magnitudes for pitch perturbations early in a phrase (160 and 250 ms after voice onset) than for perturbations later in the phrase (340 ms), particularly for phrases with planned phrase-final changes in pitch, such as question intonation. For example, in the context of a yes/no question in English, a planned phrase-final change in pitch would be to raise pitch at the end of the phrase. These results were interpreted to indicate that there may be a critical time period during phrase production in which planned changes in intonation are sensitive to mismatches in auditory feedback. The study by Liu et al. (2009) also revealed that auditory feedback control of f0 can be modulated by the intonation pattern of the phrase (i.e., question vs. statement). In the current study, we incorporate linguistic theories of intonation with the auditory feedback perturbation paradigm to better understand the role of auditory feedback for the production of phrasal prosody.

Linguistic theories of intonation that include the autosegmental–metrical (AM) framework, the enhanced autosegmental–metrical (AM+) framework, and the Rhythm and Pitch labeling system propose that there is a phonological organization for intonation that groups words into prosodic phrases, assigning greater prominence to one or more words in the phrase (Dilley & Breen, 2018; Dilley & Brown, 2005; Pierrehumbert, 1980). Intonational features such as high and low tones may be assigned to words that are designated as prominent and in order to mark the location of a prosodic phrase boundary at its left or right edge. The intonational features are implemented in speech through the modulation of f0 to produce relatively high- and low-pitch targets on designated syllables (Ladd, 2008; Pierrehumbert, 1980). The intonational features marking prominence, termed *pitch accents,* are distinct from those that mark phrase boundaries, termed *phrase accents* and *boundary tones*. Beyond f0, there are additional acoustic correlates of prosody. Prominence is signaled by acoustic enhancement with increased intensity, lengthened duration, and hyperarticulation (Cole, 2015; Ladd, 2008). Boundaries are signaled through final lengthening; lower intensity; and, for some speakers, creaky voice (Cole, 2015).

The phonological organization of prosody proposed in the AM framework and other frameworks suggests that, similar to articulation for speech sounds, there may be a neural repository of stored motor programs for prosody. The motor programs represent the phonetic specifications for the production of prominence and boundaries as well as larger programs for frequently used prominence and phrasing patterns and associated tone sequences (Friederici, 2002, 2012; Levelt et al., 1999; Pierrehumbert, 1980). These larger motor programs for phrasal prosody may include a specification of the acoustic–prosodic properties that encode relative prominence distinctions among the words in a prosodic phrase, for example, distinguishing a word that is the focus of a question or statement from other nonfocused words in the sentence (Liberman, 1975; Pierrehumbert, 1980). By merging theories of intonational phonology with models of speech production, we theorize that prominence and boundary are represented by auditory targets in the feedforward control system that are relatively dependent on the acoustic production of surrounding words in the phrase. Therefore, to achieve relative acoustic enhancement, auditory feedback control is used to scale the auditory targets for prominence and boundary based on the production of the rest of the phrase.

As of yet, no study has applied the auditory feedback perturbation paradigm to investigate the production of phrasal prominence as a feature of intonational phonology. We were interested in investigating how the production of a phrase-final prominent word is affected by earlier pitch perturbations in a phrase and whether the prominent word is still produced with relative acoustic enhancement compared to the earlier words in the phrase. In the current study, brief pitch auditory feedback perturbations (PAFPs) were applied during one of two early time points in the online production of a target phrase. We then measured both the pitch-shift reflex, which occurs rapidly after the perturbation, and the change in the acoustic production of the phrase-final prominent word due to the perturbations earlier in the phrase. For the pitch-shift reflex, we predicted that our results would replicate those of Liu et al. (2009), in that PAFPs earlier in the phrase would produce larger response magnitudes than perturbations later in the phrase. We also predicted that speakers would enhance the acoustic production of the phrase-final prominent word (increased f0, intensity, and duration) in trials with PAFPs to maintain salient prominence distinctions among the words in the phrase.

More specific predictions were also made about the timing and direction of the PAFPs in the phrase. We predicted that the specific word in which the perturbation occurred would have a differential effect on the production of the prominent word. The target phrase "You know Nina?" has the composition of pronoun + verb + proper noun.

In the default hierarchy of prominence for the phrase, the pronoun "you" receives the lowest level of prominence because it is a high-frequency pronoun with low informational content and is not the focused word in the sentence for this discourse context (the focused word being "Nina"; Shattuck-Hufnagel, 2000). Therefore, we predicted that a perturbation on "you" would be more disruptive for the production of the upcoming prominent syllable of "Nina" and result in greater enhancement of prominence than a perturbation on "know" because "you" should be produced with the least acoustic enhancement. For perturbation direction, we predicted that upward PAFPs in any location preceding "Nina" would result in greater enhancement of the prominent word ("Nina") because they would disrupt the downward-trending f0 interpolation leading to the low-tone target on "Nina."

Overall, we were interested in examining whether the presence of the PAFPs earlier in the phrase would affect the production of the phrase-final prominent word ("Nina"). The potential interaction of the PAFP timing and direction could reveal more about the prominence relations among the words and the acoustic adjustment required to achieve the tonal targets. If the effects of pitch perturbation are observed in the production of "Nina" only in f0 measures, it would reveal that the PAFP affects f0 trajectory over the phrase; however, if we also observe differences in intensity and duration, it would demonstrate that the measured effect goes beyond f0 and affects other acoustic correlates of prominence.

Method

Participants

Thirty-two participants (21 women, 11 men) between the ages of 18 and 57 years (M_{age} = 25, SD = 7) were recruited for this study. All study volunteers were administered a pure-tone hearing screening according to the American Speech-Language-Hearing Association's guidelines (American Speech-Language-Hearing Association, 2005; 30 dB HL at octave intervals between 250 and 8000 Hz). Two volunteers were excluded from the study because one participant had nonnative proficiency in English and another participant had a hearing threshold above 30 dB HL for higher frequencies. The remaining 30 participants (20 women, 10 men) reported English as their native language and denied a history of speech, language, or neurological disorders. No participant reported a proficiency in a tonal language or professional singing ability. This study was approved by the Northwestern University Institutional Review Board.

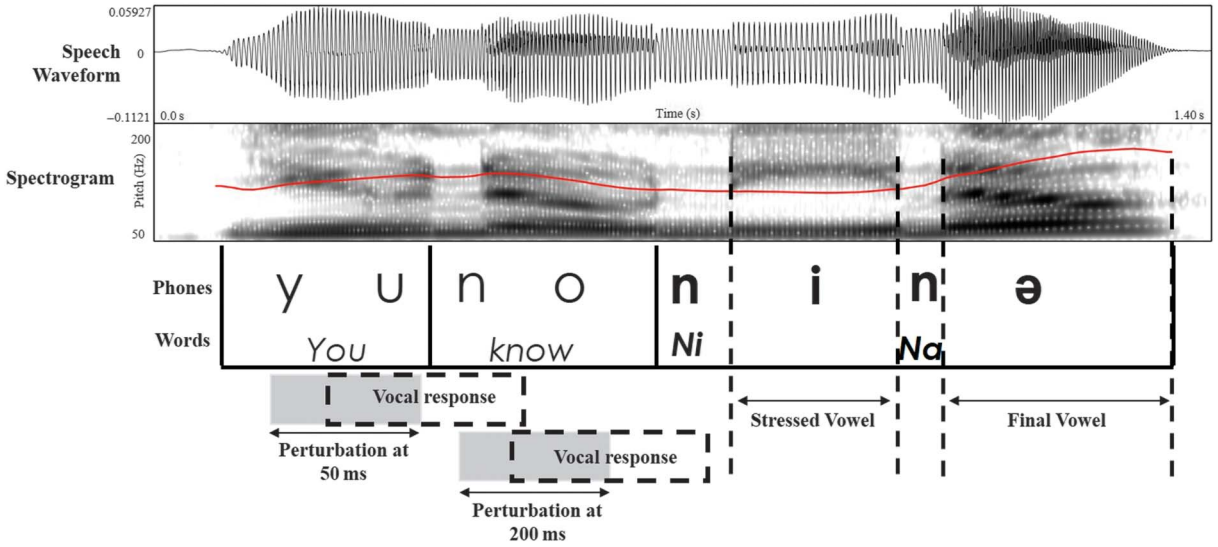
Apparatus and Procedure

After the consent process and pure-tone hearing screening, participants were seated in a double-walled, sound-treated booth (model 1201, IAC Acoustics). Participants wore insert earphones (model ER2-14A, Etymotic Research) and vocalized into an over-ear microphone (model C420, AKG) positioned approximately 1 in. from the mouth. Participants were instructed to produce a target phrase and two filler phrases according to visual prompts on a monitor screen. A simple sound level meter was also displayed on the monitor screen to aid in the maintenance of a consistent intensity level of approximately 73–75 dB SPL. The instrumentation was calibrated for intensity using the Brüel & Kjær Type 2250-S sound level meter with a 1000-Hz pure tone and a Zwislocki coupler.

Participants were prompted to produce a set of three phrases that included a target phrase ("You know Nina?") and two filler phrases ("You know Lilly?" or "You know Molly?"). The filler phrases were included to facilitate alertness during the study and to reduce the potential boredom associated with repeatedly producing the same phrase. The three phrases (the target phrase and two filler phrases) were chosen because the voiced speech sounds allowed for a continuous f0 extraction across each utterance. Additionally, the intonation of the rising declarative yes/no question allowed for less variability in the production of the intonational tones than other types of phrases such as statements (Hedberg et al., 2004). Visual prompts were presented on a computer monitor. Participants were prompted to produce the phrases as if they were introducing a friend to a hypothetical person named Nina, Lilly, or Molly. Additionally, participants were presented with a prerecorded auditory model of the desired intonation during practice trials, which is explained in depth below. All three phrases were produced during the experiment, but only the productions of the target phrase ("You know Nina?") were analyzed for this study.

As the target and filler phrases were produced, brief ±200-cent pitch perturbations were randomly applied to the participant's auditory feedback at either 50 or 200 ms after voice onset for 200 ms in duration. A control condition in which no pitch perturbation was presented was also included in the random ordering. The two perturbation time points were chosen so that the perturbations at 50 ms after voice onset would fall on "you" and perturbations at 200 ms after voice onset would fall on "know" during online production, as calculated from average productions of a small sample of five participants. The perturbations were intended to fall on their intended word target and not on the prominent word, "Nina," which occurred, on average, at 533.3 ms (SD = 40.2) after voice onset. This timing of the word "Nina" (~533.3 ms after voice onset) is well beyond the onset of the pitch perturbation on "know" (200 ms after voice onset) and the latest possible offset of the perturbation, which was intended to occur around 250–400 ms after voice onset depending on the timing of the perturbation. All perturbations had a duration of 200 ms; thus, for perturbations on "you," the offset would occur around 250 ms (with onset at 50 ms), and for perturbations on "know," the offset would occur around 400 ms (with onset at 200 ms). The timing of the prominent word, "Nina," is beyond the offset of the latest perturbation at 400 ms. Figure 1 demonstrates the timing of the onset and offset of the pitch perturbations as well as the estimated timing of the vocal responses for a rate of speech resembling the paced trials.

Figure 1. Timing of the intervals over which the pitch perturbations were applied within the target phrase "You know Nina?" for a trial from a participant in the study. Perturbations are indicated by the shaded box below the transcribed words and were 200 ms in duration, occurring at one of two different time points within the phrase. For pitch perturbations occurring at 50 ms after voice onset on the word "you," the vocal response to the perturbation occurred at approximately 110 ms after voice onset, indicated by the dashed box following the shaded box. For pitch perturbations occurring at 200 ms after voice onset on the word "know," the vocal response occurred at approximately 260 ms after voice onset.



The microphone signal was digitized with a MOTU UltraLite-mk3 and controlled by MIDI software (Max MSP 7.0, CueMix FX) to present normal and perturbed auditory feedback (Quadravox, Eventide). An algorithm within the MIDI software detected voice onset through a triggered change in voice amplitude of approximately 70 dB SPL and presented a pitch perturbation (±0 or ±200 cents) at either 50 or 200 ms after voice onset according to a randomized ordering of perturbation direction and timing conditions. When a pitch perturbation was presented, the f0 value produced by the participant was shifted ±200 cents for the entirety of the 200-ms perturbation region. The algorithm within the MIDI software applied pitch perturbations to the speaker's ongoing f0 in the prescribed magnitude and direction. The pitch extraction within this algorithm requires modal voicing to prevent errors of pitch halving or doubling. Trials in which modal voicing was not achieved and/or there were errors in pitch extraction were rejected from analysis.

In order to mask the participant's bone-conducted feedback, a gain of 10 dB SPL was applied to the headphone auditory feedback of the participant's voice, resulting in an auditory feedback of 80–85 dB SPL (HeadPod 4, Aphex). Recordings of the microphone signal, auditory feedback, and timing pulses to mark the pitch perturbation onset were obtained using a multichannel recording system (model ML880, PowerLab A/D converter, ADInstruments) and LabChart software (Version 7.0, ADInstruments) with a sampling rate of 20 kHz. Recordings of speech output and timing pulses were then time-aligned in LabChart software for off-line analysis. The timing pulses were used to differentiate pitch perturbation direction for acoustic analyses.

Participants produced the phrases in five blocks of trials by reading the phrase presented from a computer monitor. Each block consisted of 40 total trials (10 filler + 30 target phrases). For the target phrases within each block, trials were randomized such that participants were presented with ten +200 cent perturbations, ten −200 cent perturbations, and 10 control trials with no perturbations (with five of each perturbation condition occurring on "you" vs. "know"). Randomized perturbation and control trials were also included in the 10 filler phrases. Previous studies using brief pitch perturbations in phrase production have utilized 20 trials per condition and have been able to measure a substantial effect in the pitch-shift reflex (Chen et al., 2007; Liu et al., 2009). Therefore, 30 trials of the target phrase per condition, with a total of 150 trials across the five blocks in this study, were deemed sufficient to analyze the potential effect, with allowance for postprocessing and the removal of trials with errors.

The order of trials in each block was first randomly sequenced and then manually checked so that no two wholly identical phrase-and-condition combinations (i.e., Phrase × Perturbation Direction × Perturbation Timing) followed consecutively. For example, two back-to-back productions of the target phrase were allowed in the ordering as long as they differed in perturbation timing and/or direction. If two entirely identical trials followed consecutively (e.g., two back-to-back trials of "You know Nina?" with a +200 cent perturbation on "you"), the trials were randomly moved to another position in the list of trials. The order of the trials within each block was predetermined so that each participant experienced the same ordering of

trials within each block. This ordering was important so that no two experimental conditions followed consecutively. However, block order was randomized for each participant. Each block contained the same ratio of trials/experimental condition.

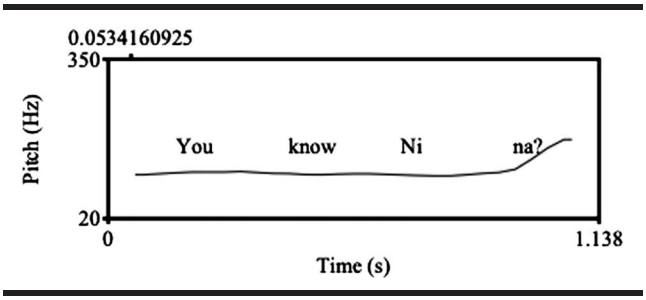
Prior to each block, participants produced 10 practice paced trials of the target phrase with an auditory model and visual pacing cue in order to model an ideal rate of speech. These practice trials did not include any pitch perturbations. The practice paced trials were included before each block of unpaced trials in an effort to prevent participants from increasing their rate of speech across the blocks of trials. The auditory model was a recording of one male speaker producing the target phrase with a flat intonation on "you" and "know," a low tone on "Ni-," and a rising high tone on "-na." The auditory model was not produced with an initial phrase accent on "you" (see Figure 2). The visual pacing cue consisted of four sequential arrows that individually turned bright green at a rate that was consistent with a natural production of the phrase.

Acoustic Analysis

Acoustic analyses were conducted to determine the temporal alignment of the produced utterances and to measure the acoustic properties of the vowels /i/ and /ə/ in "Nina." Acoustic analysis was performed using the speech analysis software Praat (Version 6.0.28; Boersma & Weenink, 2017). Target phrase productions were automatically segmented into individual words and phones using the Montreal Forced Aligner (McAuliffe et al., 2017). A final visual inspection of the `textgrids` was performed to confirm that the segmentations were correct.

Timing pulses were aligned with the segmented `textgrids` to label the onset and direction of the pitch perturbation within the production of the target phrase. Our resulting labels for the timing of the perturbations were then derived not from the original intention of the timing of the perturbation but from when the perturbation actually fell within the participant's production of the phrase. For example, if a perturbation that was intended to fall on the word "know" instead fell on the word "you," we were able to correctly label that trial as a perturbation on "you." Trials were not excluded if the intended timing of the perturbation did not match the actual timing within the phrase; instead, the trial was labeled according to the actual timing rather than the intended timing. Trials were excluded if participants paused between words due to hesitancy and/or disfluency. On average, one trial was removed per participant due to hesitancy or disfluency in the production of the target phrase. Additionally, visual and auditory inspection of the target phrases revealed that all participants produced the desired intonation of the phrase. The desired intonation of "You know Nina?" included a plateaued or downward-sloping intonation on "you" or "know," a low-tone pitch accent on "Ni-," and then a rising intonation on "-na." As we had anticipated based on the simplicity of the target phrase and the auditory model provided in the practice paced trials, participants were able to consistently produce the desired intonation across trials. Therefore, no trials were excluded for incorrect intonation. Overall, 0.58% (SD = 0.69%) of trials were removed per participant (approximately equivalent to one trial per participant).

Figure 2. Intonation contour of the auditory model of the target phrase.



To measure the production of the vowels /i/ and /ə/ in "Nina," we extracted the acoustic features of each vowel and the corresponding perturbation condition label for each trial. The acoustic features measured for vowel /i/ included vowel duration, mean f0, minimum f0, and mean intensity. Minimum f0 was chosen for vowel /i/ because, in the production of the target phrase, the /i/ in "Ni-" is produced with a low intonational tone and minimum f0. The acoustic features measured for vowel /ə/ included the same features as for vowel /i/ (i.e., vowel duration, mean f0, and mean intensity) except that, instead of measuring minimum f0, we measured maximum f0. In the production of the target phrase, the /ə/ in "-na" is produced with a high intonation and maximum f0. Minimum f0 and maximum f0 were separately chosen for /i/ and /ə/, respectively, because changes in these acoustic features are what would be expected under the predictions of AM(+) phonological theories, in that f0 arises from discrete tones plus interpolation (Dilley & Breen, 2018; Dilley & Brown, 2005; Pierrehumbert, 1980). Overall, vowel duration; mean intensity; and mean, minimum, and maximum f0 were chosen for this analysis because they have been found to distinguish prominent words from nonprominent words in English (Mahrt et al., 2012). For the vowels /i/ and /ə/ in "Nina," f0 was converted from hertz to cents using the following equation:

$$\text{Cents} = 100(39.85 \times \log_2(\text{f2}/\text{f1})) \quad (1)$$

where:

f1 = the mean f0 of the first 50 ms of the trial;

f2 = the mean f0 of the vowel /i/ or /ə/ in "Nina."

Another analysis was completed to measure the peak magnitude and latency of the pitch-shift reflex to the pitch perturbation. Here, f0 values (in cents) were sampled in 10-ms intervals starting with the onset of the perturbation in a window from 50 ms before perturbation onset to 400 ms after perturbation onset. First, the trials were sorted by response direction or whether the direction of the response (i.e., up vs. down) matched the direction of

the perturbation (i.e., up vs. down). Response direction was calculated by comparing the mean f0 of the 50-ms window before perturbation onset with the mean f0 of the 400-ms window after perturbation onset (Behroozmand & Larson, 2011). If the direction of the response and the direction of the perturbation matched (e.g., up–up or down–down), the trial was labeled as "following"; if they differed (e.g., up–down or down–up), the trial was labeled as "opposing." After the trials were sorted by response direction, difference waves were calculated to normalize any differences in the response due to the position of the response within the phrase. The difference waves were created by subtracting the average control wave for each participant (or the trials with no perturbations) from the average test wave of each condition for each participant (Chen et al., 2007).

The voice f0 response magnitude of the pitch-shift reflex was calculated by finding the maximum or minimum cent value in a window of 60–300 ms after perturbation onset based on the direction of the response. For example, if the response direction was downward in pitch, a minimum cent value was calculated, and vice versa for upward movements in response direction. The window of 60–300 ms after perturbation onset was chosen to identify the response magnitude because the minimum latency of the pitch-shift reflex is approximately 60 ms after perturbation onset, according to the timing of muscle activation and the corresponding changes in f0 (Kempster et al., 1988; Larson et al., 1987; Perlman & Alipour-Haghighi, 1988), and to avoid capturing a later volitional response that may occur in the 300- to 400-ms window (Hain et al., 2000). Although research has demonstrated that the response is sometimes immediately initiated at 50 ms after perturbation onset, the peak of the response, which is the variable of interest, reliably occurs at 60–300 ms after perturbation onset (Burnett et al., 1998). Response latency was defined as the time point of the peak of the pitch-shift response.

Statistical Analysis

To measure changes in the production of the vowels /i/ and /ə/ in "Nina" after perturbations earlier in the phrase, we constructed separate Bayesian mixed-effects multivariate regression models in RStudio for each vowel (the /i/ in "Ni-" and the /ə/ in "-na"; Version 3.5.3; RStudio Team, 2015). The Bayesian mixed-effects multivariate regression model was chosen for this analysis because of its flexibility in the model in defining fixed and random effects for the measured variables and participant variance as well as its ability to include multiple dependent variables in one model. The fixed effects of interest included the experimental perturbation conditions (direction and timing). A random effect of subject variance was also included in the model.

Because of the redundancy of the control condition across perturbation timing and direction conditions, a model with both of these factors included would not converge. To address this, we divided the data by perturbation timing and then compared our acoustic variables of interest by perturbation direction for each subset of trials for "you" and "know." Therefore, we ran four models (i.e., two models for each vowel). For vowel /i/, the first model included the subset of data when the perturbations fell on "you." This model regressed the dependent variables of vowel duration, mean f0, minimum f0, and mean intensity as a function of the independent variables of perturbation direction (upward vs. downward). The second model of vowel /i/ was identical to the first model except that the data were a subset of trials when the perturbation fell on "know." The models performed for vowel /ə/ were likewise divided into subsets for perturbations on "you" and "know," and the model setup was identical to the models for vowel /i/ except that minimum f0 was replaced with maximum f0. All dependent variables (i.e., acoustic features) in each model were normalized using standard scores.

We used the R package `brms()` (Bayesian regression models using Stan; Bürkner, 2017) to perform a Bayesian mixed-effects multivariate regression analysis that sampled coefficients from the posterior probability distribution of the mixed-effects model, which were conditioned on the data and the model's prior. The samples were used to estimate the 95% credible interval for each coefficient to assess whether the coefficients were likely to make a significant contribution to the model. The parameters of the model were assigned weakly informative priors centered around zero with an *SD* of 10, denoting that we assumed no effects of our experimental conditions. The posterior distribution was estimated using a Markov chain Monte Carlo procedure with four independent chains implemented through Stan language (Carpenter et al., 2017) by the `brms()` package in RStudio. Chain convergence was assessed using a visual inspection of the trace plots.

For each parameter of interest, we report the 95% credible interval and the posterior probability that the coefficient parameter β is greater than zero, $Pr(\beta > 0)$. Credible intervals for each dependent variable indicate a 95% probability that the true parameter value lies within the described interval as conditioned by the model, the prior, and the data (Morey et al., 2016). The credible intervals provide information about the magnitude and precision of the potential effects. A coefficient estimate of zero indicates no change in the production of that acoustic feature. Intervals greater than zero indicate a positive change in the estimate of β for the acoustic feature. Likewise, intervals less than zero indicate a negative change in the estimate of β for the acoustic feature. Credible intervals that excluded zero were interpreted as statistically robust and as contributing significantly to the model. A Bayes factor was calculated for each credible interval that excluded zero to determine the degree of evidence to reject the null hypothesis. For detailed information about using `brms()` in speech production research and for interpreting Bayesian analyses, please refer to the work of Nalborczyk et al. (2019).

Additionally, we hoped to replicate the findings from Liu et al. (2009) by comparing the magnitude and latency of the pitch-shift reflex to pitch perturbation based on the timing ("you" vs. "know") and direction (upward vs. downward) of the perturbation within the phrase as well as

the direction of the vocal response (opposing vs. following). We performed two separate linear mixed-effects regression models using the `lmer()` function from the lme4 package in RStudio for the dependent variables of response magnitude (cents) and response latency (ms) regressed on the independent variables of perturbation timing and direction as well as the direction of the response within the phrase (Bates et al., 2015). A random intercept for each subject was included as a random effect. Lastly, a logistic regression was performed using the `glm()` function from the glm2 package in RStudio to analyze potential differences in the number of opposing and following responses by perturbation timing and direction (Marschner & Donoghoe, 2018). For all `lmer()` models, post hoc comparisons were made using Tukey contrasts from the `glht` function in the multcomp package of RStudio that automatically adjusted for multiple comparisons (Hothorn et al., 2008).

Results

Pitch Accent Syllable /i/ From "Ni-"

The first set of Bayesian mixed-effects multivariate regression models was constructed to determine whether the production of the stressed vowel /i/ in "Nina" increased after upward and downward perturbations on "you" or "know." Table 1 provides summaries of unscaled coefficient estimates and 95% credible intervals for all acoustic variables and perturbation conditions, which are visually displayed in Figure 3A. In the models, both upward and downward perturbations on "you" and "know" resulted in increases in vowel duration. Downward perturbations on "you" resulted in increases in mean and minimum f0 and mean intensity.

Boundary Tone Syllable /ə/ From "-na"

The second set of multivariate mixed-effects regression models was constructed for the final vowel /ə/ from the boundary tone syllable "-na" of "Nina" by perturbation condition. Table 1 provides summaries of unscaled coefficient estimates and 95% credible intervals for all acoustic variables and perturbation conditions, which are visually displayed in Figure 3B. For vowel /ə/, both upward and downward perturbations on "you" and "know" resulted in increases in vowel duration. Upward perturbations on "you" resulted in an increase in mean f0, and downward perturbations on "you" resulted in an increase in maximum f0. Both upward and downward perturbations on "you" resulted in an increase in mean intensity.

Vowel /o/ From "know"

Lastly, we constructed a Bayesian mixed-effects multivariate regression model for the vowel /o/ from the word "know," assumed here as lacking phrasal prominence, to determine whether the effect of the perturbations on the production of the prominent word "Nina" could be interpreted as an effect of phrasal prominence or just a modification of an upcoming word in the phrase. To measure this, we compared the acoustic variables of the vowel /o/ in "know" after upward and downward pitch perturbations on the word "you." Table 2 provides summaries of coefficient estimates and 95% credible intervals for all acoustic variables and perturbation conditions. Both upward and downward perturbations resulted in an increase in vowel duration. There were no significant changes in other acoustic variables.

Pitch-Shift Reflex Magnitude and Latency

Two linear mixed-effects regression models were constructed to compare absolute reflexive response magnitude (cents) and latency (ms) by perturbation timing ("you" vs. "know"), perturbation direction (upward vs. downward perturbations), and response direction (following vs. opposing responses). In Figure 4, averaged responses are included for each experimental condition for both following and opposing responses.

Response Magnitude

A significant main effect was found for response magnitude by the timing of the perturbation within the phrase, $F(1, 188.64) = 15.03$, $p = .0001$, $\eta^2 = .49$ (see Figure 5). Vocal responses were greater after perturbations on "you" ($M = 111.8$ cents, $SE = 19.3$) than after perturbations on "know" ($M = 69.3$ cents, $SE = 17.1$). A significant three-way interaction was found for response magnitude by perturbation timing, perturbation direction, and response direction, $F(1, 188.64) = 11.4$, $p = .0009$, $\eta^2 = .37$ (see Figure 6). Multiple comparisons using Tukey contrasts revealed four significant comparisons: Response magnitudes were greater for perturbations on "you" ($M = 112.3$, $SE = 26.2$) than for perturbations on "know" ($M = 34.4$, $SE = 8.4$) when they opposed downward perturbations ($z = 3.39$, $p = .016$). When vocal responses followed upward perturbations that occurred on "you" ($M = 130.7$, $SE = 36.5$), they were greater than responses that opposed downward perturbations on "know" ($M = 34.4$, $SE = 8.4$; $z = 4.11$, $p < .01$). Opposing responses were greater for upward perturbations on "you" ($M = 105.4$, $SE = 42.1$) than for downward perturbations on "know" ($M = 34.4$, $SE = 8.4$; $z = 3.16$, $p = .03$). Lastly, following responses to upward perturbations were greater for perturbations on "you" ($M = 130.7$, $SE = 36.5$) than for perturbations on "know" ($M = 40.7$, $SE = 25.1$; $z = 3.68$, $p < .01$). Essentially, two responses drove this complex interaction: Opposing responses to downward perturbations and following responses to upward perturbations on "know" were significantly smaller in magnitude than responses to other timing and direction conditions. In other words, when the speakers moved their pitch upward on "know," whether to follow an upward perturbation or oppose a downward perturbation, the vocal responses were significantly reduced.

Response Latency

A significant main effect was found for response latency by perturbation direction, $F(1, 188.41) = 7.40$, $p = .007$, $\eta^2 = .23$ (see Figure 7). Response latencies were shorter for downward perturbations ($M = 150$ ms, $SE = 2.0$) than

Table 1. Acoustic variables for the /i/ and /ə/ vowels by experimental condition.

Variable	Vowel	Perturbation condition		Coefficient estimate	95% CI	Bayes factor
Duration	/i/	Up	You	**4.88**	**[3.66, 6.14]**	**Inf**
			Know	**5.97**	**[4.75, 7.17]**	**Inf**
		Down	You	**5.49**	**[4.24, 6.74]**	**Inf**
			Know	**3.96**	**[2.64, 5.13]**	**Inf**
	/ə/	Up	You	**1.96**	**[0.32, 3.59]**	**40.24**
			Know	**3.28**	**[1.62, 4.91]**	**999**
		Down	You	**4.80**	**[3.12, 6.51]**	**Inf**
			Know	**3.32**	**[1.63, 4.96]**	**Inf**
Mean intensity	/i/	Up	You	0.01	[–0.11, 0.14]	
			Know	**0.15**	**[0.03, 0.27]**	**306.69**
		Down	You	–0.11	[–0.24, 0.01]	
			Know	0.04	[–0.08, 0.20]	
	/ə/	Up	You	**0.13**	**[0.01, 0.25]**	**21.47**
			Know	**0.22**	**[0.09, 0.34]**	**306.69**
		Down	You	–0.05	[–0.07, 0.18]	
			Know	0.03	[–0.09, 0.15]	
Mean f0	/i/	Up	You	9.34	[–6.73, 25.42]	
			Know	**19.34**	**[3.77, 35.78]**	**47.19**
		Down	You	3.40	[–13.14, 20.26]	
			Know	10.83	[–5.60, 27.24]	
	/ə/	Up	You	**21.71**	**[3.94, 39.46]**	**46.06**
			Know	16.80	[–1.41, 35.03]	
		Down	You	4.56	[–13.48, 23.11]	
			Know	7.99	[–10.31, 26.24]	
Minimum f0	/i/	Up	You	7.09	[–10.72, 24.42]	
			Know	**18.53**	**[1.54, 36.14]**	**27.78**
		Down	You	1.94	[–16.18, 19.77]	
			Know	14.62	[–3.19, 32.92]	
Maximum f0	/ə/	Up	You	16.03	[–2.15, 33.76]	
			Know	**23.62**	**[5.32, 42.51]**	**58.7**
		Down	You	13.60	[–5.28, 32.69]	
			Know	11.06	[–7.61, 30.33]	

Note. Coefficient estimate and 95% credible interval for each acoustic variable of the /i/ and /ə/ vowels in "Nina." Credible intervals that excluded zero were considered to make nonnull contributions to the model with statistically robust changes in production. The Bayes factor provides a measure to assess the degree of evidence for the effect. Acoustic variables are listed by experimental condition. Bold data indicate robust findings where the credible interval excludes zero.

for upward perturbations (M = 180 ms, SE = 1.0). A significant two-way interaction was found for perturbation direction by response direction for response latency, $F(1, 190.89) = 23.57$, $p < .0001$, $\eta^2 = .75$ (see Figure 8). For downward perturbations, response latencies were shorter for opposing responses (M = 120 ms, SE = 2.0) than for following responses (M = 170 ms, SE = 2.0; $z = -3.71$, $p = .001$). In contrast, for upward perturbations, response latencies were shorter for following responses (M = 150 ms, SE = 3.0) than for opposing responses (M = 200 ms, SE = 1.0; $z = 3.22$, $p = .007$). Lastly, opposing responses had shorter latencies to downward perturbations (M = 120 ms, SE = 2.0) than to upward perturbations (M = 200 ms, SE = 2.0; $z = 5.50$, $p < .001$). Overall, this interaction reveals that response latencies were shorter when the response was upward in pitch, for example, opposing a downward perturbation or following an upward perturbation.

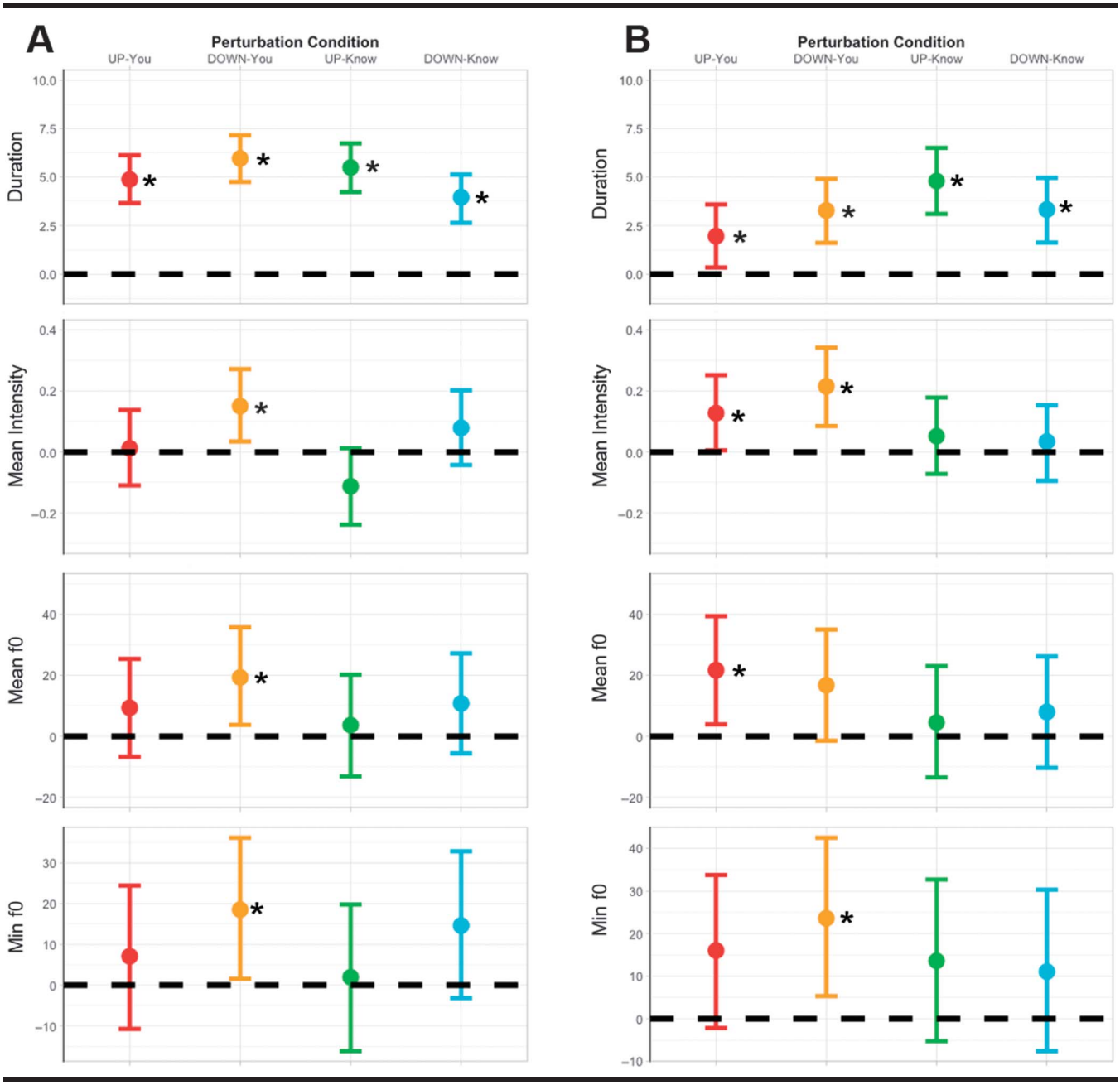
Response Direction

Lastly, we performed a logistic regression to measure the number of opposing and following responses by the timing of the perturbation in the phrase ("you" vs. "know") and the direction of the perturbation (upward vs. downward pitch perturbations). There were no significant main effects or interactions ($p > .05$), meaning that no specific perturbation condition elicited more opposing or following responses.

Discussion

In this study, we utilized an auditory feedback perturbation paradigm to investigate auditory feedback control in phrasal prosody. First, we measured the pitch-shift reflex to replicate the results from Liu et al. (2009) in an English-speaking population. Similar to Liu et al. (2009), we found that the pitch-shift reflex was larger in magnitude to earlier than later PAFPs in the phrase. This result supports their conclusion that early time points in phrasal production are more sensitive to changes in auditory feedback because these time points may be important for planning upcoming intonation targets.

Figure 3. The coefficient estimate and the 95% credible interval (CI) representing the change in production (dashed line at zero indicating no change) for each acoustic variable of (A) /i/ in "Ni-" and (B) /ə/ in "-na" by the following perturbation conditions: upward perturbation on "you" (red), downward perturbation on "you" (yellow), upward perturbation on "know" (green), and downward perturbation on "know" (blue). Black stars indicate the acoustic variables with robust differences for a perturbation condition as assessed by whether the CI overlaps with zero. In Panel A, from top to bottom, duration, mean intensity, mean f0, and minimum f0 of vowel /i/ are plotted by perturbation condition. In Panel B, from top to bottom, duration, mean intensity, mean f0, and maximum f0 of vowel /ə/ are plotted by perturbation condition.



Our next analysis measured the downstream effect of the PAFPs on the scaling and adjustment of the production of phrasal prominence and boundary. We hypothesized that the intended relative scaling of prominence and boundary in phrase production is achieved because the auditory feedback control system monitors acoustic output and sends correction signals to adjust and scale the intended acoustic output of anticipatory intonation targets. We theorized that when a mismatch in production is detected early in the phrase, the feedback-based corrective error commands are integrated not only into revised motor plans for immediate production but also into revised motor plans for downstream intonation targets in the phrase for relative phrasal prominence. We found that, in general, speakers enhanced the production of

Table 2. Acoustic variables for the /o/ vowel by experimental condition.

			Vowel /o/		
Variable	**Perturbation condition**		**Coefficient estimate**	**95% CI**	**Bayes factor**
Duration	Up	You	**8.54**	**[7.13, 9.92]**	**Inf**
	Down	You	**6.51**	**[5.10, 7.89]**	**Inf**
Mean intensity	Up	You	–0.03	[–0.15, 0.09]	
	Down	You	0.09	[–0.03, 0.22]	
Mean f0	Up	You	7.69	[–7.89, 23.15]	
	Down	You	12.80	[–2.76, 28.50]	

Note. Estimate and credible interval for each acoustic variable of the vowel /o/ of the unstressed word "know." Credible intervals that exclude zero are considered to make nonnull contributions to the model with a statistically robust effect. The Bayes factor provides a measure to assess the degree of evidence for the effect. Acoustic variables are listed by experimental condition.

both the stressed and phrase-final vowels of the prominent word "Nina" after PAFPs earlier in the phrase by increasing duration, f0, and intensity, although this effect varied by perturbation condition. Overall, this result supported our hypothesis that auditory feedback control is used not only to correct immediate errors in production (as evidenced by the pitch-shift reflex) but also to scale the production of anticipatory prosodic targets for relative phrase production.

Because phrasal prosody is produced with relative rather than absolute levels of acoustic enhancement across words in a phrase, it is a plausible interpretation that auditory feedback control is used to update the motor plans within the production of a phrase to achieve relative phrasal prominence. This hypothesis is motivated by both the AM framework and the AM+ framework (Dilley & Breen, 2018; Pierrehumbert, 1980). Both frameworks require that intonational features be acoustically realized in such a way that relative (syntagmatic) distinctions in prominence across the phrase are conveyed, along with distinctions between successive tone targets. Similarly, distinctions that maintain the (paradigmatic) contrast between high- and low-tone targets in any given syllable location are also required. A theory of speech production that is compatible with these phonological frameworks would specify that the sensorimotor mappings for the pitch targets of intonational tones may be adjusted so that the syllables associated with these tones comply with the particular relations to other words and tones in the phrase. Overall, the findings from this study support the integration of theories of speech motor production from the DIVA model (Guenther, 2016) and theories of intonation (Dilley & Breen, 2018; Dilley & Brown, 2005; Ladd, 2008; Pierrehumbert, 1980) for improving our understanding of speech production mechanisms for intonation.

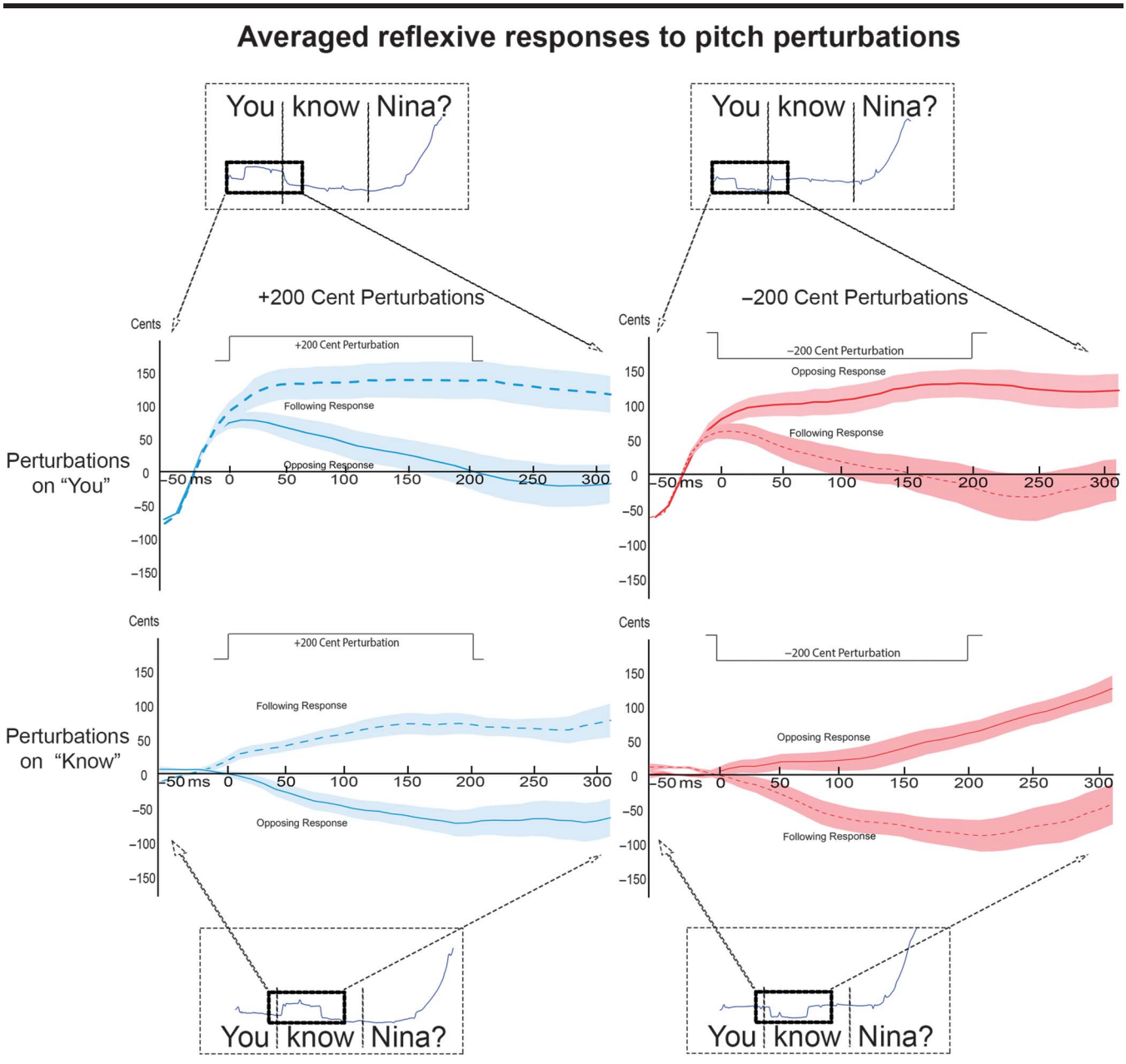
Perturbation Timing

The timing of the perturbation within the phrase modulated both the production of the prominent word "Nina" and the magnitude of the pitch-shift reflex. For the production of the prominent word "Nina," multiple acoustic variables were uniquely enhanced after perturbations on "you." Duration was lengthened for both the stressed and final vowels of "Nina" regardless of perturbation timing. Mean intensity as well as mean, minimum, and maximum f0 were only increased after perturbations on "you." Multiple interpretations could account for this timing effect. One interpretation is that perturbations on "you" were perceived as more disruptive than perturbations on "know." According to the hierarchy of prominence, the word "you" was intended to be the least prominent word in the target phrase and, therefore, should have been produced with the least amount of acoustic enhancement. This interpretation suggests that speakers consistently alter the production of words in a phrase to achieve intended relative prominence distinctions among words.

There is another possibility for why a PAFP on "you" resulted in greater enhancement of "Nina" compared to a PAFP on "know." "You" is at the start of the phrase and may be produced with an optional phrase-initial high tone (Breen et al., 2012; Dilley & Breen, 2018; Dilley & Brown, 2005) and a weaker secondary prominence in relation to the nuclear prominence on "Nina." Therefore, a PAFP on "you" may be more disruptive than that on "know" if the PAFP has the consequence of making "you" sound more prominent, threatening to invert the intended prominence relation between "you" and "Nina." Lastly, it is also possible that perturbations on "know" did not trigger as much enhancement of "Nina" because of a proximity effect. There is potential ambiguity in the production of syllables that are adjacent to stressed words (Dilley et al., 2010; Fear et al., 1995). Because "know" directly preceded "Nina," there may not have been ample time for the speaker to adjust the f0 and intensity of "Nina."

Lastly, an alternative interpretation for this timing effect is that the phrase-initial syllable could have been a reference point for calibrating all upcoming intonational effects that included prominence and boundary. This alternative interpretation is supported by the modulation of the pitch-shift reflex by the timing of the perturbation. Specifically, the pitch-shift reflex was greater in magnitude for perturbations on "you" than for perturbations on "know." These results mirror the findings from Liu et al. (2009), in which responses were greater in magnitude and later in latency for perturbations occurring at 160 and 240 ms after

Figure 4. Reflexive vocal responses to upward pitch perturbations (blue), downward pitch perturbations (red), perturbations on "you" (top), and perturbations on "know" (bottom). Opposing responses are the solid lines, and following response are the dotted lines. For context of when the perturbations and reflexive responses are occurring within the phrase, zoomed-out figures are provided above or below each graph of the estimated timing of the perturbations and responses within the production of the phrase. Note that the timescale in the graphs is normalized to a window from 50 ms before the onset of the perturbation to 300 ms after the onset of the perturbation.



voice onset than for those occurring at 360 ms after voice onset. Liu et al. (2009) attributed these findings to an early critical time point in the phrase that may be important for the planning of upcoming changes in intonation and is particularly sensitive to changes in auditory feedback. In the study by Liu et al. (2009), which used Mandarin phrases, the modulation effect of the timing of the perturbation only occurred for phrases with a phrase-final rise in intonation (i.e., question intonation) than for phrases with plateaued intonation (i.e., statement intonation). In our study using phrases in English, the target phrase was produced with a phrase-final rise in f0 to achieve a yes/no question intonation pattern. According to the interpretation from Liu et al. (2009), the word "you" in our target phrase might have been a sensitive time point for the planning of the upcoming rise in intonation. In light of this interpretation, it is possible that

Figure 5. The mean absolute response magnitude (cents) with standard error for perturbation timing ("you" vs. "know"). Response magnitude was greater for perturbations on "you" than for perturbations on "know."

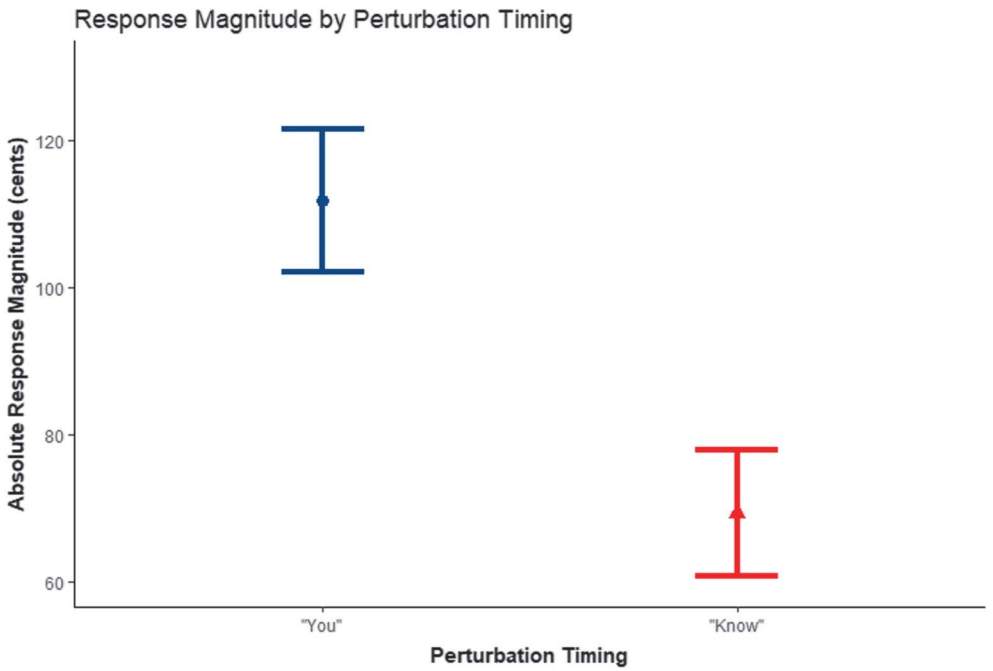


Figure 6. The mean absolute response magnitude (cents) with standard error for the three-way interaction among perturbation timing ("you" vs. "know"), perturbation direction (down vs. up), and response direction (opposing vs. following). The interaction was driven by the following response to upward perturbations on "know" and the opposing response to downward perturbations on "know."

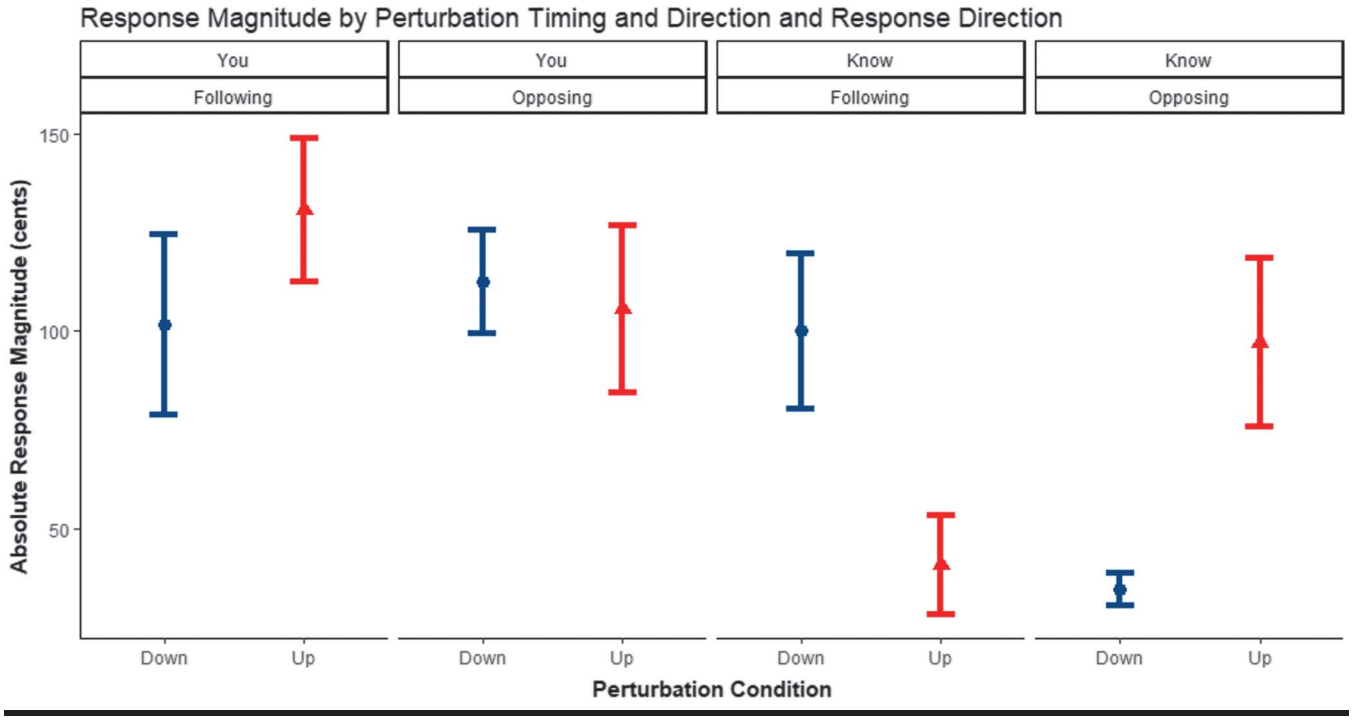


Figure 7. The mean peak response latency (ms) with standard error for perturbation direction (down vs. up). Latency was significantly shorter for downward perturbations than for upward perturbations.

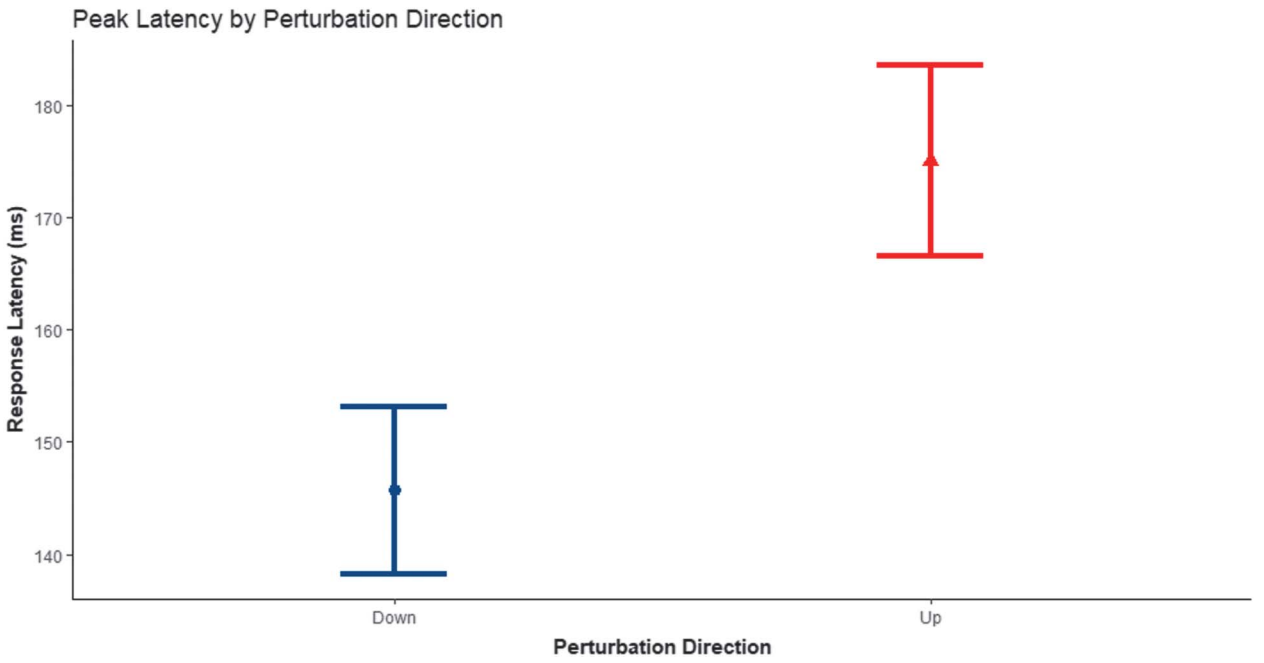
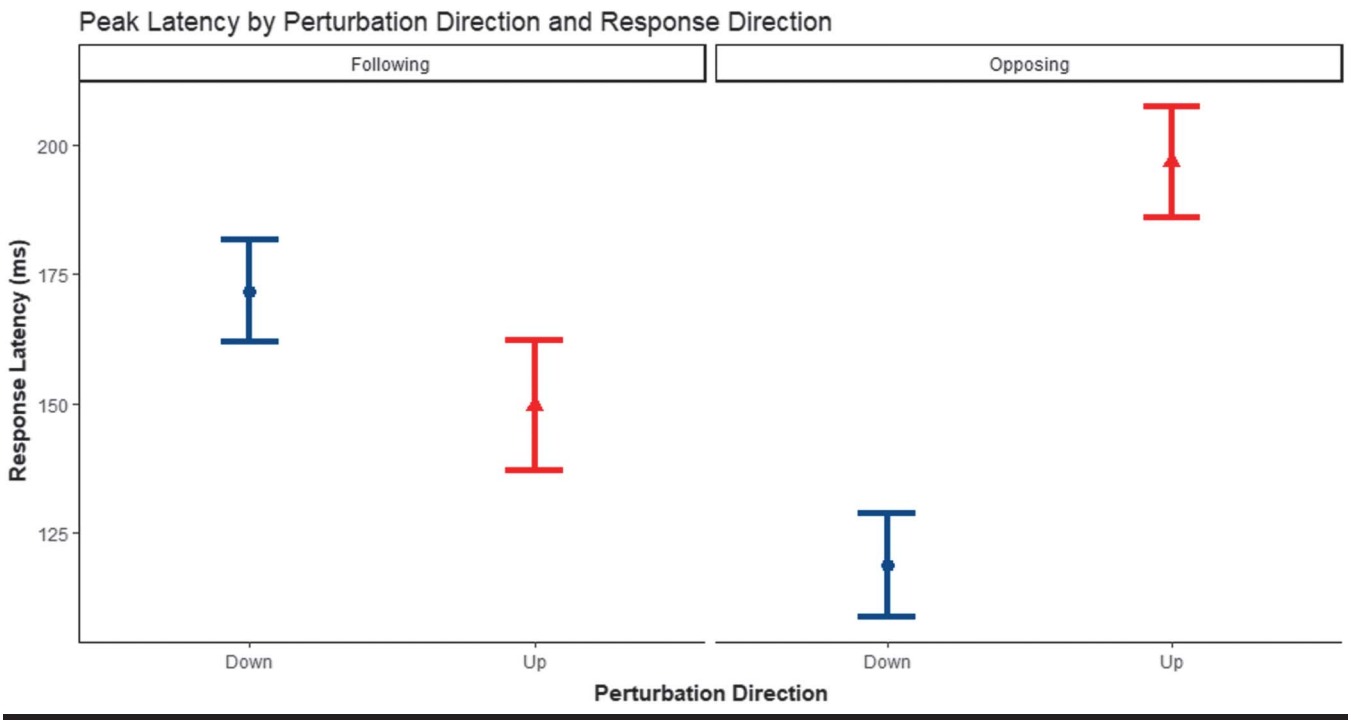


Figure 8. The mean peak response latency (ms) with standard error for the two-way interaction between perturbation direction (down vs. up) and response direction (opposing vs. following).



the acoustic enhancement of the prominent syllable and phrase boundary occurred for perturbations on "you" but not for perturbations on "know" because this earlier time point in the phrase was important for the planning of these intonation targets. This result is also supported by recent evidence that a pitch perturbation prior to a planned change in pitch results in an overshoot or undershoot of the target pitch value (Kim & Larson, 2019). The modulation effect for the timing of the perturbation may then be attributed to an effect of relative prominence and/or an early, sensitive time period in the phrase important for the planning of upcoming intonation targets. Importantly, the results from this study provide a replication for Liu et al. (2009) in English phrases, indicating that the sensitive time period of intonation planning is not specific to Mandarin but reflects a more general linguistic feature of intonation planning/phrasal prominence. Future studies are needed to better understand this phenomenon.

Perturbation Direction and Response Direction

In this study, we also tested the effects of perturbation direction (upward vs. downward pitch perturbations) on the production of the intonation targets as well as the magnitude and latency of the pitch-shift reflex. We predicted that upward PAFPs in any location preceding "Nina" would result in greater enhancement of the prominent word ("Nina") and a larger pitch-shift reflex magnitude because they would disrupt the downward-trending f0 interpolation leading to the low-tone target on "Nina." Additionally, it is possible that upward PAFPs could be particularly disruptive on "know" because "know" directly precedes the prominent word and would be typically produced with a higher pitch than "Ni-," allowing the drop in pitch to cue the low-tone pitch accent on "Nina."

Our prediction that upward PAFPs would result in greater enhancement of the prominent word was supported by an increase in mean f0 for the final vowel of "Nina" after upward perturbations on "you." However, the opposite effect was observed for the stressed vowel of "Nina," in which mean intensity as well as mean and minimum f0 increased after downward perturbations on "you" but not after upward perturbations. Additionally, maximum f0 of the final vowel of "Nina" only increased after downward perturbations on "you," and mean intensity increased more after downward than upward perturbations on "you." These results could indicate that any large and unexpected change in pitch on an unstressed word could be perceived by the speaker as an unintended pitch movement regardless of the direction of the change in pitch. This last interpretation is supported by research by Kakouros et al. (2018), in which listeners perceived the prominence of a word as a function of the statistical structure of recently perceived speech. A word that has unexpected prominence marking, relative to the listeners' expectations, will be perceived as prominent, which is an effect observed for both rising and falling f0 trajectories alike in their study. The participants in our experiment produced and heard their own productions of multiple trials of a flat or gently falling pitch trajectory over "you" and "know." Therefore, the prediction from Kakouros et al. is that a perturbation in either direction would be unexpected to the participant and could, therefore, induce the perception of prominence.

The direction of the pitch perturbation did influence the latency of the pitch-shift reflex. The pitch-shift reflex was longer in latency for upward compared to downward perturbations. These results suggest that, for the pitch-shift reflex, speakers may have perceived upward perturbations as more disruptive to their intended f0 trajectory than downward perturbations, even if they were not more disruptive to the overall acoustic implementation of the syllable prominence.

Signaling Prominence Distinctions Between Words and Syllables

To determine if the acoustic enhancement from the earlier PAFPs was truly an effect of phrasal prominence or simply a modification to any word after a perturbation, we also analyzed the change in the production of the vowel /o/ of the unstressed word "know" after perturbations on the word "you." Results showed that the production of vowel /o/ increased in vowel duration after both upward and downward perturbations, but mean intensity and mean f0 were not affected. We interpret these results in light of findings from a study by Mahrt et al. (2012), who found that some acoustic cues signaled gradient prominence distinctions and others signaled binary prominence distinctions. For example, in the study by Mahrt et al., vowel duration was found to signal gradient prominence distinctions so that words with greater prominence levels had longer vowel durations than words of lower prominence levels in a linear trend. However, intensity was found to signal binary prominence distinctions so that words that listeners perceived as prominent versus nonprominent showed distinct differences in intensity.

In application to our current findings, vowel duration was affected for all three vowels in the study: the /o/ in the unstressed word "know," the /i/ in the stressed syllable "Ni-," and the /ə/ in the phrase boundary syllable "-na." It is possible that vowel duration was used by speakers to signal a gradient prominence distinction between words. This result would explain why the /o/ in "know" was affected by pitch perturbations on the word "you" because the word "you" was lower in prominence than "know." Speakers adjusted vowel duration for the vowel /o/ in "know" to produce a gradient prominence distinction between "you" and "know." Intensity and f0 might instead be modified to signal binary prominence distinctions based on the results in our study. That the /o/ in "know" was not affected for these acoustic features could then be explained because intensity and f0 were used to signal binary prominence distinctions between prominent and nonprominent words instead of words of varying prominence levels. This finding is supported by research that has found that f0 can be an ambiguous cue for phrasal prominence (Brown et al., 2015; Dilley & McAuley, 2008; Kochanski et al., 2005). Overall, the results of our study can be seen as demonstrating that speakers use

different acoustic cues to signal gradient and binary prominence distinctions.

In studies by R. Patel and colleagues (R. Patel et al., 2011, 2015), perturbations in pitch and loudness were used to measure changes across multiple acoustic features to determine whether acoustic features of prosody are controlled in an integrated or independent channel manner. From their results, they found that the acoustic features of pitch and loudness acted in an integrated manner when pitch was manipulated but in an independent manner when loudness was manipulated (R. Patel et al., 2011, 2015). In our study, we found that the acoustic correlates of pitch, loudness, and duration worked both as integrated units and as independent units depending on the timing and direction of the PAFP. Downward perturbations on "you" triggered a more integrated response in pitch, loudness, and duration for the production of "Nina" than upward perturbations on "you." Additionally, perturbations on "know" triggered only a change in duration for the production of "Nina" and not a robust change in pitch or loudness. This may be explained by the number of interpretations described earlier when discussing the timing of the perturbation. However, this finding does demonstrate that the integrated or independent manner of acoustic features of prosody may depend on the intonationally encoded pragmatic function of the manipulated word in the phrase.

Limitations

Although the findings from this study are compelling, further research is needed on the planning and adjustment of phrasal prominence. This study had limitations due to the constrained nature of speech and language production. Participants read the phrases from a screen, which reduced the naturalness of the task. Only one intonation pattern was tested, which also reduced the generalizability of the results. Additionally, the target phrase was short, and there was a potential carryover of the duration of the perturbation to the subsequent word in the phrase, affecting the interpretation of the timing effect of our results. Future studies should investigate various pitch accent types and intonation patterns as well as incorporate longer target phrases with increased spacing between perturbation time points. Lastly, we infer that participants perceived the pitch perturbations as errors in intonation. It is possible that some participants perceived the perturbations as external manipulation and not their own error. This perception could modify how participants respond to the perturbations and adjust the production of phrasal prominence.

Conclusion

Overall, we found that the timing of the pitch perturbation during phrasal production modulated both the magnitude of the pitch-shift reflex and the production of the downstream intonation targets for prominence and boundary. These results indicate that speakers may integrate feedback-based error corrective commands into revised motor plans of anticipatory intonation targets for relative acoustic salience in phrasal production.

Acknowledgments

This research was funded by departmental funds from the Department of Communication Sciences and Disorders at Northwestern University. We would like to acknowledge Chun Liang Chan for his expertise and assistance with the experiment setup and software development and support. We would also like to thank Munirah Alkhuwaiter for her assistance with defining the analysis window for the reflexive response. Additionally, we would like to thank Timo Roettger for providing consultation on performing the Bayesian analyses in R.

References

American Speech-Language-Hearing Association. (2005). *Guidelines for manual pure-tone threshold audiometry.* ASHA.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). "Fitting linear mixed-effects models using lme4." *Journal of Statistical Software, 67*(1), 1–48. https:/doi.org/10.18637/jss.v067.i01

Bauer, J. J., & Larson, C. R. (2003). Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique. *The Journal of the Acoustical Society of America, 114*(2), 1048–1054. https://doi.org/10.1121/1.1592161

Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America, 119*(4), 2363–2371. https://doi.org/10.1121/1.2173513

Behroozmand, R., Korzyukov, O., Sattler, L., & Larson, C. R. (2012). Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control. *The Journal of the Acoustical Society of America, 132*(4), 2468–2477. https://doi.org/10.1121/1.4746984

Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience, 12*(1), Article 54. https://doi.org/10.1186/1471-2202-12-54

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer* (Version 6.0.21) [Computer software]. https://www.praat.org

Breen, M., Dilley, L., Kraemer, J., & Gibson, E. (2012). Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory, 8*(2), 277–312. https://doi.org/10.1515/cllt-2012-0011

Brown, M., Salverda, A. P., Dilley, L., & Tanenhaus, M. (2015). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 306–323. https://doi.org/10.1037/a0038689

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America, 103*(6), 3153–3161. https://doi.org/10.1121/1.423073

Bürkner, P. C. (2017). Advanced Bayesian multilevel modeling with the R package brms. arXiv preprint arXiv:1705.11123.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal*

of Statistical Software, 76(1), 24477. https://doi.org/10.18637/jss.v076.i01

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America, 121*(2), 1157–1163. https://doi.org/10.1121/1.2404624

Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience, 30*(1–2), 1–31. https://doi.org/10.1080/23273798.2014.963130

Dilley, L. C., & Breen, M. (2018). An *enhanced* autosegmental-metrical theory (AM^{+}) facilitates phonetically transparent prosodic annotation: A reply to Jun. In J. Barnes & S. Shattuck-Hufnagel (Eds.), *Prosodic theory and practice* (pp. 67–71). MIT Press.

Dilley, L. C., & Brown, M. (2005). *The RaP (Rhythm and Pitch) labeling system* (Version 1.0). http://speechlab.cas.msu.edu/rap-system.htm

Dilley, L. C., Mattys, S., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language, 63*(3), 274–294. https://doi.org/10.1016/j.jml.2010.06.003

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59*(3), 294–311. https://doi.org/10.1016/j.jml.2008.06.006

Fairbanks, G., & Guttman, N. (1958). Effects of delayed auditory feedback upon articulation. *Journal of Speech and Hearing Research, 1*(1), 12–22. https://doi.org/10.1044/jshr.0101.12

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *The Journal of the Acoustical Society of America, 97*(3), 1893–1904. https://doi.org/10.1121/1.412063

Friederic i, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences, 16*(5), 262–268. https://doi.org/10.1016/j.tics.2012.04.001

Guenther, F. H. (2016). *Neural control of speech*. MIT Press. https://doi.org/10.7551/mitpress/10471.001.0001

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., & Kenney, M. K. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Experimental Brain Research, 130*(2), 133–141. https://doi.org/10.1007/s002219900237

Hedberg, N., Sosa, J. M., & Fadden, L. (2004). *Meanings and configurations of questions in English* [Conference session]. Second International Conference on Speech Prosody, Nara, Japan (pp. 309–312).

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346–363. https://doi.org/10.1002/bimj.200810425

Kakouros, S., Salminen, N., & Räsänen, O. (2018). Making predictable unpredictable with style—Behavioral and electrophysiological evidence for the critical role of prosodic expectations in the perception of prominence in speech. *Neuropsychologia, 109,* 181–199. https://doi.org/10.1016/j.neuropsychologia.2017.12.011

Kempster, G. B., Larson, C. R., & Kistler, M. K. (1988). Effects of electrical stimulation of cricothyroid and thyroarytenoid muscles on voice fundamental frequency. *Journal of Voice, 2*(3), 221–229. https://doi.org/10.1016/S0892-1997(88)80080-8

Kim, J. H., & Larson, C. R. (2019). Modulation of auditory-vocal feedback control due to planned changes in voice f_0. *The Journal of the Acoustical Society of America, 145*(3), 1482–1492. https://doi.org/10.1121/1.5094414

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America, 118*(2), 1038–1054. https://doi.org/10.1121/1.1923349

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press. https://doi.org/10.1017/CBO9780511808814

Larson, C. R., Burnett, T. A., Kiran, S., & Hain, T. C. (2000). Effects of pitch-shift velocity on voice F_0 responses. *The Journal of the Acoustical Society of America, 107*(1), 559–564. https://doi.org/10.1121/1.428323

Larson, C. R., Kempster, G. B., & Kistler, M. K. (1987). Changes in voice fundamental frequency following discharge of single motor units in cricothyroid and thyroarytenoid muscles. *Journal of Speech and Hearing Research, 30*(4), 552–558. https://doi.org/10.1044/jshr.3004.552

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776

Liberman, M. Y. (1975). *The intonational system of English* [Doctoral dissertation]. Massachusetts Institute of Technology.

Liu, H., Auger, J., & Larson, C. R. (2010). Voice fundamental frequency modulates vocal response to pitch perturbations during English speech. *The Journal of the Acoustical Society of America, 127*(1), EL1–EL5. https://doi.org/10.1121/1.3263897

Liu, H., & Larson, C. R. (2007). Effects of perturbation magnitude and voice F_0 level on the pitch-shift reflex. *The Journal of the Acoustical Society of America, 122*(6), 3671–3677. https://doi.org/10.1121/1.2800254

Liu, H., Xu, Y., & Larson, C. R. (2009). Attenuation of vocal responses to pitch perturbations during Mandarin speech. *The Journal of the Acoustical Society of America, 125*(4), 2299–2306. https://doi.org/10.1121/1.3081523

Mahrt, T., Cole, J., Fleck, M., & Hasegawa-Johnson, M. (2012). *Modeling speaker variation in cues to prominence using the Bayesian information criterion* [Conference session]. Sixth International Conference on Speech Prosody, Shanghai, China.

Marschner, I., & Donoghoe, M. W. (2018). Package "glm2." *The R Journal, 3*(2), 12–15.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). *Montreal Forced Aligner: Trainable text-speech alignment using Kaldi* [Conference session]. 18th Annual Conference of the International Speech Communication Association (INTERSPEECH), Stockholm, Sweden (pp. 498–502). https://doi.org/10.21437/Interspeech.2017-1386

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review, 23*(1), 103–123. https://doi.org/10.3758/s13423-015-0947-8

Munhall, K. G., MacDonald, E. N., Byrne, S. K., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *The Journal of the Acoustical Society of America, 125*(1), 384–390. https://doi.org/10.1121/1.3035829

Nalborczyk, L., Batailler, C., Lœvenbruck, H., Vilain, A., & Bürkner, P.-C. (2019). An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian. *Journal of Speech, Language, and Hearing Research, 62*(5), 1225–1242. https://doi.org/10.1044/2018_JSLHR-S-18-0006

Natke, U., & Kalveram, K. T. (2001). Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *Journal of Speech, Language, and Hearing Research, 44*(3), 577–584. https://doi.org/10.1044/1092-4388(2001/045)

Patel, R., Niziolek, C., Reilly, K. J., & Guenther, F. H. (2011). Prosodic adaptations to pitch perturbation in running speech. *Journal of Speech, Language, and Hearing Research, 54*(4), 1051–1059. https://doi.org/10.1044/1092-4388(2010/10-0162)

Patel, R., Reilly, K. J., Archibald, E., Cai, S., & Guenther, F. H. (2015). Responses to intensity-shifted auditory feedback during running speech. *Journal of Speech, Language, and Hearing Research, 58*(6), 1687–1694. https://doi.org/10.1044/2015_JSLHR-S-15-0164

Perlman, A. L., & Alipour-Haghighi, F. (1988). Comparative study of the physiological properties of the vocalis and cricothyroid muscles. *Acta Oto-Laryngologica, 105*(3–4), 372–378. https://doi.org/10.3109/00016488809097021

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* [Unpublished doctoral dissertation]. Massachusetts Institute of Technology.

RStudio Team. (2015). *RStudio: Integrated development for R* [Computer software]. RStudio, Inc. http://www.rstudio.com/

Shattuck-Hufnagel, S. (2000). Phrase-level phonology in speech production planning: Evidence for the role of prosodic structure. In M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 201–229). Springer. https://doi.org/10.1007/978-94-015-9413-4_8

Sivasankar, M., Bauer, J. J., Babu, T., & Larson, C. R. (2005). Voice responses to changes in pitch of voice or tone auditory feedback. *The Journal of the Acoustical Society of America, 117*(2), 850–857. https://doi.org/10.1121/1.1849933

Xu, Y., Larson, C. R., Bauer, J. J., & Hain, T. C. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *The Journal of the Acoustical Society of America, 116*(2), 1168–1178. https://doi.org/10.1121/1.1763952