



Semantic focus mediates pitch auditory feedback control in phrasal prosody

Allison I. Hilger, Jennifer Cole & Charles Larson

To cite this article: Allison I. Hilger, Jennifer Cole & Charles Larson (2022): Semantic focus mediates pitch auditory feedback control in phrasal prosody, *Language, Cognition and Neuroscience*, DOI: [10.1080/23273798.2022.2116060](https://doi.org/10.1080/23273798.2022.2116060)

To link to this article: <https://doi.org/10.1080/23273798.2022.2116060>

 [View supplementary material](#) 

 Published online: 01 Sep 2022.

 [Submit your article to this journal](#) 

 [View related articles](#) 

 [View Crossmark data](#) 

Semantic focus mediates pitch auditory feedback control in phrasal prosody

Allison I. Hilger ^{a,b}, Jennifer Cole^c and Charles Larson^a

^aDepartment of Communication Sciences and Disorders, Northwestern University, Evanston, IL, United States; ^bCurrent department: Department of Speech, Language, and Hearing Sciences, University of Colorado Boulder, Boulder, CO, United States; ^cDepartment of Linguistics, Northwestern University, Evanston, IL, United States

ABSTRACT

This study investigated the effect of semantic focus on pitch auditory feedback control in the production of phrasal prosody through an experiment using pitch-shifted auditory feedback. We hypothesized that pitch-shift responses would be mediated by semantic focus because highly informative focus types, such as corrective focus, impose more specific constraints on the prosodic form of a phrase and require greater consistency in the production of pitch excursions compared to sentences with no such focus elements. Twenty-eight participants produced sentences with and without corrective focus while their auditory feedback was briefly and unexpectedly perturbed in pitch by ± 200 cents at the start of the sentence. The magnitude and latency of the reflexive pitch-shift responses were measured as a reflection of auditory feedback control. Our results matched our prediction that corrective focus would elicit larger pitch-shift responses, supporting our hypothesis that auditory feedback control is mediated by semantic focus.

ARTICLE HISTORY

Received 27 September 2021
Accepted 15 August 2022

KEYWORDS

Prosody; auditory feedback; pitch shifts; feedback perturbations; intonation; semantic focus

Introduction

Auditory feedback is used to make online corrections in speech production; however, the use of auditory feedback for pitch control in the presence of focus-marking prosody is not clearly defined. In this study, we investigated reflexive auditory feedback control in relation to anticipatory semantic focus as a step toward understanding complex pitch control. As speech is produced, the auditory feedback control system monitors acoustic output and makes corrections if the actual output does not match the auditory targets of the intended output. These corrections are likened to a negative feedback loop in which there is a sensory target (in this case pitch) and deviations from that target are automatically and reflexively corrected (Hain et al., 2000). The degree of correction reflects both the degree of the detected deviation from the target as well as the precision of the target. Auditory feedback correction of pitch can be experimentally studied by synthetically creating perceived deviations in pitch in real-time and measuring the resulting reflexive response, known as the *pitch-shift reflex* (Behroozmand et al., 2012; Burnett et al., 1998; Hain et al., 2000; Kim & Larson, 2019; Larson & Robin, 2016; Scheerer & Jones, 2018). The goal of this study was to measure the pitch-shift reflex to unintended changes in pitch auditory feedback in sentences that

vary in the presence/absence of a prosodically marked focus. We theorize that there is greater precision in the implementation of a pitch target for a phrase with focus-marking prosody, and therefore, perceived deviations in pitch will result in larger reflexive responses for phrases with prosodically marked focus.

The pitch perturbation technique is an established research method for measuring pitch auditory feedback control. In this experimental paradigm, auditory feedback is perturbed during an online speaking task by modifying voice pitch feedback in real-time through headphones. When pitch auditory feedback is briefly and unexpectedly perturbed, speakers automatically produce a pitch-shift response (Behroozmand et al., 2012; Burnett et al., 1998; Hain et al., 2000; Kim & Larson, 2019; Larson & Robin, 2016; Scheerer & Jones, 2018). The pitch-shift response is thought to be reflexive because of the automatic and involuntary nature of the response and the inability for speakers to suppress it (Bauer & Larson, 2003; Burnett et al., 1998; Zarate & Zatorre, 2008). The magnitude and timing of the pitch-shift response has been measured as an indication of the efficiency and sensitivity of the auditory feedback control system to correct for errors in voice.

Precision of the voice auditory pitch target. Although the correction in auditory feedback does not

fully compensate for the mismatch, the overall degree of correction is dependent on the degree of mismatch between the intended and actual output (Liu et al., 2009). A larger mismatch, then, will result in a larger correction. The degree of mismatch is calculated by the amount of error as well as the precision of the auditory target, where the precision of the auditory target is scaled by the production task. For example, an auditory target for pitch will be more precise when singing a melody than during casual speech production. When singing a melody, the sensory pitch target should precisely match the pitch in the melody, whereas in speaking, pitch target values are relative, not absolute. For instance, variation in absolute pitch across utterances can arise from contextual factors, e.g. clear vs. casual speech style or the speaker's psychological state (Cole et al., 2019), but relative pitch differences reflecting semantic focus distinctions can nevertheless be implemented, with focused words generally having larger pitch excursions and overall greater acoustic prominence than non-focused words in the same sentence (Katz & Selkirk, 2011; Breen et al., 2010).

Studies comparing the pitch-shift response between singing and speaking provide evidence for precision effects: People make larger reflexive corrections in pitch to sudden pitch-shifts in auditory feedback during singing than speech production (Natke et al., 2003). When the same magnitude pitch perturbation results in a larger pitch-shift response during singing than during speaking, it suggests that the sensory pitch target for singing is more precise than for speaking. Additional evidence of the task-dependent nature of the pitch-shift reflex is in the finding of larger and faster pitch-shift reflexes in sentence-production than in simple sustained-vowel production (Chen et al., 2007; Hilger et al., 2020; Liu et al., 2009). While most of the auditory feedback research has focused on pitch-shift reflex magnitude, the speed of the response following the perturbation also reflects the demands and precision of the vocal task. A more demanding task, such as producing intonation in speech, should require faster correction than simple sustained vowel production (Chen et al., 2007). Furthermore, the pitch-shift reflex is also modulated by language if the language requires more precise pitch differentiation, such as Cantonese vs. Mandarin (Liu et al., 2010). The *economy of effort* theory by Lindblom (1983, 1990) provides a potential explanation for this task dependency in auditory feedback control. Essentially, the motor speech system employs a strategy to scale the auditory target according to the demands of the speaking task and the communicative context as a way to retain efficiency while maintaining intelligibility (Guenther,

2016). For example, target regions for speech sounds shrink when speakers are asked to speak more clearly, resulting in more precise articulation (Perkell et al., 2002). Viewed from this perspective, the motor speech system is seen as highly efficient in the generation of sensory pitch targets for different tasks: singing requires more precise pitch control than speech and so the pitch auditory target is more precisely specified in singing tasks than in speaking.

Building on these findings, the present study investigates the effect of semantic focus on the pitch-shift reflexive response. We hypothesize that prosodic encoding of corrective focus involves more precisely specified pitch targets compared to sentences without such a focused word, and that this greater precision increases the degree of auditory feedback correction to a perceived pitch error in online speech production. Before explaining the motivation for this hypothesis, we first present a brief overview of the prosodic encoding of semantic focus in English.

Prosodic encoding of focus in English. In English, and other Germanic languages, a word that is stressed at the phrase level has greater prominence than nearby words in the same phrase and is generally associated with words that contribute information to the discourse. The stressed word includes words that introduce new entities (i.e. *referents*) and words that have semantic focus, e.g. to convey correction, express contrast with semantic alternatives, or provide the answer to a preceding question (Büring, 2016). Phrasal stress is typically marked through the assignment of a tonally specified *pitch accent* that defines a pitch target (high, low, rising or falling) for the stressed word.¹ The choice among pitch accent types is determined in part by information structure: the specification of a word as discourse-new or -given, or as conveying focus within the phrase. While high, rising pitch accents are commonly used for any word that contributes new information, the most acoustically prominent, steep-rising pitch accents are associated with focus (Breen et al., 2010; Cooper et al., 1985; Katz & Selkirk, 2011; Pierrehumbert & Hirschberg, 1990).² Prosodic marking of information structure is variable in American English (Chodroff & Cole, 2019), especially for words that convey new information, but the use of a steep-rising accent is preferred in the context of contrastive or corrective focus (Breen et al., 2010; Turnbull et al., 2015).

As an example of a *new information* context, imagine there are two speakers who see each other after a long day:

Speaker 1: What did you do today?

Speaker 2: I went to the **gym**.

In this example, “gym” is realized with phrasal stress (marked with boldface) and conveys new information simply because “gym” has not yet been uttered by either speaker in the conversation and is not inferable from the sparse preceding discourse. As the only word in the sentence with phrasal stress, “gym” will be perceived as having greater prominence than other words in the sentence (Cole et al., 2019). It is also likely to be realized with the shallow-rising pitch accent that is conventionally used for new information.

A word may also signal *corrective focus*, replacing wrong information in a previous utterance (Katz & Selkirk, 2011; Ouyang & Kaiser, 2015). Imagine a communicative scenario where one person returns after being away from a setting for a couple of hours:

Speaker 1: Were you just at the store?

Speaker 2: I went to the **gym**.

In this example, Speaker 1 makes an assumption that Speaker 2 corrects. The word “gym,” in this scenario again has phrasal stress, but here it is likely to be produced with even greater prominence relative to the surrounding words, and with a steep-rising pitch accent. A notable feature of sentences with corrective focus is that the focused word is typically and preferentially the only word in the sentence to realize a pitch accent (Gussenhoven, 2007). Preceding and following words, including content words that may otherwise be assigned phrasal stress (in longer sentences than the example above), are unaccented, and realize no salient pitch movement. This results in a distinctive prominence peak in the focused word, signaling that the listener should substitute this word for the incorrect semantic alternative, in this case the word “store” in the previous utterance.

Prosodic marking is arguably more critical for conveying corrective focus than it is for signaling the new information status of a word. New information can, in most cases, be identified solely on the basis of the prior discourse context and is especially obvious for a word that has not been previously mentioned, and for which there is no antecedent in the preceding discourse. Corrective focus is different in that it requires that the listener establish a link between the referent of the focused word and an antecedent that is present or accessible from the prior discourse. Although the speaker has the option to explicitly convey that relationship (e.g. “I didn’t go to the store, I went to the gym”), prosodic marking of corrective focus is an alternative, and in some situations may be preferred as a politeness strategy to avoid the direct negation of a previous assertion.

Hypothesis of precision scaling. We propose that sensory pitch targets are more precisely defined in sentences with corrective focus than in sentences that convey new information without corrective focus (hereafter, “new information”). This follows from the observation that corrective focus is prosodically encoded not only through the salient rising pitch accent on the focused word, but also through the absence of pitch prominence on preceding and following words. In comparison, the sensory pitch target may be less precise for new information focus, in line with the observation that the prosodic expression of new information focus is overall more variable (Stavropoulou & Baltazani, 2021). Furthermore, we reason that because prosodic encoding is often the sole expression of corrective focus and is therefore more critical compared to its lesser role in signaling new information, the pitch encoding of corrective focus requires greater auditory feedback control to ensure its acoustic salience. Therefore, we expect that the speech motor system will be more sensitive to deviations in pitch auditory feedback in sentences with corrective focus. Moreover, we expect greater sensitivity to pitch deviations not only in the corrective focus word, but across the entire sentence.³

To test this hypothesis, we used a vocal production task that elicited sentences with corrective focus and new information sentences (without corrective focus) while randomly perturbing pitch auditory feedback direction for a brief duration at the start of the sentence. We predicted that the magnitude of the pitch-shift response would be greater in sentences with corrective focus than with new information as a reflection of the task-dependent role of the auditory feedback control system. If the pitch auditory targets are more precisely defined in sentences with corrective focus than in new information, unexpected changes in pitch in a corrective focus sentence will elicit a larger reflexive response due to a greater mismatch between the auditory target and the perceived error.

Because our primary aim here is to assess the effect of phrasal prosody on reflexive auditory feedback control, we report results related to the pitch-shift reflex. An additional question is whether perturbed auditory feedback early in the sentence has further downstream effects outside of the short window of the reflexive response. For reasons of space, we do not examine such potential downstream effects here, but we glean some insight into that question from our recent work showing that speakers enhance the production of the word with phrasal stress in response to pitch-shifts earlier in the phrase (Hilger et al., 2020). In that study speakers produced the phrase, “You know Nina?” (with phrasal stress on “Nina”), while short, unexpected

pitch-shifts were applied at the start of the phrase. Overall, speakers enhanced the stressed word (i.e. “Nina”) by increasing vowel duration, intensity, and fundamental frequency. These results indicate that speakers use auditory feedback to scale anticipatory intonation targets. However, the experiment design did not manipulate the information structure of any words in the sentence, so the results do not address the present research question of whether downstream effects of perturbed auditory feedback vary in relation to the presence or absence of corrective focus on the word with phrasal stress.⁴

In the present paper, we measured an additional variable related to the pitch-shift response: the direction of the response in relation to the direction of the perturbation. Although a majority of pitch-shift responses compensate for the perceived error in voice f_0 (fundamental frequency) by opposing the direction of the perturbation (termed *opposing responses*), responses that follow the perturbation direction have also been observed (termed *following responses*) (Behroozmand et al., 2012; Burnett et al., 1998; Franken et al., 2018; Hain et al., 2000; Kim & Larson, 2019). Opposing responses are thought to reflect a negative feedback control system to compensate for vocal errors (Hain et al., 2000). Following responses, on the contrary, are not as well-understood. A potential explanation is that following responses operate as part of the feedforward control system to achieve a certain vocalization (Patel et al., 2014). Because response direction is not well-understood, we opt to measure the effect of semantic focus on response direction to observe whether speakers opposed or followed the response more under certain perturbation direction (\pm 200 cent perturbation) and semantic focus conditions (i.e. corrective vs. new information focus) or whether response magnitude or latency changed as an effect of response direction.

The theoretical contribution from this research would be greater understanding into how the auditory feedback control system monitors and corrects for errors in pitch related to prosodically marked semantic focus. If there are differences in the pitch-shift reflex based on prosodically marked semantic focus, these results would support theories in motor speech that sensory targets for pitch are scaled by the production task, and furthermore, by semantic focus. The results would also imply that sensory pitch targets are more precisely defined for corrective focus (compared to sentences without corrective focus), indicating that auditory feedback control is more sensitive to deviations in pitch when producing prosodic marking for corrective focus.

Materials and methods

This current paper is part of a larger study on auditory feedback control in cerebellar ataxia (Hilger, 2020). In this larger study, speech recordings were obtained from 27 individuals with ataxia and 28 age- and sex-matched controls, and analyses included the pitch-shift response and change in production of the pitch-accented word. For the current paper, due to space and complexity limits, only data from pitch-shift responses from the control participants were analyzed. Results for the production of the pitch-accented word as well as the participants with ataxia are the subject of separate, forthcoming papers.

Participants

Twenty-eight adults, with no reported history of speech, language, or neurological impairment, were recruited for this study (10 men, 18 women). Participants were recruited through community flyers, social media posts, and word of mouth. All participants were native speakers of American English. Ages ranged from 24–79 years ($M = 54.1$, $SD = 15.0$). The large age range is due to participants being age-matched to the other group of participants (cerebellar ataxia) not included in this study. All participants passed hearing screenings, indicating that despite differences in age, hearing status was normal for participation in this study. Years of education ranged from 12–22 years ($M = 17.3$; $SD = 2.1$). Participants had normal, or corrected to normal, visual acuity. Additionally, all participants passed a cognitive screening (Nasreddine et al., 2005).

Experiment overview

To investigate voice pitch auditory feedback control, we conducted a production study to elicit both sustained vowels and sentences. In the sustained vowel task, participants were instructed to repeatedly hold an /a/ sound for three seconds at a time. For the sentence production task, we used a visual world paradigm to elicit sentences conveying new information and those with corrective focus. For this analysis, only the sentence production task was analyzed. We used a repeated-measures within- and between-subjects design with two independent variables of *focus* (new information or corrective) and *perturbation direction* (\pm 200 cents; 100 cents = 1 semitone). Two dependent variables were analyzed, pitch-shift reflex magnitude (cents) and peak latency (milliseconds).

Experimental testing occurred in a number of places: the Speech Physiology Lab at Northwestern University, a

rented office space in Downtown Chicago, and in a quiet room in participants' homes. Our goal was to increase accessibility of this study to participants from the local community by providing more convenient locations for participation. The recording environments were similar across sites and all testing occurred prior to the COVID-19 pandemic.

Instrumentation

A similar experimental setup was used to conduct the auditory feedback perturbation paradigm as previous studies (Burnett et al., 1998; Chen et al., 2013; Kim & Larson, 2019; Liu & Larson, 2007). Participants wore Etymotic Insert Earphones (model ER2-14A) and vocalized into an over-ear microphone (AKG, model C420) positioned approximately one inch from the corner of the mouth. 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 to the participant (Quadravox, Eventide). Brief 200-msec (millisecond) pitch perturbations of +200 cents, -200 cents, or 0 cents (control trials) of the voice f_0 were presented over the headphones in real time with approximately a 12-msec delay.

In order to mask the participant's bone-conducted feedback, a gain of about 10 dB SPL (Decibel Sound Pressure Level) was applied to the headphone auditory feedback of the participant's voice resulting in auditory feedback of around 80-85 dB SPL (Aphex Headpod 4). Recordings of the microphone signal, auditory feedback, and timing pulses to mark the pitch perturbation onset were obtained using a multi-channel recording system (AD Instruments, model ML785, PowerLab A/D converter) and LabChart software (AD Instruments, v.7.0) with a sampling rate of 20 kHz. Recordings of speech output and timing pulses were then time-aligned in LabChart software for offline analysis. The timing pulses were used to differentiate pitch perturbation direction for acoustic analyses.

Design and stimuli

Participants followed instructions from a computer monitor and were told to vocalize at a comfortable but stable pitch and loudness level. To elicit corrective focus in the sentence production task, participants produced instructions within a visual world paradigm modeled from Ouyang and Kaiser (2015). They were told that they were playing a game with the computer (i.e. a computer-player), which would use the participant's verbal instructions to move the pictures on the screen accordingly. However, they were told that the

computer-player would occasionally make mistakes and move the wrong picture. Even though the trials were pre-designed, the participants were led to believe that the computer-player was listening to their instructions. Color pictures were presented on the screen in circular frames with the picture names displayed underneath each picture. Pictures were chosen from a normed set of standardized picture drawings (Duñabeitia et al., 2018).

Four pictures were presented at a time and arrows were used to indicate the instructions the participants should produce. For example, in Figure 1A, the participants are presented with four pictures and are cued to wait until the arrow appears on the screen. In Figure 1B, an arrow points from the picture of a *knee* to a picture of a *web*, so participants should say, "Lay your **knee** by your **web**." After the instruction is produced, they see a picture move on the screen that responds to their instruction either correctly or incorrectly. For example, in Figure 1C, the *whale* is moved next to the *web*, which is an incorrect response. Participants were instructed that when a picture is moved incorrectly, they are first to explain which picture was moved incorrectly and then to repeat the original instruction. For example, in Figure 1C, the *whale* is incorrectly moved by the *web*, so participants should say, "Not your **WHALE** by your **web**." Then, in Figure 1D, the participants should repeat the correct instruction with corrective emphasis, saying, "Lay your **KNEE** by your **web**." Finally, in Figure 1E, the correct picture is moved, and the next trial is initiated.

Figure 1B represents a trial that elicits sentences conveying new information, i.e. without corrective focus. The pictures presented on the screen within Figures 1A and 1B are new within the discourse context because they are within a new set of pictures. Therefore, when the instruction is produced, the name of the picture within the instruction cannot be inferred from the previous discourse context. Figures 1C and 1D both represent productions using corrective focus.

The carrier phrase used in this task was, "Lay/not your OBJECT by your LOCATION." This phrase was chosen because voicing is continuous across the production of the phrase (apart from the break in voicing for the /b/ sound in "by" and the /t/ sound in "not"). Continuous voicing was essential to implement pitch perturbations within the phrase and to measure a pitch-shift response, which both use pitch tracking analyses that require modal voicing. The target word that was manipulated in this task was always the word in the OBJECT position. Words in the OBJECT position occur in the middle of the phrase where modal voicing is frequently used. On the contrary, words in the LOCATION position occur at

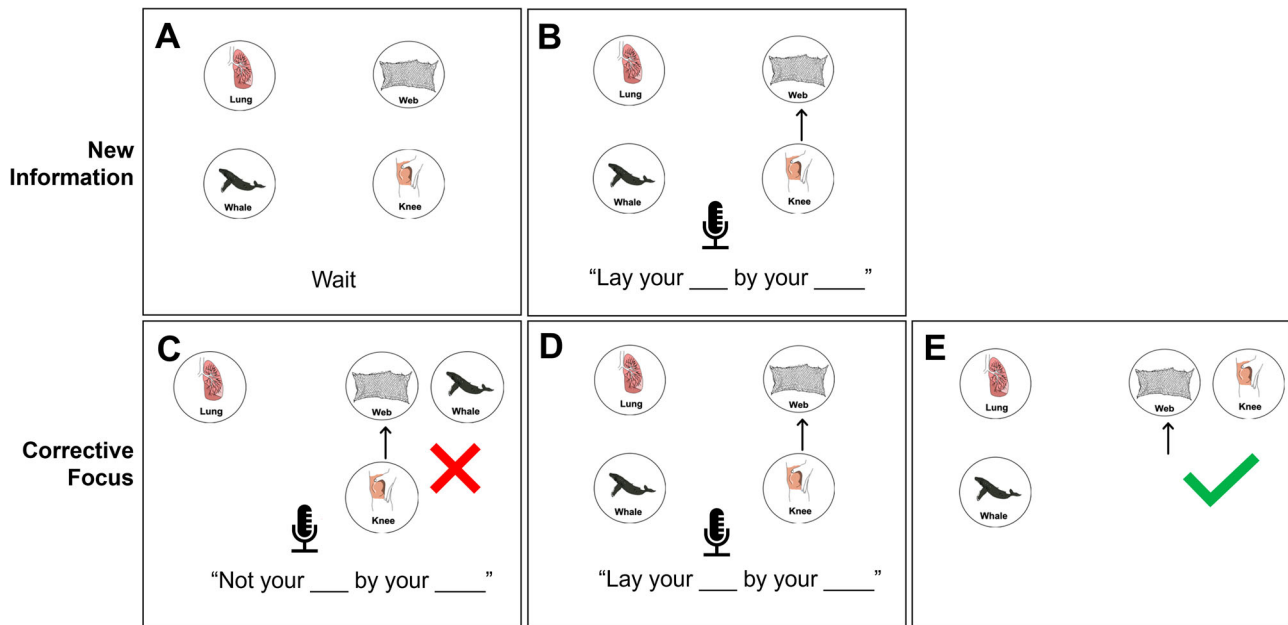


Figure 1. Sample display of the task. The first screen presented is 1A in which four pictures are presented with a cue to wait. In 1B, an arrow appears between *knee* and *web*, cueing the participant to produce the instruction (new focus), “Lay your **knee** by your **web**.” In 1C, an incorrect picture is moved, and the participant is cued to provide a corrective statement, “Not your **WHALE** by your **web**.” In 1D, the participant is cued to repeat the original instruction with corrective emphasis, “Lay your **KNEE** by your **web**.” In 1E, the correct picture is moved.

the end of the phrase, which is a position highly vulnerable to creaky voice (Kreiman, 1982). By manipulating words in the OBJECT position, we hoped to elicit productions conveying new information and corrective focus for these as our target words, with modal voicing. Target words were chosen from the MultiPic pictures that were monosyllabic and contained all voiced sounds. Participants produced a total of 250 instructive phrases that were subdivided into five blocks of 50 trials each. Within each block, there were around 20 trials each of new information and corrective phrases, depending on the ordering of the pictures within a trial.

The elicited information structure distinctions in this study are based on the participant’s perspective of the listener’s perspective (in this case the computer-player). It is the prior discourse context, which in our study is shared between the participant and the computer-player, that establishes the status of a word as conveying new information or corrective focus. Prior work shows that focus is influenced by listener-oriented processes (Lam & Watson, 2010; Watson et al., 2008). Therefore, we used the listener’s perspective (i.e. the computer-player) to define new information and corrective focus. Figure 2 displays pitch tracks from individual productions that illustrate the distinct pitch patterns of new information and corrective focus utterances. In the corrective focus production (2B) there is a large

pitch excursion on the object and small excursions (or relatively flat pitch) on the portions of the utterance preceding and following the object. In the new information production, the pitch excursion on the object is not as prominent in relation to the preceding and following words in the utterance.

The distinction between the two pitch patterns can be measured in terms of the difference in the mean f_0 values in the word “lay” and the object: for corrective focus productions this difference is predicted to be larger than for new information productions. Figure 3 shows the distribution of by-participant mean f_0 difference measures over all new information and corrective focus trials, confirming that participants in this study did indeed produce the predicted distinction in the pitch patterns for new information and corrective focus utterances.

To study the effect of semantic focus on pitch auditory feedback control, brief pitch perturbations were applied in random trials of sentence production. The perturbation magnitude used in this study was +/- 200 cents. Pitch perturbations were applied 50 msec after voice onset on the first word in the phrase (i.e. *lay* or *not*) on random trials. Perturbations were 200 msec in duration before auditory feedback was switched back to normal (i.e. unperturbed). We chose to apply the perturbation on the first word in the phrase for two reasons:

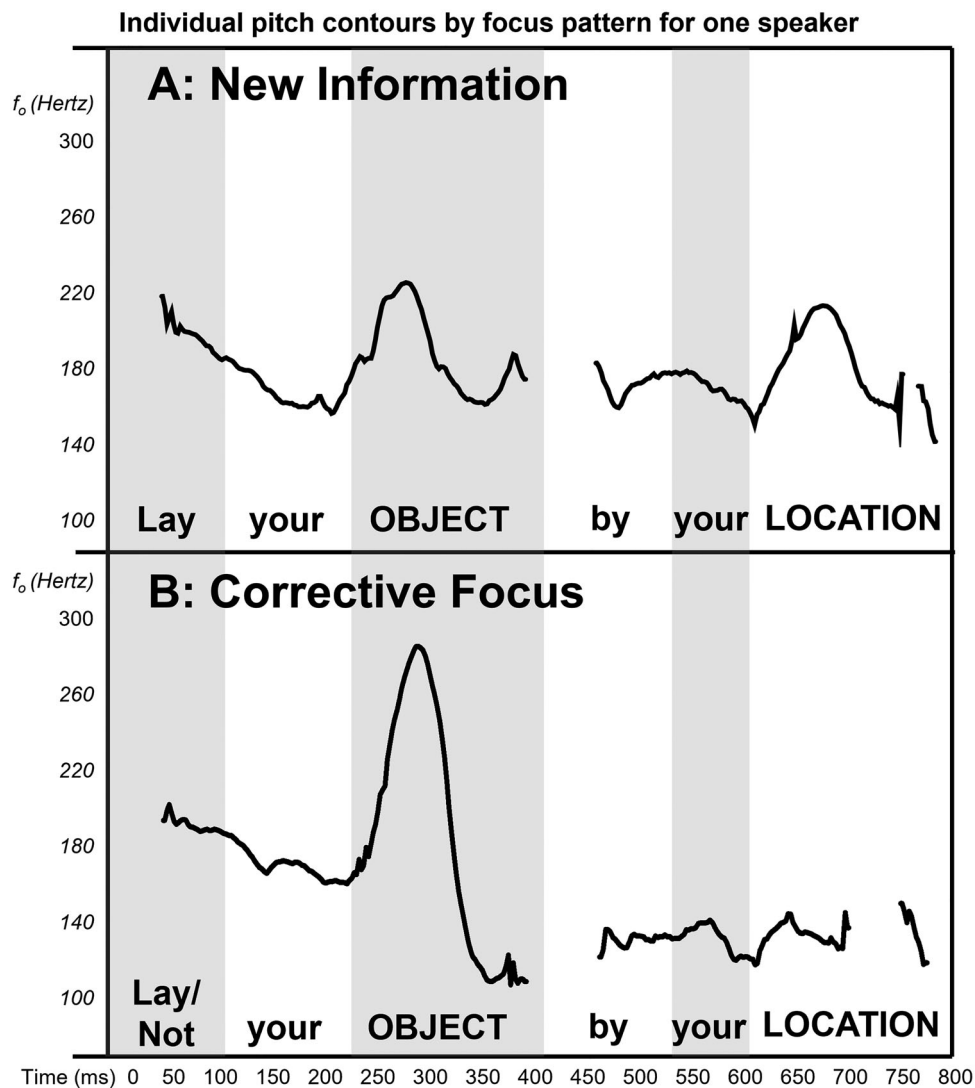


Figure 2. Pitch contours from individual productions for one female participant. The pitch contours are segmented by the words used in the experimental phrases. The word in the “OBJECT” position was the focus-bearing word in this study. In the corrective focus production, for example, “Not your LAMB by your WHALE,” (2B) there is a large pitch excursion on the object and small excursions (or relatively flat pitch) on the portions of the utterance preceding and following the object. In the new information production, for example, “Lay your KNEE by your WHALE,” (2A), the pitch excursion on the object is not as prominent in relation to the preceding and following words in the utterance.

(1) there is evidence that auditory feedback control is more sensitive at the start of the phrase, possibly because the acoustic features at the start of the phrase are used as a reference to calibrate the relative acoustic production of the rest of the phrase (Hilger et al., 2020; Liu et al., 2007), and (2) we were interested in how auditory feedback control is utilized at the start of the phrase to prepare for anticipatory prosodic marking of semantic focus on the word with phrasal stress. Figure 4 displays an example production from a study participant producing corrective focus for the phrase, “Lay your **well** by your van” with *well* as the target word. In all of the elicited utterances, the target word (“well”, in this example) has phrase-level stress, marked by a pitch

accent. Both the pitch perturbation and the pitch-shift response occur well before the onset of the stressed word (i.e. before the object of the verb, “well”). At least thirty trials of each perturbation condition (i.e. +200 cents, –200, and 0 cents) were included for each sentence focus type (i.e. new vs. corrective focus), which has been shown to be sufficient for the signal averaging technique used in the pitch-shift response analysis (Bauer & Larson, 2003).

Acoustic analysis

Acoustic analyses were conducted (1) to compare acoustic correlates of phrasal stress (i.e. f_0 , intensity, and duration)

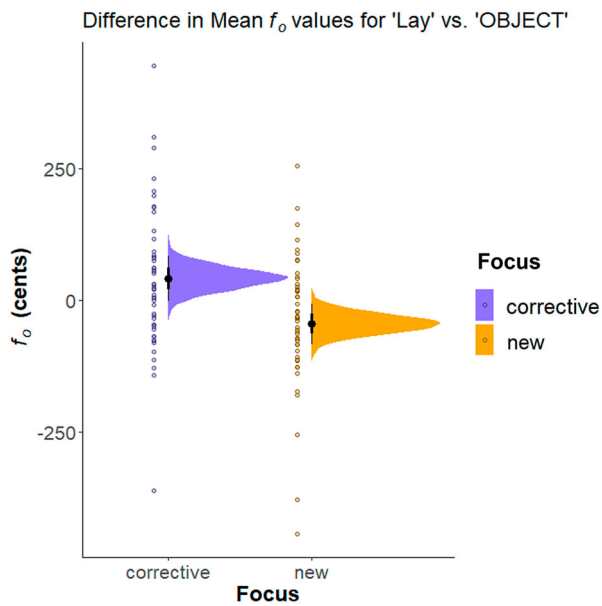


Figure 3. Median estimate and 95% credible interval for the difference in the mean f_0 values overall all new information and corrective focus trials between the word “lay” and the object word. Individual subject means are plotted along with the posterior distributions.

for new information vs. corrective focus of the target word, and (2) to measure the pitch-shift response (magnitude and latency) to the pitch perturbation stimuli as a

function of the type of semantic focus. Audio data from the voice recordings for each trial were first analyzed using autocorrelation-based pitch tracking in Praat software to transform the raw data into time-course measures of pitch (Praat Version 6.0.28; Boersma & Weenink, 2019). The phrase productions were then automatically segmented into individual words and phones using the Montreal Forced Aligner (McAuliffe et al., 2017). A final visual inspection of the segmentations was performed to confirm accuracy.

The recorded timing pulses for perturbation onset were aligned with the segmented audio files to label the onset and direction of the pitch perturbation within the production of each phrase. Trials were excluded if the onset of the pitch perturbation did not fall during the production of the first word in the phrase. Trials were also excluded that contained pitch tracking errors, hesitancy, disfluency, or mis-timings in the onset of the pitch perturbations. These exclusion criteria resulted in 25% of the trials being removed, the majority of which (approximately 90%) were due to mis-timings of the pitch perturbation, with the remaining 10% due to pitch tracking errors, hesitancy, or disfluency.⁵ The final set of 3,800 segmented audio files allowed us to measure (1) the production of phrasal stress (i.e. the target word), and (2) the pitch-shift response.

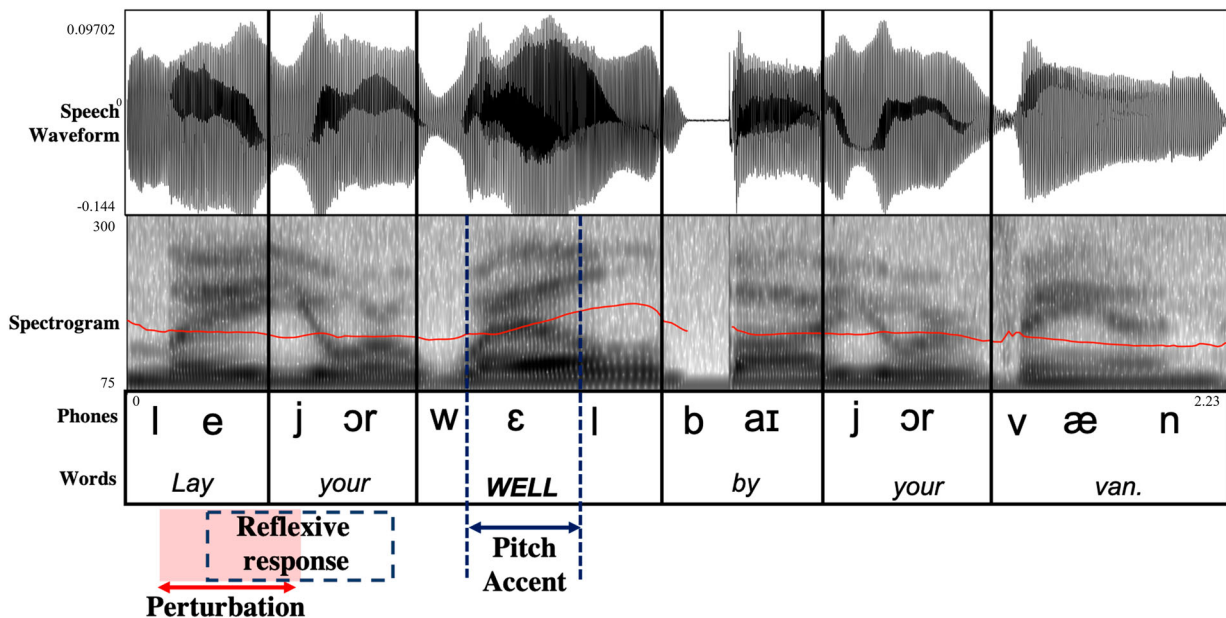


Figure 4. Example timing of the pitch perturbation in the phrase, “Lay your well by your van” for one study participant. The speech waveform (top) and the spectrogram (middle) are segmented by words and phonemes (bottom). The pitch track is displayed as a red line within the spectrogram. In this example, the target word, *well*, is the stressed word, termed “pitch accent.” The pitch perturbation occurs on the word *lay*, indicated by the red box and arrow at the bottom, and the pitch-shift response occurs shortly after the onset of the perturbation, indicated by the dashed box.

Analysis of phrasal stress

Duration, mean f_o , and mean intensity were extracted for each phone segment, and then grouped so that only the vowels in the target words (the object in each sentence production) were analyzed. f_o was converted from Hertz to cents using the following equation: $\text{Cents} = 1200 (\log_2(f_2/f_1))$ where f_1 equals the mean f_o of the first 50 ms of the trial and f_2 equals the mean f_o of the vowel in the target word.

Analysis of the pitch-shift reflex

Measurements of the pitch-shift responses (PSR) were also performed in Praat software. The voice f_o contours were epoched into segments from 50 ms before the perturbation onset (the baseline section) to 400 msec after the perturbation onset (post-perturbation window). The voice f_o contours in Hertz were extracted using Praat software and converted to the cent scale using the following formula: $\text{Cents} = 1200 (\log_2(f_2/f_1))$ where f_1 equals the mean f_o of the baseline section and f_2 equals the mean f_o of the post-perturbation window.

To isolate the PSR from the pitch movement due to phrasal intonation, we completed a difference wave analysis. Without the difference wave analysis, we would not be able to determine if a change in pitch was due to the pitch perturbation or from natural changes in intonation. The difference wave analysis was accomplished by subtracting out the average intonation contour per participant and focus pattern from each individual experimental trial. First, the control trials per participant per focus condition were averaged together to calculate the average intonation contour each participant produced. Then, the average intonation contour was subtracted from the individual experimental trials (i.e. trials with perturbations) for that participant and focus condition. The resulting pitch contours reflected changes in pitch from the pitch perturbation. By completing this analysis for each individual trial, we were able to subtract out variability in intonation that may occur trial-by-trial. This analysis technique has been successfully utilized to analyze the PSR in phrase production for a variety of intonation patterns (Chen et al., 2007; Patel et al., 2019).

The resulting difference waves were then sorted by response direction, i.e. whether the response opposed or followed the direction of the pitch perturbation. Response direction was calculated by comparing the mean f_o of the 50-msec window before perturbation onset with the mean f_o of the 400-msec 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.” We decided to include response direction in this analysis because it is not currently well-understood why speakers occasionally follow the perturbation instead of correcting (i.e. opposing) the unexpected change in pitch (Behroozmand et al., 2012; Franken et al., 2018). It is possible that semantic focus could be an important factor for learning more about following responses in auditory feedback control. Therefore, we included both opposing and following responses in our analyses.

After the voice contours were sorted by perturbation direction and response direction, an event-related averaging was completed by participant that reduced the noise in the audio signal and allowed for extraction of the response (Bauer & Larson, 2003). Essentially, the individual trials were grouped by participant, perturbation direction, and response direction, and then averaged together to compute a final averaged waveform. Response magnitude was then calculated by finding the maximal point (for upward responses) or the minimal point (for downward responses) in a window 60msec–300 msec post perturbation-onset. This analysis window was chosen to identify the response magnitude because the minimum latency of the pitch-shift reflex is approximately 60 msec after perturbation-onset, according to the timing of muscular activation and corresponding changes in f_o (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 300–400 ms window (Hain et al., 2000). Response latency was defined as the time-point of the peak (i.e. the maximal or minimal point) of the PSR.

Statistical analysis

All code for the statistical analyses is included in an RMarkdown file at <https://osf.io/3bhaq/>. Statistical analyses were conducted with R version 4.0.5 (R Core Team, 2022) using RStudio version 1.4.1103 (RStudio Team, 2020). Three Bayesian mixed effects models were run using the Stan modeling language (Carpenter et al., 2017) and the R package brms (Bürkner, 2017). A detailed description of Bayesian statistics is beyond the focus for this paper, however, please refer to Nalborczyk et al. (2019) for a guide to applying Bayesian statistics to speech acoustic research. Bayesian modeling was chosen in contrast to frequentist modeling because of the flexible ability to define hierarchical models that include the maximal random effect structure as recommended by Barr et al. (2013). For all three models,

weakly informative priors were specified for all model parameters. All models included maximal random effect structures, including a random intercept for participants and random slopes allowing the fixed effects to vary by participant.

The first model assessed the production of phrasal stress under the two conditions of semantic focus to verify that the production task successfully elicited distinct corrective and new information focus patterns. We fit this first model to mean f_o , mean intensity, and duration of the vowel of the stressed word, predicted by semantic focus (new information vs. corrective focus). For the model predictors of semantic focus, we used regularizing Gaussian priors adjusted by the dependent variable ($\mu=0$, $\sigma=10$ for mean intensity and duration; $\mu=0$, $\sigma=100$ for mean f_o), signifying that we assumed no effect of semantic focus on the dependent variables. For the random effects, a half Cauchy distribution was used for the standard deviation ($\mu=0$, $\sigma=0.1$ for mean intensity and duration, $\mu=0$, $\sigma=1$ for mean f_o) and an LKJ(2) distribution for the correlation. For the residual standard deviation, a half Cauchy distribution was used ($\mu=0$, $\sigma=1$).

The second model assessed changes in PSR magnitude and latency by semantic focus, perturbation direction, and response direction. We fit this second model to PSR magnitude (the absolute value of the PSR) and peak latency predicted by semantic focus (corrective vs. new focus), perturbation direction (+/−200 cent perturbation), and response direction (opposing vs. following response). For the model predictors, we used regularizing Gaussian priors ($\mu=0$, $\sigma=10$) for all variables, signifying that we assumed no effect of the predictors on PSR magnitude and latency. For the random effects, a half Cauchy distribution was used for the standard deviation ($\mu=0$, $\sigma=0.1$) and an LKJ(2) distribution for the correlation. For the residual standard deviation, a half Cauchy distribution was used ($\mu=0$, $\sigma=1$).

Finally, a generalized linear model was used for the third model to assess the effects of semantic focus (new information vs. corrective focus) and perturbation direction (+/−200 cent perturbation) on the number of opposing or following responses. For the model predictors, we used regularizing gaussian priors ($\mu=0$, $\sigma=10$), signifying that we assumed no effect of the predictors on the number of opposing or following responses. For the random effects, a half Cauchy distribution was used for the standard deviation ($\mu=0$, $\sigma=0.1$) and an LKJ(2) distribution for the correlation.

Four sampling chains with 2,000 iterations were run for each model, with a warm-up period of 1,000

iterations. We report 95% credible intervals (CI's) and probability of direction (pd) for each effect. Probability of direction is the probability that a parameter is positive or negative (Makowski et al., 2019). Given that a value of zero indicates no effect, a higher pd value indicates a greater probability that the effect is greater than zero. The 95% CI means that we are 95% certain that the true value lies within the specified interval. We determine whether there is compelling evidence for an effect by whether the 95% interval overlaps with zero, and pd is greater than 95%.

Results

Production of phrasal stress by semantic focus

Figure 5 and Table 1 display the median estimate and 95% credible interval for vowel duration, mean f_o , and mean intensity (dB) of the stressed word by semantic focus. Contingent on the data and model, there is compelling evidence that vowel duration was greater for the phrase-level stressed word in the new information condition ($\beta_{duration} = 232.71$ ms, 95%CI = [211.61, 254.54]) compared to corrective focus ($\beta_{duration} = 207.02$ ms, 95% CI = [185.67, 227.96]), but mean intensity of the phrase-level stressed word was greater for corrective focus ($\beta_{mean\ intensity} = 66.98$ dB, 95% CI = [65.69, 68.16]) than new information ($\beta_{mean\ intensity} = 66.57$ dB, 95% CI = [65.29, 67.75]). There was evidence (though not compelling) that mean f_o of the phrase-level stressed word increased for corrective focus ($\beta_{mean\ f_o} = -4.50$ cents, 95% CI = [−69.44, 60.16]) compared to new focus ($\beta_{mean\ f_o} = -30.37$ cents, 95% CI = [−88.40, 28.08]). Overall, mean f_o and mean intensity were increased for corrective focus, but duration was increased for new information focus.

Pitch-shift response magnitude and latency

Figure 6 and Table 2 show the 95% credible intervals and mean estimates for absolute response magnitude and response latency of the PSR by semantic focus (corrective vs. new focus), perturbation direction (+/−200 cents), response direction (opposing vs. following response), and the interactions among them. Contingent on the data and model, there is compelling evidence that PSR response magnitude was greater in sentences with corrective focus ($\beta = 114.09$ cents, 95% CI = [94.91, 134.51]) than new information ($\beta = 85.27$ cents, 95% CI = [72.36, 98.41]).

There were also robust two-way and three-interactions for PSR response latency that were driven by

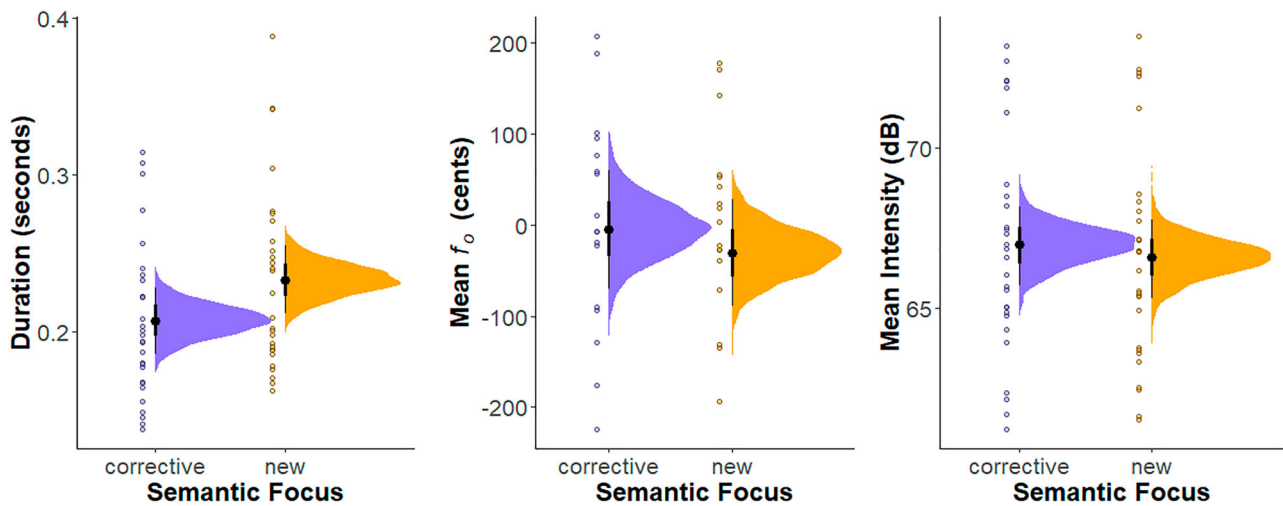


Figure 5. Median estimate and 95% credible interval for vowel duration (left), mean f_0 (middle), and mean intensity (right) of the stressed word by semantic focus. Individual subject means are plotted along with the posterior distributions.

Table 1. Median estimate and 95% credible interval for the Bayesian multivariate mixed effects regression model on the effect of semantic focus on the production of vowel duration, mean f_0 , and mean intensity of the stressed word. Probability of direction (pd) indicates the probability that the parameter is strictly positive or negative. Bolded parameters indicate compelling evidence for the effect. Random effects estimates are included for σ^2 , between-subject variance (τ_{00}), intra-class coefficient (ICC), number of subjects, and total number of observations.

Predictors	Duration			Mean f_0			Mean Intensity		
	Estimates	CI (95%)	pd	Estimates	CI (95%)	pd	Estimates	CI (95%)	pd
Intercept	0.21	0.19–0.23		–4.50	–69.44–60.16		66.98	65.69–68.16	
New Focus	0.03	0.02–0.03	100%	–26.32	–55.47–3.78	95%	–0.40	–0.63 – –0.18	100%
Random Effects									
σ^2	1.14								
τ_{00}	0.00								
ICC	0.73								
N_{subj}	28								
Observations	1451								

two responses with a later latency than all other responses: (1) opposing responses to -200 cent perturbations in sentences with new focus ($\beta = 244.04$ ms, 95% CI = [205.22, 290.53]) and (2) following responses to $+200$ cent perturbations in sentences with new focus ($\beta = 251.36$ ms, 95% CI = [212.81, 299.13]). Essentially, these two responses represent the PSR that moved upward in pitch in sentences with new information (e.g. opposing a downward perturbation or following an upward perturbation). In Figure 6, these responses can be viewed in the third row as the two distributions that are higher than the other distributions (indicating later peak latencies). The contrasts in the three-way interaction are shown in Figure 7 where robust contrasts are signified by the purple median dot indicating that the contrast does not overlap with zero. As can be seen by the bolded labels, the two responses described earlier were involved in every robust interaction. Overall, these interactions demonstrate that PSR's that move

upward in pitch in sentences with new focus had a later peak latency than all other responses.

Response direction

The final analysis compared the number of opposing and following responses by perturbation direction and focus condition to determine if the likelihood of an opposing vs. following response was conditioned by perturbation direction (upward vs. downward), or by focus condition (new information vs. corrective). Table 3 lists the average number of responses by experimental condition. A Bayesian generalized linear model was run to determine if the number of opposing or following responses differed by sentence focus or perturbation direction. Contingent on the data and the model, there were no robust differences in the average number of responses for any condition, indicating that speakers opposed and followed the perturbation for all perturbation direction and sentence focus conditions.

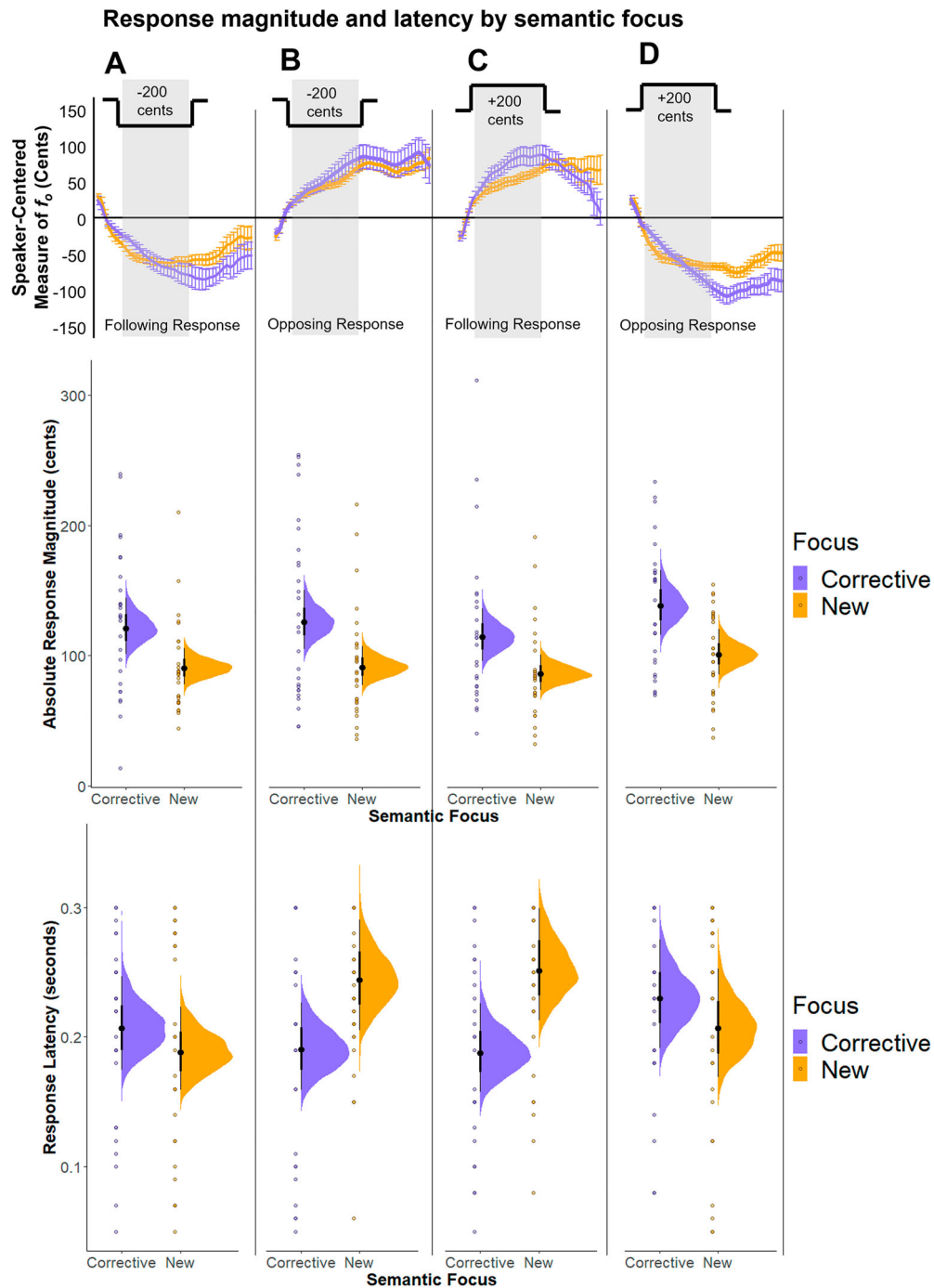


Figure 6. Pitch-shift reflex response magnitude and peak latency by semantic focus, perturbation direction, and response direction. The top row shows the averaged response curves with error bars representing standard error for corrective focus (purple) and new focus (orange). The grey bar within the top row indicates the onset and offset of the pitch perturbation. Columns A and B (i.e. the first two columns) show responses to -200 cent perturbations and columns C and D (i.e. the last two columns) show responses to $+200$ cent perturbations. Column A displays following responses to -200 cent perturbations and column B shows opposing responses to -200 cent perturbations. Column C displays following responses to $+200$ cent perturbations and column D shows opposing responses to $+200$ cent perturbations. The second row in the figure shows absolute response magnitude (cents), and the third row shows peak response latency (cents). For absolute response magnitude and peak response latency, the median estimate and 95% credible interval are shown alongside each posterior distribution. The perturbation conditions for the second and third rows correspond to the column descriptors at the top of the figure (i.e. Columns A, B, C, and D).

Table 2. Median estimate and 95% credible interval for the Bayesian multivariate mixed effects regression model on the effect of semantic focus, perturbation direction, response direction, and their interactions on absolute response magnitude and peak response latency of the pitch-shift reflex. Probability of direction (pd) indicates the probability that the parameter is strictly positive or negative. Bolded parameters indicate compelling evidence for the effect. Random effects estimates are included for σ^2 , between-subject variance (τ_{00}), intra-class coefficient (ICC), number of subjects, and total number of observations.

Predictors	Response Magnitude			Response Latency		
	Estimates	CI (95%)	pd	Estimates	CI (95%)	pd
Intercept	4.73	4.56–4.90	100%	–1.68	–1.85 – –1.50	100%
Focus	–0.29	–0.52 – –0.06	99.38%	–0.09	–0.34–0.14	78.40%
Perturbation Direction	–0.06	–0.25–0.13	72.47%	–0.10	–0.33–0.15	77.40%
Response Direction	0.04	–0.15–0.23	65.05%	–0.08	–0.32–0.15	74.52%
Focus: Perturbation Direction	0.00	–0.27–0.27	52.48%	0.39	0.05–0.72	98.83%
Focus: Response Direction	–0.03	–0.29–0.23	59.40%	0.34	0.01–0.67	97.95%
Perturbation Direction: Response Direction	0.15	–0.11–0.42	87.42%	0.28	–0.04–0.63	95.39%
Focus: Perturbation Direction: Response Direction	0.00	–0.36–0.38	50.05%	–0.74	–1.21 – –0.28	99.95%
Random Effects						
σ^2	0.01					
τ_{00}	0.01					
ICC	0.37					
N_{subj}	28					
Observations	220					

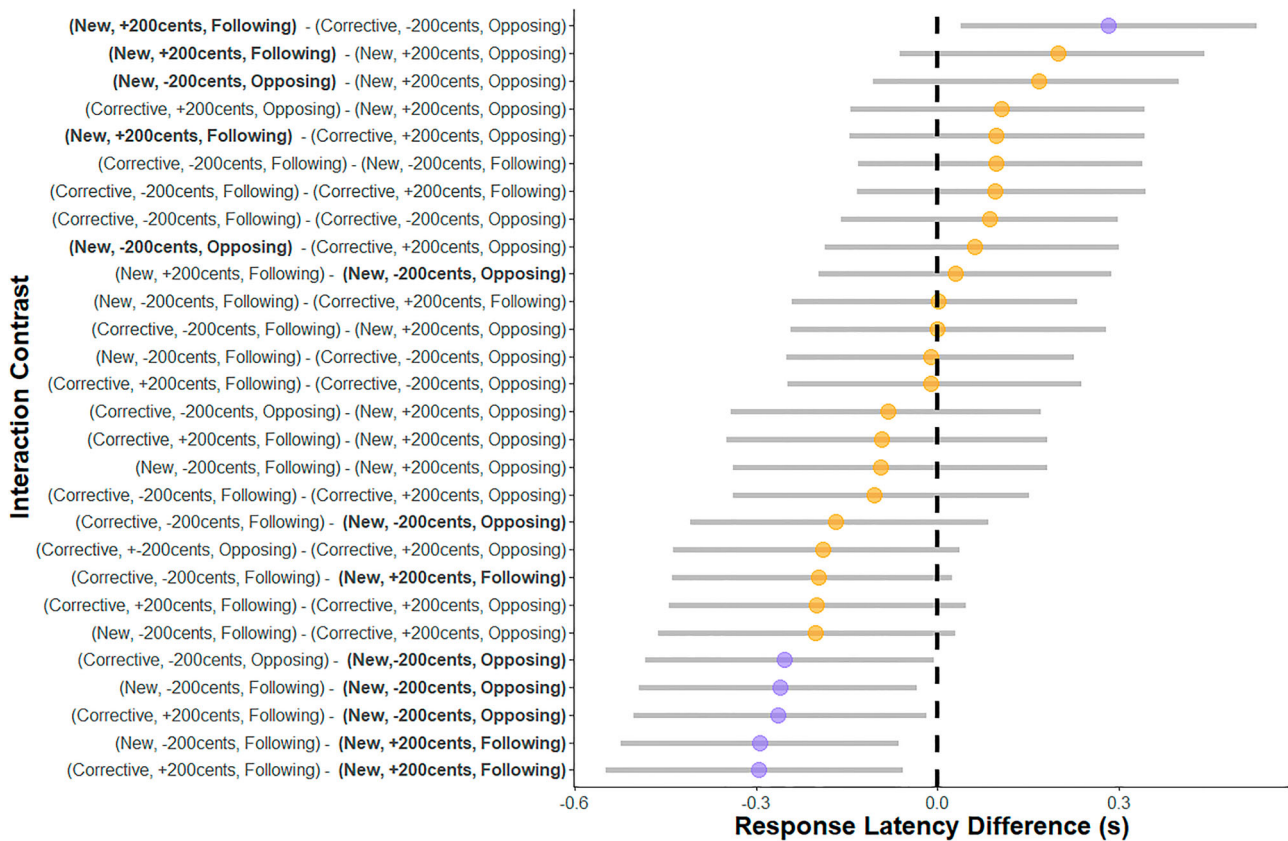


Figure 7. Median estimate and 95% credible interval for the individual contrasts in the three-way interaction among the fixed effects of semantic focus, perturbation direction, and response direction. Compelling evidence for a contrast is interpreted as a distribution that does not overlap with zero (represented by the purple vs. orange median estimate marks). Two combinations of factors are bolded (New, +200 cents, Following; New, –200 cents, Opposing) to visualize their role in driving these interactions.

Discussion

In this study, we proposed that sentences with corrective focus have more precise pitch auditory targets than sentences that convey new information (without

corrective focus). The pitch targets reflect the conventionalized prosodic encoding of corrective focus with a rising pitch excursion localized on the focus word and relatively flat pitch preceding and following the

Table 3. The average number of opposing and following responses by sentence focus and perturbation direction. Response direction (opposing and following) are listed on the left-hand column, followed by sentence focus and perturbation direction.

Response Direction	Sentence Focus	Perturbation Direction	Average Responses per Category
Opposing	New	+200 cents	36.1
		-200 cents	33.3
	Corrective	+200 cents	37.0
		-200 cents	34.6
Following	New	+200 cents	37.0
		-200 cents	38.8
	Corrective	+200 cents	28.5
		-200 cents	32.5

focused word, compared to the more variable prosodic encoding of new information focus with a smaller pitch excursion on the target (new information) word. Accordingly, we predicted that the reflexive pitch-shift response, a behavioral marker of auditory feedback control, would be larger in magnitude in sentences with corrective focus than in sentences conveying new information. This finding would indicate a greater need for precise pitch control with corrective focus, resulting in turn in a larger mismatch in auditory feedback between the sensory pitch target and the perceived deviation in pitch.

To measure pitch auditory feedback control for semantic focus, we elicited utterances with verbal objects expressing new information and sentences with verbal objects expressing corrective focus using a visual-world paradigm. We found that participants increased mean f_0 and mean intensity of the phrasally stressed word marked for corrective focus compared to those conveying new information without corrective focus. Additionally, the difference in mean f_0 between the stressed word (i.e. OBJECT word) and the first word of the phrase was greater for corrective focus than when conveying new information. These findings indicate that the study task successfully elicited differential pitch accent and prominence patterns marking the information structure contrast. Vowel duration showed the opposite effect, with longer duration in the new information condition than in corrective focus. This was a surprising finding and merits further study.

After it was established that prosodically marked distinctions related to semantic focus were elicited in this task, we then measured the pitch-shift response to the randomized pitch perturbations that occurred on the first word in the phrase (prior to the word with phrasal stress) and found a robustly larger response magnitude in sentences conveying corrective focus than those conveying new information. This result supports our hypothesis that auditory feedback control is more

sensitive to pitch errors in sentences with corrective focus, which further supports the claim that corrective focus requires a highly salient acoustic expression of pitch. More specifically, these findings indicate that prosodic phrases that contain a word with corrective focus have more precisely defined auditory targets for pitch across the phrase than do prosodic phrases that convey new information (without corrective focus). In relation to the theory of economy of effort (Lindblom, 1983), sensory pitch targets conveying distinctions in information structure are scaled for efficiency and communicative intent. When producing a sentence conveying new information, and lacking corrective focus, greater variation in pitch is allowed, possibly reflecting the fact that the information status of words in the phrase is relayed through the discourse context. In contrast, more precisely specified pitch movements are required for the expression of corrective focus, which we take to be related to the fact that corrective focus cannot necessarily be inferred from the discourse context alone, and which in the absence of an explicit statement of correction (e.g. "It's not X, but Y") requires additional acoustic prosodic cues. Therefore, the motor speech system specifies the sensory target for pitch more precisely for corrective focus than new information focus, maintaining efficiency in production while also achieving the communicative intent.

Looking more closely at pitch-shift response magnitude, we found no interaction among perturbation direction, semantic focus, or response direction. Although responses were larger for corrective focus than for new information, neither the direction of the perturbation nor the direction of the response influenced the magnitude of the response. However, we did observe a complicated interaction for peak response latency. Responses that were upward in pitch (i.e. following an upward perturbation or opposing a downward perturbation) were later in latency in sentences conveying new information compared to these same responses in sentences with corrective focus.

There are a few potential explanations for this result. First, there may be a physiological explanation: rising pitch movements generally take longer to produce than falling pitch movements (Xu & Sun, 2002). Therefore, the longer latency for upward pitch-shift responses may be due to physiological constraints in the speed in which a speaker can produce the pitch movements. However, this pattern was only observed for sentences with new information focus; there were no robust differences in response latency by response direction for sentences with corrective focus. The delay in peak latency was due to a combination of both response direction and sentence focus. It is possible that the faster response

is required in corrective focus despite the physiological constraint because the acoustic expression of corrective focus is more critical than is the acoustic expression of new information. Therefore, a more rapid response is required to move the f_0 closure to the intended target. A faster pitch-shift response is indicative of quicker neural processing of the error for correction (Chen et al., 2007). When pitch control is more important for a task, such as in sentence-production compared to simple sustained-vowel production, speakers produce a faster pitch-shift response. We make the assumption, then, that the motor speech system employs an economy of effort in the speed of the response as well as the magnitude. Faster correction requires more neural resource but is required when a specific pitch target is necessary for production. So, although upward pitch movements take longer to implement, the motor speech system will employ more articulatory effort to produce a faster corrective response where the communicative goal depends more critically on the acoustic encoding, as in the prosodic expression of corrective focus.

A surprising finding from this study was that participants produced an equal number of opposing and following responses across the experimental conditions. We predicted that speakers would oppose the perturbation direction more in sentences with corrective focus, however, our prediction was not born out because there were no interactions with focus. A possible explanation for the higher number of following responses than usual is that the speaker may have subconsciously attempted to anticipate and preemptively correct the feedback perturbations. There is evidence that people produce more following responses when the perturbation is more predictable (Behroozmand et al., 2012). Although the perturbations were randomized across trials by perturbation direction, when a perturbation did occur in a trial, it always occurred 50 msec after voice-onset. Therefore, the timing of the perturbation may have had higher predictability. Because the timing of the perturbation didn't vary (only the perturbation direction was varied), it is possible that speakers subconsciously anticipated a change in auditory feedback around the 50-msec time-point and utilized the feedforward system to preemptively change their pitch. This explanation would account for the movement in f_0 prior to the onset of the perturbation, as seen in the grand averages. Therefore, a limitation in this study is that extra trials should be included with randomized timing within the phrase to prevent an anticipation of the perturbation. By including this randomization, we may have elicited more opposing responses. However, this is speculative and should be further investigated.

Despite these limitations, clear pitch-shift responses were measured with a robustly larger response magnitude in sentences with corrective focus than in sentences conveying new information without corrective focus. These results demonstrate that pitch auditory feedback control is mediated by semantic focus. We theorize that the auditory targets for production of semantic focus differ according to the presence or absence of corrective focus, and thus, affect the degree of correction by the auditory feedback system. The theoretical implications of this research are that sensory pitch targets for intonation are scaled by semantic focus so that the motor speech system maintains efficiency while achieving the production of the focus pattern for the communicative message. Overall, the findings of this study demonstrate that auditory feedback control is mediated by semantic focus as a reflection of efficiency within the motor speech system to specify sensory pitch targets for intonation.

Conclusion

Overall, the findings from this study show that voice pitch auditory feedback control is mediated by semantic focus. We propose that auditory feedback control is more sensitive in sentences in which pitch plays an important role in marking information structure, such as corrective focus. Therefore, sensory errors in pitch elicit larger responses in sentences with corrective focus because of a greater mismatch between auditory feedback and the auditory target. These results have important implications for understanding the control of voice f_0 for intonation in speech production. It is evident that semantic focus has a strong effect on pitch auditory feedback control.

Notes

1. In the prevailing framework of Autosegmental-Metrical theory (Ladd, 2008), pitch accents in American English are in terms of L(ow), H(igh) and downstepped high (!H) tones: L*, H*, L+H*, L+H*, H+!H*, where the * indicates the tone that is anchored to the stressed syllable, and consequently, the pitch target for that syllable.
2. Within the literature in intonational phonology, there have been claims that corrective and contrastive focus are produced with an L+H* pitch accent, while words introducing new information are produced with an H* pitch accent. However, empirical support for that claim is disputed (Chodroff & Cole, 2019; Ladd & Morton, 1997). We do not address the tonal specification of focus-marking pitch accents in this paper.
3. In the framework of alternative semantics (Rooth, 1992), focus is understood as a semantic notion which references a set of semantic alternatives to the focused

expression and triggers a set of propositions which contrast in the focused element. Our hypothesis of precision scaling is grounded in this alternatives-based account of focus, and accordingly we predict precision scaling in the “wide” domain of the proposition (i.e., the sentence, in our materials). To the extent that we find evidence for precision scaling in a domain wider than the focused word, and within the domain bounded by the proposition, it can be taken as support for Rooth’s alternatives-based analysis.

4. Our work with data from the present experiment is ongoing, and the analysis of downstream effects of perturbed auditory feedback on the production of the focus-marked word is the subject of a manuscript in preparation.
5. The experimental software was set up to elicit the pitch perturbation 50 msec after voice onset. If the participant cleared their throat, clicked their tongue, or breathed too loudly, the software could mistake these noises as the onset of voice production. Therefore, the pitch perturbation was sometimes elicited before the phrase was produced or at the very onset of production. Every trial was manually checked and removed if the pitch perturbation did not occur 50 msec after voice onset. We anticipated the possibility of mistiming of the perturbation and, therefore, included many trials to have an adequate number after exclusion.

Acknowledgements

This research was funded by the NIH NIDCD F31 DC017877-01A1 and the Council of Academic Programs in Communication Sciences and Disorders. We would like to thank the participants for their time and effort to participate in this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Council of Academic Programs in Communication Sciences and Disorders: [Grant Number]; the National Institutes of Health National Institute of Deafness and Other Communication Disorders: [Grant Number DC017877-01A1].

Declaration of Interest Statement

The authors have no financial or non-financial relationships to disclose.

ORCID

Allison I. Hilger  <http://orcid.org/0000-0001-5507-2042>

References

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- 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>
- 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. <https://doi.org/10.1186/1471-2202-12-54>
- Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer [Computer program], Version 6.0.46. *Retrievable Online at [Http://www. Praat. Org](http://www.praat.org)*
- Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, 25(7–9), 1044–1098. <https://doi.org/10.1080/01690965.2010.504378>
- Büring, D. (2016). *Intonation and meaning*. Oxford University Press.
- Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 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>
- 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), 1–32. <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>
- Chen, X., Zhu, X., Wang, E. Q., Chen, L., Li, W., Chen, Z., & Liu, H. (2013). Sensorimotor control of vocal pitch production in Parkinson’s disease. *Brain Research*, 1527, 99–107. <https://doi.org/10.1016/j.brainres.2013.06.030>
- Chodroff, E., & Cole, J. (2019). THE PHONOLOGICAL AND PHONETIC ENCODING OF INFORMATION STRUCTURE IN AMERICAN ENGLISH NUCLEAR ACCENTS. 4–10.
- Cole, J., Hualde, J. I., Smith, C. L., Eager, C., Mahrt, T., & Napoleão de Souza, R. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics*, 75, 113–147. <https://doi.org/10.1016/j.wocn.2019.05.002>
- Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *The Journal of the Acoustical Society of America*, 77(6), 2142–2156. <https://doi.org/10.1121/1.392372>

- Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). Multipic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*, 808–816. <https://doi.org/10.1080/17470218.2017.1310261>.
- Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., & Eisner, F. (2018). Opposing and following responses in sensorimotor speech control: Why responses go both ways. *Psychonomic Bulletin & Review*, *25*(4), 1458–1467. <https://doi.org/10.3758/s13423-018-1494-x>
- Gunther, F. H. (2016). *Neural control of speech*. MIT Press.
- Gusenhoven, C. (2007). Types of focus in English. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation* (pp. 83–100). Springer. https://doi.org/10.1007/978-1-4020-4796-1_5
- 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>
- Hilger, A., Cole, J., Kim, J. H., Lester-Smith, R. A., & Larson, C. (2020). The effect of pitch auditory feedback perturbations on the production of anticipatory phrasal prominence and boundary. https://doi.org/10.1044/2020_JSLHR-19-00043
- Hilger, A. I. (2020). Impaired Sensorimotor Integration for Prosodic Production in Ataxic Dysarthria.
- Katz, J., & Selkirk, E. (2011). Contrastive focus vs. Discourse-new: Evidence from phonetic prominence in English. *Language*, *87*(4), 771–816. <https://doi.org/10.1353/lan.2011.0076>
- 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](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>
- Kreiman, J. (1982). Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics*, *10*(2), 163–175. [https://doi.org/10.1016/S0095-4470\(19\)30955-6](https://doi.org/10.1016/S0095-4470(19)30955-6)
- Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.
- Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, *25*(3), 313–342. <https://doi.org/10.1006/jpho.1997.0046>
- Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition*, *38*(8), 1137–1146. <https://doi.org/10.3758/MC.38.8.1137>
- 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, Language, and Hearing Research*, *30*(4), 552–558. <https://doi.org/10.1044/jshr.3004.552>
- Larson, C. R., & Robin, D. A. (2016). Sensory processing: Advances in understanding structure and function of pitch-shifted auditory feedback in voice control. *AIMS Neuroscience*, *3*(1), 22–39. <https://doi.org/10.3934/Neuroscience.2016.1.22>
- Lindblom, B. (1983). Economy of speech gestures. In Peter MacNeilage (Ed.), *The production of speech* (pp. 217–245). Springer. https://doi.org/10.1007/978-1-4613-8202-7_10
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In William Hardcastle & Alain Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Springer. https://doi.org/10.1007/978-94-009-2037-8_16
- Liu, H., & Larson, C. R. (2007). Effects of perturbation magnitude and voice F0 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., Wang, E. Q., Chen, Z., Liu, P., Larson, C. R., & Huang, D. (2010). Effect of tonal native language on voice fundamental frequency responses to pitch feedback perturbations during sustained vocalizations. *The Journal of the Acoustical Society of America*, *128*(6), 3739–3746. <https://doi.org/10.1121/1.3500675>
- 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>
- Liu, H., Zhang, Q., Xu, Y., & Larson, C. R. (2007). Compensatory responses to loudness-shifted voice feedback during production of mandarin speech. *The Journal of the Acoustical Society of America*, *122*(4), 2405–2412. <https://doi.org/10.1121/1.2773955>
- Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, *10*, 2767–2767. <https://doi.org/10.3389/fpsyg.2019.02767>.
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). *Montreal forced aligner [computer program]*. Version.
- Nalborczyk, L., Batailler, C., Loevenbruck, 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 Sciences*, *62*(5), 1225–1242. https://doi.org/10.1044/2018_JSLHR-S-18-0006.
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>
- Natke, U., Donath, T. M., & Kalveram, K. T. (2003). Control of voice fundamental frequency in speaking versus singing. *The Journal of the Acoustical Society of America*, *113*(3), 1587–1593. <https://doi.org/10.1121/1.1543928>
- Ouyang, I. C., & Kaiser, E. (2015). Prosody and information structure in a tone language: An investigation of mandarin Chinese. *Language, Cognition and Neuroscience*, *30*(1–2), 57–72. <https://doi.org/10.1080/01690965.2013.805795>
- Patel, S., Nishimura, C., Lodhavia, A., Korzyukov, O., Parkinson, A., Robin, D. A., & Larson, C. R. (2014). Understanding the mechanisms underlying voluntary responses to pitch-shifted auditory feedback. *The Journal of the Acoustical Society of America*, *135*(5), 3036–3044. <https://doi.org/10.1121/1.4870490>
- Patel, S. P., Kim, J. H., Larson, C. R., & Losh, M. (2019). Mechanisms of voice control related to prosody in autism spectrum disorder and first-degree relatives. *Autism Research*, *12*(8), 1192–1210. <https://doi.org/10.1002/aur.2156>

- Perkell, J. S., Zandipour, M., Matthies, M. L., & Lane, H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *The Journal of the Acoustical Society of America*, 112(4), 1627–1641. <https://doi.org/10.1121/1.1506369>
- 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., & Hirschberg, J. B. (1990). *The Meaning of Intonational Contours in the Interpretation of Discourse*, 271–311. <https://doi.org/10.7916/D8KD24FP>
- R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1(1), 75–116. <https://doi.org/10.1007/BF02342617>
- RStudio Team. (2020). *RStudio: Integrated Development for R. RStudio*. Boston, MA: PBC. <http://www.rstudio.com/>.
- Scheerer, N. E., & Jones, J. A. (2018). Detecting our own vocal errors: An event-related study of the thresholds for perceiving and compensating for vocal pitch errors. *Neuropsychologia*, 114, 158–167. <https://doi.org/10.1016/j.neuropsychologia.2017.12.007>
- Stavropoulou, P., & Baltazani, M. (2021). The prosody of correction and contrast. *Journal of Pragmatics*, 171, 76–100. <https://doi.org/10.1016/j.pragma.2020.10.004>
- Turnbull, R., Burdin, R. S., Clopper, C. G., & Tonhauser, J. (2015). Contextual predictability and the prosodic realisation of focus: A cross-linguistic comparison. *Language, Cognition and Neuroscience*, 30(9), 1061–1076. <https://doi.org/10.1080/23273798.2015.1071856>
- Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, 106(3), 1548–1557. <https://doi.org/10.1016/j.cognition.2007.06.009>
- Xu, Y., & Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America*, 111(3), 1399–1413. <https://doi.org/10.1121/1.1445789>
- Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage*, 40(4), 1871–1887. <https://doi.org/10.1016/j.neuroimage.2008.01.026>