




Research Article

Controlling Pitch for Prosody: Sensorimotor Adaptation in Linguistically Meaningful Contexts

Kimberly L. Dahl,^a  Manuel Díaz Cádiz,^a Jennifer Zuk,^a  Frank H. Guenther,^{a,b}
and Cara E. Stepp^{a,b,c} 

^aDepartment of Speech, Language and Hearing Sciences, Boston University, MA ^bDepartment of Biomedical Engineering, Boston University, MA ^cDepartment of Otolaryngology–Head and Neck Surgery, Boston University School of Medicine, MA

ARTICLE INFO**Article History:**

Received August 2, 2023

Revision received October 9, 2023

Accepted November 2, 2023

Editor-in-Chief: Julie A. Washington

Editor: Raymond D. Kent

https://doi.org/10.1044/2023_JSLHR-23-00460

ABSTRACT

Purpose: This study examined how speakers adapt to fundamental frequency (f_0) errors that affect the use of prosody to convey linguistic meaning, whether f_0 adaptation in that context relates to adaptation in linguistically neutral sustained vowels, and whether cue trading is reflected in responses in the prosodic cues of f_0 and amplitude.

Method: Twenty-four speakers said vowels and sentences while f_0 was digitally altered to induce predictable errors. Shifts in f_0 (± 200 cents) were applied to the entire sustained vowel and one word (emphasized or unemphasized) in sentences. Two prosodic cues— f_0 and amplitude—were extracted. The effects of f_0 shifts, shift direction, and emphasis on f_0 response magnitude were evaluated with repeated-measures analyses of variance. Relationships between adaptive f_0 responses in sentences and vowels and between adaptive f_0 and amplitude responses were evaluated with Spearman correlations.

Results: Speakers adapted to f_0 errors in both linguistically meaningful sentences and linguistically neutral vowels. Adaptive f_0 responses of unemphasized words were smaller than those of emphasized words when f_0 was shifted upward. There was no relationship between adaptive f_0 responses in vowels and emphasized words, but adaptive f_0 and amplitude responses were strongly, positively correlated.

Conclusions: Sensorimotor adaptation occurs in response to f_0 errors regardless of how disruptive the error is to linguistic meaning. Adaptation to f_0 errors during sustained vowels may not involve the exact same mechanisms as sensorimotor adaptation as it occurs in meaningful speech. The relationship between adaptive responses in f_0 and amplitude supports an integrated model of prosody.

Supplemental Material: <https://doi.org/10.23641/asha.25008908>

Prosody is an important element of human communication. It includes the acoustic features of fundamental frequency (f_0), amplitude, and duration (Cole, 2015), which are perceived as variations in pitch, loudness, and syllable length. These features serve important functions, such as allowing the speaker to convey meaning, express emotion, and speak intelligibly and naturally. When someone has difficulty speaking with effective prosody—a common result of motor speech disorders (Kent & Rosenbek,

1982)—any of these critical functions may be impaired, leading to reduced intelligibility (Bunton et al., 2000, 2001; de Bodt et al., 2002; Lares & Weismer, 1999) and speech naturalness (Anand & Stepp, 2015; Klopfenstein, 2015; Vojtech et al., 2019).

A speaker with impaired prosody may also struggle to express their intended meaning. Certain meanings, such as emphasizing one word over others, rely on prosodic contrasts. The type of emphasis relevant to the present study—narrow focus (Ladd, 2008)—serves to direct the listener's attention to a particular element of a sentence. A production with such emphasis is often considered a response to a *wh*-question (e.g., “Who did you see on

Correspondence to Kimberly L. Dahl: dahl@bu.edu. **Disclosure:** The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.

Monday?”), though the question may be either explicit or implicit (Breen et al., 2010; Ladd, 2008; Roettger et al., 2019). Prosodic impairment may prevent a speaker from creating sufficient contrast to convey, for example, the difference between “I saw *Mel* on Monday” and “I saw Mel on *Monday*” (and thus the differences in the underlying *wh*-questions, “Who did you see on Monday?” and “When did you see Mel?”). The acoustic realization of emphasis may vary depending on the speaker’s preferences, the location of emphasis within a sentence (Breen et al., 2010), and whether the emphasis serves not just to direct the listener’s focus but also to contrast with information previously given (Breen et al., 2010; Roettger et al., 2019). Speakers can emphasize a word by increasing its f_o , amplitude, duration, or any combination of these acoustic cues. Speakers may also decrease their f_o to emphasize a word, though there is some evidence that narrow focus in declarative sentences is likely to elicit increased f_o (Breen et al., 2010; Roettger et al., 2019).

The flexibility a speaker has to select and combine cues is captured by the cue-trading theory of prosody (Lieberman, 1960). Cue trading may, for some speakers, simply reflect individual tendencies (Howell, 1993). For others, like speakers with motor speech disorders, it allows the speaker to replace an impaired acoustic cue with a spared one (Martens et al., 2011; R. Patel, 2002). That is, a speaker with impaired control of f_o can still effectively emphasize a word by increasing its amplitude or duration. Despite the flexibility granted by cue trading, f_o is the most salient marker of phrase-level emphasis in English for most speakers (Howell, 1993; O’Shaughnessy, 1979). Understanding how speakers control f_o to convey meaning may shed light on the mechanisms underlying effective prosody and inform treatments aimed at restoring prosodic function.

Our understanding of f_o control comes largely from studies that use altered auditory feedback techniques. In these studies, researchers digitally alter a person’s voice and transmit the altered signal back to that person over headphones in near real time. This manipulation induces an error in the intended production, and speakers usually respond to the error by opposing the manipulation; if f_o is digitally increased, speakers lower their f_o , and vice versa (Burnett et al., 1997; Jones & Munhall, 2000). Responses to experimentally induced f_o errors show that speakers quickly correct *intermittent* errors in what is termed the pitch reflex. It also shows that speakers adapt over time to *predictably recurring* errors, a motor learning process known as sensorimotor adaptation. These reflexive and adaptive responses reflect the interacting mechanisms of feedback and feedforward control that are the major components of a prevailing theory of speech motor control—directions into velocities of articulators (DIVA; Guenther

et al., 2006). Though feedback control is important for quickly addressing occasional speech errors, the DIVA model posits that the fully developed speech system of typical adults relies primarily on feedforward control.

The speech task commonly used to study feedforward control of voice—a sustained vowel—is a notable limitation of this body of work. Sustained vowels differ substantially from the connected speech used in everyday communication. They are simpler, less natural, and devoid of linguistic meaning. Although sensorimotor adaptation occurs even in such speech tasks, speakers may be particularly sensitive to f_o errors that could disrupt their intended message. This is true of feedback control of f_o —reflexive responses are larger when f_o is important for conveying meaning (Chen et al., 2007; Hilger et al., 2023).

Establishing a relationship between adaptive responses during sustained vowels and during linguistically meaningful modulation of f_o would have important methodological implications for the study of vocal motor control. The widespread reliance on sustained vowels in prior work suggests that most researchers assume the adaptation observed during sustained vowels relates to adaptation in more complex tasks such as running speech, but this relationship has yet to be established. If no such relationship exists, we cannot generalize vowel-based findings to draw conclusions about f_o control in running speech.

There is limited evidence of sensorimotor adaptation to f_o errors in running speech. R. Patel et al. (2011) altered f_o during emphasized words in short sentences and found that speakers did compensate for these induced f_o errors. Specifically, speakers adapted by increasing the f_o contrast between emphasized and unemphasized words. This was true whether f_o was shifted downward or upward, but not to the same degree. If f_o is raised during an emphasized word, a downward shift in f_o would reduce the intended emphasis and an upward shift would enhance it. Unsurprisingly, then, speakers exposed to the more disruptive downward shift increased the prosodic contrast more so than those exposed to the less disruptive upward shift. This finding hints at the importance of linguistic meaning in determining how speakers adapt to predictable f_o errors.

Yet, key questions remain unanswered by this work. First, Patel and colleagues exposed participants to f_o perturbations only in a running speech task. The study therefore cannot establish whether the observed responses would relate to the same speakers’ responses to f_o errors during sustained vowels. Again, confirming this relationship is critical for allowing researchers to generalize vowel-based findings to running speech. Second, f_o was extracted from the entire perturbed word. The acoustic outcomes thus likely reflect a mix of feedforward and

feedback control, since responses to auditory feedback begin 100–150 ms after voicing onset (Burnett & Larson, 2002; Burnett et al., 1997; Larson et al., 2001). Third, f_0 perturbations were applied only to emphasized words, even though the importance of f_0 for conveying meaning may differ between emphasized and unemphasized words. Determining the effect linguistic meaning has on adaptation may require a nuanced approach in which adaptation is measured in contexts that vary in the degree to which an f_0 error is disruptive to an intended meaning.

Yet another lingering question is whether corrections of f_0 errors are achieved through integrated or independent control of the acoustic cues of prosody. An integrated model of prosody posits that prosodic cues are combined to reach a single prosodic target, whereas an independent model states that speakers control f_0 , amplitude and duration independently to reach separate targets for each prosodic cue (Guenther, 2016). Both are consistent with the cue-trading theory (Lieberman, 1960), which does not specify the control mechanisms involved in trading cues. R. Patel et al. (2011) found that speakers adjusted both f_0 and amplitude in response to f_0 perturbations, which points to integrated control. However, they found in a later study that amplitude perturbations elicited responses only in amplitude, not f_0 (R. R. Patel et al., 2015), suggesting independent control. However, the perturbation magnitudes between these two studies may not have been equivalent and thus the disruptions to the intended productions also not comparable. These are important differences that preclude confidently embracing the independent control hypothesis based on the findings of this later study.

One way to clarify the question of integration versus independence is to measure the relationship between f_0 and amplitude responses when only one cue is manipulated. If these cues are indeed integrated, a larger response in one would correlate with a smaller response in the other, with individual tendencies (Howell, 1993) dictating which cue shows the larger response. With this integrated approach, the relative contributions of each cue would combine to reach a single prosodic target.

To address these fundamental gaps in knowledge, this study aimed to (a) evaluate sensorimotor adaptation to f_0 errors that affect linguistic meaning in running speech, (b) examine relationships between this adaptation and that observed during sustained vowels, and (c) identify relationships between f_0 and amplitude responses when only f_0 is manipulated. Speakers' voices were digitally manipulated in near-real time to alter the auditory feedback of f_0 during sentences and sustained vowels. In sentences, f_0 was shifted downward or upward during emphasized and unemphasized words. Shifts in f_0 were

also applied during sustained vowels. These multiple conditions allowed us to examine how responses differ based on the degree to which perturbations disrupted the intended meaning. Our expectation was that f_0 shifts that were more disruptive would elicit larger responses.

We therefore hypothesized that f_0 perturbations would elicit larger adaptive responses in (a) down-shifted emphasized words compared to up-shifted emphasized words and (b) in emphasized words compared to unemphasized words, regardless of shift direction. We also hypothesized that (c) the magnitude of adaptive responses in down-shifted emphasized words would correlate with that of sustained vowels. Finally, in accordance with the cue-trading model of prosody, we hypothesized that (d) adaptive responses in f_0 and amplitude would be negatively correlated, indicating that speakers may resolve recurring f_0 errors by adjusting amplitude to achieve the prosodic target, consistent with an integrated model of prosody.

Method

Participants

Participants were 24 adults (12 cisgender women, 12 cisgender men) with an average age of 23 years ($SD = 4$ years, range: 19–40 years) and no history of neurological, speech, voice, or language disorders. All participants spoke North American English as a first language, spoke no tonal languages, and had no formal singing experience. Participants underwent a pulsed-tone hearing screening at octaves from 125 Hz to 4 kHz with a Grason-Stadler GSI 18 audiometer. Most participants (18/24) met the screening criterion of a 25-dB HL maximum threshold at all frequencies presented.¹ All participants provided written consent in accordance with the Boston University Institutional Review Board.

Three additional participants (one cisgender woman, two cisgender men; $M = 23$ years) attempted the study but were excluded because of speech behaviors that disrupted our perturbation technique. Two of these participants prevoiced the initial consonant of emphasized words, and one voiced continuously between the words of the sentences. In both cases, this prevented the f_0

¹Three participants had unilaterally elevated thresholds (30 dB HL) and two had bilaterally elevated thresholds (30–35 dB HL) at a single frequency (250 or 500 Hz). One participant had slightly elevated thresholds at two frequencies (30–35 dB HL at 250 Hz bilaterally; 30 dB HL at 500 Hz unilaterally). These thresholds were not expected to invalidate adaptation data from these participants.

perturbation from being reliably applied to only the first word of each sentence as intended.

Amplification and Calibration

The amplitude of the auditory feedback signal was amplified 5 dB above the microphone signal to account for relative dB differences in the signal at the participant's mouth and ear (Cornelisse et al., 1991). This amplification level was calibrated with a 2-cc coupler (Brüel & Kjaer 4192) attached to a handheld sound level meter (Brüel & Kjaer 2250-L). A 1-kHz pure tone was played at ~75 dB from a digital recorder (Olympus LS-10) placed 7 cm from an omnidirectional condenser earset microphone (Shure MX153). The gain was adjusted such that the amplitude of the signal transmitted to a pair of circumaural headphones (Sennheiser HD 280 Pro) was ~80 dB as measured with an artificial ear (Brüel & Kjaer 4153).

Experimental Setup

Participants completed the experiment in a sound-attenuating booth at Boston University in a single session lasting approximately 2.5 hr. Participants wore an omnidirectional condenser earset microphone (Shure MX153) positioned 7 cm from the corner of the mouth at a 45° angle from midline (R. R. Patel et al., 2018). They sat before a computer monitor that displayed prompts for each trial throughout the session. Stimulus presentation and audio settings were controlled via MATLAB scripts (Version 2018a; MathWorks, Inc.).

The microphone signal was amplified with an RME QuadMic II preamplifier and digitized with an RME Fireface UCX sound card with a 32-bit resolution and a 44.1-kHz sampling rate. The signal was transmitted through an Eventide Eclipse V4 Harmonizer for full-spectrum frequency shifts without formant correction.² The Eclipse-processed, amplified signal was transmitted in near-real time to the participant's headphones (Sennheiser HD 280 Pro) through a Behringer Xenyx Q802USB mixer. The auditory feedback signal was transmitted with minimal delay (~11 ms; Heller Murray et al., 2019) through headphones that attenuated air-conducted sound by 32 dB. The headphone signal thus served as the primary source of auditory feedback.

Stimuli and Training

The adaptation procedure consisted of two task types—those in which f_0 conveyed no linguistic meaning

(vowel task) and those in which f_0 was linguistically meaningful (sentence tasks). The vowel task entailed 3-s productions of /a/. Participants were instructed to hold a steady /a/ at a comfortable pitch and loudness for as long as the trial prompt appeared on the screen.

Sentence tasks were four three-word sentences³ with either the first or second word emphasized (e.g., “*Bev* builds doors,” “Bev *builds* doors”). All sentences contained single-syllable words with voiced phonemes. Single-syllable words eliminated any effect of lexical stress on within-word f_0 contours, and voiced phonemes allowed for continuous f_0 tracking.

Participants were familiarized with the sentences, which were described as coming from “a conversation with someone having difficulty hearing or understanding everything you say,” before beginning any experimental trials. They were told to emphasize certain words to aid their imaginary conversation partner's understanding. For each trial, participants were prompted on the screen with a question (e.g., “Who builds doors?” or “Bev does what with doors?”) that was designed to elicit the target sentence with a particular type of emphasis—namely, narrow focus (Ladd, 2008). The target sentence then appeared with the emphasized word in bold, italicized text. The interstimulus interval in all tasks was jittered from 2 to 4 s to encourage the participants' sustained attention to the task. Participants were not informed that the auditory feedback of their voice would be perturbed during the session.

Experimental Procedure

Participants completed the experiment under eight conditions. The two vowel task conditions were defined by shift type (control, down), and the six sentence task conditions were defined by shift type (control, down, or up) and first-word emphasis (emphasized or unemphasized). Condition order was counterbalanced across participants, first by task and emphasis (for sentences), then shift, without consecutive f_0 -shifted conditions.

The vowel task included only control and shift-down conditions. We eliminated a shift-up condition to shorten the study session and avoid participant fatigue. Reflexive auditory feedback studies show that downward shifts elicit larger reflexive responses than upward shifts (Liu & Larson, 2007). It is unclear if the same is true of adaptive responses, but if so, a downward shift during sustained vowels would provide the best comparison for adaptive responses during sentences.

²The f_0 shifts in the present study were not sufficient to alter the perception of vowel identity.

³“Bev builds doors,” “Jove boils beans,” “Maeve brews beer,” and “Dave buys beds.”

Before beginning experimental trials for a given speech task, eight practice trials of that task were recorded. These practice trials served to (a) familiarize the participant with the prompts for each task; (b) allow the researcher to give feedback if the participant did not use the intended emphasis; and (c) provide the data needed to calculate amplitude thresholds that controlled the onset and offset of f_o perturbations, described below.

Perturbations of f_o were applied to the entire sustained vowel, but only the first word of sentences. This approach allowed us to more effectively disrupt intended f_o contours, which are defined by relative f_o differences between words, not absolute f_o values (Tang et al., 2017); manipulating f_o throughout the entire sentence, rather than a single word within it, would have no effect on relative f_o differences. Specifically selecting the *first* word as the perturbation target, regardless of which word was emphasized, offered both scientific and technical benefits. The scientific benefit was the ability to manipulate f_o in ways that varied in how much the induced error disrupted the use of f_o to convey an intended meaning. For example, an emphasized word could be heard as unemphasized during downshifted conditions, thus maximally disrupting meaning. An unemphasized word under the same downshifted condition, on the other hand, could be simply rendered more unemphasized and thus the meaning minimally disrupted. The technical benefit was that the boundaries of the first word, which had a clearly defined onset, could be more reliably identified in pilot testing than later words in the sentence.

We implemented these targeted f_o perturbations using participant- and task-specific amplitude thresholds. Specifically, the root-mean-square (RMS) of the amplitude of the microphone signal was calculated across sliding 60-ms windows with 90% overlap over the first word in eight practice trials of the target sentences (two repetitions per sentence). Thresholds were then set at 22% of the maximum RMS level for a given participant for conditions in which the first word was emphasized and 20% when it was not.⁴ When the RMS of the signal at the microphone rose above this threshold, the f_o manipulation was triggered and remained in place until the RMS fell below that threshold. The trigger was only active for the first two threshold crossings, thus preventing the perturbation from being applied again later in the sentence. This approach applied the f_o perturbation to the first word with 98% accuracy for the included participants.

Each condition consisted of 64 trials divided equally across four phases. During f_o -shifted conditions, auditory feedback was unaltered in the baseline phase. Manipulation

of f_o was then implemented during the ramp phase, such that f_o of the auditory feedback signal increased by 12.5 cents per trial in shift-up conditions or decreased by 12.5 cents per trial in shift-down conditions. The maximum perturbation of ± 200 cents was maintained for all trials of the hold phase. Auditory feedback then returned to its unaltered state in the after-effect phase. During control conditions, auditory feedback remained unaltered across all phases. Sentences were pseudorandomized within each phase such that all sentences were produced 4 times within each phase without consecutive repetitions.

Data Analysis

The acoustic waveform, spectrogram, and f_o trace of the microphone signal for each trial were visualized with a MATLAB script (Version 2022a; MathWorks, Inc.) using data extracted from Praat (Boersma & Weenink, 2015). The timing of f_o perturbations was also displayed on the spectrogram. A trained research assistant first confirmed that the trial was usable by playing the recording of the microphone signal to rule out missed trials, speech errors, or nonspeech vocalizations (e.g., yawn, laughter). They then visually inspected the f_o tracking and confirmed that the f_o perturbation was applied to the first word as intended. Trials with perturbation or speech errors were excluded (1.8% and 0.3%, respectively), as were trials during which the participant began speaking before recording started (0.3%). Trials with f_o tracking errors were manually analyzed in Praat after adjusting pitch settings to achieve accurate tracking. Three trials with f_o tracking errors could not be manually corrected and so were excluded. A total of 145 trials (2.5%) were excluded, leaving a final data set of 12,288 trials.

For usable trials, the research assistant marked the onset and offset of sustained vowels or first word of the sentence. Both f_o and amplitude were extracted from the period 40–120 ms after the marked onset, a window that always included the vowel. This analysis window (“early”) excluded f_o fluctuations at voicing onset and minimized responses of the auditory feedback control system, which occur 100–150 ms after voicing onset (Burnett & Larson, 2002; Burnett et al., 1997; Larson et al., 2001). However, f_o and amplitude were also extracted from the entire vowel or first word (“full”) to compare our findings to relevant prior work that used this longer analysis window (i.e., R. Patel et al., 2011).

The mean f_o of all baseline trials in each condition served as the reference frequency to convert f_o of each trial in that condition from Hz to cents.⁵ Similarly,

⁴Pilot testing revealed these thresholds to identify the first word of the sentence most reliably in both conditions.

⁵ $f_o(\text{cents}) = 3986 \times \log_{10}\left(\frac{f_1}{f_2}\right)$, where f_1 is the f_o of a given trial in Hz and f_2 is the reference frequency (i.e., mean f_o of baseline trials).

amplitude was normalized by subtracting the mean amplitude of all baseline trials in each condition from each trial in that condition. The f_o (cents) and amplitude (dB) of each trial in the control condition were subtracted from each corresponding trial in f_o -shifted conditions for a given speech task. This trial-by-trial normalization accounted for natural f_o and amplitude variability across repeated productions. For data visualization, normalized f_o was averaged over every four trials.

There remains no consensus on how to quantify adaptation, so we used two common approaches—the mean during the hold phase and the mean across the first three trials of the after-effect phase (“early after-effect”). Responses from all participants were included in these calculations, whether compensatory (opposing the perturbation), following (in the same direction as the perturbation), or nonresponsive (no adjustment to the perturbation). A third common approach requires the use of masking noise to remove auditory feedback during some trials of the hold phase. Masking noise, however, could induce the Lombard effect, leading participants to increase their amplitude in the presence of noise (Lombard, 1911). Since we intended to measure changes in amplitude as a possible response to f_o perturbations, the use of masking noise was considered unsuitable for this study.

Statistical Analysis

Statistical analyses were run in Minitab (Version 21; Minitab Inc.) with significance set a priori at $\alpha = .05$. To test our hypotheses on the effects of emphasis and shift direction on the magnitude of adaptive f_o responses in linguistically meaningful speech, we constructed a repeated-measures analysis of variance (ANOVA) for each analysis window approach (early and full). The outcome was the mean normalized f_o (cents) across the baseline, hold, and early after-effect of sentence tasks. The sign was inverted for all values in up-shifted trials, such that a larger positive value indicated a larger compensatory response in all conditions. Each model included fixed effects of emphasis (emphasized, unemphasized), shift type (down, up), and phase (baseline, hold, early after-effect); all interactions; and a random effect of speaker. Effect sizes for significant effects and interactions were calculated as partial curvilinear correlations (η_p^2) and designated as small ($\sim .01$), medium ($\sim .09$), and large ($> .25$; Witte & Witte, 2009). Significant effects were further evaluated with post hoc Tukey’s tests. Effect sizes for post hoc tests were measured as Cohen’s d and interpreted as small (.2), medium (.5), or large ($> .8$; Cohen, 1988).

To confirm adaptation also occurred during the vowel task, we constructed two repeated-measures ANOVAs to determine the effect of phase on the mean normalized

f_o (cents) of vowels using early and full analysis windows. Again, speaker was entered as a random factor, significant effects evaluated with post hoc Tukey’s tests, and all effect sizes calculated and interpreted as above.

To test our hypotheses regarding relationships between responses in different speech tasks and in different prosodic cues, we calculated Spearman rank-order correlations. Based on the ANOVA results showing equivalence for the early and full windows and evidence of adaptation in both the hold phase and early after-effect trials, we calculated these correlations using data derived from the early window in the early after-effect. This reduced the number of correlations statistically analyzed and thus minimized the likelihood of a Type I error. To evaluate the relationship between adaptive responses in productions with and without linguistically meaningful f_o , we calculated the correlation between f_o responses (cents) during down-shifted emphasized words and during sustained vowels. To test our hypothesis on the relationship between adaptive responses in different prosodic cues, we calculated the correlations between normalized f_o (cents) and normalized amplitude (dB) of down-shifted emphasized and unemphasized words.

We conducted additional analysis, as in R. Patel et al. (2011), to examine the physiological link between f_o and amplitude (Titze, 1989) and thus provide important context for findings on the relationship between adaptive responses in f_o and amplitude. That is, we calculated the Pearson correlation between f_o (Hz) and amplitude (dB) of all trials in all sentence tasks for each participant. We then applied a Fisher Z transformation (i.e., the inverse hyperbolic tangent) to each r value⁶ and averaged the Z values across participants. This analysis quantified the trial-by-trial relationship between f_o and amplitude to determine whether a correlation between adaptive responses in f_o and amplitude should be attributed to behavior or physiology.

Results

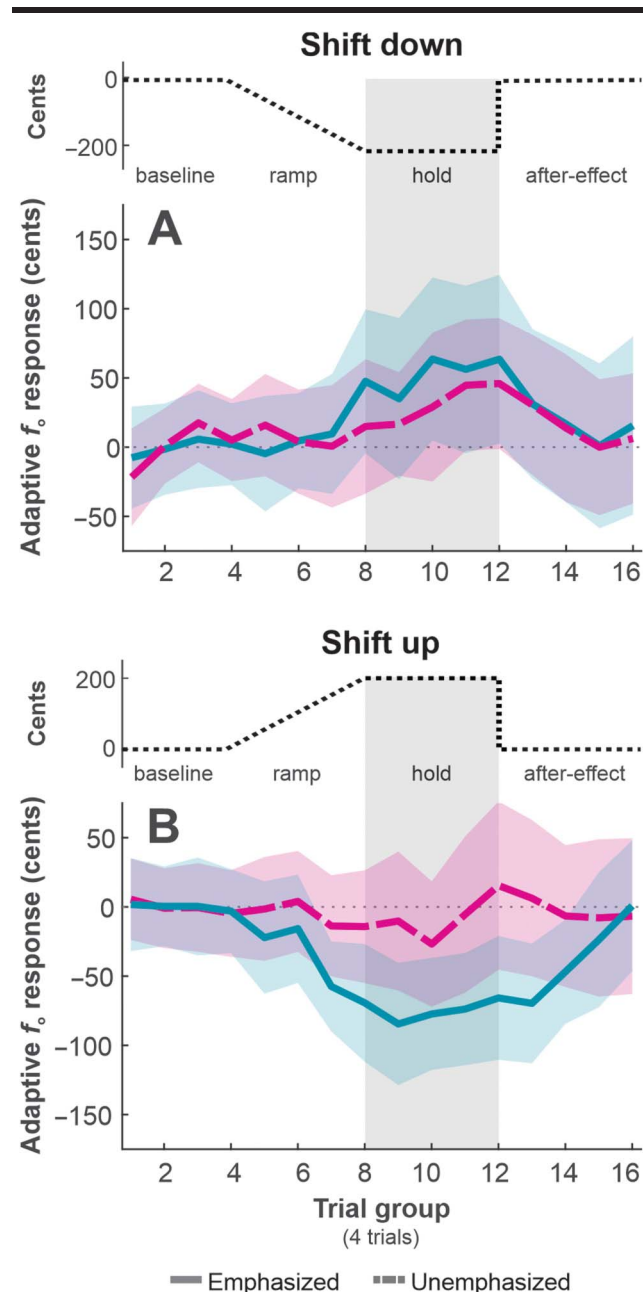
Statistical results were qualitatively equivalent whether data from the early or full analysis window were used. We therefore report below the results of the early window, which better capture feedforward responses. Results from the full-window analysis are in Supplemental Material S1.

⁶Pearson’s correlations are bounded between -1 and 1 , which results in nonnormal distributions. The Fisher transformation results in variables that are normally distributed and thus allows for tests of significance.

f_0 Adaptation During a Prosody Task

As a group, speakers responded to f_0 manipulations during sentences by opposing the ± 200 -cent perturbation (see Figure 1). There was a significant effect of phase during sentences, with significantly greater f_0 in the hold phase (43.7 cents, $d = 8.54$, $p_{\text{adj}} = .001$) and early after-

Figure 1. Group mean adaptive response in fundamental frequency (f_0) to downward (Panel A) and upward (Panel B) shifts in auditory feedback of f_0 (± 200 cents) during emphasized (blue) and unemphasized (dashed pink) words in running speech. Shaded areas are 95% confidence intervals.



effect trials (33.9 cents, $d = 5.64$, $p_{\text{adj}} = .012$) than the baseline (see Table 1). There was no significant difference between the magnitudes of f_0 responses in the hold phase and early after-effect trials.

There was also a significant effect of emphasis on f_0 response magnitudes, which were greater in emphasized words (39.2 cents) than unemphasized words (12.6 cents, $d = 5.04$, $p_{\text{adj}} = .006$), across all phases. This difference appears to be driven by responses to upward perturbations (see Figure 1B). That is, there was also a significant interaction between emphasis and shift direction; f_0 response magnitudes were greater across all phases in emphasized words (48.5 cents) than unemphasized words (-0.5 cents, $d = 6.41$, $p_{\text{adj}} = .002$) when f_0 was shifted upward.

f_0 Adaptation During a Vowel Task

As a group, speakers responded to downward f_0 shifts during sustained vowels by opposing the -200 -cent perturbation (see Figure 2). There was a significant effect of phase, with greater f_0 response magnitudes in the hold phase (109.0 cents, $d = 31.54$, $p_{\text{adj}} < .001$) and early after-effect trials (68.7 cents, $d = 14.14$, $p_{\text{adj}} < .001$) than at baseline. The f_0 response magnitude was also significantly greater in the hold phase than in the early after-effect trials ($d = 14.14$, $p_{\text{adj}} < .001$).

Relationships Between Adaptation in Different Speech Tasks and Prosodic Cues

The magnitude of speakers' adaptive f_0 responses to downward shifts during sustained vowels did not significantly correlate with their responses during emphasized words ($r = .23$, $p = .281$; see Figure 3A). There were strong, positive correlations between adaptive f_0 and amplitude responses during downward f_0 shifts in both emphasized words ($r = .49$, $p = .016$) and unemphasized words ($r = .67$, $p < .001$; see Figure 3B). The average Fisher Z -transformed correlation between f_0 and amplitude was $Z = -0.29$ ($SD = 0.34$), which was back-converted as the hyperbolic tangent to an average $r = -0.28$ ($SD = 0.33$). Thus, f_0 and amplitude were weakly correlated on a trial-by-trial basis during sentence tasks. See Supplemental Materials S2 and S3 for adaptive amplitude responses across all trials for each speech task.

Discussion

The purpose of this study was to determine how the use of f_0 to convey meaning through prosody would affect sensorimotor adaptation to f_0 errors. We also examined how f_0 adaptation in these linguistically meaningful

Table 1. Results of repeated-measures analyses of variance for normalized fundamental frequency (f_0) during sustained vowels and sentences with emphasized and unemphasized words, based on data extracted from an early window 40–120 ms after voicing onset.

Effect	<i>df</i>	<i>F</i>	<i>P</i>	η_p^2	Effect size
Sustained vowel					
Phase	2	24.85	< .001*	.52	Large
Sentences					
Emphasis	1	7.64	.006*	.03	Small
Shift	1	0.16	.691	—	—
Phase	2	7.52	.001*	.06	Medium
Emphasis × Shift	1	5.34	.022*	.02	Small
Emphasis × Phase	2	1.99	.139	—	—
Shift × Phase	2	0.05	.952	—	—
Emphasis × Shift × Phase	2	1.54	.217	—	—

*Significant at $\alpha = .05$. — = not applicable for nonsignificant results.

contexts relates to adaptation in a commonly used but linguistically neutral sustained vowel task. Finally, we evaluated the two models of control that are both possible under the cue-trading theory of prosody by measuring relationships between responses to f_0 errors in the distinct prosodic cues of f_0 and amplitude.

Adapting f_0 for Prosody

Altering auditory feedback is a well-established technique to study sensorimotor adaptation to f_0 errors. With

few exceptions, however, this technique has been applied to f_0 in sustained vowels, during which f_0 carries no linguistic meaning. This task eschews a primary function of f_0 in communication—to convey meaning through prosody. When control of f_0 for prosody has been studied, it has largely been through a reflexive paradigm (Chen et al., 2007; Hilger et al., 2020, 2023). A single study, to our knowledge, has evaluated sensorimotor adaptation of f_0 in running speech (R. Patel et al., 2011), though their f_0 extraction method likely also captured contributions of feedback control.

Our findings thus add to the limited evidence of f_0 adaptation during the production of sentences with meanings that relied on prosodic contrasts. Like R. Patel et al. (2011), we found that speakers did adapt to f_0 errors in running speech. Our results diverge from and expand upon this prior research in two ways.

First, R. Patel et al. (2011) found downward f_0 shifts to elicit a larger response in emphasized words than did upward shifts. They interpreted this difference as linguistically motivated. Downward f_0 shifts were more disruptive to the intended meaning and thus required more correction to restore that meaning. However, we found no such difference in f_0 adaptation during emphasized words in the present study, suggesting that linguistic meaning had no effect on speakers' responses. These contradictory findings may be explained by methodological differences. R. Patel et al. (2011) measured adaptation as changes in f_0 contrast between emphasized and unemphasized words within a sentence. We measured adaptation based on f_0 of the perturbed word. Our measurement approach is consistent with that taken in most other f_0 adaptation studies, thus allowing for integration with a larger body of work. However, our approach does not account for the importance of relative f_0 differences between words over absolute f_0 of any given word when it comes to effective

Figure 2. Group mean adaptive response in fundamental frequency (f_0) to a 200-cent downward shift in auditory feedback of f_0 during sustained vowels. Shaded area is 95% confidence interval.

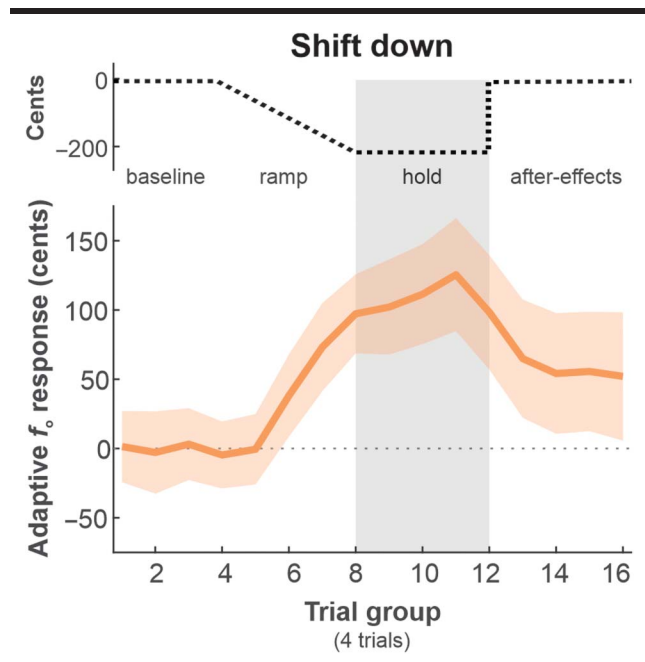
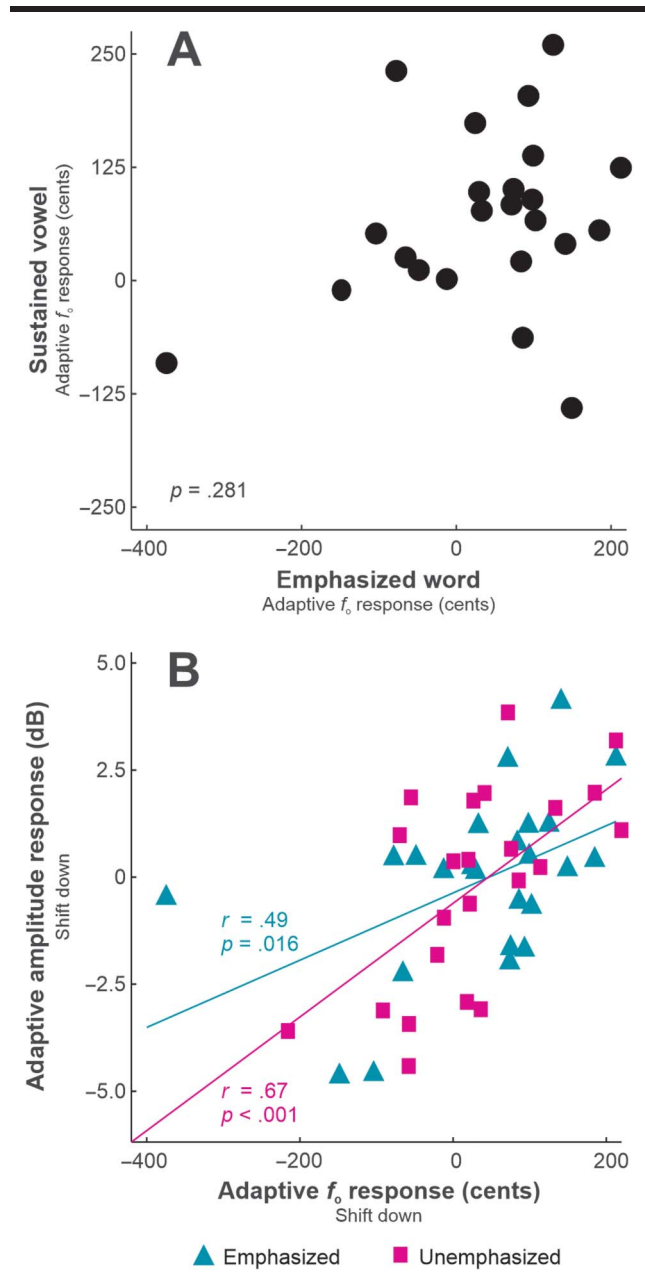


Figure 3. Panel A: Mean fundamental frequency (f_0) in the early after-effect trials (i.e., adaptive f_0 responses) during 200-cent downward shifts in f_0 during emphasized words and sustained vowels. Panel B: Mean adaptive responses in f_0 and amplitude during 200-cent downward f_0 shifts in emphasized (blue triangles) and unemphasized (pink squares) words.



prosody (Tang et al., 2017). This makes it difficult to directly compare our result with that of R. Patel et al. (2011). Quantifying responses using Patel and colleagues' approach (see the Appendix) would allow such a comparison, but this analysis no longer isolates feedforward responses and so cannot address the primary objective of this study.

Second, Patel and colleagues manipulated f_0 during emphasized words only. Here, we have also characterized how speakers respond to f_0 errors in unemphasized words, when f_0 is less important for conveying a given meaning. We found that speakers did adapt to f_0 shifts in unemphasized words. However, f_0 responses during upwards shifts were smaller in unemphasized words than emphasized words. In fact, the mean f_0 response to upward shifts in unemphasized words was only 7.6 cents in the hold phase, or 3.8% of the perturbation. Speakers without communication disorders typically correct for 35%–91% of an f_0 perturbation (Abur et al., 2018; Abur, Subacutute, Daliri, et al., 2021; Lester-Smith et al., 2020; Scheerer et al., 2016; Stepp et al., 2017; Weerathunge et al., 2020). Thus, although the adaptive response to upward f_0 shifts in unemphasized words was statistically significant, it is substantially smaller than adaptation in either emphasized words or isolated vowels.

If responses to f_0 shifts were linguistically motivated, we would expect f_0 responses to upward shifts in unemphasized words to be larger than those in emphasized words. That is, shifting f_0 of an unemphasized word upward could push it toward an emphasized word. Crossing this hypothetical prosodic boundary would disrupt the meaning of the utterance in a way that a downward shift would not and thus warrant a stronger response. This interpretation is contingent upon speakers using an elevated f_0 to indicate emphasis, though speakers may also emphasize a word in other ways, like decreasing f_0 or increasing amplitude or duration. There is therefore the potential for considerable variability in how speakers use f_0 to contrast emphasized and unemphasized words. Nevertheless, prior work does provide some support for the expectation of increased f_0 in the context of narrow focus (Breen et al., 2010; Roettger et al., 2019), and visual inspection of the f_0 contours of the speakers in the present study reveals elevated f_0 as the most common approach to emphasis. Thus, linguistic motivation does not offer a compelling explanation here.

This finding may instead be the result of a floor effect. The unemphasized words in our stimuli always preceded an emphasized word. In preparation for this second-word emphasis, speakers may already have lowered the f_0 of the first unemphasized word. This could have left little room to compensate for the upward shift with a further lowering of f_0 . This small response to upward f_0 shifts in unemphasized words may therefore be physiologically, not linguistically, motivated.⁷ Taken

⁷Recall that f_0 could already be elevated during an emphasized word, leaving more room for an opposing response to an upward f_0 shift. This physiological limitation is thus likely only applicable to unemphasized words.

together with the equivalent responses to both upward and downward shifts during emphasized words, these findings do not support the hypothesis that linguistic meaning would affect the magnitude of adaptive responses in f_0 .

Adapting f_0 in Sustained Vowels: Limitations of a Common Approach

We found strong evidence that speakers adapted to induced f_0 errors during sustained vowels. This finding remained true across different methods of analyzing f_0 (early vs. full windows) and quantifying adaptation (hold phase vs. early after-effect trials). This finding is consistent with a substantial body of research using sustained vowels to document f_0 adaptation in adults without speech or voice impairment (Jones & Munhall, 2000; Scheerer et al., 2016; Weerathunge et al., 2020), children without speech or voice impairment (Heller Murray & Stepp, 2020; Scheerer et al., 2016), singers (Abur, Subaciute, Kapsner-Smith, et al., 2021; Jones & Keough, 2008; Keough & Jones, 2009), and individuals with communication disorders (Abur et al., 2018; Abur, Subaciute, Daliri, et al., 2021; Abur, Subaciute, Kapsner-Smith, et al., 2021; Stepp et al., 2017).

It is often implicitly or explicitly assumed that adaptation during sustained vowels reflects motor learning processes that operate in the more complex speech that speakers use to communicate. In line with this assumption, we hypothesized that adaptive f_0 responses in a sustained vowel task would relate to those in a running speech task. Specifically, we tested for a relationship that would indicate that speakers with a large f_0 response during sentences, relative to the range of responses for the sentence task, would also show a large response during vowels, relative to the range of responses for the vowel task. There was, in fact, no such relationship. It may be that the mechanisms of adaptation in running speech and sustained vowels are not identical. Determining the exact ways in which they differ and whether those differences affect behavioral responses requires more research. However, researchers should consider the possibility of such differences when interpreting adaptation during sustained vowels as a reflection of how adaptation occurs in everyday communication.

Importantly, we cannot determine, based on this finding, what speech task might offer a more valid approach. In this study, we were primarily interested in the effect of linguistic meaning on f_0 adaptation, given evidence that such meaning affects reflexive f_0 responses (Chen et al., 2007) and potentially also adaptive responses (R. Patel et al., 2011). However, we cannot conclude that it was the absence of linguistic meaning in sustained vowels that precluded a relationship with f_0 adaptation in running speech. These tasks differed not just in the

meaningfulness of f_0 , but also in perturbation duration. The mean perturbation duration during the vowel task was 2.4 s. This is considerably longer than the 442 ms per trial in emphasized words in running speech (and 367 ms per trial in unemphasized words). The vowels also differed between the two tasks, with participants sustaining an /a/ but producing /ε/ or /ə/ in the sentence task. Thus, to better determine the effect of linguistic meaning on f_0 adaptation and to identify a speech task that better reflects adaptive f_0 control in everyday communication, future work should compare adaptation across speech tasks in which perturbation duration and vowel identity are well controlled.

Co-Occurring Responses Across Prosodic Cues

The cue-trading theory of prosody proposes that speakers can emphasize a word by using any of the acoustic cues of prosody— f_0 , amplitude, and duration—alone or in various combinations (Lieberman, 1960). The theory does not specify whether a speaker controls each cue independently to reach individual targets or integrates control of all cues toward an overall prosodic target. Research to date leaves the question unsettled. Evidence from both reflexive (Larson et al., 2007) and adaptive (R. Patel et al., 2015) studies points to an independent model of prosody. These studies have shown that speakers compensate for simultaneous but opposing manipulations of f_0 and amplitude during sustained vowels (Larson et al., 2007) and that speakers respond to amplitude manipulations in running speech by adapting amplitude but no other prosodic cue (R. Patel et al., 2015).

In contrast, R. Patel et al. (2011) previously found that speakers responded to f_0 manipulations in running speech by adjusting both f_0 and amplitude, which they interpreted as supporting an integrated model. However, their approach to quantifying adaptive responses—changes in f_0 and amplitude contrasts between words—captured both feedforward and feedback responses, and so we cannot confidently conclude, based on their finding alone, that integration of cues is the operative mechanism in prosodic adaptation. The strong positive correlation between f_0 and amplitude responses during emphasized words in the present study, however, does provide additional support for this model.

However, the physiological link between f_0 and amplitude—both increase with higher subglottal pressure (Titze, 1989)—could undermine this interpretation. To be certain that the observed changes in amplitude were specifically related to reaching a prosodic target and not incidental to volitional changes in f_0 , we followed the confirmatory approach of R. Patel et al. (2011). The

result showed that f_0 and amplitude were weakly correlated on a trial-by-trial basis. We therefore conclude that the strong correlation between f_0 and amplitude *adaptive responses* does represent a behavioral, not physiological, relationship.

Of note, however, is that the observed correlation between adaptive responses was in the opposite direction of our hypothesis. We expected that speakers with a strong f_0 response would have a weak amplitude response, and vice versa. This would allow a speaker to reach an overall prosodic target without overshooting it. Instead, we identified a significant *positive* relationship, such that speakers with larger f_0 responses had larger amplitude responses in the same direction; this relationship held regardless of whether the f_0 responses were compensatory or following (see Figure 3B). The positive relationship we identified suggests that the same degree of target precision does not apply to prosodic emphasis, as it may to articulation (i.e., formant targets). Rather, hitting a prosodic target for emphasizing a word is evaluated binarily; a single boundary distinguishes emphasized words from unemphasized, and anything beyond that boundary is a successful production. Speakers may therefore recruit multiple, integrated cues— f_0 and amplitude—to cross that boundary and correct the missed target in future productions.

Limitations

This study applied well-established altered auditory feedback techniques to speech tasks that more closely approximated everyday communication. As such, some of the usual means of experimental control were adjusted. For example, the targets of perturbations in the sentence tasks were relatively common names in the study region (e.g., “Bev,” “Dave”), which we hoped would allow for more natural productions. These varied stimuli did not control for inherent f_0 differences due to vowel identity (Peterson & Barney, 1952) and onset consonant (Xu & Xu, 2021). These differences are small, however, and thus unlikely to have affected the outcomes.

Perturbations of f_0 were applied only during the first word of the sentence. According to research on feedback control of f_0 , the timing of a perturbation within an utterance affects how speakers respond (Hilger et al., 2020; Liu et al., 2009; Ning, 2022). The location of emphasis within a sentence (i.e., on subject, verb, or object) may also affect the acoustic realization of emphasis (Breen et al., 2010). Because we only manipulated f_0 at the start of the sentence, on the subject of the sentence, we cannot say whether the timing or target of an f_0 error also affects f_0 adaptation.

This study evaluated only f_0 and amplitude, though duration is also an important cue of emphasis. Measuring

changes in word duration, however, requires an analysis window that would incorporate both feedforward and feedback responses, and thus changes could not solely be attributed to sensorimotor adaptation. There are also many ways in which f_0 is used to convey meaning, and this study evaluated only one—a type of emphasis known as narrow focus (Ladd, 2008). Future research should confirm if results are similar when f_0 is important for constructing other meanings (e.g., contrastive or corrective focus).

Finally, the difference in perturbation duration noted above (see the Adapting f_0 in Sustained Vowels: Limitations of a Common Approach section) prevented direct comparison of f_0 responses between vowel and sentence tasks. Qualitatively, adaptive f_0 responses appeared largest during the vowel task, but this may not prove true if the duration of the vowels and words were equivalent. No study, to our knowledge, has tested the effect of perturbation duration on sensorimotor adaptation of f_0 . However, research on limb motor control showed that error rates decreased with longer trial durations in a cursor movement task (Hardwick et al., 2017). If a similar effect is true of speech, differences between the vowel and sentence tasks in the present study could not necessarily be attributed to the presence or absence of linguistic meaning. A vowel task entailing naturally short rather than sustained vowels would facilitate a useful comparison between linguistically meaningful and neutral speech while controlling for perturbation duration. Such future work would allow for additional conclusions regarding the effect of linguistic meaning on f_0 adaptation and the suitability of vowel tasks for investigating f_0 control.

Conclusions

This study used an altered auditory feedback approach to evaluate how speakers adapt to f_0 errors that affect linguistic meaning in running speech, how that f_0 adaptation relates to responses to f_0 errors in the linguistically neutral context of a sustained vowel, and how that f_0 adaptation relates to responses in the unaltered prosodic cue of amplitude. We found robust evidence that speakers adapted to f_0 errors in all contexts, except when f_0 was shifted upward in unemphasized words in running speech. There were no differences according to shift direction in emphasized words. These findings are inconsistent with a hypothesis that a disruption of intended meaning may affect adaptive responses. However, methodological and physiological factors may have affected results, and so further work to clarify the effect of linguistic meaning on feedforward control of f_0 is warranted. We also found no relationship between f_0 adaptation in running speech and during sustained vowels, revealing a potential limitation of the

latter, more common approach to studying f_0 adaptation. Finally, we found f_0 and amplitude responses to f_0 shifts were positively correlated, with no evidence that this correlation was driven by the physiological link between these measures, thus supporting an integrated model of prosody.

Data Availability Statement

The data sets generated and/or analyzed during the current study are not publicly available due to commitments to protect participant confidentiality but are available from the corresponding author on reasonable request.

Acknowledgments

This work was supported by Grants DC021080 (K. L. D.), DC016270 (C. E. S. and F. H. G.), DC015446 (R. E. H.), and T32 DC013017 (C. E. S.) from the National Institute on Deafness and Other Communication Disorders; an ASH Foundation New Century Scholars Doctoral Scholarship (K. L. D.); and a PhD Scholarship from the Council of Academic Programs in Communication Sciences and Disorders (K. L. D.). The authors thank Sarah Cocroft, Taylor Feaster, and Jen Weston for their help with data analysis.

References

- Abur, D., Lester-Smith, R. A., Daliri, A., Lupiani, A. A., Guenther, F. H., & Stepp, C. E. (2018). Sensorimotor adaptation of voice fundamental frequency in Parkinson's disease. *PLOS ONE*, 13(1), Article e0191839. <https://doi.org/10.1371/journal.pone.0191839>
- Abur, D., Subaciute, A., Daliri, A., Lester-Smith, R. A., Lupiani, A. A., Cilento, D., Enos, N. M., Weerathunge, H. R., Tardif, M. C., & Stepp, C. E. (2021). Feedback and feedforward auditory-motor processes for voice and articulation in Parkinson's disease. *Journal of Speech, Language, and Hearing Research*, 64(12), 4682–4694. https://doi.org/10.1044/2021_JSLHR-21-00153
- Abur, D., Subaciute, A., Kapsner-Smith, M., Segina, R. K., Tracy, L. F., Noordzij, J. P., & Stepp, C. E. (2021). Impaired auditory discrimination and auditory-motor integration in hyperfunctional voice disorders. *Scientific Reports*, 11(1), Article 13123. <https://doi.org/10.1038/s41598-021-92250-8>
- Anand, S., & Stepp, C. E. (2015). Listener perception of monopitch, naturalness, and intelligibility for speakers with Parkinson's disease. *Journal of Speech, Language, and Hearing Research*, 58(4), 1134–1144. https://doi.org/10.1044/2015_JSLHR-S-14-0243
- Boersma, P., & Weenink, D. (2015). *Praat: Doing phonetics by computer*. <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>
- Bunton, K., Kent, R. D., Kent, J. F., & Duffy, J. R. (2001). The effects of flattening fundamental frequency contours on sentence intelligibility in speakers with dysarthria. *Clinical Linguistics & Phonetics*, 15(3), 181–193. <https://doi.org/10.1080/02699200010003378>
- Bunton, K., Kent, R. D., Kent, J. F., & Rosenbek, J. C. (2000). Perceptuo-acoustic assessment of prosodic impairment in dysarthria. *Clinical Linguistics & Phonetics*, 14(1), 13–24. <https://doi.org/10.1080/026992000298922>
- Burnett, T. A., & Larson, C. R. (2002). Early pitch-shift response is active in both steady and dynamic voice pitch control. *The Journal of the Acoustical Society of America*, 112(3), 1058–1063. <https://doi.org/10.1121/1.1487844>
- Burnett, T. A., Senner, J. E., & Larson, C. R. (1997). Voice F0 responses to pitch-shifted auditory feedback: A preliminary study. *Journal of Voice*, 11(2), 202–211. [https://doi.org/10.1016/S0892-1997\(97\)80079-3](https://doi.org/10.1016/S0892-1997(97)80079-3)
- 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>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- 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>
- Cornelisse, L. E., Gagné, J.-P., & Seewald, R. C. (1991). Ear level recordings of the long-term average spectrum of speech. *Ear and Hearing*, 12(1), 47–54. <https://doi.org/10.1097/00003446-199102000-00006>
- de Bodt, M. S., Hernández-Díaz Huici, M. E., & van de Heyning, P. H. (2002). Intelligibility as a linear combination of dimensions in dysarthric speech. *Journal of Communication Disorders*, 35(3), 283–292. [https://doi.org/10.1016/S0021-9924\(02\)00065-5](https://doi.org/10.1016/S0021-9924(02)00065-5)
- Guenther, F. H. (2016). *Neural control of speech*. MIT Press. <https://doi.org/10.7551/mitpress/10471.001.0001>
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3), 280–301. <https://doi.org/10.1016/j.bandl.2005.06.001>
- Hardwick, R. M., Rajan, V. A., Bastian, A. J., Krakauer, J. W., & Celnik, P. A. (2017). Motor learning in stroke: Trained patients are not equal to untrained patients with less impairment. *Neurorehabilitation and Neural Repair*, 31(2), 178–189. <https://doi.org/10.1177/1545968316675432>
- Heller Murray, E. S., Lupiani, A. A., Kolin, K. R., Segina, R. K., & Stepp, C. E. (2019). Pitch shifting with the commercially available eventide eclipse: Intended and unintended changes to the speech signal. *Journal of Speech, Language, and Hearing Research*, 62(7), 2270–2279. https://doi.org/10.1044/2019_JSLHR-S-18-0408
- Heller Murray, E. S., & Stepp, C. E. (2020). Relationships between vocal pitch perception and production: A developmental perspective. *Scientific Reports*, 10(1), Article 1. <https://doi.org/10.1038/s41598-020-60756-2>
- 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. *Journal of Speech, Language, and Hearing Research*, 63(7), 2185–2201. https://doi.org/10.1044/2020_JSLHR-19-00043
- Hilger, A., Cole, J., & Larson, C. (2023). Semantic focus mediates pitch auditory feedback control in phrasal prosody. *Language, Cognition and Neuroscience*, 38(3), 328–345. <https://doi.org/10.1080/23273798.2022.2116060>

- Howell, P. (1993). Cue trading in the production and perception of vowel stress. *The Journal of the Acoustical Society of America*, 94(4), 2063–2073. <https://doi.org/10.1121/1.407479>
- Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research*, 190(3), 279–287. <https://doi.org/10.1007/s00221-008-1473-y>
- Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, 108(3), 1246–1251. <https://doi.org/10.1121/1.1288414>
- Kent, R. D., & Rosenbek, J. C. (1982). Prosodic disturbance and neurologic lesion. *Brain and Language*, 15(2), 259–291. [https://doi.org/10.1016/0093-934X\(82\)90060-8](https://doi.org/10.1016/0093-934X(82)90060-8)
- Keough, D., & Jones, J. A. (2009). The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing. *The Journal of the Acoustical Society of America*, 126(2), 837–846. <https://doi.org/10.1121/1.3158600>
- Klopfenstein, M. (2015). Relationship between acoustic measures and speech naturalness ratings in Parkinson's disease: A within-speaker approach. *Clinical Linguistics & Phonetics*, 29(12), 938–954. <https://doi.org/10.3109/02699206.2015.1081293>
- Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511808814>
- Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., & Hain, T. C. (2001). Comparison of voice F0 responses to pitch-shift onset and offset conditions. *The Journal of the Acoustical Society of America*, 110(6), 2845–2848. <https://doi.org/10.1121/1.1417527>
- Larson, C. R., Sun, J., & Hain, T. C. (2007). Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control. *The Journal of the Acoustical Society of America*, 121(5), 2862–2872. <https://doi.org/10.1121/1.2715657>
- Laures, J. S., & Weismer, G. (1999). The effects of a flattened fundamental frequency on intelligibility at the sentence level. *Journal of Speech, Language, and Hearing Research*, 42(5), 1148–1156. <https://doi.org/10.1044/jslhr.4205.1148>
- Lester-Smith, R. A., Daliri, A., Enos, N., Abur, D., Lupiani, A. A., Letcher, S., & Stepp, C. E. (2020). The relation of articulatory and vocal auditory-motor control in typical speakers. *Journal of Speech, Language, and Hearing Research*, 63(11), 3628–3642. https://doi.org/10.1044/2020_JSLHR-20-00192
- Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, 32(4), 451–454. <https://doi.org/10.1121/1.1908095>
- 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., 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>
- Lombard, E. (1911). Le signe de l'elevation de la voix [The sign of the elevation of the voice]. *Annales Des Maladies de l'Oreille, Du Larynx, Du Nez et Du Pharynx*, 37, 101–110.
- Martens, H., Van Nuffelen, G., Cras, P., Pickut, B., De Letter, M., & de Bodt, M. (2011). Assessment of prosodic communicative efficiency in Parkinson's disease as judged by professional listeners. *Parkinson's Disease*, 2011, Article 129310. <https://doi.org/10.4061/2011/129310>
- Ning, L.-H. (2022). The effect of stimulus timing in compensating for pitch perturbation on flat, rising, and falling contours. *The Journal of the Acoustical Society of America*, 151(4), 2530–2544. <https://doi.org/10.1121/10.0010237>
- O'Shaughnessy, D. (1979). Linguistic features in fundamental frequency patterns. *Journal of Phonetics*, 7(2), 119–145. [https://doi.org/10.1016/S0095-4470\(19\)31045-9](https://doi.org/10.1016/S0095-4470(19)31045-9)
- Patel, R. (2002). Prosodic control in severe dysarthria: Preserved ability to mark the question-statement contrast. *Journal of Speech, Language, and Hearing Research*, 45(5), 858–870. [https://doi.org/10.1044/1092-4388\(2002\)069](https://doi.org/10.1044/1092-4388(2002)069)
- Patel, R., Niziolek, C., Reilly, K., & 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\(2010\)0162](https://doi.org/10.1044/1092-4388(2010)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
- Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology*, 27(3), 887–905. https://doi.org/10.1044/2018_ajslp-17-0009
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184. <https://doi.org/10.1121/1.1906875>
- Roettger, T. B., Mahrt, T., & Cole, J. (2019). Mapping prosody onto meaning – The case of information structure in American English. *Language, Cognition and Neuroscience*, 34(7), 841–860. <https://doi.org/10.1080/23273798.2019.1587482>
- Scheerer, N. E., Jacobson, D. S., & Jones, J. A. (2016). Sensorimotor learning in children and adults: Exposure to frequency-altered auditory feedback during speech production. *Neuroscience*, 314, 106–115. <https://doi.org/10.1016/j.neuroscience.2015.11.037>
- Stepp, C. E., Lester-Smith, R. A., Abur, D., Daliri, A., Pieter, N. J., & Lupiani, A. A. (2017). Evidence for auditory-motor impairment in individuals with hyperfunctional voice disorders. *Journal of Speech, Language, and Hearing Research*, 60(6), 1545–1550. https://doi.org/10.1044/2017_JSLHR-S-16-0282
- Tang, C., Hamilton, L. S., & Chang, E. F. (2017). Intonational speech prosody encoding in the human auditory cortex. *Science*, 357(6353), 797–801. <https://doi.org/10.1126/science.aam8577>
- Titze, I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *The Journal of the Acoustical Society of America*, 85(2), 901–906. <https://doi.org/10.1121/1.397562>
- Vojtech, J. M., Noordzij, J. P., Jr., Cler, G. J., & Stepp, C. E. (2019). The effects of modulating fundamental frequency and speech rate on the intelligibility, communication efficiency, and perceived naturalness of synthetic speech. *American Journal of Speech-Language Pathology*, 28(2S), 875–886. https://doi.org/10.1044/2019_AJSLP-MS18-0052
- Weerathunge, H. R., Abur, D., Enos, N. M., Brown, K. M., & Stepp, C. E. (2020). Auditory-motor perturbations of voice fundamental frequency: Feedback delay and amplification. *Journal of Speech, Language, and Hearing Research*, 63(9), 2846–2860. https://doi.org/10.1044/2020_JSLHR-19-00407
- Witte, R. S., & Witte, J. S. (2009). *Statistics* (9th ed.). Wiley.
- Xu, Y., & Xu, A. (2021). Consonantal F perturbation in American English involves multiple mechanisms. *The Journal of the Acoustical Society of America*, 149(4), 2877–2895. <https://doi.org/10.1121/10.0004239>

Appendix (p. 1 of 2)

Analysis of f_0 Contrast

The perturbation of f_0 during one word in a sentence may have consequences for f_0 at another point in the sentence (Hilger et al., 2020, 2023). Downstream effects of f_0 perturbations may be especially likely in productions with emphasis since emphasis relies heavily on prosodic contrasts. If that contrast were disrupted by an f_0 error in one word, the speaker may maintain the contrast by adjusting the f_0 of a neighboring word.

Patel et al. (2011) accounted for this phrase-level aspect of prosody by measuring responses to f_0 perturbations as changes in the f_0 contrast between neighboring emphasized and unemphasized words. This approach captured both feed-forward and feedback responses and was thus incompatible with the primary objective of our study. However, to better situate our findings within the context of this related work, we also analyzed changes in f_0 contrast between the first and second word of each sentence as a potential effect of the f_0 perturbations.

Four participants (3 women, 1 man) were excluded from this supplemental analysis because glottal fry precluded f_0 extraction from over 30% of hold and after-effect trials. For the remaining 20 participants, a trained research assistant extracted f_0 (Hz) from the second word of each sentence (full analysis window) and converted it to cents. The f_0 (cents) of the second word was subtracted from the first to quantify the f_0 contrast. The f_0 contrast of each trial in the control condition was subtracted from each corresponding trial in f_0 -shifted conditions for a given speech task. The normalized f_0 contrast was averaged over every four trials for data visualization. Results are shown in Figure A1 and Table A1.

Table A1. Results of a repeated-measures analysis of variance for fundamental frequency (f_0) contrast between the first two words of sentences.

Effect	<i>df</i>	<i>F</i>	<i>p</i>	η_p^2	Effect size
Emphasis	1	3.16	.077	—	—
Shift	1	6.20	.014*	.03	Small
Phase	2	0.56	.575	—	—
Emphasis × Shift	1	2.92	.089	—	—
Emphasis × Phase	2	0.81	.448	—	—
Shift × Phase	2	1.81	.166	—	—
Emphasis × Shift × Phase	2	0.93	.395	—	—

*Significant at $\alpha = .05$. — not applicable for nonsignificant results.

Figure A1. Group mean contrast between f_o of the first (perturbed) and second (unperturbed) words in a three-word sentence. Contrasts are plotted during downward (Panel A) and upward (Panel B) shifts (± 200 cents) in auditory feedback of f_o during emphasized (blue) and unemphasized (dashed pink) words in running speech. Shaded areas are 95% confidence intervals.

