JSLHR

Research Article

Responses to Intensity-Shifted Auditory Feedback During Running Speech

Rupal Patel,[a,b] Kevin J. Reilly,[c] Erin Archibald,[a]
Shanqing Cai,[d] and Frank H. Guenther[b,d]

Purpose: Responses to intensity perturbation during running speech were measured to understand whether prosodic features are controlled in an independent or integrated manner.
Method: Nineteen English-speaking healthy adults (age range = 21–41 years) produced 480 sentences in which emphatic stress was placed on either the 1st or 2nd word. One participant group received an upward intensity perturbation during stressed word production, and the other group received a downward intensity perturbation. Compensations for perturbation were evaluated by comparing differences in participants' stressed and unstressed peak fundamental frequency (F0), peak intensity, and word duration during perturbed versus baseline trials.
Results: Significant increases in stressed–unstressed peak intensities were observed during the ramp and perturbation phases of the experiment in the downward group only. Compensations for F0 and duration did not reach significance for either group.
Conclusions: Consistent with previous work, speakers appear sensitive to auditory perturbations that affect a desired linguistic goal. In contrast to previous work on F0 perturbation that supported an integrated-channel model of prosodic control, the current work only found evidence for intensity-specific compensation. This discrepancy may suggest different F0 and intensity control mechanisms, threshold-dependent prosodic modulation, or a combined control scheme.

Communication is dependent as much on what one says (segmental speech) as on how one says it (prosody). It is widely agreed that changes in fundamental frequency (F0), duration, and amplitude convey a variety of linguistic contrasts (e.g., Bolinger, 1958, 1989; Fry, 1955, 1958; Kochanski, Grabe, Coleman, & Rosner, 2005; Lehiste, 1976; Shattuck-Hufnagel & Turk, 1996; Turk & Sawusch, 1996; Welby, 2003). The relative importance of these prosodic cues, however, remains contested. For example, whereas Fry (1958) and Turk and Sawusch (1996) found that duration is the prominent prosodic cue to indicate sentential stress, Kochanski et al. (2005) have argued that vocal intensity, or loudness, is a stronger predictor. Yet others have suggested that pitch cues are key to the perception of focus or stress (Bolinger, 1958; 't Hart, Collier, & Cohen, 1990; Lieberman, 1960; Welby, 2003). Thus, a persistent question is how these cues interact and whether they are manipulated in coordination or in isolation in order to convey linguistic contrasts in connected speech.

Despite the importance of prosody, computational models of speech production have focused almost exclusively on segmental aspects (Guenther, 1994, 1995; Guenther, Ghosh, & Tourville, 2006; Hickok, 2012). The current study has implications for incorporating prosodic control into computational models of speech production such as the Directions Into Velocities of Articulators model (Guenther et al., 2006; Guenther, 1994, 1995), which presently focuses on segmental aspects of speech. As a first step in modeling prosodic control, Patel, Niziolek, Reilly, and Guenther (2011) proposed two competing models: the *integrated-channel model*, in which prosodic components (e.g., pitch, intensity, duration) are modulated in coordination with each other, and the *independent-channel model*, in which each component is controlled individually. Though several studies have described cue-trading relations, or speakers' natural tendency to use and perceive intensity, duration, and F0 interchangeably to convey linguistic contrasts (Howell, 1993; Lieberman, 1960), the phenomenon itself does not shed light on how prosodic cues are represented and programmed during running speech.

[a]Northeastern University, Boston, MA
[b]Harvard–MIT Division of Health Sciences and Technology, Cambridge, MA
[c]University of Tennessee Health Science Center, Knoxville
[d]Boston University, MA
Correspondence to Rupal Patel: r.patel@neu.edu

Editor: Jody Kreiman
Associate Editor: Kate Bunton
Received May 2, 2015
Revision received August 11, 2015
Accepted August 12, 2015
DOI: 10.1044/2015_JSLHR-S-15-0164

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

The auditory-perturbation paradigm is an approach that can inform the nature of the relationship between prosodic cues by assessing a speaker's responses to shifts in a particular cue. Studies of F0 perturbation during isolated syllables or words have found that a majority of speakers alter their F0 in the opposite direction of the perturbation (Burnett, Freedland, Larson, & Hain, 1998; Chen, Liu, Xu, & Larson, 2007; Jones & Munhall, 2002). This protocol can be extended to connected speech in which the targeted F0 perturbation of a select word can alter a meaningful linguistic contrast (e.g., word stress, question/statement contrasts) in the sentences speakers hear themselves producing. In one such study, Patel et al. (2011) selectively perturbed the F0 of the word containing emphatic stress in sentences produced by speakers. One group of speakers received an upward F0 perturbation during their stressed-word productions, and a second group received a downward F0 perturbation. Analysis of the stressed–unstressed contrast yielded two primary findings: (a) Speakers increased F0 during stressed versus unstressed words in the group that received the downward F0 perturbation but not the upward perturbation; and (b) responses to the F0 perturbation in this group were not limited to F0 but also included increases in the intensity of the stressed versus unstressed word in each sentence. Taken together, these findings provide support for an integrated-channel model of prosodic control that is tuned to optimizing linguistic contrasts.

The current study extends this line of inquiry to determine whether intensity perturbations will also elicit compensatory responses. Although it has been demonstrated that intensity perturbations at the isolated vowel, syllable, or word level result in compensation in the opposite direction (Bauer, Mittal, Larson, & Hain, 2006; Heinks-Maldonado & Houde, 2005; Larson, Sun, & Hain, 2007; Liu, Zhang, Xu, & Larson, 2007), it is unclear whether intensity perturbations during connected speech will produce compensations in other prosodic variables as well, and whether these compensations will be modulated by linguistic intent. Both upward and downward perturbations are examined to determine the impact of linguistic goals on direction and magnitude of compensation.

The specific research questions are the following:

1. Do speakers adapt to perturbations of the intensity difference between stressed and unstressed words during connected speech?
2. Do speakers exhibit greater compensations for perturbations that violate stressed–unstressed contrasts compared with perturbations that do not violate this contrast?
3. Which prosodic features do speakers use when compensating for intensity-shifted auditory feedback?

Method

Participants

Nineteen healthy, monolingual adult speakers of American English were recruited (mean age = 25.8 years; range = 21–41 years). None of the participants reported the presence of speech-language disorders, neurological disorders, or hearing loss. All participants passed hearing screenings at 25 dB at 250, 500, 1000, 2000, 4000, and 8000 Hz in both ears using binaural headphones. Participants were randomly assigned to groups (upward-shifted or downward-shifted perturbation) as they were accepted into the study. The upward-shifted (UP) group consisted of six men and five women (mean age = 25.9 years). The downward-shifted (DN) group consisted of three men and five women (mean age = 25.6 years).

Experimental Stimuli

To simulate a linguistic goal during connected speech, speakers produced four pairs of short sentence, with one sentence in each pair containing emphatic stress on the first word and the other containing emphatic stress on the second word. In this way, each sentence pair conveyed contrastive meanings but contained the same speech segments. Each sentence comprised four monosyllabic words; within a sentence, the vowel was held constant to control for vowel-dependent changes in F0 that are unrelated to prosodic control. To elicit contrastive meanings, each sentence was preceded by a contextual question presented visually, which cued the speaker to place stress on either the first or second word of the sentence. For example, the question "What did Doug do to a bud?" prompted stress on "cut" in the sentence "Doug CUT a bud." On the converse, the question "Who cut a bud?" elicited emphasis on the word "Doug" in the same sentence. The four sentence stimuli are listed in Table 1.

Procedure

Participants were seated in a sound-treated booth with visual access to a monitor that displayed the information for each trial. At the beginning of a trial, participants were presented with a question that was followed 1 s later by the sentence they were to produce for that trial. The question was presented in brackets and cued the participants to the word with lexical stress in the sentence that followed. In addition, the sentence was presented with the stressed word capitalized and displayed in a bold, red font to distinguish it further from the unstressed words in the sentence. Participants were instructed to read the prompt silently and then produce the sentence with the appropriate stress placement to convey the intended meaning. Speech signals were obtained using a head-worn directional microphone (C520, AKG,

Table 1. Experimental stimuli by location of linguistic stress (denoted by all caps).

Stress 1	Stress 2
DAD tagged a cat.	Dad TAGGED a cat.
DOUG cut a bud.	Doug CUT a bud.
BOB caught a dog.	Bob CAUGHT a dog.
DICK bit a kid.	Dick BIT a kid.

Vienna, Austria) placed at a fixed distance of approximately 5.5 cm from the participant's lips. Speech auditory feedback was delivered to each participant using calibrated, noise-isolating insert earphones (ER-4 microPro, Etymotic Research, Elk Grove Village, IL). After producing the target sentence for a given trial, participants began the next trial by pressing the space bar on a keyboard in front of them.

A brief testing session was conducted immediately prior to the experiment to determine each participant's perturbation threshold. The testing session consisted of two productions of each of the eight speech stimuli, for a total of 16 trials. The setup for the testing session was identical to that of the experiment except that the participants' auditory feedback was not perturbed. A custom graphical user interface was designed in MATLAB to analyze the speech recordings of the testing session. A trained user marked the onset and offset of the stressed word in each utterance. The interface then displayed histograms of the decibel values in the stressed and unstressed word productions and identified threshold as the lowest decibel value that occurred more frequently in stressed words than unstressed words.

The study consisted of four experimental phases varying in perturbation magnitude: a baseline (BASE) phase in which no perturbation was applied, a ramp (RAMP) phase in which perturbation was applied in gradually increasing increments, a full perturbation (PERT) phase in which the perturbation was maintained at a maximum value, and a postperturbation (POST) phase in which the perturbation value was again set to zero. The scaling factor for transforming the input (microphone) intensity to output (feedback) was derived using the following formulae:

$$\text{Up} : \text{intensityscale} = 1 + ([\text{dB}/\text{threshold} - 1] \times \text{pertval})$$
$$\text{Down} : \text{intensityscale} = 1 - ([\text{dB}/\text{threshold} - 1] \times \text{pertval}).$$

In these formulae, the magnitude of the scaling factor intensityscale changed according to the instantaneous speech intensity dB and the values of the parameters pertval and threshold. The parameter pertval controlled the perturbation magnitude associated with each phase of the experiment. This coefficient was set to 0 during the BASE phase, gradually increased to .5 during the RAMP phase, held at .5 during the PERT phase, and reset to 0 during the POST phase. The parameter threshold was set to the intensity level that best separated a participant's stressed and unstressed word productions during the preexperiment testing session (see earlier). As a result, the intensity level of the feedback heard by the participant through the headphones was proportional to both pertval and the amount by which the participant's instantaneous intensity exceeded threshold.

Perturbation of speech-intensity feedback was accomplished by running custom software routines on a DSK 6713 digital signal-processing (DSP) board (Texas Instruments, Dallas, TX) that introduced only a minimal processing delay of ~16 ms between the microphone and headphone channels. The values for threshold and pertval were sent via USB to the DSP board prior to every trial. The microphone signal was amplified and then split into two channels using a commercial analog mixer (Mackie 402VLZ3, LOUD Technologies, Inc., Woodinville, WA). One channel was sent to the computer for recording, and the other was sent to the DSP board for processing. Speech input to the DSP board was acquired in 8-ms blocks at a sampling rate of 8 kHz and then double buffered. The intensity of each 16-ms block of speech data was calculated, and the speech data were then scaled using the formulae listed previously. The output of the DSP board was sent to the sound mixer, where it was split again and sent both to the participant's headphones and to the recording computer. As a result, both the microphone input signal and the intensity-transformed output signal were recorded for each trial.

Acoustic Analysis

As in previous work (Patel et al., 2011), a custom software tool called CLAAS (CadLab Acoustic Analysis Software) was used to perform the acoustic analysis. First, word boundaries were manually marked for each of the 480 sentences per participant (interrater reliability for 10% of trials = 94.3%). For each sentence, F0 and intensity contours were extracted in CLAAS (using the Praat autocorrelation algorithm) and used along with the word-boundary annotations to determine peak F0, peak intensity, and duration of each word.

Preliminary data analysis resulted in the exclusion of one participant from the study due to inability to follow the experimental task (i.e., inaccurate and inconsistent stress placement throughout the protocol). Two additional participants with greater than 10% of trials with failed F0 tracking were not included. In all, 13,440 utterances were collected (28 participants × 480 utterances). Manual correction of F0 tracking was required on 15.2% of trials. There were 232 utterances (2%) removed from analysis due to failed F0 tracking as a result of glottal fry or due to either speech errors or hesitations. Although mean F0, mean intensity, and word duration were calculated for all four words in each utterance, analyses were limited to Words 1 and 2 (W1 and W2), as contrastive stress placement occurred on only one of these two words in each trial.

Statistical Analysis

Responses to intensity perturbation were measured across three acoustic measures of prosody: peak intensity (I), peak F0 (F), and word duration (D). Response magnitude was measured using the contrast distance between stressed and unstressed words. To calculate the contrast distance, the unstressed word (W1 or W2) was subtracted from the stressed word (W2 or W1). For intensity, F0, and duration, the respective contrast distances were denoted C_I, C_F, and C_D. They were defined as

$$\begin{aligned} C_I &= I_{\text{S}} - I_{\text{U}} \\ C_F &= F_{\text{S}} - F_{\text{U}} \\ C_D &= D_{\text{S}} - D_{\text{U}}, \end{aligned} \tag{1}$$

where I_{S}, F_{S}, and D_{S} are the peak intensity, peak F0, and duration extracted from the stressed word and I_{U}, F_{U}, and

D_U are the values of the three measures extracted from the unstressed word. The mean contrast distance in the BASE phase was calculated for each measure, leading to the mean BASE contrast distances $\overline{C}_I^b$, $\overline{C}_F^b$, and $\overline{C}_D^b$ for intensity, pitch, and duration, respectively. In order to quantify the compensatory changes in the participants' productions, these mean-based contrast distances were used to normalize the contrast distances obtained from the individual trials, leading to normalized contrast distances for peak intensity, peak F0, and word duration, defined respectively as

$$\begin{cases} M_I = \left(\dfrac{C_I}{\overline{C}_I^b} - 1\right) = \left(\dfrac{I_S - I_U}{\overline{C}_I^b} - 1\right); \\ M_F = \left(\dfrac{C_F}{\overline{C}_F^b} - 1\right) = \left(\dfrac{F_S - F_U}{\overline{C}_F^b} - 1\right); \\ M_D = \left(\dfrac{C_D}{\overline{C}_D^b} - 1\right) = \left(\dfrac{D_S - D_U}{\overline{C}_D^b} - 1\right). \end{cases} \quad (2)$$

Note that a value of 0 in a normalized contrast distance indicates no change from the mean BASE value; a value greater than 0 or less than 0 indicates an increase or a decrease from the mean BASE value, respectively. In addition, to form an integrative measure of compensatory response to the perturbation that includes changes in all three prosodic cues, we defined the composite prosody alteration (CPA) score as the arithmetic mean of the three normalized contrast distances:

$$P = (1/3)(M_I + M_F + M_D). \quad (3)$$

For each normalized contrast distance (M_I, M_F, and M_D) and the CPA score (P), comparisons were made for all three phases after BASE. This leads to the issue of multiple comparisons. To address this issue, we used Monte Carlo permutation tests to control for Type I errors. The principles of this method were described by Westfall and Young (1993). In brief, for each between-groups comparison (DN vs. UP) of a dependent variable (M_I, M_F, M_D, P), the group labels of the participants were randomly reshuffled, the statistical test (the nonparametric Wilcoxon rank-sum test) was performed on the reshuffled data for the three phases (RAMP, PERT, and POST), and the minimum of the three p values was recorded. Over 20,000 iterations, the reshuffling leads to an approximate null distribution of the minimum p value, with which the uncorrected p values from the original, nonshuffled data were compared to generate the corrected p values. For each within-group test, random reassignment of the sign and the nonparametric Wilcoxon signed-rank test were used during each iteration. For these tests, all possible combinations of positive and negative signs were evaluated for each permutation test.

Results

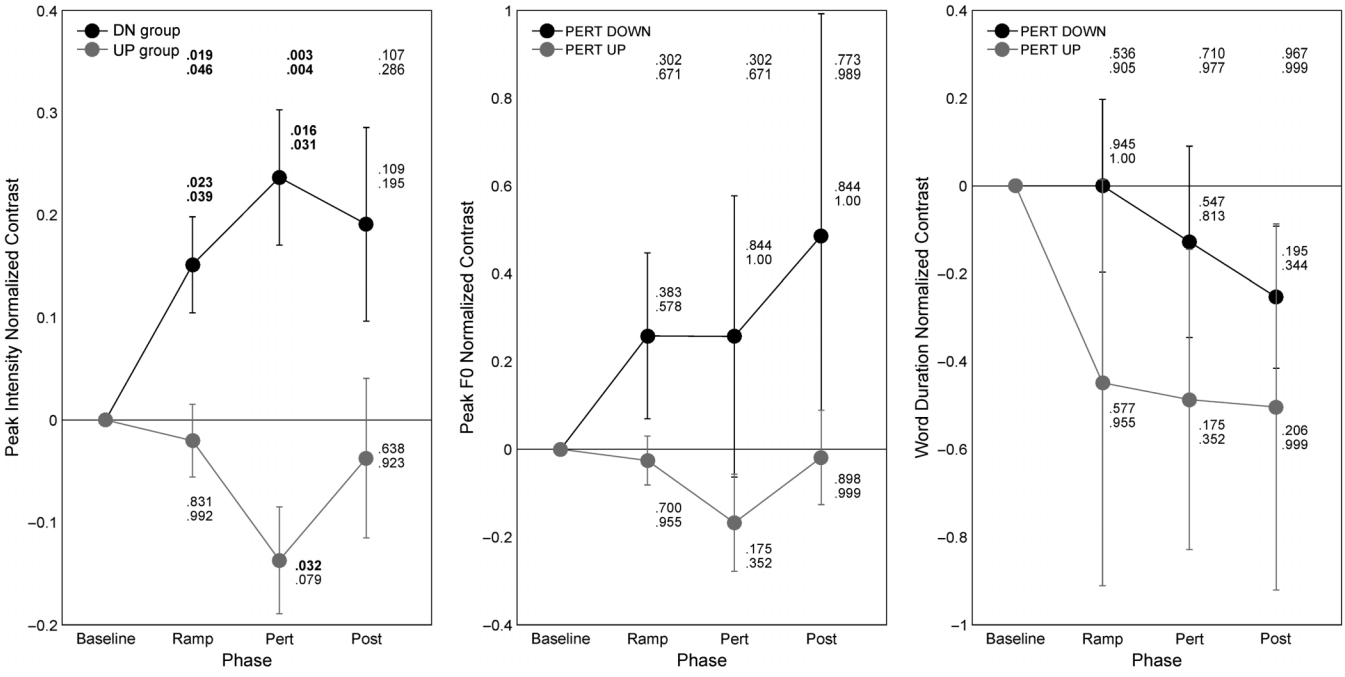
The present study examined compensations for perturbations of auditory feedback that either increased or decreased the speech intensity of the word containing emphatic stress in short sentences. Responses to the auditory perturbations were quantified by measuring the peak intensity, peak F0, and word duration of the stressed word (W1 or W2) or the unstressed word (W2 or W1) on each trial. The differences between the stressed and unstressed value for each variable were calculated and then normalized by the mean stressed–unstressed difference produced during the BASE phase. The resulting normalized measures quantified the stressed–unstressed contrast distance for each individual prosodic variable, and an additional measure, CPA score, quantified the integrative contrast distance across all three variables.

Normalized Contrast Distances

The three panels of Figure 1 show the means and standard errors of the normalized contrast distances for peak intensity (left panel), peak F0 (middle panel), and word duration (right panel). Within-group analyses of normalized intensity contrast distances were performed to evaluate whether participants in either group altered peak-intensity differences between stressed and unstressed words in response to the stressed-word perturbations of speech intensity. In the DN group, significant increases in intensity contrast distance were present in the RAMP ($p < .05$; mean increase = .15) and PERT ($p < .005$; mean increase = .24) phases compared with the BASE phase. No significant difference between the POST and BASE phases was detected. At the top of each panel in Figure 1 are displayed the uncorrected (top) and corrected (bottom) p values denoting the significance of each between-groups comparison. The numbers located near each data point are the uncorrected (top) and corrected (bottom) p values for the differences from the BASE mean observed for each participant group (0; corrected for multiple comparisons with the permutation tests; see Method for details). The differences that are significant ($p < .05$, corrected) are highlighted by the bold font. The pattern of findings in the DN group indicates that participants compensated for the intensity perturbation but that these compensations were not maintained during the POST phase, when the perturbation was turned off. In contrast to the findings for the DN group, intensity contrast distances in the UP group did not deviate significantly from their BASE values during any phase of the experiment.

A between-groups analysis of normalized intensity contrast distances revealed significant differences between the DN and UP groups during the RAMP ($p < .05$) and PERT ($p < .01$) phases, but not the POST phase. The average difference between the two participant groups was .17 during the RAMP phase and .37 during the PERT phase. The findings of the between-groups analysis indicate that intensity responses to the perturbation were dependent on the direction of the perturbation. To be specific, participants in the DN group compensated for perturbations that reduced the intensity

Figure 1. Normalized contrast distance for the three prosodic cues: peak intensity (A), peak fundamental frequency (F0; B), and duration (C). The error bars show ±1 standard error of the mean. The corrected *p* values for the within-group difference between the normalized contrast distances and the BASE mean (0) are listed beside the data points. The corrected *p* values for the within-phase, between-groups differences are shown at the top of each panel. Pert = perturbation.



contrast between stressed and unstressed words in an utterance, whereas participants in the UP group did not compensate for perturbations that augmented that same contrast.

Analysis of F0 and duration contrast distances failed to identify significant deviations from BASE values for either variable during the RAMP, PERT, or POST phases of the experiment. Between-groups comparisons similarly did not reveal any significant differences in the F0 and duration contrast distances of the UP and DN groups during the RAMP, PERT, or POST phases of the experiment, although all three phases showed nonsignificant trends in the direction expected for partial compensation for the intensity perturbation (i.e., higher F0 and longer duration for the DN group compared with the UP group). These findings indicate that compensation responses, when present, were primarily for the prosodic feature that was perturbed (i.e., intensity).

CPA

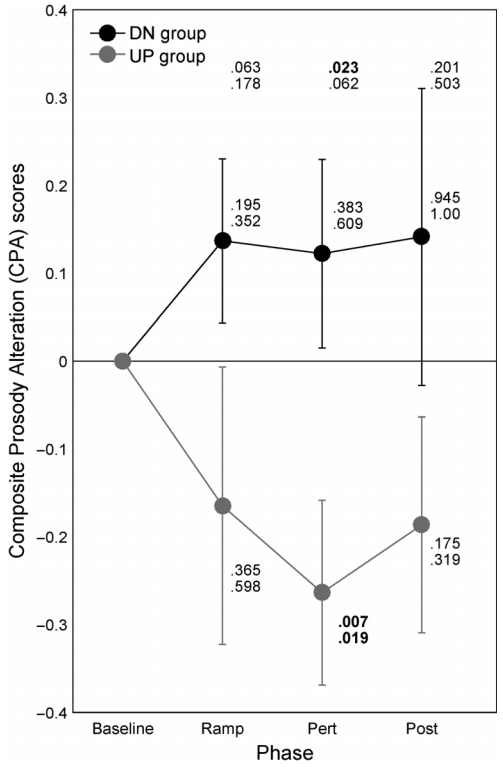
For each participant, CPA scores were calculated as a measure of the combined contrast distance between stressed and unstressed words across the three prosody variables (see Equation 3 in Methods for an operational definition of CPA). Figure 2 displays the means and standard errors of the CPA scores by phase for the DN (black) and UP (gray) groups. Judging by the permutation-corrected *p* values, significant changes from BASE were seen in the PERT phase ($p < .05$) in the UP group. In this group, CPA scores during PERT were on average 0.26 less than those observed during BASE. Significant deviations from BASE were not present for any phase in the DN group. These within-group findings were due, at least in part, to a tendency for participants in both groups to reduce normalized word-duration contrasts over the course of the experiment (Figure 1, right panel). This reduction in duration contrasts resulted in CPA scores that were less positive in the DN group and more negative in the UP group.

An analysis of between-groups differences in CPA scores with permutation-corrected *p* values did not reveal significant differences between the DN and UP groups during any phase of the experiment. Again, this finding was due in part to the observation that normalized duration contrasts changed in the same direction (i.e., downward) in both groups of participants, which reduced between-groups differences in CPA. Although no significant differences in CPA were found between the two groups, a nonsignificant trend in the expected direction (higher CPA for the DN group than the UP group) is evident in the RAMP, PERT, and POST phases.

Discussion

Numerous studies have demonstrated that auditory perturbations elicit rapid compensations in speech output that correct for the perturbation-induced errors in auditory

Figure 2. Composite prosody alteration scores for the DN and UP groups. The error bars show ±1 standard error of the mean. As in Figure 1, the corrected *p* values for the within-group difference from the BASE mean are shown beside each data point; the corrected *p* values for the between-groups differences are shown at the top.



feedback (Bauer et al., 2006; Chen et al., 2007; Heinks-Maldonado & Houde, 2005; Larson et al., 2007; Liu et al., 2007). These studies not only confirm the importance of auditory feedback to speech motor control, but also provide insight into feed-forward representations for speech. In the present study, responses to the perturbation of stressed and unstressed word intensities were examined to elaborate feed-forward mechanisms of speech prosody control. Significant compensations were observed for participants in the DN group but not in the UP group, indicating that compensations were associated with perturbations that reduced the contrast between the stressed versus unstressed words in sentences and not with perturbations that increased this contrast. Moreover, compensations consisted primarily of changes to participants' speech intensity level; changes in nonperturbed prosodic cues were not statistically significant.

The failure to observe significant compensation in the UP group could be attributable to the fact that there was not a linguistically meaningful reason for participants to compensate for an upward shift in intensity. In particular, upward shifts in intensity during the production of a stressed word only serve to bring the participant closer to his or her linguistic goal, and compensation for this shift would be counterproductive (Fry, 1955, 1958; Junqua, 1996; Kochanski et al., 2005; Turk & Sawusch, 1996). As a result, the present findings indicate that linguistic goals interact with an auditory perturbation in determining participants' compensation responses. Goal-dependent modulation in speech compensation has been observed in response to perturbations of both segmental-level (Mitsuya, MacDonald, Purcell, & Munhall, 2011; Mitsuya, Samson, Ménard, & Munhall, 2013; Niziolek et al., 2013; Reilly & Dougherty, 2013) and sentential-level stimuli (Chen et al., 2007; Patel et al., 2011). For example, the companion study by Patel et al. (2011) reported significantly greater compensation for auditory perturbations that decreased the pitch of stressed words in a sentence than to perturbations that increased the pitch. The finding of these and other studies indicate that compensations for perturbations of auditory feedback are not uniform but are instead tuned to feed-forward linguistic goals.

The current findings are somewhat at odds with our related previous study on F0 perturbation in that significant compensations here were limited to the perturbed prosodic feature, speech intensity, and did not involve other, nonperturbed prosodic features. In contrast, our companion study (Patel et al., 2011) found that speakers compensated for perturbations of pitch during stressed words in a sentence by modifying multiple prosodic cues (F0, intensity, and duration). There are several possible explanations for this discrepancy. First, it is possible that control of intensity differs from control of F0 in that the latter involves responding to perceived F0 errors with adjustments to multiple prosodic cues (F0, intensity, and duration) in an integrated fashion, whereas the former involves adjustments only to the perturbed feature (intensity) as in the independent-channel model. As noted in the Results section, however, both F0 and duration showed nonsignificant trends in the direction of compensation in the RAMP, PERT, and POST phases when comparing the responses to UP and DN intensity perturbations (Figure 1).

This raises a second possibility, which is that the discrepancy may be due to differences in perturbation magnitude across the studies. Perhaps the magnitude of pitch perturbation was perceptually more salient, and thus the compensation required a more integrated response compared with the level of intensity perturbation. In fact, related studies on speech in noise (a naturally occurring intensity perturbation) provide corroborating evidence that at soft noise levels, speakers increase intensity to overcome the competing noise level, yet at more extreme noise levels, multiple prosodic cues are altered (Junqua, 1996; Lane & Tranel, 1971; Patel & Schell, 2008). For example, Patel and Schell (2008) found that speakers shifted intensity upward but maintained F0 and durational contrasts when speaking in 60-dB noise versus quiet. However, when speaking in 90-dB noise, speakers increased not only intensity, but also duration and F0. Moreover, compensations were linguistically modulated in that content words within a sentence were lengthened and heightened in pitch to a greater extent than function words. These findings support a threshold model of prosodic control, such that all three cues are recruited (integrated-channel model) for conveying prosodic

changes above a certain threshold, but only the perturbed cue is adjusted for subthreshold perturbations. Perhaps the magnitude of the intensity perturbation in the current study did not meet the threshold necessary to recruit all three cues (independent-channel model) in a compensatory response.

A third possibility is that auditory targets exist for the three stress cues individually (as in the independent-channel model) as well as for a combined stress cue (as in the integrated-channel model). Interpreted within the framework of the Directions Into Velocities of Articulators model, perturbation of intensity would lead to an error signal for the intensity cue as well as the overall stress cue. Both of these error signals would lead to the generation of motor commands that increase intensity. The stress-cue error would also generate compensatory motor commands for F0 and duration. However, as F0 and duration change in response to these commands, they will move out of their target regions, leading to error signals and corrective movements that counteract the compensatory response for the stress-cue error. The end result would be relatively small compensatory responses in F0 and duration that might not reach statistical significance, as in the current findings.

Future Directions

The experimental protocol used in the present study allowed for the incorporation of a meaningful linguistic goal within an intensity-perturbation paradigm. However, the stimuli used called for alternating stress placement on one of two words in a sentence. An upward shift in intensity inherently brings speakers closer to this goal, and therefore participants may have been less motivated to compensate for an upward perturbation. To further explore the significance of the directional bias observed here, future work may consider stimuli that allow for meaningful compensation in both directions (i.e., contrasts such as question vs. statement). In addition, the role of perturbation magnitude in observing integrated versus independent channel responses warrants further investigation. Nonetheless, it is evident that the use of linguistically salient stimuli within an auditory-perturbation framework can shed light on the neural control of prosody and merits further exploration.

Acknowledgments

This research was supported in part by National Institute on Deafness and Other Communication Disorders Grants R01 DC002852 (F. Guenther, P. I.) and R03 DC011159. (K. Reilly, P. I.)

References

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

Bolinger, D. (1958). Stress and information. *American Speech, 33,* 5–20.

Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse.* Stanford, CA: Stanford University Press.

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,* 3153–3161.

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F_0 responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America, 121,* 1157–1163.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America, 27,* 765–768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1,* 126–152.

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics, 72,* 43–53.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech. *Psychological Review, 102,* 594–621.

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,* 280–301.

Heinks-Maldonado, T. H., & Houde, J. F. (2005). Compensatory responses to brief perturbations of speech amplitude. *Acoustics Research Letters Online, 6,* 131–137.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13,* 135–145.

Howell, P. (1993). Cue trading in the production and perception of vowel stress. *The Journal of the Acoustical Society of America, 94,* 2063–2073.

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: Studies of Mandarin tone production. *Journal of Phonetics, 30,* 303–320.

Junqua, J.-C. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication, 20,* 13–22.

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

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research, 14,* 677–709.

Larson, C. R., Sun, J., & Hain, T. C. (2007). Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F_0 and amplitude control. *The Journal of the Acoustical Society of America, 121,* 2862–2872.

Lehiste, I. (1976). Suprasegmental features of speech. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 225–239). New York, NY: Academic Press.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America, 32,* 451–454.

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,* 2405–2412.

Mitsuya, T., MacDonald, E. N., Purcell, D. W., & Munhall, K. G. (2011). A cross-language study of compensation in response to real-time formant perturbation. *The Journal of the Acoustical Society of America, 130,* 2978–2986.

Mitsuya, T., Samson, F., Ménard, L., & Munhall, K. G. (2013). Language dependent vowel representation in speech production. *The Journal of the Acoustical Society of America, 133,* 2993–3003.

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,* 1051–1059.

Patel, R., & Schell, K. G. (2008). The influence of linguistic content on the Lombard effect. *Journal of Speech, Language, and Hearing Research, 51,* 209–220.

Reilly, K. J., & Dougherty, K. E. (2013). The role of vowel perceptual cues in compensatory responses to perturbations of speech auditory feedback. *The Journal of the Acoustical Society of America, 134,* 1314–1323.

Shattuck-Hufnagel, S., & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25,* 193–247.

't Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody.* Cambridge, United Kingdom: Cambridge University Press.

Turk, A. E., & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America, 99,* 3782–3790.

Welby, P. (2003). Effects of pitch accent position, type, and status on focus projection. *Language and Speech, 46,* 53–81.

Westfall, P. H., & Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for* p-*value adjustment*. New York, NY: Wiley.