Article

Prosodic Adaptations to Pitch Perturbation in Running Speech

Rupal Patel,[a,b] Caroline Niziolek,[b] Kevin Reilly,[a] and Frank H. Guenther[b,c]

Purpose: A feedback perturbation paradigm was used to investigate whether prosodic cues are controlled independently or in an integrated fashion during sentence production.

Method: Twenty-one healthy speakers of American English were asked to produce sentences with emphatic stress while receiving real-time auditory feedback of their productions. The fundamental frequency (F0) of the stressed word in each 4-word sentence was selectively shifted in a sensorimotor adaptation protocol. Speakers experienced either an upward or a downward shift of the stressed word, which gradually altered the perceived stress of the sentence.

Results: Participants in the Up and Down groups adapted to F0 shifts by altering the contrast between stressed and unstressed words differentially, such that the two groups deviated from each other in the perturbation phase. Furthermore, selective F0 perturbation in sentences with emphatic stress resulted in compensatory changes in both F0 and intensity.

Conclusions: Present findings suggest that F0 and intensity are controlled in an integrated fashion to maintain the contrast between stressed and unstressed words. When a cue is impaired through perturbation, speakers not only oppose the perturbation but enhance other prosodic cues to achieve emphatic stress.

Key Words: prosody, fundamental frequency, auditory feedback, perturbation, sensorimotor adaptation

Despite the importance of prosody in conveying numerous linguistic and attitudinal contrasts, models of speech production largely focus on segmental and not prosodic control (Guenther, Ghosh, & Tourville, 2006; Saltzman & Munhall, 1989). One such model of speech acquisition and production is known as *Directions Into Velocities of Articulators* (DIVA; Guenther, 1994, 1995; Guenther et al., 2006). DIVA is a biologically plausible adaptive neural network in which acoustic feedback is used to acquire sensory and motor targets for speech sounds. Currently, DIVA lacks a representation of prosodic control, limiting its scope as a comprehensive model of spoken communication. Furthermore, modeling prosody may lead to improved assessment and intervention of neuromotor speech disorders that are characterized by prosodic deficits (Darley, Aronson, & Brown, 1969, 1975; Duffy, 2005).

The present study is designed to extend the DIVA model to include the control of speech prosody. Minimally, this requires representations of the acoustic cues associated with prosody: fundamental frequency (F0), intensity, and syllable duration, perceived by listeners as pitch, loudness, and length, respectively (Bolinger, 1989; Lehiste, 1970, 1976; Shattuck-Hufnagel & Turk, 1996). It is unclear, however, whether these cues should be represented in an independent or integrated fashion. An *independent channel model* would posit that F0, intensity, and duration are controlled separately, whereas in an *integrated model,* two or more acoustic cues would be jointly controlled. The present study aims to distinguish between these opposing models as a first step toward representing the complex phenomenon of prosody.

To study prosody without the influence of segmental variables, experimental stimuli were constructed to differ only in the location of emphatic stress within an utterance. Although many researchers agree that F0 is the primary cue for signaling stress (Atkinson, 1978; Morton & Jassem, 1965; O'Shaughnessy, 1979), some have argued that duration and intensity cues are also important and may be "traded" for F0 cues (cf. Cooper, Eady, & Mueller, 1985; Eady & Cooper, 1986; Fry, 1955, 1958; Huss, 1978; Kochanski, Grabe, Coleman, & Rosner, 2005; Sluijter

[a]Northeastern University, Boston, MA

[b]Harvard–MIT Division of Health Sciences and Technology, Cambridge, MA

[c]Boston University, MA

Correspondence to Rupal Patel: r.patel@neu.edu

& van Heuven, 1996a, 1996b; Weismer & Ingrisano, 1979). This transfer of informational cues among prosodic features has been referred to as *cue trading* (Howell, 1993; Lieberman, 1960). Listeners appear to be able to leverage the cue-trading phenomenon to perceive stress even when the speaker's cue patterns differ from their own (in healthy speakers, see Howell, 1993; Peppé, Maxim, & Wells, 2000; in speakers with dysarthria, see Patel, 2002, 2003, 2004; Patel & Campellone, 2009; Patel & Watkins, 2007; Wang, Kent, Duffy, & Thomas, 2005; Yorkston, Beukelman, Minifie, & Sapir, 1984).

Such cross-speaker cue trading is consistent with both an integrated model and an independent channel model of prosodic feedback control. The two models can be differentiated by examining the effects of auditory perturbations during speech production. Perturbation paradigms show the importance of auditory feedback for online vocal control during speaking tasks. Numerous studies have investigated gradual or sudden perturbations to F0 (Burnett, Freedland, Larson, & Hain, 1998; Chen, Liu, Xu, & Larson, 2007; Jones & Munhall, 2002, 2005; Larson, Burnett, Kiran, & Hain, 2000; Xu, Larson, Bauer, & Hain, 2004), as well as to intensity (Bauer, Mittal, Larson, & Hain, 2006; Chang-Yit, Pick, & Siegel 1975; Heinks-Maldonado & Houde, 2005) and to vowel formant frequencies (Houde & Jordan, 1998; Tourville, Reilly, & Guenther, 2008; Villacorta, Perkell, & Guenther, 2007). A consistent finding in perturbation studies is a compensatory response: Speakers alter their production of the perturbed feature in the direction opposite to the perturbation. This opposing response is noted both for adaptation paradigms and for paradigms that use brief, unexpected perturbations to auditory feedback. Adaptation paradigms involve persistent exposure to the same perturbation, allowing subjects to adapt their feedforward commands ("adaptation") such that they continue to respond to the perturbation even after it has been removed. In contrast, unexpected perturbation studies use one or more brief, unpredictable perturbations to elicit a compensatory response within a given trial ("rapid compensation").

Most F0 perturbation studies have examined rapid compensations during sustained vowel phonation rather than in linguistic contexts (Burnett et al., 1998; Larson et al., 2000; Xu et al., 2004). Although recent work has examined linguistically relevant perturbations to tones and tone sequences in Mandarin (Jones & Munhall, 2002, 2005; Xu et al., 2004), meaningful prosodic contrasts remain largely unexplored in English. A notable exception is the work of Chen et al. (2007), which examined brief, unexpected upward and downward F0 perturbations as speakers produced the question "You know Nina?" The authors noted that upward perturbations, which were not at odds with the rising intonation contour of the target question, resulted in a smaller compensatory response than downward perturbations. Although the perturbation

had linguistic relevance, the use of an imitation paradigm may have influenced speaker responses. Further work on eliciting a range of prosodic contrasts in linguistically motivated communicative contexts is warranted. Additionally, speakers tend to use multiple acoustic cues to signal prosodic contrasts, yet compensatory responses have only been examined within the perturbed parameter—for example, measuring compensations in F0 for pitch-shifted feedback.

The present study extended the F0 auditory perturbation literature in two main directions. First, meaningful prosodic contrasts in English were elicited by providing contextual scenarios that cued the location of stress within each utterance. Thus, during perturbed trials, speakers were required to compensate for F0 shifts of the stressed word to preserve the intended prosodic contrast. This linguistically motivated task may better resemble auditory feedback control during running speech. Second, compensatory responses to F0 perturbation were examined across multiple cues. In light of cue-trading relations, changes in intensity and duration may also contribute to the compensatory response, which would be consistent with the integrated model. Alternatively, compensatory responses limited to F0 alone would be evidence for an independent channel model.

In summary, the present study aimed to investigate the prosodic cues used to convey emphatic stress under conditions of near-real-time pitch perturbation. Specifically, the following research questions were addressed:

1. Do speakers adapt to targeted F0 perturbations of stressed words within an utterance?

2. Does this adaptation response occur in other features besides F0 (e.g., intensity, duration)?

Method
Participants

Twenty-five monolingual speakers of American English with normal hearing and no known speech, language, and neurological disorders between the ages of 20 and 28 (12 men, 13 women, M_{age} = 22;0 [years;months]) were recruited. Participants were randomly assigned to either the upward shift protocol (Up, hereafter; 6 men, 6 women, M_{age} = 22;2) or the downward shift protocol (Down, hereafter; 6 men, 7 women, M_{age} = 21;11). All participants passed a hearing screening with thresholds at or below 25 dB in at least one ear for 250, 500, 1000, 2000, 4000, and 8000 Hz tones and reported having vision within correctable limits.

Procedure

Participants were seated in a sound-treated booth and wore a head-mounted cardioid microphone (AKG C420) and over-the-ear headphones (AKG K240), which were

used to record productions and present auditory feedback, respectively. The microphone-to-mouth distance was held constant at 1 in., measured from the left-hand corner of the mouth. A customized graphical interface presented stimuli that participants read aloud. Four sentences were used, each consisting of four monosyllabic words. To control for vowel-dependent differences in F0, vowel nuclei were kept relatively constant across the sentence (Lehiste & Peterson, 1961; Peterson & Barney, 1952). In each trial, participants produced the four-word sentence with stress on either the first or the second word. The stressed word was cued visually (i.e., using a capitalized, red font) and by providing a contextual scenario. For example, the context sentence "Who caught a dog?" would prompt the target sentence "BOB caught a dog" on the screen. Conversely, "What did Bob do to a dog?" prompted the sentence "Bob CAUGHT a dog." (The remaining three sentences were "Dick bit a kid," "Doug cut a bud," and "Dad pat a cat.") Participants were instructed to produce emphatic stress such that a naive listener could identify the intended stress location.

Given that stressed words tend to have a higher F0 than unstressed words (e.g., Cooper, et al., 1985; Eady & Cooper, 1986; Morton & Jassem, 1965; O'Shaughnessy, 1979), participant-specific F0 thresholds allowed for selective F0 perturbation of stressed words alone. A brief pre-test consisting of 16 practice sentences, identical to the experimental stimuli, was used to determine the perturbation threshold for each participant. The threshold was operationally defined as the F0 value that optimally separated stressed words from unstressed words across all 16 trials. The experimenter visually determined the lowest F0 value that exceeded all unstressed F0 values. F0 values below the threshold value were never perturbed.

In the experimental protocol, each participant produced a total of 480 sentences across four phases: a *baseline* phase with no perturbation, a *ramp* phase during which the perturbation was applied to the auditory feedback in increments, a *perturbation* phase involving full feedback perturbation on the stressed word, and a *post* phase with no perturbation. In the ramp and perturbation phases, F0 of the stressed word was scaled in proportion to the amount it exceeded the threshold. The formulae used to calculate the scaling factors that transformed input F0 to output F0 were

Up: pitchscale = 1 + [(F0/threshold − 1) × pertval];
Down: pitchscale = 1 − [(F0/threshold − 1) × pertval].

The coefficient pertval, reflecting the perturbation value, was set to 0 during the baseline phase, gradually increased to .5 during the ramp phase, held constant at .5 during the perturbation phase, and reset to 0 during the post phase.

For example, if a subject were assigned to the Down group and her threshold was 200 Hz, a 220-Hz production

during the perturbation phase would result in a scaling factor of 1 − [(220/200 − 1) × .5], or 0.95. Scaling the input F0 of 220 Hz by 0.95 would result in an output F0 of 209 Hz, an apparent decrease in F0 that would cause the stressed word to sound less stressed. On the other hand, if the same subject were assigned to the Up group, the scaling factor for the same utterance would be 1.05 and would increase the perceived F0 to 231 Hz, thereby increasing the apparent F0 contrast between the stressed word and the unstressed words (see Figure 1).

Perturbation was implemented using a Texas Instruments (TI DSK 6713) digital signal processing (DSP) board with only minimal processing delay (~26 ms). An audio mixer split the subjects' speech signal into two channels, one sent to a computer for recording and one sent to the DSP board. The DSP board used a near-real-time autocorrelation algorithm to track and shift the F0 of each participant. This F0-shifted output was further split and sent both to the subjects' headphones and to the recording computer. Thus, each experimental session produced a stereo waveform consisting of one channel of microphone-recorded data (i.e., what the participant produced) and one channel of feedback-perturbed data (i.e., what the participant heard). The two channels were compared with and without perturbation to ensure that the F0 shift had no effect on intensity.

Acoustic Analysis

Customized software implemented in MATLAB (CadLab acoustic analysis suite [CLAAS]) was used to derive estimates of F0, relative intensity, and duration for each word across all utterances. Each utterance was

Figure 1. Example input–output fundamental frequency (F0) curve for a 200-Hz stress threshold. For values below the threshold, feedback is unaltered, even during the perturbation (pert) phase.
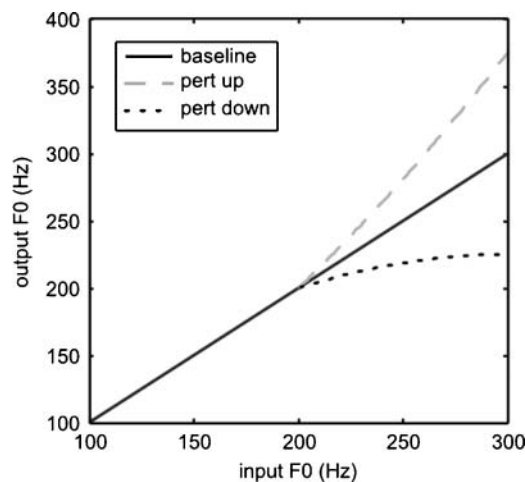
manually annotated to demarcate word boundaries (r = .984 interlabeler reliability for 10% of the data). CLAAS used the Praat autocorrelation algorithm to estimate time-stamped F0 values (Boersma & Weenink, 2009). Similarly, time-stamped intensity values were derived via a root-mean-square calculation of the acoustic waveform. The software operated on the annotations and the time-stamped pitch and intensity values to calculate word duration, average F0, and average intensity across stressed and unstressed words. All analyses were performed on the original spoken utterance, not on the F0-perturbed feedback. The perturbed signal was compared with the microphone-recorded signal to ensure perturbation occurred on the intended trials.

A total of 12,000 utterances were acoustically analyzed (480 trials × 25 participants). A subset of the utterances was examined by hand to ensure correct pitch tracking of all words. Pitch tracking errors, when found, were manually corrected. Errors in pitch tracking were especially problematic for women, particularly for the third and fourth words, which were often in the glottal fry register. Manual correction of automatically generated F0 values was required on 8.3% of the total data set; 2.7% were excluded. Two female subjects had more than 100 mistracked trials (> 20%) and were excluded from further analysis. Furthermore, one male subject was excluded due to corrupted acoustic data, and one female subject was excluded because she produced incorrect stress on more than 40% of trials. Incorrect stress for all other subjects ranged from 0 to 13 tokens (0%–2.7%), with an average of 2.46 (0.5%) errors in the Down group and 1.33 (0.3%) errors in the Up group. Trials with incorrect stress were discarded from the analysis. The resultant data set after exclusions was 9,752 utterances from 21 participants (Up: 6 men, 5 women, M_{age} = 22;0; Down: 5 men, 5 women, M_{age} = 22;2).

Although acoustic measures were obtained for all four words within an utterance, analyses were restricted to the first and second word (W1 and W2) for two main reasons. First, W1 and W2 were the only two word locations that were counterbalanced in both stressed and unstressed conditions. Second, the word length and sentence position of W3 and W4, respectively, led to variable and imprecise acoustic measurements. Specifically, W3, which was the word a in all stimuli, was often reduced or even omitted, whereas W4 was often glottalized or excessively lengthened owing to phrase-final boundary effects.

Results

Speaker responses to F0 perturbation were examined in three acoustic variables: mean F0, mean intensity, and word duration. For each trial, the dependent measure was the difference in a given acoustic variable between the stressed word (W1 or W2) and the unstressed word (W2 or W1). This difference was normalized by the mean stressed–unstressed difference in baseline. For simplicity, this normalized value is referred to as the *contrast distance* because it represents the degree to which speakers contrasted the stressed and unstressed words within an utterance.

Broadly, the Up and Down groups responded differently to the F0 perturbation, altering their contrast distances in opposite directions with only a short delay from perturbation onset. However, individual speakers were noted to use differing strategies to compensate for perturbations and had individualized time courses for adaptation. Additionally, there was a groupwise trend of a slow increase in F0 and intensity of stressed words across the experiment. The covariance of F0 and intensity could be explained by a cross-featural compensation, as described in the integrated model, or may be an artifact of the physiological coupling between F0 and intensity. Covariance of F0 and intensity due to tracheal pressure changes would result in a per-trial correlation of these two variables because this physiological coupling would necessarily be present for every trial. To ensure that variations in intensity were not simply the result of a passive physiological correlation with F0, the correlation of these two measures was calculated on a trial-by-trial basis for each participant, and the resulting r scores were Fisher z–transformed before averaging across the group. This analysis yielded weak correlations (z = .14 averaged across participants, back-converted to r = .139), suggesting that changes in subglottal pressure required to modulate intensity had little direct influence on F0 changes in this study.

To quantify the changes in contrast distance between and within subject groups, paired and independent-samples t tests were conducted on conditions of interest. Between-group (Up vs. Down) differences were compared at all four experimental phases. Taking into account the upward drift of both F0 and intensity over the course of the experiment, which affected interpretation of within-group changes, the analysis focused on these between-group differences. Additionally, within each perturbation direction (Up or Down), differences between all phases (baseline, ramp, perturbation, and post) were compared; therefore, there were six comparisons for each perturbation direction, or 12 within-subject comparisons. In total, 16 t tests were carried out for each acoustic variable. To account for multiple comparisons, the Bonferroni correction factor was used to adjust the α level to .003.
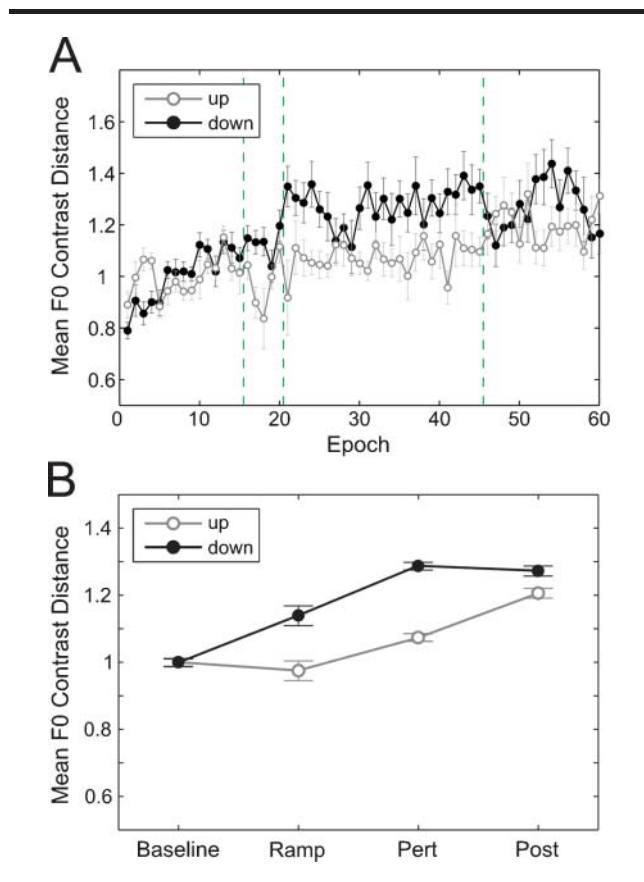
Mean F0

Between-group comparisons show evidence of adaptation to the upward and downward F0 perturbations

(see Figure 2). In the baseline phase, in which no perturbation was applied, there was no significant difference between the Up and Down groups ($p = .45$). However, the two groups diverged in the ramp phase ($p = .0014$) and remained significantly different in the perturbation phase ($p < .0001$) before falling back below the adjusted significance level in the post phase ($p = .02$). Thus, the perturbation resulted in a difference in F0 contrast distance between the two groups. Specifically, speakers altered F0 to enhance or reduce emphatic stress, with the Down group increasing the F0 difference between stressed and unstressed words as compared to the Up group.

Contrasts between phases were used to examine the time course of adaptation within a perturbation direction group. In the Down group, there was a difference between baseline and every other phase (ramp, $p < .0001$; perturbation, $p < .0001$; post, $p < .0001$) as well as between ramp and every other phase (perturbation, $p = .0002$; post, $p = .003$). There was no difference between perturbation and post phases ($p = .6$). Thus, the F0 contrast distance increased from the baseline through to the perturbation phase and then stabilized in the post phase. In the Up group, only differences between the post phase and every other phase (baseline, $p < .0001$; ramp, $p < .0001$; perturbation, $p < .0001$) were statistically significant. In other words, the F0 contrast distance did not change from baseline values until the upward perturbation was removed in the post phase. Patterns of adaptation within groups may be due in part to the overall upward drift of F0 during the course of the experiment.

Mean Intensity

As with F0, there was evidence of adaptation in intensity (see Figure 3). Speakers who received a downward perturbation increased the intensity contrast between stressed and unstressed words more than speakers who received an upward perturbation, even though speakers' intensities were unaffected by the perturbation. The two perturbation-direction groups significantly differed in intensity contrast distance during the perturbation phase ($p < .0001$); however, they were not significantly different in any other phase (baseline, $p = .5$; ramp, $p = .018$; post, $p = .006$).

Figure 2. Groupwise mean F0 contrast distance by (A) epoch and (B) phase. The contrast distance is a measure of the difference between stressed and unstressed words, normalized by the average of that difference in the baseline phase. Error bars show 95% CI.

**Figure 3.** Groupwise mean intensity contrast distance by (A) epoch and (B) phase. Error bars show 95% CI.
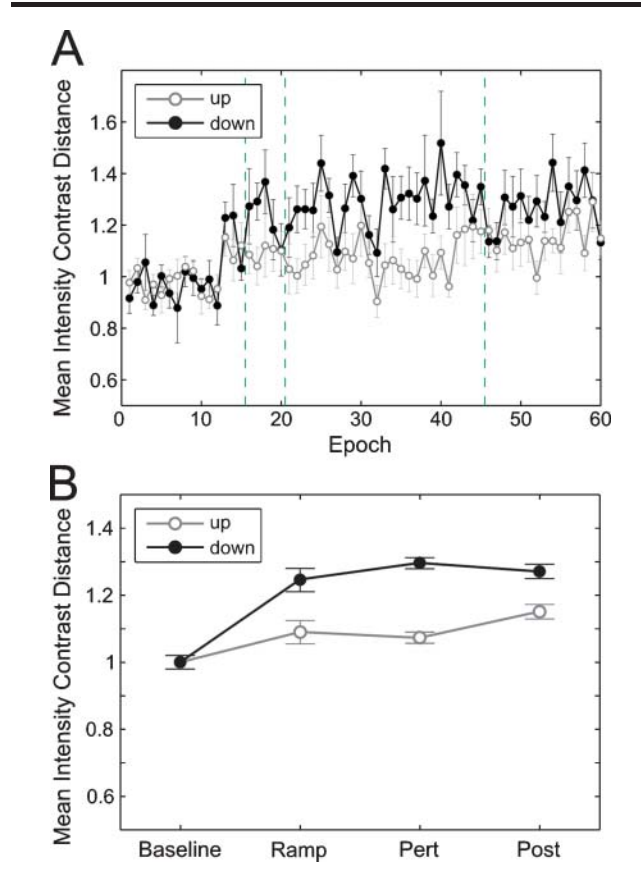
With regard to within-group contrasts, in the Down group only, the baseline phase was significantly different from other phases (ramp, $p = .0001$; perturbation, $p < .0001$; post, $p < .0001$). In other words, the intensity contrast distance increased during the ramp phase and remained increased throughout the experiment. In the Up group, the only significant phase contrast was that between the baseline and post phases ($p = .0004$), again suggesting a slow drift in intensity contrast distance over the course of the experiment.

Word Duration

Unlike in F0 and intensity, there were no significant between-subject differences (Up vs. Down) in duration contrast distance (see Figure 4). In other words, the perturbation did not effect a durational change between the stressed and unstressed words.

Although there were no between-subject differences, there was a difference in the Down group between the baseline and perturbation phases ($p = .0004$) as well as between the baseline and post phases ($p < .0001$). In the Up group, however, experimental phase had no effect on duration contrast difference.

Figure 4. Group-wise word duration contrast distance by (A) epoch and (B) phase. Error bars show 95% CI.
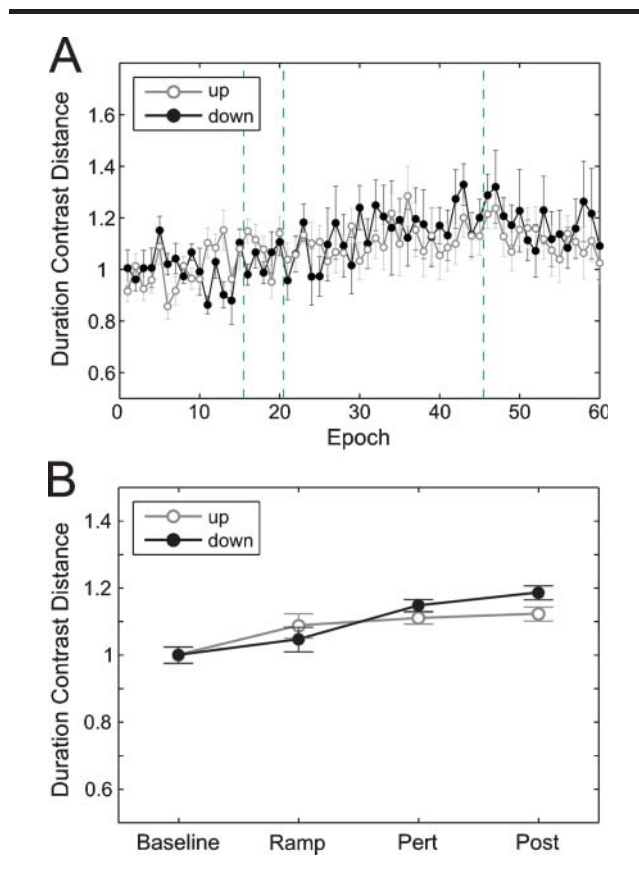
Discussion

A major aim of the present study was to distinguish between two representations of speech prosody: one involving independent control of pitch, loudness, and duration, and one involving integrated control of these cues. The emphatic stress task required speakers to modulate prosody while keeping segmental units constant. Introducing a gradual F0 perturbation altered the influence of F0 as a stress-bearing cue. To maintain the appropriate degree of contrast between stressed and unstressed words, speakers might alter only F0, consistent with the independent channel model, or they might alter a combination of prosodic cues to oppose the F0 shift, consistent with the integrated model.

Speaker responses to pitch-shifted auditory feedback were measured in three acoustic variables: F0, intensity, and duration. Results indicated that the Up and Down groups adapted to shifts in F0 by altering the contrast between stressed and unstressed words. Specifically, when participants heard their stressed F0 shifted downward, they increased F0 contrast compared with when they heard their stressed F0 shifted upward. The interaction between the two F0 manipulations supports the conclusion that speakers are sensitive to upward and downward shifts of F0 in a meaningful prosodic context. Furthermore, compensatory effects were not restricted to F0 but also extended to intensity: Speakers altered intensity contrast distance during perturbation, making stressed words relatively louder in the Down group compared with the Up group. This change occurred even though (a) the auditory feedback preserved intensity differences between stressed and unstressed words, and (b) intensity and F0 were only very weakly correlated in individual trials. These findings provide support for the integrated model in that speakers modified both F0 and intensity (although not duration) in response to F0 shifts.

Although perturbation-induced changes in word duration were not observed in the present experiment, it is possible that syllable timing may be affected by F0 shifts. Duration effects may be more apparent in other prosodic contexts or in other measures of duration (e.g., in the length of the pauses between words).

In both experimental groups, Up and Down, speakers' F0s in the perturbation phase were the same or higher than their initial baseline values. This is consistent with the gradual increase in F0 noted in past sensorimotor adaptation studies (Jones & Munhall, 2000, 2002; Villacorta et al., 2007). During F0 perturbations to single-word productions, subjects were found to increase F0 over many trials (Jones & Munhall 2000, 2002). The authors attributed this increase to vocal fatigue; however, there may be additional factors involved, given that the increase was present even in the baseline phase. In the

present sentence production task, only the stressed word showed a similar drift, resulting in an apparent increase in the contrast distance over the course of the experiment. In the context of a gradual increase in contrast distance, it appears that the Up group was in fact opposing the perturbation by decreasing F0 relative to an upward drift.

Unlike Jones and Munhall (2002), who found no significant differences in intensity between Up and Down groups, the results of the present study show an increased intensity for the Down group as well as a trend toward increasing intensity over many trials. This may be due to the prosodic nature of the emphatic stress task. Loudness can be used as a stress cue, and thus participants had reason to manipulate intensity in line with F0 to maintain emphatic stress (Fry, 1955; Kochanski et al., 2005).

Although the gradual upward drift in F0 and intensity is consistent with an adaptive response from both Up and Down groups, an alternative explanation is that the two groups adapted to different degrees. The increase in F0 by the Down group may be indicative of a larger compensation than that of the Up group, whose F0 did not decrease from baseline values. F0 shifts may have been perceptually greater in the Down condition than in the Up condition because of the use of the Hertz scale, which is acoustically and not psychophysically defined. However, the differences in this region of acoustic space are small relative to the cross-subject variation in threshold values and in upward deviations from those thresholds, and are unlikely to explain the magnitude of the difference. An additional explanation is that differences between Up and Down groups are linguistically motivated. The perturbation experienced by the Up group did not interfere with the planned intonation pattern of a stressed–unstressed contrast; thus, there may have been less need for a corrective response. That is, there may be more incentive to restore a stress contrast that has been attenuated (Down group) than to decrease a stress contrast that has been enhanced (Up group) because the enhancement is aligned with the speaker's goals. This interpretation is also supported by previous work in which downward perturbations elicited a larger rapid compensation response than upward perturbations during question productions (Chen et al., 2007).

Although the compensatory changes in both F0 and intensity are in line with the integrated model, they might also be attributed to a correlation between F0 and intensity (Dromey & Ramig, 1998; Gramming, Sundberg, Ternström, Leanderson, & Perkins, 1988), as F0 has been found to increase at higher than habitual loudness levels associated with a high tracheal pressure (though not at soft levels). However, our analysis of the trial-by-trial F0–intensity correlation suggests this was not the case. The correlation explains less than 2% of the variance observed in F0; thus, physiological dependencies alone are unlikely to explain the commensurate intensity increase in the group data.

In adverse listening conditions, speakers enhance prosodic cues to optimize communication (Lane & Tranel, 1971; Letowski, Frank, & Caravella, 1993; Lombard, 1911; Patel & Schell, 2008; Rivers & Rastatter, 1985; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). Downward F0 perturbation in the present study served as a targeted "adverse" condition that speakers had to overcome in order to convey meaningful differences. Similar to the Lombard effect, our targeted F0 shift led speakers to alter contrast distance in multiple cues, using both F0 and intensity in an integrated fashion to signal stress contrasts.

Future Directions

The present results support an integrated model of prosodic control in which the motor system modulates F0 and intensity in combination to convey stress. However, extending this protocol to perturbation of other prosodic cues will help to generalize the findings. In a planned follow-up to the present study, subjects will undergo intensity perturbations rather than F0 perturbations, allowing a direct comparison of the F0–intensity interaction in each direction. Future experiments will also help disambiguate the roles of word type and word position within an utterance. All of the stimuli in the present study followed the same pattern of word types across the sentence (*Name verbed* a *noun*). It will be important to vary this pattern to assess the influence of different parts of speech and sentence positions on adaptation responses. Finally, given that perturbation paradigms can induce enhanced linguistic contrasts in healthy talkers, it may be possible to leverage this paradigm for therapeutic benefit in disordered populations.

Acknowledgments

References

Atkinson, J. E. (1978). Correlation analysis of the physiological features controlling fundamental voice frequency. *The Journal of the Acoustical Society of America, 63,* 211–222.

Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice

loudness feedback: An automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America, 119*(4), 2363–2371.

Boersma, P., & Weenink, D. (2009). *Praat: Doing phonetics by computer* (Version 5.1.05). [Computer software]. Amsterdam, the Netherlands: Institute of Phonetic Sciences.

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*(6), 3153–3161.

Chang-Yit, R., Pick, H. L., & Siegel, G. M. (1975). Reliability of sidetone amplification effect in vocal intensity. *Journal of Communication Disorders, 8*(4), 317–324.

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.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *The Journal of the Acoustical Society of America, 77*(6), 2142–2156.

Darley, F., Aronson, A., & Brown, J. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12,* 246–269.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Motor speech disorders*. Philadelphia, PA: W. B. Saunders.

Dromey, C., & Ramig, L. O. (1998). Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation, and articulation. *Journal of Speech, Language, and Hearing Research, 41*(5), 1003–1018.

Duffy, J. R. (2005). *Motor speech disorders: Substrates, differential diagnosis, and management*. St. Louis, MO: Elsevier-Mosby.

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America, 80*(2), 402–415.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America, 27*(4), 765–768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1*(2), 126–152.

Gramming, P., Sundberg, J., Ternström, S., Leanderson, R., & Perkins, W. H. (1988). Relationship between changes in voice pitch and loudness. *Journal of Voice, 2*(2), 118–126.

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 production. *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*(3), 131–137.

Houde, J. F., & Jordan, M. I. (1998, February 20). Sensorimotor adaptation in speech production. *Science, 279*(5354), 1213–1216.

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.

Huss, V. (1978). English word stress in the post-nuclear position. *Phonetica, 35,* 86–105.

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.

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: Studies of Mandarin tone production. *Journal of Phonetics, 30*(3), 303–320.

Jones, J. A., & Munhall, K. G. (2005). Remapping auditory-motor representations in voice production. *Current Biology, 15*(19), 1768–1772.

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

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research, 14*(4), 677–709.

Larson, C. R., Burnett, T. A., Kiran, S., & Hain, T. C. (2000). Effects of pitch-shift velocity on voice F0 responses. *The Journal of the Acoustical Society of America, 107*(1), 559–564.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

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.

Lehiste, I., & Peterson, G. E. (1961). Transitions, glides, and diphthongs. *The Journal of the Acoustical Society of America, 33,* 268–277.

Letowski, T., Frank, T., & Caravella, J. (1993). Acoustical properties of speech produced in noise presented through supra-aural earphones. *Ear and Hearing, 14*(5), 332–338.

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

Lombard, E. (1911). Le signe de l'élévation de la voix [The sign of voice rise]. *Annales Maladies Oreilles Larynx Nez Pharynx, 37,* 101–119.

Morton, J., & Jassem, W. (1965). Acoustic correlates of stress. *Language and Speech, 8,* 159–181.

O'Shaughnessy, D. (1979). Linguistic features in fundamental frequency patterns. *Journal of Phonetics, 7,* 119–145.

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.

Patel, R. (2003). Acoustic differences in the yes-no question-statement contrast between speakers with and without dysarthria. *Journal of Speech, Language, and Hearing Research, 46*(6), 1401–1415.

Patel, R. (2004). Contrastive prosody in adults with cerebral palsy. *Journal of Medical Speech Pathology, 12*(4), 189–193.

Patel, R., & Campellone, P. (2009). Acoustic and perceptual cues to contrastive stress in dysarthria. *Journal of Speech, Language, and Hearing Research, 52*(1), 206–222.

Patel, R., & Schell, K. W. (2008). The influence of linguistic content on the Lombard effect. *Journal of Speech, Language, and Hearing Research, 51*(1), 209–220.

Patel, R., & Watkins, C. (2007). Stress identification in speakers with dysarthria due to cerebral palsy: An initial report. *Journal of Medical Speech Language Pathology, 15*(2), 149–159.

Peppé, S., Maxim, J., & Wells, B. (2000). Prosodic variation in southern British English. *Language and Speech, 43*(3), 309–334.

Peterson, G. E., & Barney, H. E. (1952). Control methods used in a study of vowels. *The Journal of the Acoustical Society of America, 24,* 175–184.

Rivers, C., & Rastatter, M. P. (1985). The effects of multi-talker and masker noise on fundamental frequency variability during spontaneous speech for children and adults. *Journal of Auditory Research, 25*(1), 37–45.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333–382.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193–247.

Sluijter, A., & van Heuven, V. (1996a). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of the ICSLP 96* (pp. 630–633). Philadelphia, PA: Applied Science and Engineering Laboratories, Alfred I. duPont Institute.

Sluijter, A., & van Heuven, V. (1996b). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America, 100*(4), 2471–2485.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America, 84*(3), 917–928.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage, 39*(3), 1429–1443.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America, 122*(4), 2306–2319.

Wang, Y.-T., Kent, R. D., Duffy, J. R., & Thomas, J. E. (2005). Dysarthria associated with traumatic brain injury: Speaking rate and emphatic stress. *Journal of Communication Disorders, 38,* 231–260.

Weismer, G., & Ingrisano, D. (1979). Phrase-level timing patterns in English: Effects of emphatic stress location and speaking rate. *Journal of Speech and Hearing Research, 22*(3), 516–533.

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

Yorkston, K. M., Beukelman, D. R., Minifie, F., & Sapir, S. (1984). Assessment of stress patterning. In M. McNeil, J. Rosenbek, & A. Aronson (Eds.), *The dysarthria: Physiology, acoustics, perception, management* (pp. 131–162). Austin, TX: ProEd.

Prosodic Adaptations to Pitch Perturbation in Running Speech

Rupal Patel, Caroline Niziolek, Kevin Reilly, and Frank H. Guenther

AMERICAN
SPEECH-LANGUAGE-
HEARING
ASSOCIATION