Acoustic and Perceptual Cues to Contrastive Stress in Dysarthria

Rupal Patel
Pamela Campellone
Northeastern University, Boston, MA

Purpose: In this study, the authors sought to understand acoustic and perceptual cues to contrastive stress in speakers with dysarthria (DYS) and healthy controls (HC).
Method: The production experiment examined the ability of 12 DYS (9 male, 3 female; *M* = 39 years of age) and 12 age- and gender-matched HC (9 male, 3 female; *M* = 37.5 years of age) to signal contrastive stress within short sentences. Acoustic changes in fundamental frequency (F0), intensity, and duration were studied. The perceptual experiment explored whether 48 unfamiliar listeners (24 male, 24 female; *M* = 23.4 years of age) could identify the intended stress location in DYS and HC productions.
Results: Although both speaker groups used all 3 prosodic cues, DYS relied more heavily on duration. Despite reduced F0 and intensity variation within DYS utterances, listeners were highly accurate at identifying both DYS (> 93%) and HC (> 97%) productions. Acoustic predictors of listener accuracy included heightened prosodic cues on stressed words along with marked decreases in these variables for neighboring nonstressed words.
Conclusions: Speakers signaled contrastive stress using relative changes in one or more prosodic cue. Although individual speakers employed different cue combinations, listeners were highly adept at discerning the intended stress location. The communicative potential of prosody in speakers with congenital dysarthria is discussed.

KEY WORDS: prosody, dysarthria, acoustics, listener perception, contrastive stress

Prosody serves various grammatical, semantic, social, and psychological roles. Healthy speakers employ prosody to convey a multitude of functions, including signaling questions versus statements, conveying contrastive meanings, and expressing emotions and attitudinal states. Prosody also plays a role in marking lexical boundaries (Liss, Spitzer, Caviness, Adler, & Edwards, 1998) and in sentence disambiguation (Schafer, Speer, Warren, & White, 2000), thereby aiding in listener comprehension. The acoustic cues commonly associated with prosody include fundamental frequency (F0), intensity, and duration (cf. Bolinger, 1961, 1989; Lieberman, 1967). In adverse listening conditions, these same cues are further enhanced to optimize communicative success (Lombard, 1911; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). Thus, prosody is a multifaceted aspect of the speech signal on which speakers and listeners must rely to accurately transfer information.

Prosodic control in dysarthria has typically been characterized in terms of its deviations from nonimpaired speech, with the focus on reduced precision and flexibility of the dysarthric vocal apparatus (Canter, 1963; Darley, Aronson, & Brown, 1969, 1975; Hardy, 1983; Le Dorze, Oulette, & Ryalls, 1994; Wit, Maassen, Gabreels, & Thoonen, 1993). F0 range and its modulation within that range, however, are separate issues (Abberton,

1092-4388/09/5201-0206

Fourcin, & Hazan, 1991). Although speakers with dysarthria are often described as monopitch and/or monoloud (Darley et al. 1969, 1975), recent investigations suggest that many individuals have preserved ability to mark prosodic contrasts across various spoken tasks (Le Dorze et al., 1994; Liss et al., 1998; Liss & Weismer, 1994; Patel, 2002a, 2003, 2004; Puyuelo & Rondal, 2005; Vance, 1994; Wang, Kent, Duffy, & Thomas, 2005; Wit et al., 1993; Yorkston, Beukelman, Minifie, & Sapir, 1984). For example, Patel (1999, 2002a; Patel & Salata, 2006) found that children and adults with severe dysarthria could control the F0 and duration for sustained vowel productions despite reduced speech sound intelligibility. Accurate modulation of prosodic cues has also been documented at the sentence level through the question–statement contrast (Le Dorze et al., 1994; Patel, 2002b, 2003).

Although these previous studies suggest that some speakers with dysarthria may have preserved ability to signal global sentence-level changes in prosody, less is known about their ability to mark local, word-, or syllable-level contrasts. To further examine this issue, we chose a production task in which the information conveyed is altered by the location of syllabic stress, commonly referred to as *contrastive stress*. For example, the meaning of "SHE hit me here" differs from "She hit me HERE." In such utterances, speech sounds (segmental units) are held constant while manipulating prosody. This task provides a means for examining whether speakers with dysarthria can modulate prosody along a fine-grained time scale. This ability would allow speakers to encode affective, semantic, and syntactic information and thus convey more diverse communicative functions. Furthermore, these cues may aid listeners in segmenting and parsing the dysarthric speech stream into words and phrases (see Cutler, 1990, 1994; Cutler, Dahan, & van Donselaar, 1997; Morgan & Demuth, 1996, for segmentation of healthy speech).

The acoustics of contrastive stress are somewhat controversial even in nonimpaired speech. There are three principle acoustic cues associated with stress: F0, duration, and intensity (cf. Bolinger, 1961; Lehiste, 1970, 1976; Lieberman, 1960). Although many researchers would agree that F0 is the primary cue (Atkinson, 1978; Morton & Jassem, 1965; O'Shaughnessy, 1979), some have argued that duration and/or intensity cues also convey stress and may be "traded" for F0 cues (cf. Cooper, Eady, & Mueller, 1985; Eady & Cooper, 1986; Fry, 1955, 1958; Huss, 1978; Sluijter & van Heuven, 1996a, 1996b; Weismer & Ingrisano, 1979). This transfer of informational cues between prosodic features has been referred to as *cue trading* (Howell, 1993; Lieberman, 1960). Cue trading relations may arise partly from the fact that the same prosodic cues (F0, intensity, and duration) are used for various functions. For example, syllable lengthening is a sign for utterance termination in addition to syllable prominence (Cooper & Danly, 1981). Additionally, healthy speakers usually increase F0 and intensity at the beginning of an utterance, and both F0 and intensity fall at the end of an utterance irrespective of syllabic prominence (Cooper & Sorensen, 1981; Lehiste, 1970; Lieberman, 1967). Thus, the degree to which F0, intensity, and duration cues are used to mark contrastive stress may depend on the location of the stressed word within a sentence (Chueng, Holden, & Minifie, 1973; Cooper et al., 1985; Eady & Cooper, 1986; Huss, 1978).

Cooper et al. (1985) found that the F0 contour of sentences varied by the location of the stressed word in the sentence. Contrastively, stressed words in sentence-initial position were marked by a sharp F0 drop after the focused word and with low F0 values for all subsequent words. Similarly, sentence-medial focused words were marked by an F0 rise prior to, and a substantial fall after, the stressed word. The F0 contour of sentences with contrastive stress placed at the end of the sentence, however, was relatively flat, rather like neutral productions. Cooper et al. (1985) also noted that stress location impacted the amount of duration increase evident on the contrasted word. Stressed words in sentence-initial and medial positions were 30%–40% longer than when they were unstressed, compared with only 10%–15% longer in the sentence-final position. Further, Huss (1978) noted that increases in duration and intensity on the stressed word relative to unstressed words were smaller in magnitude when stress was placed toward the end of declarative sentences.

Sentence position factors also seem to influence the perception of stress placement. Pierrehumbert (1979) found that listeners perceive F0 peaks later in the sentence as being the same pitch as those earlier in the sentence even if the absolute frequency value was lower at the end of the sentence. It appears that listeners' expectations about pitch declination within an utterance play a role in their ability to estimate F0 peaks. Listeners appear to estimate prominence based on F0 range and its contour (Rietveld & Gussenhoven, 1985; Terken, 1991). Thus, the absolute differences in prosodic cues needed to perceive a word as being stressed may differ based on its position within an utterance.

We were interested in examining whether speakers with dysarthria could use F0, intensity, and/or durational cues to mark contrastive stress within a sentence and how sentence positional factors would interact with their ability. We hypothesized that speakers with dysarthria may be able to employ cue trading relations to compensate for parameters over which they had little or no control (Brewster, 1989; Crystal, 1979; Liss & Weismer, 1994; Patel, 1999, 2002a, 2002b; Vance, 1994). For example, speakers with restricted F0 range may be able to indicate contrastive stress using elongated syllable duration cues rather than, or in addition to, modest increases in F0. They may also use multiple redundant cues to mark stress, such as increasing loudness to heighten pitch contrasts (Patel, 2002b, 2003, 2004; Vance, 1994; Yorkston et al., 1984). How these cue trades would interact with physiological constraints of their speech production mechanism and positional influences of stress location within an utterance remains unknown.

A second goal was to assess whether naïve listeners could successfully decode the possible remapping of prosodic cues produced by speakers with dysarthria. Previous studies have noted that increased physiological effort required for marking stress often leads to perceptions of inaccurate, inconsistent, exaggerated, and bizarre stress patterning in dysarthria (Liss & Weismer, 1992, 1994; Netsell, 1973; Yorkston et al., 1984). Although it appears that listeners are able to attune to the idiosyncratic stress patterning in healthy speech even if it differs from their own methods of signaling stress (Howell, 1993; Peppe, Maxim, & Wells, 2000), the cue(s) used by speakers with dysarthria may be inconsistent or confusing. Ansel and Kent (1992) found that even when speakers with dysarthria were able to acoustically mark phonetic contrasts, if they used a different strategy than healthy controls, listeners had difficulty interpreting these differences and tended to judge dysarthric productions as less intelligible. In contrast, in a prosodic task, listeners were able to successfully identify question-statement contrasts produced by speakers with severe dysarthria, even though some speakers used different cues than the healthy controls (Patel, 2002b). Here, we explore whether listeners can identify the locus of contrastive stress in dysarthric productions, even if the cues used appear to be nonstandard.

Although there is some literature on contrastive stress in dysarthria, most studies have included speakers with mild to moderate dysarthria from varying etiologies. Among such speakers, reduced rate and restricted range in F0 and intensity variation have been thought to negatively impact listener perception (Pell, Cheang, & Leonard, 2006; Yorkston et al., 1984). Yorkston et al. (1984) found that deviations in expected prosody led listeners to judge contrastive stress produced by 3 speakers with mild ataxic dysarthria as "bizarre" compared with that of a healthy control. Furthermore, Pell et al. (2006) found that speakers with mild to moderate hypokinetic dysarthria (associated with Parkinson's disease) failed to optimize contrasts within their prosodic range, and thus listeners had difficulty identifying the locus of stress, particularly for the sentence-final position. These findings differ from our preliminary evidence on speakers with severe congenital dysarthria. In a pilot study of the present investigation, we found that 3 speakers with severe dysarthria due to cerebral palsy were able to exploit their narrowed prosodic range to mark contrastive stress using all three cues of increased F0, intensity, and duration (Patel, 2004). Although these speakers used cues similar to those used by healthy controls, they tended to rely more heavily on intensity. A follow-up perceptual study indicated that although listeners were highly accurate (94%–100%) at identifying stress placement in the dysarthric productions, they applied their a priori biases, which tended to rely on F0 cues more than intensity and duration (Patel & Watkins, 2007). Thus, when speakers with dysarthria marked contrastive stress using intensity cues alone, listeners were more prone to errors. An understanding of the acoustic correlates to accurate perception of contrastive stress in dysarthria is needed for designing interventions that align speaker abilities with listener expectations in order to harness preserved prosodic function in dysarthria.

The present study aimed to better understand the production abilities of speakers with congenital dysarthria across a range of severities. Additionally, we sought to examine the interplay between acoustic and perceptual factors underlying their prosodic control. The following research questions were addressed:

1. Can speakers with dysarthria use F0, intensity, and/or duration cues to signal contrastive stress on different words within a sentence? If so, how does their acoustic patterning differ from that used by healthy speakers?
2. Can listeners accurately identify words intended to receive contrastive stress in the productions of speakers with dysarthria? If so, how does listener accuracy on dysarthric productions compare to that for healthy controls?
3. What are the acoustic predictors underlying perceptual judgments of contrastive stress in dysarthria? How do these predictors differ for healthy control productions?

Method

Speaker Participants

Twelve speakers with dysarthria (DYS) secondary to cerebral palsy[1] and 12 healthy controls (HC) were recruited for the production experiment (see Table 1). Speakers were age- and gender-matched (DYS: 21–64 years, M = 39 years, 3 female, 9 male; HC: 22–59 years, M = 37.5 years, 3 female, 9 male). All participants were monolingual speakers of American English, with hearing thresholds at or below 25 dB HL for 0.25, 0.5, 1.0, 2.0, 4.0, and 8.0 kHz in at least one ear. DYS speakers had a primary speech diagnosis of spastic or mixed spastic-flaccid

---

[1] Note that none of the 12 DYS or 12 HC had participated in previous pilot studies on prosodic control (Patel, 2004; Patel & Watkins, 2007).

Table 1. Description of speakers with dysarthria.

Speaker	Age	Gender	Speech intelligibility on modified AIDS	Reported modes of communication	Motor control
DYS_1	21	F	83%	Speech, gestures, and some signs	In wheelchair
DYS_2	34	M	84%	Speech and gestures	In wheelchair; use of both arms but not fine motor
DYS_3	38	M	85%	Speech and gestures	In wheelchair; use of left arm and limited use of right arm
DYS_4	21	M	87%	Speech, typing on laptop	In wheelchair
DYS_5	61	M	60%	Speech, gestures, and writing	Ambulatory
DYS_6	44	M	63%	Vocalizations, gestures, and typing on laptop	In wheelchair
DYS_7	29	M	67%	Vocalizations, gestures, and typing on laptop	Ambulatory
DYS_8	31	F	71%	Vocalizations, gestures, some writing	In wheelchair
DYS_9	21	M	40%	Vocalizations, gestures, and alphabet board	In wheelchair
DYS_10	54	M	47%	Vocalizations and alphabet board	In wheelchair; limited use of arms
DYS_11	50	M	43%	Vocalizations, Vantage communication aid, and alphabet board	In wheelchair; limited use of arms
DYS_12	64	F	36%	Vocalizations, gestures, and alphabet board	In wheelchair; able to use arms but limited fine motor control

Note. AIDS = Assessment of Intelligibility of Dysarthric Speech.

dysarthria as determined by the first author, a certified speech-language pathologist, using a battery of formal and informal evaluations and as confirmed by an independent speech-language clinician. To minimize vocal fatigue, dysarthria severity was assessed using a modified version of the Assessment of Intelligibility of Dysarthric Speech (AIDS; Yorkston & Beukelman, 1981), consisting of 25 single-word productions that were rated by three unfamiliar listeners. Four DYS speakers were judged as mild, four were judged as moderate, and four were judged to be relatively severe. These classifications were based on operational definitions of intelligibility, in which scores above 80% were judged to be mild, 50%–80% were judged to be moderate, and under 50% were considered severe. All speakers demonstrated grossly adequate receptive language and cognitive skills necessary for completing the experimental task. None of the speakers had received speech therapy that focused on contrastive stress drills.

Listener Participants

Forty-eight monolingual speakers of American English (21–51 years, M = 23.4 years, 24 female, 24 male) participated as listeners. All participants underwent an audiometric screening to ensure that pure-tone thresholds at 0.5, 1.0, 2.0, and 4.0 kHz were at or below 25 dB HL in at least one ear. The listeners were unfamiliar with dysarthric speech and with the presented stimulus materials. Each listener was randomly assigned to an age- and gender-matched DYS–HC speaker pair. Four listeners heard each speaker pair: Two heard the speaker with DYS speaker first, and 2 heard the HC speaker first.

Speech Production Task

Speech recordings were collected in either a sound-treated audiometric booth or in a quiet, isolated room when speakers had limited mobility or resided in clinical facilities that required in-house data-collection. Identical procedures and equipment were used in both settings. Recordings were made using a MiniDisc recorder (HHB MDP500) and a head-mounted microphone (Shure, SM10A) at a sampling rate of 44100 Hz and 16-bit linear quantization. Constant mouth-to-microphone distance and recording settings were maintained throughout each recording session, and conditions were replicated across participants. Each speaker produced five phrases, each consisting of four monosyllabic words (see Appendix). Vowel height was controlled within each phrase because of intrinsic differences in F0 (Lehiste & Peterson, 1961; Peterson & Barney, 1952). Three phrases consisted of high front vowels, and two phrases had low front vowels. Speakers produced each phrase with contrastive stress on either the first, second, third, or fourth word, or neutrally. Contextual scenarios were used to elicit each stress location. Speakers produced each phrase and stress location at least five times, resulting in 125 recordings per speaker. These samples were then subjected to acoustic analyses.

Acoustic Analyses

The Praat speech analysis software package (Boersma & Weenink, 2005) was used to derive estimates of F0, relative intensity, and duration for each word within utterances. Acoustic analyses followed a multistep process. First, the beginning and end of individual words within each spoken utterance was manually labeled (r = .992 interlabeler reliability for 10% of the data) in order to generate a series of relative intensity values (dB) and frequency values (Hz) across the duration of each word. Manual correction of automatically generated F0 values was required on 83 of the 3,000 DYS and HC samples due to faulty pitch-tracking that could not be verified auditorily. Adjusting the upper and lower F0 limits and frame duration parameters in Praat typically led to improved tracking. These new F0 values were verified through visual and auditory inspection and were confirmed using direct calculation of the pitch period from the waveform. When Praat-derived F0 values continued to be judged as errors (this occurred in 14 productions), these values were replaced by manually derived values obtained from the waveform. A customized program operated on the Praat-generated values to calculate the following acoustic features per word: duration, peak F0 (F0peak), average F0 (F0ave), peak intensity (INTpeak), and average intensity (INTave), as well as pauses between adjacent words within each phrase (pause).

Listener Judgments of Stress

The listening task was conducted in a sound-treated audiometric booth with the participant seated in front of a computer displaying a graphical interface that guided them through the experiment and recorded their identification judgments. Stimuli were presented through headphones (AKG K240 Studio) at a comfortable listening level. A subset of the original recordings served as stimuli in the perceptual experiment. Specifically, three randomly selected trials of each phrase and stress location (A, B, C, D, N) were utilized for each speaker, resulting in a total of 75 productions per speaker. The task consisted of two segments: listening to productions from a speaker with dysarthria and listening to productions from that speaker's age- and gender-matched healthy control. The order of the segments was counterbalanced across all listeners. Each segment consisted of 75 productions (5 phrases × 5 stress locations × 3 repetitions) and 8 additional randomly selected reliability tokens. A short break was provided between segments and when requested by the participant.

For each utterance, listeners were instructed to choose whether they felt the speaker was placing stress on the first, second, third, or fourth word or whether they felt the phrase was produced neutrally. To ensure that listeners were not primed to only consider stressed productions, they were explicitly told that some trials were neutral in tone. Participants could choose to listen to an utterance up to two times prior to making their decision and proceeding to the next trial. The order of the recordings was randomized across the 4 listeners who heard a given speaker pair.

To determine intrarater reliability for each listener, 8 of the 75 tokens were randomly repeated. A listener was considered reliable if he/she responded consistently for both the original token and the reliability token, regardless of the response's accuracy. Listeners with reliability scores of at least 80% on the DYS and HC productions were included in the final analysis of the data. Two listeners who were initially recruited were excluded as a result of this criterion. Their data were replaced by those of 2 additional listeners who met the selection and reliability criteria.

Statistical Analysis

All data analyses were conducted in SAS (Version 9.1). To examine the acoustic correlates of contrastive stress in the production experiment, separate repeated measures analyses of variance (ANOVAs) were conducted for each speaker group and each dependent variable (average F0, peak F0, average intensity, peak intensity, duration, and pause). For each acoustic feature, the data were pooled over all five phrases, and analyses were conducted using two within-subjects factors. Factor 1 represented stress location—that is, whether contrastive stress was requested on the first (A), second (B), third (C), or fourth (D) word or if neutral production (N) was requested. Factor 2 represented the sentential word position—Word 1 (W1), Word 2 (W2), Word 3 (W3), or Word 4 (W4). In addition to the main effects, the Stress Location × Word Position interaction effects were of particular interest. The interaction would be statistically significant if speakers were able to temporally align the acoustic cues for contrastive stress at the appropriate word position within an utterance. The response variables were all continuous; F0 in Hz, intensity in dB, and duration and pause in seconds. Although pause is not typically associated with contrastive stress, our pilot data (Patel, 2004, Patel & Watkins, 2007) suggested that DYS may signal stress with an additional pause either before or after the target word as a consequence of reduced respiratory drive. Analyses on the pause feature, however, did not yield reliable results, given large variability among and between speakers in whether or not they placed a notable pause between words. As a result, there were many cells with missing data. Thus, this feature was not examined further in the present study. The F statistic was used to test the null hypothesis with α = .05. Additionally, for each of the remaining five acoustic features, 19 pair-wise contrasts (with Bonferroni-corrected α levels of .002) were conducted to examine differences within and across stress location and word position.

For the perception experiment, listener accuracy scores across phrases were analyzed using an ANOVA with two within-subjects factors: stress location, which had five levels (A, B, C, D, and N), and speaker group, which had two levels (DYS and HC).[2] For each listener, percentage accuracy scores were calculated across 15 tokens per stress location (3 trials × 5 phrases) for the DYS speaker and 15 tokens per stress location for the HC speaker that they heard. The F statistic was used to test the null hypothesis at an alpha level of .05. Post hoc, five pair-wise contrasts were conducted to examine differences in listener accuracy across speaker group for each stress location. To account for multiple comparisons, the Bonferroni correction factor was applied when interpreting these results with an adjusted alpha level of $p < .01$.

For each speaker group, separate logistic regression analyses were conducted for each stress location (A, B, C, D, and N) to determine which, if any, of five acoustic feature(s)[3] (average F0, peak F0, average intensity [dB], peak intensity, and duration) were predictive of listener identification accuracy. Before fitting the model, we performed a stepwise regression analysis with all five acoustic predictor variables, assuming statistical independence among the observations and employing a conservative entry/exit criterion of $p < .10$. In this way, we addressed the potential for multicollinearity in the final model by including only those acoustic parameters that were most likely to be predictive of listener identification accuracy. We constructed a binary dependent variable based on the average listener accuracy across all 4 listeners who heard a given spoken sample. When all 4 listeners identified the stress location accurately, we considered the token to be an agreement; otherwise, it was considered a disagreement. Each speaker was a clustering variable, given that we expected variability in the agreement between listeners who heard the same speaker as well as variability in listener responses across speakers and speaker groups. Small values in the working correlation matrix ($r < .05$) of the final model allowed us to assume an independence working correlation in the estimation of the regression parameters and standard errors. Acoustic predictor variables were considered significant if they met the criterion alpha level of .05.

Results

The findings are presented in four sections. The first section examines the acoustic cues for signaling contrastive stress in DYS and HC speakers by stress location. The second section presents the differences in listener accuracy scores between speaker groups and stress locations. The third section relates to the acoustic predictor variables for identifying contrastive stress placement within DYS and HC productions. The last section provides a descriptive analysis of the acoustic profile that may have led to listener errors in identifying DYS productions.

Acoustic Cues for Signaling Contrastive Stress

Recall that separate repeated measures ANOVAs were conducted for each acoustic feature (average F0, peak F0, average dB, peak dB, and duration) for the DYS and HC data sets. Each analysis examined the effects of two within-subject factors: stress location (A, B, C, D, and N) and sentential word position (W1, W2, W3, and W4). Table 2 provides a statistical summary of the main effects and interactions for each speaker group. The significance of each main effect was assessed at $\alpha = .05$, whereas contrasts were evaluated using an adjusted level of $\alpha = .002$. Findings for each speaker group are further elaborated in the sections that follow.

Acoustic cues in DYS productions. A statistically significant main effect in duration was found for stress location, $F(3, 33) = 20.70$, $p < .0001$. The main effect for sentential word position was not significant; however, the two-way interaction between sentential word position and stress location was significant, $F(12, 132) = 28.15$, $p < .0001$. Speakers with dysarthria elongated word duration to indicate contrastive stress. Within an utterance, the stressed word was approximately 25–30 ms longer than nonstressed words. Additionally, the same word was significantly longer in the stressed conditions compared with when it was produced in the neutral condition (A vs. N, $p = .0001$; B vs. N, $p = .0067$; C vs. N, $p = .0043$; D vs. N, $p = .0095$). The degree of lengthening on stressed words exceeded the natural tendency for final syllable lengthening within an utterance. It should also be noted that DYS productions were markedly slower than HC productions. This generalized slowing did not appear to impact the degree to which DYS speakers used word duration as an acoustic marker for contrastive stress. Figure 1 illustrates differences in word duration for each word position by stress location for DYS and HC productions.

Similar patterns of marking contrastive stress were evident in analyses of average and peak F0 in DYS productions (see Figure 2). Differences in peak F0 by stress location and word position are plotted to minimize redundancy and given that the perceptual literature tends to focus on the importance of F0 maxima for identifying contrastive stress placement (Gussenhoven, Repp, Rietveld, Rump, & Terken, 1997; Pierrehumbert, 1979; Rietveld & Gussenhoven, 1985; see also Terken, 1991, for stress patterning in nonsense syllables).

Statistically significant main effects in average and peak F0 were found for sentential word position ($p < .0001$)

---

[2] Initial analyses also examined phrase type as an additional within-subjects factor; however, the lack of statistical significance led to averaging listener inaccuracies across all five phrases.

[3] These acoustic cues were selected based on previous investigations on contrastive stress in DYS (Patel, 2004; Pell et al., 2006; Yorkston et al., 1984) and healthy speakers (cf. Eady & Cooper, 1986; Lieberman, 1960).

Table 2. Summary of ANOVA analyses of acoustic cues associated with contrastive stress in DYS and HC productions.

Speaker group	Acoustic cue	Main effects and interactions	*df*	*F*	*p*
DYS	Ave F0	Word position	3, 33	9.77	< .0001
		Stress location	4, 44	24.32	< .0001
		Word Position × Stress Location	12, 132	114.31	< .0001
	Peak F0	Word position	3, 33	14.06	< .0001
		Stress location	4, 44	216.27	< .0001
		Word Position × Stress Location	12, 132	157.04	< .0001
	Ave dB	Word position	3, 33	23.33	< .0001
		Stress location	4, 44	15.93	< .0001
		Word Position × Stress Location	12, 132	41.38	< .0001
	Peak dB	Word position	3, 33	14.26	< .0001
		Stress location	4, 44	19.72	< .0001
		Word Position × Stress Location	12, 132	46.36	< .0001
	Duration	Word position	3, 33	4.11	.0139
		Stress location	4, 44	20.70	< .0001
		Word Position × Stress Location	12, 132	28.15	< .0001
HC	Ave F0	Word position	3, 33	19.29	< .0001
		Stress location	4, 44	16.86	< .0001
		Word Position × Stress Location	12, 132	35.31	< .0001
	Peak F0	Word position	3, 33	21.21	< .0001
		Stress location	4, 44	21.88	< .0001
		Word Position × Stress Location	12, 132	50.78	< .0001
	Ave dB	Word position	3, 33	57.16	< .0001
		Stress location	4, 44	6.89	.0002
		Word Position × Stress Location	12, 132	32.55	< .0001
	Peak dB	Word position	3, 33	63.94	< .0001
		Stress location	4, 44	12.87	< .0001
		Word Position × Stress Location	12, 132	98.95	< .0001
	Duration	Word position	3, 33	7.78	.0005
		Stress location	4, 44	4.50	.0039
		Word Position × Stress Location	12, 132	77.63	< .0001

Note. ANOVA = analysis of variance; DYS = speakers with dysarthria; HC = healthy controls; Ave = average.

Figure 1. Mean word duration (sec) at each word position and stress location for DYS (left) and HC (right) speakers.

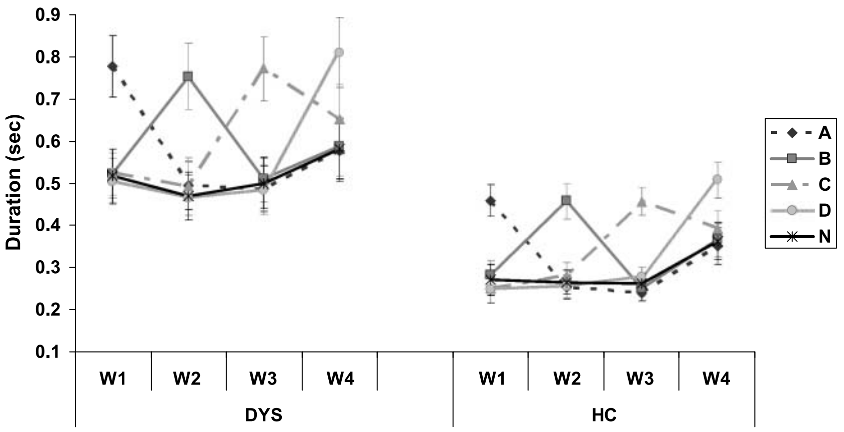
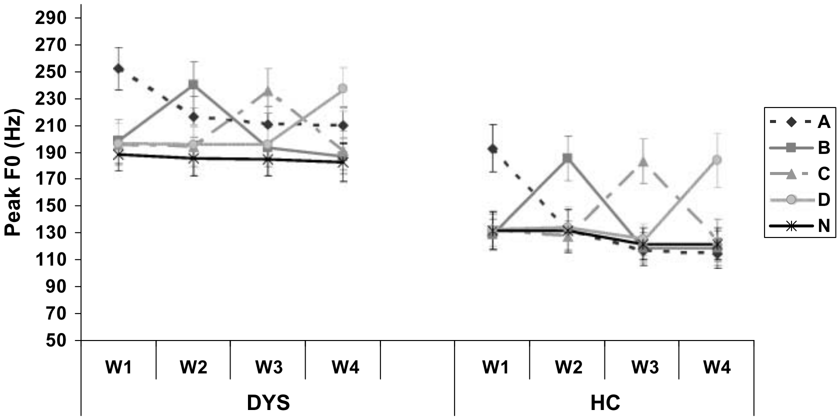


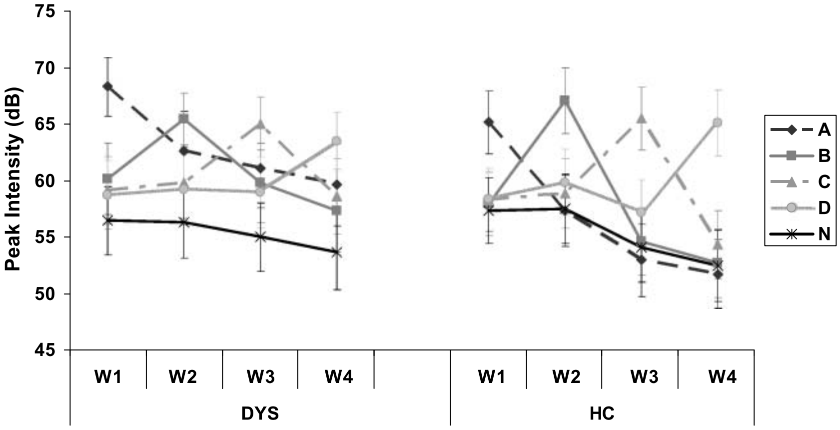
Figure 2. Mean peak F0 (Hz) at each word position and stress location for DYS speakers (left) and HC (right) speakers.



and stress location ($p < .0001$). In both cases, the two-way interactions between word position and stress location were also significant ($p < .0001$). DYS speakers produced neutral utterances with a relatively flat F0 contour rather than a gently falling contour, which is typical for English declarative sentences. Also noteworthy, F0 was elevated for all stress locations across the entire phrase in DYS productions. For stress A, DYS increased F0 on W1 and kept it relatively high even for nonstressed words compared with these same words when produced in the neutral condition. In all other stress conditions (B, C, and D), nonstressed words were characterized by similar F0 as when produced in the neutral condition. In the contrastive conditions, stressed words were approximately 25–30 Hz higher in average F0 compared with nonstressed words within the utterance. Regardless of sentential position, words were significantly higher in average F0 (approximately 40 Hz) in the stressed conditions versus in the neutral condition (A vs. N, $p < .0001$; B vs. N, $p = .0041$; C vs. N, $p = .0015$; D vs. N, $p = .0007$). Similarly, stressed words were approximately 40–45 Hz higher in peak F0 compared with nonstressed words within an utterance and approximately 60 Hz higher in peak F0 in contrastive productions versus neutral productions (A vs. N, $p < .0001$; B vs. N, $p = .0021$; C vs. N, $p = .0006$; D vs. N, $p = .0001$).

As in F0, analyses of average and peak intensity revealed similar patterns for marking contrastive stress. The results of differences in peak dB are presented in Figure 3. As indicated in Table 2, statistically significant main effects in average and peak intensity were found for sentential word position ($p < .0001$) and stress location ($p < .0001$). In both cases, the two-way interactions between word position and stress location were also

Figure 3. Mean peak intensity (dB) at each word position and stress location for DYS speakers (left) and HC (right) speakers.



significant ($p < .0001$). Neutral productions were characterized by a falling intensity contour across an utterance. Although contrastively stressed utterances also followed this general falling contour, they were characterized by increased average intensity (approximately 3–4 dB) and peak intensity (approximately 5–6 dB) on the stressed word compared to nonstressed words. For each stress location, stressed words were significantly higher in average intensity (A vs. N, $p = .0002$; B vs. N, $p = .0015$; C vs. N, $p = .0002$; D vs. N, $p = .0002$) and peak intensity (A vs. N, $p < .0001$; B vs. N, $p = .0003$; C vs. N, $p < .0001$; D vs. N, $p < .0001$) compared with when produced in the neutral condition. Also noteworthy is the fact that neutral productions were markedly softer in intensity than nonstressed words in the contrastive conditions. In other words, in the contrastive conditions regardless of sentential position, average and peak intensity were elevated even for nonstressed words.

It is important to note that not only did DYS use duration, F0, and intensity cues to mark contrastive stress within a sentence but that they did so in a time-aligned manner. In other words, they selectively altered these cues on the target word to convey stress.

Acoustic cues in HC productions. All five acoustic features generated statistically significant main effects for stress location and sentential word position as well as two-way interactions (see Table 2). Thus, HC speakers appeared to be marking contrastive stress with one or more of these cues. Similar to DYS speakers, HC marked contrastive stress with elongated word duration. Stressed words were approximately 15–20 ms longer than nonstressed words within an utterance. At each word position, stressed words were longer than when produced in the neutral condition (A vs. N, $p = .005$; B vs. N, $p = .0002$; C vs. N, $p = .0004$; D vs. N, $p = .0017$). This difference was least pronounced for stress D given the tendency for speakers to lengthen the final word within phrase.

With regard to average and peak F0, HC speakers produced the neutral condition with a phrase-final falling contour. In the contrastive conditions, speakers were able to selectively increase average F0 by approximately 35–40 Hz for stressed words compared with nonstressed words. In all but condition B, stressed words were significantly higher in average F0 (55–60 Hz) compared with when produced neutrally (A vs. N, $p = .0005$; B vs. N, $p = .08224$; C vs. N, $p = .0018$; D vs. N, $p = .0017$). Similarly, stressed words were approximately 50–60 Hz higher in peak F0 compared with nonstressed words within an utterance and approximately 55–60 Hz higher in peak F0 in contrastively stressed productions versus neutral productions (A vs. N, $p = .0001$; B vs. N, $p = .0018$; C vs. N, $p = .0002$; D vs. N, $p = .0004$).

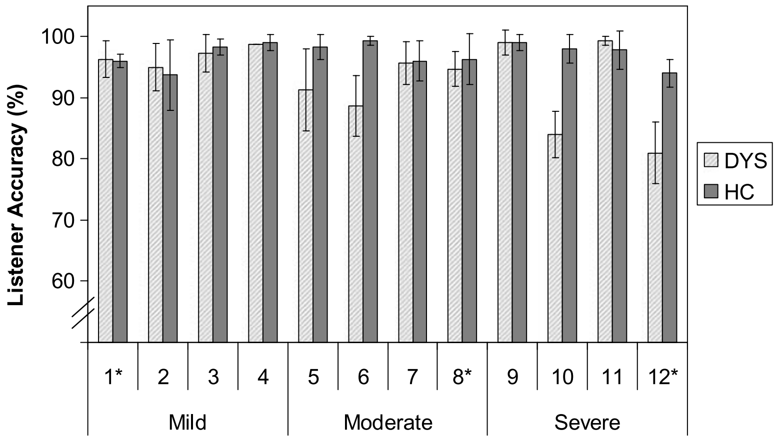
As in the DYS analyses, average and peak intensity in HC productions followed similar patterns for marking contrastive stress. A sharply falling intensity contour was noted in neutral and contrastive conditions. Although stressed words in the contrastive conditions A, B, and C were elevated in average intensity (approximately 4–6 dB) compared with nonstressed words within an utterance, there were no significant differences compared with when produced in the neutral condition (A vs. N, $p = .0295$; B vs. N, $p = .2741$; C vs. N, $p = .0402$). In contrast, although the final word in stress D was approximately 8 dB higher in average intensity ($p = .003$) in the contrastive condition versus the neutral condition, relative differences between neighboring words were not evident. Peak intensity was elevated by approximately 5–6 dB for stressed words relative to nonstressed words in all four contrastive conditions. At each word position, stressed words were higher in peak intensity for contrastive conditions versus the neutral condition (A vs. N, $p = .0003$; B vs. N, $p = .0018$; C vs. N, $p = .0003$; D vs. N, $p = .0001$).

In summary, DYS and HC speakers both utilized increases in duration, F0, and intensity to signal stress on a word within an utterance. DYS speakers increased the duration of stressed words, as compared with unstressed words, to a greater extent than HC (25–30 ms in DYS; 15–20 ms in HC). In contrast, DYS and HC produced similar changes in average F0 (DYS: 40 Hz; HC: 35–40 Hz), peak F0 (DYS: 60 Hz; HC: 55–60 Hz), average intensity (DYS: 3–4 dB; HC: 4–6 dB), and peak intensity (DYS: 5–6 dB; HC: 5–6 dB) to signal stressed versus unstressed words. It should be noted that within an utterance, the F0 and intensity contours of DYS productions were flatter than those of HC.

Listener Accuracy by Speaker Group and Stress Location

Listener accuracies for the DYS group were initially examined by severity level (mild, moderate, and severe). Average listener accuracies were 96.83% (*SD* = 1.55) for speakers with mild impairment, 92.58% (*SD* = 3.20) for speakers with moderate impairment, and 90.83% (*SD* = 9.70) for speakers with severe impairments. Figure 4 displays overall listener accuracies for each DYS–HC speaker pair clustered by severity. Accuracies differed by speaker even within each severity level, suggesting individual variation between speaker abilities and listener perception. Overall accuracy scores for speakers with mild DYS were similar to those for HC. Interestingly, listeners were more accurate on mild DYS productions than HC, when judging stress locations C, D, and neutral productions. Although accuracy tended to decrease as severity level increased, these differences across severity groups were not found to be statistically significant. As a result, subsequent analyses were conducted on the DYS group as a whole rather than considering subgroups.

Figure 4. Listener accuracy results by DYS and HC speaker pair clustered by dysarthria severity. Female speaker pairs are denoted by an asterisk (*).



The ANOVA yielded statistically significant main effects for speaker group, $F(1, 47) = 16.53$, $p = .0002$, and stress location, $F(4, 188) = 14.30$, $p < .001$. Additionally, the two-way interaction between speaker group and stress location was also significant, $F(4, 188) = 3.32$, $p = .0119$. It should be noted that although listener accuracy was statistically significantly lower for DYS productions compared with HC productions, the magnitude of the difference was quite small. On average, listeners were 93.42% (*SD* = 6.00%) accurate in identifying the locus of stress in DYS productions across all phrases and stress locations compared with 97.14% (*SD* = 1.93%) accurate for HC productions. The range of average listener accuracy per DYS speaker extended from 81.00% to 99.33%, compared with 93.67% to 99.33% for HC speakers.

Post hoc comparisons across speaker groups were conducted for each stress location. Listener accuracy for DYS productions was significantly lower than HC productions for stress locations A ($p < .0001$) and B ($p = .0019$) and less so for neutral productions ($p = .0135$). There were no differences between speaker groups for stress locations C ($p = .0876$) and D ($p = .3392$). For both speaker groups, listeners were least accurate in identifying neutral productions (DYS mean = 88.19%; HC mean = 92.92%) and most accurate in identifying stress B (DYS mean = 97.64%; HC mean = 99.72%).

Tables 3 and 4 display listener accuracy scores across all DYS speakers and all HC speakers as separate confusion matrices. Each matrix summarizes the number of accurate responses for each stress location across all 48 listeners, thus capturing the speakers' intentions and the listeners' responses simultaneously.

For the DYS productions, listeners made a total of 237 errors, with the majority occurring on neutral productions (35.86%), which were most often misjudged as stress A or B (34.12%, 35.29%, respectively) and infrequently misjudged as stress C or D (11.76%, 18.82%, respectively). Of the remaining 152 errors on the other stress locations, listeners chose neutral stress for 70.39% of trials. Furthermore, listeners mistook stress B on 38.46% of stress C trials.

Listeners displayed similar error patterns when judging HC productions. Of 103 total errors, 49.51% occurred on neutral productions. Listeners most frequently misjudged neutral productions as stress A (50.98%) and stress B (33.33%). The 52 remaining errors on other stress locations were most frequently (92.3%) misjudged as neutral stress.

Acoustic Predictors of Listener Accuracy in DYS and HC Productions

To determine which acoustic cues were predictive of listener identification accuracy, separate logistic

Table 3. Confusion matrix of observed versus expected listener responses for DYS productions. Each column sums to 720 total responses (12 speakers × 4 listeners × 5 phrases × 3 repetitions).

		Expected				
		A	B	C	D	N
Observed	A	652	4	1	4	29
	B	6	703	10	7	30
	C	2	1	694	1	10
	D	7	0	2	679	16
	N	53	12	13	29	635

Table 4. Confusion matrix of observed versus expected listener responses for HC productions. Each column sums to 720 total responses (12 speakers x 4 listeners x 5 phrases x 3 repetitions).

		Expected				
		A	B	C	D	N
Observed	A	707	0	0	1	26
	B	1	718	0	2	17
	C	0	0	711	0	4
	D	0	0	0	692	4
	N	12	2	9	25	669

regression analyses were conducted for each speaker group. The acoustic parameters retained for the final logistic models for each stress location are provided in Table 5. Odds ratios of less than 1 indicate that listeners were more likely to be accurate with lower values of the given acoustic variable, whereas odds ratios of greater than 1 indicate greater likelihood of listener accuracy with higher values of that acoustic variable. The standard regression coefficient is reported to summarize the magnitude and directionality of the change in a particular acoustic cue. The C statistic is a measure of the ability of the final model to discriminate between correct and incorrect listener responses with optimal values of 0.7 or higher. Recall, acoustic predictor variables were considered significant if they met the criterion alpha level of .05.

When listening to DYS productions, listeners were accurate at identifying stress A when the average intensity of W1 was increased ($p = .0004$) while the duration of W2 ($p = .002$) and peak intensity of W3 ($p = .0005$) were decreased. Stress B in DYS productions was predicted only by shortened duration of W3 ($p = .02$). Listeners were likely to correctly identify stress C in DYS productions when the average F0 ($p = .009$) and duration ($p = .01$) of W3 were increased along with reductions in duration of W1 ($p = .002$) and peak F0 of W2 ($p = .001$). The stringent entry/exit criteria led to no predictive acoustic variables for stress D. Listeners were more likely to accurately identify neutral DYS productions when the peak F0 ($p = .049$) and duration ($p = .008$) of W2 were decreased. In general, listener accuracy could be predicted by increases in one or more acoustic variable on the intended word and/or decreases on neighboring words.

For HC productions, stress A was predicted by increased peak intensity ($p = .003$) of W1. There were no reliable acoustic predictors of stress B. Listeners were more likely to accurately identify stress C in HC productions when the duration of W3 was increased ($p = .04$) but the duration of W2 ($p = .03$) was reduced. In contrast, increased peak intensity of W4 ($p = .001$) along with decreased average intensity ($p = .003$) and duration of W1 ($p = .03$) were predictive of accurate identification of stress D. Listeners were more likely to be accurate at identifying neutral HC productions when the average F0 of W1 was low ($p = .004$).

Table 5. Summary of logistic regression analyses aimed at determining acoustic predictors of listener accuracy for DYS and HC productions at each intended stress location (A–N).

Speaker group	Stress location	Acoustic predictor variable	Odds ratio (95% CI)	Standard regression coefficient	*p*	C statistic
DYS	A	W1-ave dB	1.85 (1.31, 2.59)	3.53	.0004	0.93
		W2-dur	0.84 (0.76, 0.94)	−1.76	.002	
		W3-peak dB	0.63 (0.48, 0.82)	−3.06	.0005	
	B	W3-dur	0.96 (0.92, 0.99)	−0.45	.02	0.77
	C	W1-dur	0.91 (0.86, 0.97)	−0.96	.002	0.90
		W2-peak F0	0.93 (0.89, 0.97)	−2.14	.001	
		W3-ave F0	1.06 (1.01, 1.10)	1.51	.009	
		W3-dur	1.18 (1.04, 1.34)	1.57	.01	
	D	None met .10 entry criterion				
	N	W2-peak F0	0.98 (0.97, 1.00)	−0.44	.049	0.82
		W2-dur	0.94 (0.90, 0.98)	−0.65	.008	
HC	A	W1-peak dB	1.29 (1.19, 1.40)	1.25	.003	0.75
	B	None met .10 retention criterion				
	C	W2-dur	0.92 (0.85, 0.99)	−0.50	.03	0.82
		W3-dur	1.23 (0.99, 1.40)	0.60	.04	
	D	W1-ave dB	0.71 (0.56, 0.89)	−1.62	.003	0.82
		W1-dur	0.93 (0.87, 0.99)	−0.45	.03	
		W4-peak dB	1.38 (1.13, 1.68)	1.78	.001	
	N	W1-ave F0	0.98 (0.96, 0.99)	−0.62	.004	0.73

Descriptive Analysis of Listener Error Patterns Among DYS Productions

To follow up on the interaction between production and perceptual variables, acoustic analyses were used to determine which cues may have led listeners to misidentify stress placement in DYS productions. This analysis was not carried out on HC productions given that our main focus was on DYS productions and that listeners were highly accurate in identifying stress placement in HC productions. Thus, the errors present were more likely to be idiosyncratic to individual speakers/listeners rather than indicative of group trends.

When DYS speakers intended to mark contrastive stress, listeners incorrectly chose N when the F0 range, intensity range, and/or duration range were markedly restricted. In particular, reduced intensity range appeared to have the greatest impact on listener errors. When speakers held a relatively constant intensity level across words (less than 8 dB variation), listeners tended to judge the production as neutral, even when F0 and duration cues were increased on the target stress location. In particular, when speakers DYS_5, DYS_6, DYS_10, and DYS_12 made accurate increases in F0 and duration but failed to differentially increase intensity, listeners tended to judge their productions as neutral. Conflicting cues within an utterance also led listeners to judge a production as neutral. For example, if separate cues were increased at two or more different stress locations or if the same cue was heightened at two or more points within the phrase, the production was misclassified as neutral. Errors of this nature were noteworthy for speakers DYS_2, DYS_3, and DYS_8.

Another predominant pattern among listener errors was to choose stress on neighboring words. For stress B and C, some speakers (particularly DYS_5, DYS_6, DYS_7, and DYS_12) had difficulty placing one or more prominence cues on the target word and instead misplaced the cue(s) on preceding or following words leading to listener errors. Perceptual errors on stress D occurred when F0 and intensity were relatively stable throughout the phrase. Additionally, listeners most frequently misjudged neutral sentences as stress A or B due to increased intensity, F0, and/or word duration at these locations.

Discussion

The present study sought to better understand the acoustic and perceptual cues to contrastive stress in the productions of DYS and HC speakers. The production experiment aimed to identify the acoustic cues used by DYS speakers and determine how the patterns differed from those used by HC. Results indicated that both speaker groups increased the F0, intensity, and duration of the stressed word. DYS relied more heavily on duration cues than HC when marking stress. Although the change in F0 and intensity on a given word between stressed and unstressed conditions was similar across speaker groups, the F0 and intensity *contours* of DYS productions were flatter than those of HC, implying somewhat restricted phonatory flexibility (similar to Liss & Weismer, 1994; Pell et al., 2006; Yorkston et al., 1984). Furthermore, F0 and duration were notably higher at each word position across all stress locations for DYS compared to HC.

Reduced range of F0 and intensity in DYS productions led to differences in sentence contours where the rise prior to a stressed word and the fall after the locus of stress tended to be more gradual in DYS productions and steeper in HC productions. Additionally, HC speakers produced nonstressed words in the contrastive conditions with similar F0, intensity, and duration as when produced in the neutral condition (see Cooper et al., 1985, and Eady & Cooper, 1986, for similar F0 and duration findings). In contrast, DYS speakers tended to produce an elevated F0 for the remainder of the phrase when stress was requested on the first word (A). Intensity of nonstressed words in all contrastive conditions was also elevated, relative to when the same words were produced in the neutral condition. These group differences may be attributable to the fact that varying F0 and intensity within a phrase requires fine grained adjustments that may be more difficult for DYS to achieve. In other words, once DYS speakers elevate F0 on the first word, they may have greater difficulty declining F0 quickly and to the same degree as HC for the remaining words within the utterance. Although this reduced flexibility for modulating F0 and intensity in DYS has been suggested in the literature (cf. Darley et al., 1969, 1975; Le Dorze et al., 1984; Yorkston et al., 1984), the present findings provide a clear account of the acoustic pattern across a set of spoken stimuli and speakers.

As expected, DYS productions were slower than HC productions across all word positions and stress locations. Despite this generalized slowing, DYS speakers were more likely to exaggerate changes in duration on stressed words compared with HC. Moreover, DYS used elongated word duration to mark contrastive stress on all four word positions including the final word, despite phrase-final lengthening. This increased reliance on duration cues to mark contrasts has been noted in previous studies on prosodic control in dysarthria (Patel, 2002, 2003, 2004; Patel & Salata, 2006; Yorkston et al., 1984). Greater reliance on duration cues may be a compensatory cue trading mechanism that DYS speakers utilize to convey contrastive meaning in light of restricted range and flexibility in F0 and intensity.

In addition to marking contrastive stress by heightening prosodic cues on the stressed word, DYS and HC speakers also markedly reduced these cues on unstressed words. Similar nonlocal effects of contrastive stress have been observed in previous production studies (Cooper, et al., 1985; Eady & Cooper, 1986; Weismer & Ingrisano, 1979). Given that DYS speakers have a restricted range available for signaling stress, such nonlocal effects may be an additional facilitating strategy for DYS speakers to convey contrastive meanings (Darley et al., 1969, 1975).

It was hypothesized that F0, intensity, and duration would vary depending on the position of the stressed word within an utterance, given that these cues are used to convey multiple linguistic functions (Cooper et al., 1985; Cooper & Danly, 1981; Eady & Cooper, 1986; Huss, 1978; Lehiste, 1970; Leiberman, 1967). We found evidence of this hypothesis in both the HC and DYS productions. For example, F0 and intensity tended to fall toward the end of an utterance, whereas duration increased to indicate phrasal boundaries. To compensate for these inherent tendencies, DYS and HC speakers further elongated the final word for stress D and heightened F0 and intensity on the first word relative to other words for stress A.

Sentence position has also been noted to impact the magnitude of change in F0 (Cooper et al., 1985), as well as in duration and intensity (Huss, 1978), in that absolute differences between stressed and unstressed words tend to be least pronounced toward the end of utterances. In the present study, however, we did not find evidence of this phenomenon in HC or DYS productions. Given that the stimuli used in previous studies were considerably longer (14–21 words in Cooper et al., 1985; 9–17 words in Huss, 1978) compared with those used in the present study (4 words), the ability with alter prosodic cues may have been adversely impacted by diminishing breath support.

The perception experiment sought to determine whether unfamiliar listeners could accurately identify contrastive stress placement in utterances produced by DYS and HC speakers. Given that reductions in F0 and intensity variation as well as slowed overall articulation rate are thought to be hallmarks of dysarthria (cf. Darley et al., 1969; Rosenbek & La Pointe, 1978), we hypothesized that listeners would have greater difficulty identifying stress placement for DYS speakers. However, the findings indicate that listeners were highly accurate at identifying stress placement in DYS productions (M = 93.42%) as well as HC productions (M = 97.14%). Moreover, there were no differences in listener accuracy across speakers with varying levels of DYS severity. Even if the combination of acoustic cues differed between individual speakers and speaker groups, listeners were able to attune to consistencies within the productions to make accurate stress identification judgments (similar to findings by Wang et al., 2005; Yorkston et al., 1984).

Although listeners were somewhat less accurate at identifying DYS productions compared with HC productions, error patterns were similar across stress locations. Listeners were most accurate in judging stress on the second word (stress B) for both DYS and HC productions and least accurate in judging neutral productions. Because the grammatical role of W2 varied across phrases, it is not likely that word class impacted listener perception. Examination of the acoustic data does not provide any obvious explanation for why listeners were most accurate at identifying stress B. In fact, speakers produced contrastive conditions B and C with similar changes in F0, intensity, and duration. For DYS productions, listeners had greater difficulty identifying stress A versus stress B. The acoustic data support this perceptual result in that the degree of change in F0 and intensity between stressed and unstressed words was limited for stress A versus other stress locations. With regard to the prevalence of errors on neutral productions, previous findings indicate that healthy speakers (Weismer & Ingrisano, 1979) and speakers with DYS (Wang et al., 2005) have difficulty maintaining relatively stable prosodic patterns throughout phrases, thus leading to listener errors.

There were no differences in listener accuracy between HC and DYS productions for stress conditions C and D. This may be due in part to Pierrehumbert's (1979) explanation that listeners expect F0 declination toward the end of a phrase and, thus, the reduced F0 range for DYS speakers would not have as much of an adverse effect on listener accuracy for these word positions.

Overall, production and perception of contrastive stress appear to be tied to local and nonlocal changes in acoustic cues across an utterance. Listener accuracy for identifying contrastive stress was best predicted by increases in F0, intensity, and/or duration on the stressed word and also by decreases in these cues on neighboring nonstressed words, suggesting that relative differences are more important than absolute targets (Pierrehumbert, 1979; Rietveld & Gussenhoven, 1985; Terken, 1991).

Three main perceptual error patterns emerged in the present study. First, listeners tended to perceive stress on neighboring words when prosodic changes were not temporally aligned with the target word. This finding is not surprising, given that DYS have difficulty coordinating the various subsystems of speech production.

Second, when speakers failed to maintain stable control of acoustic variables across an utterance, listeners identified the production as stressed even when it was intended to be neutral. Similarly, Wang et al. (2005) found that listeners had more difficulty identifying neutral sentences compared with stressed sentences regardless of whether the sentences were produced by healthy controls or speakers with dysarthria. Although they attributed listener errors in neutral productions to relative differences in intensity, we noted that errors were common when F0, intensity, and/or duration were inappropriately increased.

The third pattern of perceptual errors observed in the present study suggests that listeners may misjudge stressed productions as neutral when speakers mark stress with relatively small differences in acoustic cues and/or use conflicting cues. Similarly, Patel and Watkins' (2007) pilot study also found that listeners were more likely to choose neutral for a stressed token when there was little intensity variation across the phrase. In fact, the present study suggests that in some instances where F0 and duration cues were increased, listeners still chose neutral if intensity was not modulated. As Kochanski, Grabe, Coleman, and Rosner (2005) suggest, intensity may in fact be a strong predictor of prominence even in contrastive stress.

Although the present data set of speakers with dysarthria due to cerebral palsy had a somewhat restricted F0 and intensity range, they appear to have exploited control within this range to successfully convey contrastive stress to unfamiliar listeners. This finding stands in contrast to previous studies on speakers with acquired dysarthrias (Pell, et al., 2006; Yorkston et al., 1984). Pell et al. (2006) found that speakers with mild to moderate dysarthria due to Parkinson's disease had a reduced range of prosodic variation, which negatively impacted their ability to mark contrasts and, thus, lessened the chance for listeners to perceive stress in their productions. Similarly Yorkston et al. (1984) noted that "bizarre" stress patterning among a group of 3 speakers with mild ataxic dysarthria due to brain injury led to errors in listener perception. Perhaps the ability to exploit a narrowed range and convey contrastive meaning is impacted by etiology in that speakers with congenital dysarthria may have learned, over time, to compensate for reduced intelligibility by exaggerating residual prosodic ability. Perhaps etiology is a strong predictor of residual and compensatory prosodic abilities in speakers with dysarthria.

Limitations and Future Directions

The communicative potential of prosodic control in dysarthria is highlighted by findings that speakers were able to signal contrastive stress using F0, intensity, and duration cues and that unfamiliar listeners were highly adept at accurately identifying the intended stress location. In the present sample, DYS speakers appeared to use patterns of acoustic cues similar to those used by HC. A more extensive investigation on a larger sample of speakers with dysarthria encompassing a broader range of etiologies and severity levels is required to generalize these findings to stress patterning and its identification in dysarthria. Further research into the influence of etiology on the degree and nature of prosodic control is warranted. Perhaps speakers with congenital dysarthria may have established ways to exploit prosodic contrasts in comparison to those with acquired dysarthria, thus advocating for leveraging these residual abilities for communication. Although DYS severity did not impact listener accuracy in the present study, this result should be interpreted with caution given the limited sample size. A comparison of acoustic and perceptual differences in prosodic control among a larger sample of speakers with mild versus severe impairment may shed light on the impact of learned compensatory behaviors in speech motor control. In other words, when severity is high, the need to employ prosody for successful communication may be greater and, thus, listeners may give greater weight to prosody when segmental contrasts are reduced (cf. Cutler & Butterfield, 1992; Smith et al., 1989; Mattys, 2004; Mattys, White, & Melhorn, 2005; Spitzer, Liss, & Mattys, 2007).

Furthermore, it would be beneficial to assess prosodic control in more varied speech tasks where the interaction between prosodic and segmental control, in DYS can be explored. Examination of the effects of phrase length on ability to mark prosodic contrasts would elucidate how speaker fatigue may alter the ability to mark prosodic contrasts. Additionally, assessing prosody in more natural tasks and in conversation would reveal whether speakers with dysarthria routinely harness prosody to facilitate communicative success.

Further investigation of prosody in dysarthria would provide empirical evidence to challenge current clinical practices that typically address prosodic control in the final stages of intervention. If prosodic control is, in fact, relatively preserved in speakers with moderate to severe dysarthria, as the present findings suggest, incorporating word-, syllable-, and sentence-level prosodic tasks at the earliest stages of intervention may provide the scaffolding to achieve improved segmental intelligibility (Engstrand, 1988; Kent & Netsell, 1971) and more typical stress patterning (Kearns & Simmons, 1988). Designing comprehensive interventions that focus on speaker and listener variables would aid speakers in appropriately signaling intentions and would help listeners align expectations of intentional signals in dysarthria, thereby optimizing communication success.

Acknowledgments

This work was supported by NIH Grant DC-06118 from the National Institute on Deafness and Other Communication Disorders. This study was conducted in the Department of Speech-Language Pathology and Audiology at Northeastern University. We would like to thank Eldar Sadikov for developing the listener perception interface, Howard Cabral for his invaluable guidance with the statistical analysis and interpretation, and the participants for their time and enthusiasm for the study.

References

Abberton, E., Fourcin, A., & Hazan, V. (1991). Fundamental frequency range and the development of intonation in a group of profoundly deaf children. *Proceedings of the XII International Congress of Phonetic Sciences, 5,* 142–145.

Ansel, B. M., & Kent, R. D. (1992). Acoustic–phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *Journal of Speech and Hearing Research, 35,* 296–308.

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.

Boersma, P., & Weenink, D. (2005). *Praat: Doing phonetics by computer* (Version 4.3) [Computer program]. Retrieved July 30, 2005, from http://www.praat.org.

Bolinger, D. (1961). Contrastive accent and contrastive stress. *Language, 37,* 83–96.

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

Brewster, K. (1989). The assessment of prosody. In K. Grundy (Ed.), *Linguistics in clinical practice* (pp. 186–202). London: Taylor & Francis.

Canter, G. J. (1963). Speech characteristics of patients with Parkinson's disease: Intensity, pitch and duration. *Journal of Speech and Hearing Disorders, 28,* 221–228.

Chueng, J. Y., Holden, A. D., & Minifie, F. D. (1973). *Computer estimation and modeling of linguistic stress patterns in speech* (Technical Report No. 108). Seattle, WA: University of Washington.

Cooper, W. E., & Danly, M. (1981). Segmental and temporal aspects of utterance-final lengthening in temporal aspects of speech production and perception. *Phonetica, 31*(1–3), 106–115.

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,* 2142–2156.

Cooper, W. E., & Sorenson, J. M. (1981). *Fundamental frequency in sentence production*. New York: Springer-Verlag.

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G. T. M. Altman (Ed.), *Cognitive models of speech processing* (pp. 105–121). Cambridge, MA: MIT Press.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua, 92,* 81–104.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31,* 218–236.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40,* 141–201.

Crystal, D. (1979). Prosodic development. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 174–197). Cambridge, England: Cambridge University Press.

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.

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,* 402–415.

Engstrand, O. (1988). Articulatory correlates of stress and speaking rate in Swedish VCV utterances. *The Journal of the Acoustical Society of America, 83,* 1863–1875.

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

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

Gussenhoven, C., Repp, B. H., Rietveld, A., Rump, H. H., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *The Journal of the Acoustical Society of America, 102,* 3009–3022.

Hardy, J. C. (1983). *Cerebral palsy*. Englewood Cliffs, NJ: Prentice-Hall.

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

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

Kearns, K. P., & Simmons, N. N. (1988). Interobserver reliability and perceptual ratings: More than meets the ear. *Journal of Speech and Hearing Research, 31,* 131–136.

Kent, R. D., & Netsell, R. (1971). Effects of stress contrasts on certain articulatory parameters. *Phonetica, 24,* 23–44.

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.

Le Dorze, G., Ouellet, L., & Ryalls, J. (1994). Intonation and speech rate in dysarthric speech. *Journal of Communication Disorders, 27,* 1–18.

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: Academic Press.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America, 32,* 419–425.

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

Lieberman, P. (1967). *Intonation, perception and language*. Cambridge, MA: MIT Press.

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthria. *The Journal of the Acoustical Society of America, 104,* 2457–2466.

Liss, J. M., & Weismer, G. (1992). Qualitative acoustic analysis in the study of motor speech disorders. *The Journal of the Acoustical Society of America, 92,* 2984–2987.

Liss, J. M., & Weismer, G. (1994). Selected acoustic characteristics of contrastive stress in control geriatric, apraxic and ataxic dysarthric speakers. *Clinical Linguistics & Phonetics, 8,* 43–66.

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

Mattys, S. L. (2004). Stress versus coarticulation: Towards an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 397–408.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134,* 477–500.

Morgan, J. L., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition.* Mahwah, NJ: Erlbaum.

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

Netsell, R. (1973). Speech physiology. In F. Minifie, T. J. Hixon, & F. Williams (Eds.), *Normal aspects of speech, hearing, and language* (pp. 211–234). Englewood Cliffs, NJ: Prentice-Hall.

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

Patel, R. (1999). Prosody conveys information in severely impaired speech. In *Proceedings of the European Speech Community Association Workshop of Dialogue and Prosody* (129–135). Veldhoven, the Netherlands: De Koningshof.

Patel, R. (2002a). Phonatory control in adults with cerebral palsy and severe dysarthria. *Alternative and Augmentative Communication, 18,* 2–10.

Patel, R. (2002b). Prosodic control in severe dysarthria: Preserved ability to mark the question-statement contrast. *Journal of Speech, Language, and Hearing Research, 45,* 858–870.

Patel, R. (2003). Acoustic characteristics of the question-statement contrast in severe dysarthria due to cerebral palsy. *Journal of Speech, Language, and Hearing Research, 46,* 1401–1415.

Patel, R. (2004). The acoustics of contrastive prosody in adults with cerebral palsy. *Journal of Medical Speech-Language Pathology, 12,* 189–193.

Patel, R., & Salata, A. (2006). Using computer games to mediate caregiver-child communication for children with severe dysarthria. *Journal of Medical Speech Language Pathology, 14,* 279–284.

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,* 149–159.

Pell, M. D., Cheang, H. S., & Leonard, C. L. (2006). The impact of Parkinson's disease on vocal–prosodic communication from the perspective of listeners. *Brain and Language, 97,* 123–134.

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

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

Pierrehumbert, J. B. (1979). The perception of fundamental frequency declination. *The Journal of the Acoustical Society of America, 66,* 363–369.

Puyuelo, M., & Rondal, J. A. (2005). Speech rehabilitation in 10 Spanish-speaking children with severe cerebral palsy: A 4-year longitudinal study. *Pediatric Rehabilitation, 8,* 113–116.

Rietveld, A. C. M., & Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics, 13,* 299–308.

Rosenbek, J., & La Pointe, L. (1978). The dysarthrias: Description, diagnosis, and treatment. In D. Johns (Ed.), *Clinical management of neurogenic communicative disorders* (pp. 251–310). Boston: Little-Brown and Company.

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research, 29,* 169–182.

Sluijter, A., & van Heuven, V. (1996a, October). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of the International Conference on Spoken Language Processing, 96* (pp. 630–633). Philadelphia: 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,* 2471–2485.

Smith, M., Cutler, A., Butterfield, S., & Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech and Hearing Research, 32,* 912–920.

Spitzer, S. M., Liss, J. M., & Mattys, S. L. (2007). Acoustic cues to lexical segmentation: A study of resynthesized speech. *The Journal of the Acoustical Society of America, 122,* 3678–3687.

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,* 917–928.

Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *The Journal of the Acoustical Society of America, 89,* 1768–1776.

Vance, J. (1994). Prosodic deviation in dysarthria: A case study. *European Journal of Disorders of Communication, 29,* 61–76.

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 emphastic stress location and speaking rate. *Journal of Speech and Hearing Research, 22,* 516–533.

Wit, J., Maassen, B., Gabreels, F. J. M., & Thoonen, G. (1993). Maximum performance tests in children with developmental spastic dysarthria. *Journal of Speech and Hearing Research, 36,* 452–459.

Yorkston, K. M., & Beukelman, D. R. (1981). *The assessment of intelligibility of dysarthric speakers*. Austin, TX: Pro-Ed.

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: Pro-Ed.

Received April 7, 2007

Revision received November 11, 2007

Accepted April 16, 2008

DOI: 10.1044/1092-4388(2008/07-0078)

Contact author: Rupal Patel, Bouvé College of Health Sciences, Department of Speech-Language Pathology and Audiology, Northeastern University, 360 Huntington Avenue, Room 102 Forsyth Building, Boston, MA 02115. E-mail: r.patel@neu.edu.

Appendix. Target phrases and contextual scenarios.

Target sentence	Word of emphasis	Scenario
He lives near me.		
	He	Who lives near you?
	Lives	Does he work near you? No …
	Near	Does he live far from you? No …
	Me	Who does he live near?
Sam had fat cats.		
	Sam	Who had fat cats?
	Had	Does Sam have fat cats now? No …
	Fat	Did Sam have skinny cats? No …
	Cats	Did Sam have fat dogs? No …
She hid his key.		
	She	Who hid his key?
	Hid	What did she do with his key?
	His	Whose key did she hide?
	Key	What did she hide of his?
That man ran fast.		
	That	Which man ran fast?
	Man	Did that woman run fast? No …
	Ran	Did that man walk fast? No …
	Fast	Did that man run slowly? No …
She hit me here.		
	She	Who hit you here?
	Hit	What did she do to you here?
	Me	Who did she hit, here?
	Here	Where did she hit you?

Acoustic and Perceptual Cues to Contrastive Stress in Dysarthria

Rupal Patel, and Pamela Campellone

J Speech Lang Hear Res 2009;52;206-222; originally published online Aug 11, 2008;

DOI: 10.1044/1092-4388(2008/07-0078)