Prosodic Control in Severe Dysarthria: Preserved Ability to Mark the Question-Statement Contrast

Rupal Patel
Department of Biobehavioral Sciences
Teachers College
Columbia University
New York, NY

Speakers with severe dysarthria are known to have reduced range in prosody. Consistent control within that range, however, has largely been ignored. In earlier investigations speakers with severe dysarthria were able to control pitch and duration for sustained vowel production despite reduced flexibility of control (Patel, 1998). The present experiment examined whether 8 speakers with severe dysarthria due to cerebral palsy used prosodic parameters of pitch contour and syllable duration for phrase-level productions. Speakers with dysarthria ($N = 8$) produced 3-syllable phrases as questions and statements. Naïve listeners ($N = 48$) classifed dysarthric productions as either questions or statements. Listeners were able to distinguish questions from statements with accuracy levels ranging from 81% to 98 %. We were also interested in studying how dysarthric speakers marked the question-statement contrast. Prosodic features of pitch contour and syllable duration were systematically removed from the original recorded vocalizations to examine the salience of these features on listener classification. Removal of pitch contour cues dramatically reduced listener accuracy scores to almost chance performance. Listeners found pitch contour cues to be information-bearing cues in dysarthric vocalizations even though the range of frequency control in these speakers may be reduced. That speakers with dysarthria were able to exert sufficient control to signal the question-statement contrast has implications for diagnostic and intervention practices aimed to optimally exploit prosodic control for enhancing communication efficiency.

KEY WORDS: dysarthria, prosody, question-statement, cerebral palsy, fundamental frequency

Individuals with severe dysarthria are unable to communicate effectively using speech alone. Many individuals, however, continue to use vocalizations when interacting with familiar communication partners (cf. Allaire, Gressard, Blackman, & Hostler, 1991; Ferrier, 1991; Fried-Oken, 1985; Shein, Brownlow, Treviranus, & Parnes, 1990; Smith, 1994). These individuals exploit any available residual speech to express emotions, gain attention, and signal danger (Beukelman & Mirenda, 1992).

It is unclear which features of the dysarthric speech signal convey information to communication partners. Information within multiple contextual channels (e.g., facial expressions, body language, and situational cues) may facilitate the communication exchange. Successful communication may also be attributed to familiarity and experience with the speaker (McGarr, 1983; Monsen, 1983). Semantic and syntactic context provide additional cues that influence listeners' abilities to understand dysarthric speech (Dongilli, 1993; Yorkston, Hammen, & Dowden, 1991). However, it is possible that familiar communication partners become attuned to subtle acoustic features that facilitate disambiguation of dysarthric vocalizations. Which acoustic features, if any, are salient for uncovering the intended message within dysarthric vocalizations?

Characterization of dysarthria due to cerebral palsy has typically focused on segmental aspects of the speech signal (cf. Ansel & Kent, 1992; Hardy, 1983; Irwin, 1955; Platt, Andrews, & Howie, 1980; Platt, Andrews, Young, & Quinn, 1980). Slow and imprecise speech motor control results in reduced vowels, imprecise consonants, and changes in resonance (Ansel & Kent, 1992; Platt, Andrews, & Howie, 1980; Platt, Andrews, Young, & Quinn, 1980; Rosenbek & La Pointe, 1978). In contrast to the numerous investigations of segmental quality, few previous studies have examined the communicative role of prosody in dysarthric speech.

Prosody as an Information Carrier

Prosody refers to aspects of the speech signal that mark stress, rhythm, intonation, and pause structure (Lehiste, 1976; Netsell, 1973). Acoustic parameters associated with prosody include fundamental frequency (F_0), amplitude, duration, and segment quality (Shattuck-Hufnagel & Turk, 1996). Prosodic cues supplement the linguistic structure of the spoken message (Brewster, 1989; Kent & Read, 1992), indicate the speaker's emotional state (cf. Streeter, Macdonald, Apple, Krauss, & Galotti, 1983; Williams & Stevens, 1972), and are important for distinguishing between grammatical forms such as questions and statements (cf. Eady & Cooper, 1986; Grant & Walden, 1996).

Previous Approaches to Studying Prosodic Control in Dysarthria

Numerous studies have examined prosodic differences between normal and dysarthric speech (Darley, Aronson, & Brown, 1969, 1975; Hardy, 1983; Le Dorze, Ouellet, & Ryalls, 1994; Rosenbeck & La Pointe, 1978; Yorkston, Beukelman, Minifie, & Sapir, 1984). Relative to normal speech, dysarthria is characterized by monopitch, monoloudness, and slowed rate. Similar differences between dysarthric and normal controls have been reported for phrase-level control of prosody. Le Dorze et al. (1994) compared the differences between 10 French speakers with dysarthria and normal controls on their ability to mark question-statement contrasts. They found that speakers with dysarthria produced smaller intonational differences and that the overall rate of the utterances was slower than that of normal speakers. Yorkston et al. (1984) studied stress patterning in three speakers with mild dysarthria, and they also noted reduced variation of frequency, intensity, and duration among speakers with dysarthria compared with normal speakers.

Although these findings show that dysarthric speakers differ from normal speakers, the question of whether dysarthric speakers can in fact communicate the difference between statements and questions using prosodic cues remains unexplored. Speakers with markedly reduced control over segmental aspects may nevertheless have sufficiently consistent control to convey aspects of their intentions by varying suprasegmental, prosodic cues. Prosodic parameters generally vary at slower time scales than segmental features. Given that dysarthria is a motor speech impairment characterized by slow, weak, and imprecise movement, it follows that segmental articulatory gestures will be difficult to control. We hypothesize, however, that relatively slow and gradually varying prosodic features may nonetheless convey information.

In this experiment we investigated the prosodic control abilities of individuals with severe dysarthria to mark question-statement contrasts. A rising F_0 contour has been associated with yes-no questions in English (cf. Eady & Cooper, 1986; Morton & Jassem, 1965; Shattuck-Hufnagel & Turk, 1996) as well as other languages (Le Dorze et al., 1994; Lieberman, 1967). Durational cues have also been found to mark yes-no question-statement contrasts (Fry, 1958).

We were interested in understanding whether speakers with severe dysarthria could communicate this contrast to listeners despite reduced flexibility and control and, if so, how they were doing this. A group of speakers with severe dysarthria were asked to produce phrases as questions and statements. Fundamental frequency and syllable-duration cues were systematically removed from the original vocalizations to examine the role of these features in communicating the question-statement contrast. A group of naïve listeners were then asked to classify the original and modified recordings as either questions or statements. An analysis of listener categorization accuracy was performed to uncover the information-bearing prosodic cues in the dysarthric vocalizations.

Method

Participants

Participants in this investigation consisted of two groups: 8 speakers with severe dysarthria and 48 naïve listeners. Participants with dysarthria were recruited from speech and language clinics in the Greater Toronto Area and through word of mouth. Only those individuals capable of fully and freely consenting were included. Listener participants were recruited through flyers and by word of mouth from the Greater Toronto Area.

Speakers With Dysarthria

Eight individuals with primarily spastic dysarthria were selected to participate in the investigation. Speakers ranged in age from 27 to 44 years, with a mean age of 36 years (see Table 1 for speaker characteristics). Speakers met a set of four selection criteria. First, dysarthria was the primary speech diagnosis for all participants. A battery of formal and informal evaluations was performed to determine the nature and degree of speech impairment. An oral peripheral examination and a motor speech exam were conducted to differentially diagnose dysarthria from concomitant verbal apraxia and/or aphasia. Second, a modified version of the Assessment of Intelligibility of Dysarthric Speech (AIDS; Yorkston & Beukelman, 1981) was administered to determine the level of severity of dysarthria. To minimize speaker fatigue, the modified assessment consisted of 25 rather than 50 isolated word productions. The average of three unfamiliar raters' scores was used as a measure of the speaker's intelligibility. On the 25 single-word production task, all of the speakers' speech intelligibility ratings fell below 25%. These ratings underscore the severe extent of speech impairment in this group. Third, speakers were required to pass a pure tone audiometric evaluation with thresholds at or below 25 dB HL in at least one ear. Fourth, all speakers had grossly adequate receptive language and cognitive skills necessary for completing the experimental task.

Listeners

A total of 48 normal-hearing English monolingual speakers between the ages of 22 and 50 years with a mean age of 28 years served as listeners. All listeners were unfamiliar with dysarthric speech and with the stimulus materials. Each listener was randomly assigned to a speaker group, resulting in 8 groups, each with 6 listeners.

Adequate hearing function was required of all listener participants. An audiometric screening evaluation was completed to ensure that average pure tone thresholds (at 500, 1000, and 2000 Hz) were at or below 25 dB HL in at least one ear.

Materials and Apparatus

Speech recordings were made using a digital audio recorder (DAT) (Sony, model PCM-2300) and a unidirectional head-mounted cardioid dynamic microphone (Shure, model SM10A) and were collected in a sound-treated audiometric booth. The microphone was placed 2 cm from the left corner of the speaker's mouth.

The listening task was conducted in a controlled noise environment, and all stimuli were presented through a set of headphones (Telephonics, model TDH-50P). Listeners used a computer interface to play dysarthric vocalizations and to record their categorization responses.

Statistical analyses were performed on the data using the SAS (SAS Institute, Version 8.01) software package.

Table 1. Description of speakers with dysarthria.

Speaker	Age	Gender	Speech intelligibility	Reported modes of communication	Motor control
S1	27	M	12%	gesture, vocalizations, sign language	able to sign , unable to write, uses a wheelchair
S2	33	F	18%	head pointer, vocalizations	uses a wheelchair, head control
S3	35	M	20%	communication board with bliss symbols, vocalizations	uses a wheelchair, pointing gestures
S4	32	M	22%	picture symbols, vocalizations	uses a wheelchair, pointing gestures
S5	44	F	16%	alphanumeric communication board, vocalizations	uses a wheelchair, pointing gestures
S6	36	M	22%	alphanumeric communication board, some vocalizations	ambulatory, pointing gestures
S7	40	M	18%	alphanumeric and phrase communication board	uses a wheelchair, pointing gestures
S8	44	M	24%	alphanumeric communication board, vocalizations with familiar people	ambulatory, pointing gestures

Procedure

The experiment consisted of collecting speech samples from the 8 speakers, preparing and manipulating the vocalizations, and administering a perceptual listening task completed by the 48 listeners using original and manipulated vocalizations as stimuli.

Collecting the Original, Unmodified Sample

Each speaker produced vocalizations from a master stimulus list (see Appendix). The phrases were adapted (i.e., shortened to three syllables and consonant clusters minimized) from the contrastive stress drills in *The Source for Dysarthria* (Swigert, 1997) and *The Dysarthria Sourcebook* (Robertson, Tanner, & Young, 1989). The list consisted of 10 unique phrases (referred to as phrase types hereafter). Speakers were asked to produce each phrase type as a statement 5 times and as a question 5 times, resulting in a total of 100 recordings (referred to as tokens hereafter) per speaker. Contextual scenarios were provided to elicit the question versus statement contrast. The order of the 10 phrase types and the order of question and statement tokens within each type were randomized across speakers.

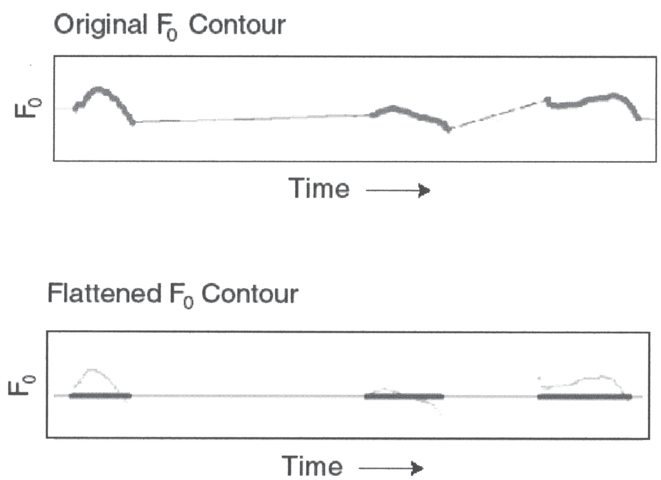
Stimulus Sets

For each speaker, a subset of 60 original utterances (SS_{orig}) consisting of 3 question tokens and 3 statement tokens per phrase type were randomly selected as representative samples. SS_{orig} was manipulated in three ways: (1) *flattened* F_0 contour (SS_p), (2) equalized syllable duration (SS_d), and (3) both flattened F_0 and equalized syllable duration (SS_{pd}).

The data sets were manipulated through a multistage procedure. First, the beginning and end of each syllable of each phrase token was marked by listening to the waveform and using the intensity envelope as a guide. The Praat speech analysis software package (Boersma & Weenik, 2000) was used to flatten the F_0 contour to the level of the average F_0 of all 6 tokens of each phrase type. F_0 contours were flattened by performing local F_0 shifts for each period of the waveform so that the overall F_0 contour was approximately constant (see Figure 1). The pitch synchronous overlap and add (PSOLA)[1] technique was used to achieve F_0 shifts without affecting the tempo of the recording (Moulines & Charpentier, 1990). This process resulted in a set of 60 flattened F_0 samples (SS_p).

[1] The PSOLA algorithm was found to introduce some audible artifacts producing a “tinny” mechanical sounding quality in the resulting F_0 flattened files. This artifact may have been due to spectral discontinuities introduced at pitch period boundaries. As an alternative, we experimented with performing pitch modification based on linear predictive coding (LPC) (see Laures & Weismer, 1999 for similar manipulations), but we found that PSOLA resulted in fewer artifacts for dysarthric speech.

Figure 1. Pitch flattening using Praat. The upper plot shows the estimated F_0 contour of an original 3-syllable vocalization. The lower plot shows the F_0 contour after applying the PSOLA- based F_0 flattening procedure.



A second stimulus set (SS_d) was created by equalizing syllable duration. For each phrase type, the average duration of all 6 phrase tokens was calculated. Each syllable within a given token was stretched or compressed to the average duration of a syllable for that phrase type. Using the PSOLA algorithm, Praat was used to modify duration without modifying pitch (see Figure 2). This resulted in a set of 60 duration-equalized vocalizations (SS_d).

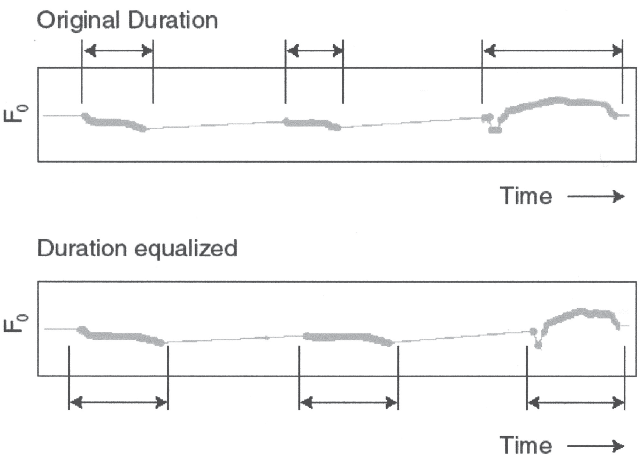
A third stimulus set (SS_{pd}) was created by first flattening F_0 and then equalizing syllable duration using the methods described above.

The Listening Task

Eight groups of 6 listeners each completed a listening task that involved classifying vocalization samples from the four stimulus sets described above: original, flattened F_0, duration equalized, and both F_0 and duration altered. Each listener classified 280 vocalizations produced by one of the 8 speakers. The total number of tokens heard corresponded to 60 SS_{orig}, 60 SS_p, 60 SS_d, 60 SS_{pd}, and 40 random repeats. The 40 random repeats consisted of 10 repeats from each of the four stimulus sets and were used to judge listener reliability. Listeners with reliability scores of at least 90% were included in the final analysis of the data. Only two listeners were excluded as a result of this reliability criterion. These listeners had reliability ratings of 76% and 80%.

The order of presentation of the listening stimuli was randomized across all listeners in each speaker

Figure 2. Equalizing syllable duration. The top plot shows the F_0 contour of an original 3-syllable vocalization. The duration of the third syllable is significantly longer than the first and second. In the lower plot, the syllable durations are equal after applying the duration equalization procedure.



block. Order effects were also controlled by randomizing playback of all 280 vocalizations across all four stimulus sets.

A computer interface was used to present the stimuli and record listener responses. Listeners pressed the "Play" button to hear each vocalization in turn. For each vocalization, the text corresponding to the target vocalization was displayed on a monitor as a statement (i.e., the text followed by a period) and as a question (i.e., the text followed by a question mark). A selection button was displayed next to each text prompt. Listeners were instructed to categorize each vocalization as either a question or a statement by selecting the appropriate button. When unsure, listeners were asked to make their best possible judgment. They had the option to repeat vocalization playback at any time. Participants were able to proceed at their own pace (using a button labeled "Next") and could pause and resume as they wished. To help the listener gauge his/her progress, a counter indicated the number of remaining trials in the experiment.

After completing the listening study, each participant completed a post-experimental questionnaire. The questionnaire was designed to probe listeners about the cues they had used in making their judgments.

Design and Analysis

A split-plot factorial design (for details, see Corey, 1998; Kirk, 1982) was used to analyze the results of this experiment. Heterogeneity among speakers with severe dysarthria has been documented thoroughly, and thus a blocked design was used in this investigation. The 8 speakers with severe dysarthria formed the main plot. Split plots consisted of 6 listeners in each of the main plots. The 6 listeners who were exposed to stimuli from a given speaker were expected to behave similarly and thus formed a block. The treatments consisted of the original and three manipulated stimulus sets, which had been randomized among all speaker blocks.

Two within-subject factors were examined. Factor 1 represented the manipulation and had 4 levels (SS_{orig}, SS_p, SS_d, SS_{pd}). Factor 2 represented the target phrase types and had 10 levels (GIB, HLT, ITD, IWH, PIL, PSS, SLD, SSN, SWH, and USS are the abbreviations for each phrase type). One between-subject factor consisting of the speaker block was also present.

The response variable in this experiment was binary; listener judgments were either correct or incorrect; thus logistic regression was used for analysis (cf. Pagano & Gauvreau, 1993). The Wald chi-square statistic was used to test the null hypothesis. The CATMOD procedure in SAS was performed because the manipulation was a categorical variable.

Results

For each listener a classification accuracy score, a proportion of correct responses out of 60 total utterances, was calculated for each of the four stimulus sets. The proportion of correct classifications in each stimulus set for each phrase was also determined.

Statistical Analysis to Determine the Effect of the Manipulation

A logistic model was built using SAS to assess whether the manipulation (SS_{orig}, SS_p, SS_d, SS_{pd}) and phrase type (GIB, HLT, ITD, IWH, PIL, PSS, SLD, SSN, SWH, and USS) affected the listeners' ability to correctly discriminate between statements and questions. The model built was of the following form:

Proportion correct responses = Intercept + Manipulation + Speaker + Phrase + Manipulation * Speaker + Manipulation * Phrase + Speaker * Phrase

The Manipulation * Phrase interaction term was not found to be significant and was removed from the model. In other words, response accuracy for the four manipulations did not differ across the 10 phrase types. The model was then refitted using the remaining terms.

Proportion correct responses = Intercept + Manipulation + Speaker + Phrase + Manipulation * Speaker + Speaker * Phrase

The results of the logistic regression revealed a statistically significant difference in the proportion of correct responses due to manipulation ($\chi^2 = 818.5$, $df = 3$, $p < 0.0001$). Pairwise comparisons revealed that all prosodic manipulations (SS_p, SS_d, SS_{pd}) produced results that differed significantly from the accuracy levels of SS_{orig}. Statistically significant main effects were also noted for phrase type. The speaker main effect was a consequence of its being a blocking factor. The main effects of the logistic regression are presented in Table 2.

Examination of the data from all 8 speakers revealed a significant phrase effect ($\chi^2 = 20.6$, $df = 9$, $p = 0.0145$). That is, some phrases were more difficult to categorize than others. To assess whether the same phrase types were difficult for all 8 speakers, eight logistic regressions were performed on an individual speaker basis. Significant phrase effects on an individual speaker level were not found for the same phrase type across all speakers. In other words, a particular phrase type was not uniformly more difficult or easy to categorize than other phrase types across speakers.

Pairwise contrasts between the manipulations were also carried out (see Table 3). Listener accuracy scores for SS_{orig} ranged between 80% to 99%, with a mean of 87% across all speakers. Accuracy scores on SS_p ranged from 49% to 65%, with a mean of 55% across all speakers. Thus, flattening the F_0 in SS_p resulted in significantly reduced listener accuracy scores as compared with scores for SS_{orig} ($\chi^2 = 466.4$, $df = 1$, $p < 0.0001$). Accuracy scores for SS_{pd} ranged between 51% and 63%, with a mean of 54% across all speakers—again a significant reduction compared to SS_{orig} ($\chi^2 = 519.8$, $df = 1$, $p < 0.0001$). Listener accuracy scores for SS_d ranged between 79% to 99%, with a mean of 86% across all speakers. Although this drop in accuracy was statistically significant ($\chi^2 = 154.8$, $df = 1$, $p < 0.0001$), the effect size of dropping from a mean of 87% with SS_{orig} to 86% in SS_d is considered marginal. The accuracy scores of SS_p and SS_d were found to differ significantly from each other ($\chi^2 = 324.0$, $df = 1$, $p < 0.0001$), as were the scores obtained from SS_d and SS_{pd} ($\chi^2 = 341.9$, $df = 1$, $p < 0.0001$). No significant differences, however, were found between SS_p and SS_{pd} ($\chi^2 = 0.9$, $df = 1$, $p = 0.3332$).

The direction of differences in accuracy between manipulations was determined by examining the total number of correct responses for each stimulus set (see Table 4). Listeners were most often correct when no manipulation was performed. Accuracy was dramatically reduced when F_0 cues were removed. Scores dropped slightly further when both F_0 and duration cues were removed together. In contrast to the F_0 manipulation, response accuracy was only slightly reduced when duration cues were removed. This pattern of reduction in accuracy scores as a result of the manipulation was found across all 8 speaker groups.

Figure 3 illustrates the change in accuracy as a function of the stimulus set for each speaker. The accuracy scores are reported as a proportion correct out of a total of 60 stimuli. The proportion is the mean for all 6 listeners in the speaker group. The standard error bars depict the variation among listeners for a given stimulus set.

An error analysis revealed that listeners in all speaker groups made more errors when listening to question tokens rather than statement tokens across all manipulations (Table 5). When duration cues were removed, listeners made a similar proportion of errors on questions as in the original stimulus set. In contrast, when F_0 cues were removed, listener accuracy was reduced dramatically, and this reduction was mainly attributable to difficulty with classifying question tokens.

Table 2. Main effects and interactions of the logistic regression analysis.

Effect	Degrees of freedom	Wald chi-square value	Probability > chi-square
Speaker	7	58.1	< .0001
Manipulation	3	818.4	< .0001
Phrase	9	20.6	0.0145
Speaker * Manipulation	21	119.7	< .0001
Speaker * Phrase	63	248.0	< .0001

Table 3. Results of all pairwise contrasts between all manipulations.

Contrast	Degrees of freedom	Chi-square value	Probability > chi-square
SS_{orig} vs SS_p	1	466.4	< .0001
SS_{orig} vs SS_d	1	154.8	< .0001
SS_{orig} vs SS_{pd}	1	519.8	< .0001
SS_p vs SS_d	1	324.0	< .0001
SS_d vs SS_{pd}	1	341.9	< .0001
SS_p vs SS_{pd}	1	0.9	0.3332

Table 4. Comparison of accuracy scores across all speakers for each stimulus set.

Manipulation	Sum of correct responses across all speaker groups	Proportion of correct responses
SS_{orig}	2506	87.0
SS_p	1596	55.4
SS_d	2482	86.2
SS_{pd}	1561	54.2

Table 5. Error analysis of question vs. statement tokens across manipulation.

Manipulation	Percentage error for statements	Percentage error for questions
SS_{orig}	2.6	23.2
SS_p	4.0	85.0
SS_d	2.8	24.7
SS_{pd}	3.0	88.2

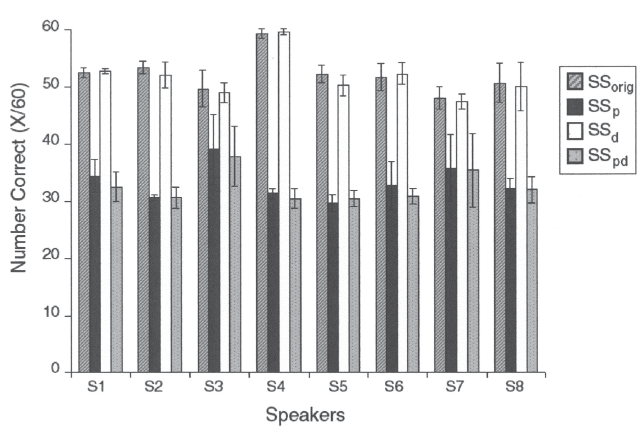
It is important to note that listeners in speaker groups S3 and S7 had proportionately fewer errors on question tokens in the SS_p and SS_{pd} conditions than listeners in other speaker groups.

Figure 4 illustrates accurately identified question and statement tokens (left column) and poorly identified question and statement tokens (right column). Acoustic analysis of how speakers mark stress in this data set is underway.

Post-experimental questionnaires revealed that most listeners (41 of the 48) reported attending to "pitch/intonational contour" cues. Some listeners (19 out of 48) also reported using duration cues such as rate of speech and the length of the last syllable, whereas others (18 of the 48) reported attending to loudness cues as well. Interestingly, 10 of the 18 listeners who responded that loudness was an important cue for them were in either speaker S3's or speaker S7's group. Listener accuracy scores for these two groups did not drop as significantly as those of other groups when F_0 cues were removed.

Other comments used by listeners to describe how they differentiated questions from statements included "effortful for questions" (3 out of 48), "emphasis for questions" (2 out of 48), "strained for questions" (2 out of 48), "urgency and abruptness of questions" (2 out of 48), and "emotional for questions" (1 out of 48).

Figure 3. Effect of manipulation on listener accuracy for all eight speaker groups.



Discussion

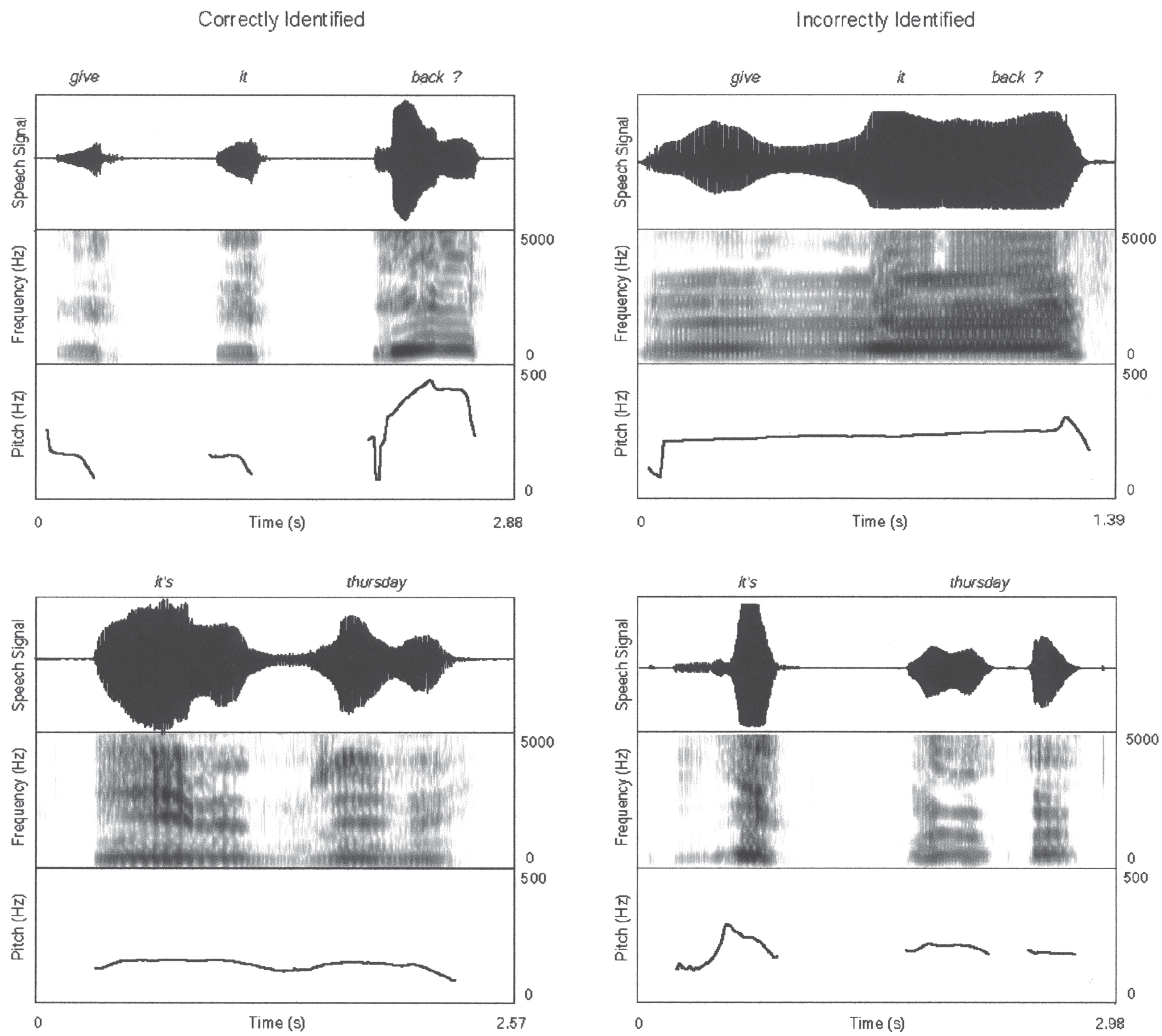
Our goal was to determine whether individuals with severe dysarthria employ prosodic control to indicate the intent of utterances to a degree sufficient for human listeners unfamiliar with their speech to accurately classify their productions. If speakers were found to vary prosodic parameters in a consistent and reliable manner, we would have an existence proof of vocal control to signal a linguistic contrast. Perhaps, then, residual prosodic control could be shaped and exploited to help the speaker with severe dysarthria convey an increased amount of information using vocalizations.

Listeners' Ability to Classify Dysarthric Vocalizations

Listeners were able to accurately classify dysarthric vocalizations using prosodic cues alone. Listener accuracy scores for SS_{orig} far exceeded chance performance (i.e., 50%) and were similar across all speaker groups. Accuracy scores in the 8 speaker groups varied from 80% to 99%, with a mean of 87%. Such high levels of accuracy demonstrate that speakers were able to encode information within the prosodic channel to signal the difference between questions and statements.

Because the same phrases were repeated as both questions and statements, listeners could rely only on prosodic cues when making classification judgments. Contextual and supplementary cues—such as facial expressions, visual information of orofacial movements, gestures, semantic context (Dongilli, 1993), word predictability (Ansel & Kent, 1992), and linguistic context (Beliveau, Hodge, & Hagler, 1995; Hustad, 1999)—were not available to listeners in the experiment. Using acoustic information alone, listeners were able to extract and utilize prosodic cues to make accurate classifications. The present results cannot be attributed to idiosyncratic cues associated with any particular production because multiple productions of each phrase were included in

Figure 4. Sample utterances from the data set: The left column depicts tokens that were correctly identified by listeners; the right column depicts incorrectly identified tokens. For each token, the elements illustrated are the acoustic wave (top), the spectrogram (middle), and the F_0 contour (bottom).



each stimulus set. Moreover, results were similar across all 8 speaker groups, demonstrating the robustness of the findings.

Effect of Removing Particular Prosodic Cues on Listener Accuracy

F_0 and duration cues are prominent components of prosody and play an important role in normal speech (Eady & Cooper, 1986; Fry, 1958; Morton & Jassem, 1965; Shattuck-Hufnagel & Turk, 1996). Given this, F_0 and syllable duration cues were manipulated to determine their importance in dysarthric speech. The expectation was that classification accuracy would deteriorate if the prosodic cue removed was salient for differentiating questions from statements. Results indicate that listeners found the F_0 contour to be a far more informative cue than syllable duration for classifying dysarthric utterances. Flattening the F_0 contour reduced classification accuracy from SS_{orig} to SS_p by 32%—from 87% to 55% ($p < 0.0001$). Removal of durational cues on the other hand only reduced accuracy from SS_{orig} to SS_d by 0.8% ($p < 0.0001$). Although this change in accuracy was statistically significant, duration cues were clearly not as important to listeners as F_0. Removal of F_0 and duration cues together reduced accuracy from SS_{orig} to SS_{pd} by 33% ($p < 0.0001$). We were unable to verify an additive effect of both flattening F_0 and equalizing syllable duration, because the difference in accuracy between SS_p and SS_{pd} was not statistically significant (p = 0.3332).

The effect of duration manipulation may have been diluted in part by the methodology used to equalize syllable duration. Recall that in the equalized syllable duration condition, all 6 tokens of a particular phrase type (3 question tokens and 3 statement tokens) were stretched or compressed to the average duration of each syllable across all 6 tokens. Any existing differences between questions and statements in terms of absolute duration were thus eliminated by this manipulation. For example, if questions were longer than statements in general, that information was removed from the signal.

Similarly for the F_0 manipulations, contours were flattened to the average F_0 of all 6 tokens of a given phrase type. Observations of the data indicated that questions appeared to have an overall higher F_0 than statements. Once again, this overall cue of heightened F_0 for questions may have been an additional information-bearing cue that was inadvertently eliminated.

Similar Trends in Deterioration of Listener Classification Across Speaker Groups and Phrases

There were no statistically significant differences among listeners within speaker groups that were due to manipulation. That is, the effect of manipulation on performance was similar for all listeners within each speaker group. Variance in accuracy among the 6 listeners in each speaker group was, in fact, very low.

There were no phrases that were found to be particularly difficult for listeners in all 8 speaker groups. Some phrases were more difficult for listeners in some speaker groups than others. For example, the phrase "play it loud" (PIL) was problematic for listeners in S7's group, yet easy for listener groups of S2, S4, S5, S6, and S8. Results from a previous pilot study (Patel & O'Keefe, 2000) with a different list of target phrases were also similar. Although some phrases were more difficult, no specific sound sequence patterns within the chosen phrases were found to account for these differences. Nonetheless, the phonetic context of some phrases influenced the production of prosodic contrasts and in turn influenced the listeners' abilities to categorize those phrases. These findings highlight the importance of phonetic context and the interplay between prosodic and phonetic features in dysarthric speech (Yorkston, Beukelman, & Bell, 1988).

Although the overall trend in degradation of classification accuracy due to removal of F_0 and duration cues was found in all 8 speaker groups, listeners performance across different speaker groups varied. The adverse effect of equalizing syllable duration was stronger for listeners of S2 and S5 than for other speakers. In addition, the effect of flattening F_0 did not affect listeners of S3 and S7 to the same extent as listeners of other speakers. In the post-experimental questionnaire, these listeners reported that they were using loudness cues to differentiate questions and statements. It appears that speakers employed different strategies for marking the statement-question contrast.

These finding support the observations of Yorkston et al. (1984), Vance (1994), and Brewster (1989) that speakers with dysarthria may remap control of specific prosodic contrasts onto other prosodic features to optimize communication. For example, if a speaker cannot raise pitch, he or she may substitute pitch rises with increased loudness or syllable lengthening in order to signal a given contrast.

In nonimpaired speakers, a similar transfer of informational cues between prosodic features has been referred to as *cue trading* (Howell, 1993; Lieberman, 1960). The notion is that certain cue combinations are equivalent to other cue combinations and can thus be used interchangeably. Thus, it seems that speakers can vary the relative salience of prosodic markers within equivalent cue combinations to communicate the perception of stress. In fact, Howell (1993) states that different speakers mark stress using a different order of acoustic features (i.e., some may use F_0, others may use loudness as a prominent feature, etc.). In addition, listeners are able to tune to the speakers' idiosyncratic stress patterning even if it differs from their own methods of signaling stress. Cue trading phenomena seem to be present in our severely dysarthric recordings, similar to that exhibited in normal and mildly dysarthric speakers. Somewhat surprisingly, naïve listeners are able to leverage this phenomenon in spite of severe dysarthria.

Implications

Identification of consistent cues within dysarthric utterances is an essential first step toward using these consistencies for the purpose of communication. Treatment programs could be developed to use prosodic contrasts for signaling phonemic differences, for distinguishing words in a way that is analogous to the role of F_0 in tonal languages. Specific prosodic features could be targeted within therapy to improve consistency of productions. For example, if the speaker could reliably mark a contrastive F_0 contour, then he or she would be able to increase his or her vocalization repertoire by superimposing the contour on different messages. Our results also hint that loudness may be an additional information-bearing prosodic cue for some speakers. If speakers could independently control multiple prosodic parameters, they could increase their vocalization repertoires by exploiting the combinatorial space of parameters.

Yorkston et al. (1988) have argued that prosodic and phonetic parameters are intertwined in their effect on speech intelligibility. Flattening F_0 contours has been shown to reduce speech intelligibility for normal speech (Laures & Weismer, 1999; Wingfield, Lombardi, & Sokol, 1984) and dysarthric speech (Bunton, Weismer, & Kent, 2000), suggesting that F_0 cues may be useful for signaling segmental as well as prosodic information. If this is true, then targeting prosodic aspects in therapy may also improve overall speech intelligibility. Speech therapy with dysarthric speakers typically focuses on segmental aspects to improve intelligibility. Exaggerated articulation is often used to improve phonetic contrasts. Similarly, the results of this investigation indicate that perhaps teaching exaggerated prosody may be an effective strategy for achieving improved prosodic contrasts and, in turn, improving overall intelligibility. Training prosody may also improve the naturalness of dysarthric vocalizations, and thus speakers may be more inclined to use vocalizations as a communication mode with unfamiliar communication partners.

Future Directions

There are several caveats that must be considered when interpreting the results of this investigation. The number of syllables in the target phrases may have influenced our results. For example, Grant and Walden (1996) and Le Dorze et al. (1994) used phrases that were 8–11 and 5–7 syllables in length, respectively, in their studies of question/statement contrasts. In our investigation, all target phrases were only 3 syllables long. Using shorter phrases may have enabled speakers to more easily mark differences between questions and statements. Longer phrases may tax the production system more than shorter phrases. Given that reduced breath support for speech is a concern within this population, increasing the target-phrase length may cause the production system to have more difficulty marking the contrast. Future studies could explore interactions between prosodic control and phrase length. It will also be important to further explore how prosodic control varies with changes in the phonetic context of target phrases, grammatical categories of words in the phrases, and number of contrastive stress categories being signaled.

Physiological and acoustic studies are needed to corroborate our perceptual findings. These studies can help determine whether speakers raise pitch throughout the utterance, marking particular words within the utterance with rising intonation and/or whether they raise pitch and then drop it toward the end of the utterance. Similarly for duration, acoustic analyses may shed light on whether speakers were lengthening the entire utterance, lengthening particular syllables, and/or altering the number of syllables. These investigations are currently underway and the results will follow.

Conclusions

To date, the capacity of individuals with severe dysarthria to convey information through prosody has not been investigated. We have demonstrated that 8 speakers with severe dysarthria due to cerebral palsy were able to mark the question-statement contrast such that listeners were approximately 87% accurate in classifying the vocalizations. Although we have only begun to learn about how speakers with dysarthria may be marking the difference between questions and statements the information seems, at least partially, to be encoded in F_0 contour and to a lesser extent in syllable duration. In contrast to previous studies that underscore the diminished control of prosody available to individuals with severe dysarthria, we have demonstrated that speakers are nonetheless able to exert sufficient control to communicate intentions.

Acknowledgments

This research was conducted in the Department of Speech-Language Pathology at the University of Toronto toward partial fulfillment of the requirements for my doctoral dissertation, *Identifying Information-Bearing Prosodic Parameters in Severely Dysarthric Vocalizations, 2000.* I am grateful to Bernard O'Keefe for his mentorship, support, and guidance. I would also like to thank Hans Kunov, Christopher Dromey, Luc De Nil, and Deb Roy for their contributions to this work. I am grateful to John H. Saxman and Stephanie Shattuck-Hufnagel for reading the manuscript and offering invaluable suggestions. Last, this study would not have been possible without the involvement of the participants and support from the National Science and Engineering Research Council of Canada, Bloorview Childrens Hospital Foundation, Whipper Watson Foundation, Donalda MacGeachy Foundation, and the Harmonize of Speech Foundation of Ontario.

References

Allaire, J. H., Gressard, R. P., Blackman, J. A., & Hostler, S. L. (1991). Children with severe speech impairment: Caregiver survey of AAC use. *Augmentative and Alternative Communication, 7,* 248–255.

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.

Beliveau, C., Hodge, M. M., & Hagler, P. H. (1995). Effects of supplemental linguistic cues on the intelligibility of severely dysarthric speakers. *Augmentative and Alternative Communication, 11,* 176–186.

Beukelman, D. R., & Mirenda, P. (1992). *Augmentative and alternative communication: Management of severe communication disorders in children and adults.* Baltimore: Paul H. Brookes.

Boersma, P., & Weenik, D. (2000). *Praat, A system for doing phonetics by computer, version 3.4* (Technical Report 132). Institute of Phonetic Sciences of the University of Amsterdam. www.praat.org.

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

Bunton, K., Weismer, G., & Kent, R. (2000). *The effects of flattened fundamental frequency contours on sentence intelligibility in dysarthric speakers.* Conference on Motor Speech: Motor Speech Disorders and Speech Motor Control. San Antonio, TX.

Corey, P. (1998). *Analysis of variance.* Unpublished manuscript, Department of Community Health and Living, University of Toronto, Toronto, Canada.

Darley, F. L., Aronson, A. E., & Brown, J. R. (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: W. B. Saunders Co.

Dongilli, P. A., Jr. (1993). *Semantic context and speech intelligibility.* Doctoral dissertation, University of Nebraska, Lincoln.

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

Ferrier, L. J. (1991). Clinical study of a dysarthric adult using a Touch Talker with words strategy. *Augmentative and Alternative Communication, 7,* 266–274.

Fried-Oken, M. (1985). Voice recognition device as a computer interface for motor and speech impaired people. *Archives of Physical Medicine and Rehabilitation, 66,* 678–681.

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

Grant, K. W., & Walden, B. E. (1996). Spectral distribution of prosodic information. *Journal of Speech and Hearing Research, 39,* 228–238.

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

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

Hustad, K. C. (1999). *Effects of context on intelligibility and comprehensibility of severely dysarthric speech.* Doctoral dissertation. University of Nebraska, Lincoln.

Irwin, J. (1955). Phonetic equipment of spastic and athetoid children. *Journal of Speech and Hearing Disorders, 20,* 54–67.

Kent, R. D., & Read, C. (1992). *The acoustic analysis of speech.* San Diego, CA: Singular Publishing Group.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences.* Belmont, CA: Brooks/Cole.

Laures, J. S., & Weismer, G. (1999). The effects of a flattened fundamental frequency contour on intelligibility at the sentence level. *Journal of Speech, Language, and Hearing Research, 42,* 1148–1156.

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

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

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

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

McGarr, N. S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research, 26,* 451–458.

Monsen, R. B. (1983). The oral speech intelligibility of hearing impaired talkers. *Journal of Speech and Hearing Disorders, 48,* 286–296.

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

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9,* 453–467.

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.

Pagano, M., & Gauvreau, K. (1993). *Principles of biostatistics.* Belmont, CA: Duxbury Press.

Patel, R. (1998). Acoustic analysis of vocalizations produced by individuals with severe dysarthria. *Proceedings of the International Society of Alternative and Augmentative Communication Biennial Conference, Dublin, Ireland,* 489–491.

Patel, R., & O'Keefe, B. (2000). *Prosodic control in severe dysarthria: Preliminary results.* Conference on Motor Speech: Motor Speech Disorders and Speech Motor Control. San Antonio, TX.

Platt, L., Andrews, G., & Howie, F. M. (1980). Dysarthria of adult cerebral palsy II. Phonemic analysis of articulation errors. *Journal of Speech and Hearing Research, 23,* 41–55.

Platt, L., Andrews, G., Young, M., & Quinn, P. (1980). Dysarthria of adult cerebral palsy: I. Intelligibility and articulatory impairment. *Journal of Speech & Hearing Research, 23,* 28–40.

Robertson, S. J., Tanner, B., & Young, F. (1989). *The dysarthria sourcebook: Exercises to photocopy* (pp. 59–60 and 110–115). Oxon, U.K.: Winslow Press.

Rosenbek, J., & LaPointe, 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 Co.

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

Shein, F., Brownlow, N., Treviranus, J., & Parnes, P. (1990). Climbing out of the rut: The future of interface technology. In B. A. Mineo (Ed.), *1990 ASEL Applied Science and Engineering Laboratories: Augmentative and alternative communication in the next decade* (pp. 17–19). Wilmington: University of Delaware Press.

Smith, M. M. (1994). Speech by any other name: The role of communication aids in interaction. *European Journal of Disorders of Communication, 29,* 225–240.

Streeter, L., Macdonald, N., Apple, W., Krauss, R., & Galotti, K. (1983). Acoustic and perceptual indicators of emotional stress. *Journal of the Acoustical Society of America, 73,* 1354–1360.

Swigert, N. B. (1997). *The source for dysarthria* (pp. 210–235). East Moline, IL: LinguiSystems.

Vance, J. E. (1994). Prosodic deviation in dysarthria: A case study. *European Journal of Disorders of Communication, 29*(1), 61–76.

Williams, C., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America, 52,* 1238–1250.

Wingfield, A., Lombardi, L., & Sokol, S. (1984). Prosodic features and the intelligibility of accelerated speech: Syntactic versus periodic segmentation. *Journal of Speech and Hearing Research, 27,* 128–134.

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

Yorkston, K. M., Beukelman, D. R., & Bell, K. R. (1988). *Clinical management 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: ProEd.

Yorkston, K. M., Hammen, V. L., & Dowden, P. (1991). Index of contextual intelligibility: A perceptual analysis of intelligible versus unintelligible productions in severe dysarthria. In C. A. Moore, K. M. Yorkston, & D. R. Beukelman (Eds.), *Dysarthria and apraxia of speech: Perspectives on management* (pp. 55–64). Baltimore: Paul Brooks Publishing.

Received September 21, 2001

Accepted May 7, 2002

DOI: 10.1044/1092-4388(2002/069)

Contact author: Rupal Patel, PhD, Department of Biobehavioral Sciences, Teachers College Columbia University, 525 West 120th Street, Box 180, New York, NY 10027. E-mail: rp564@columbia.edu

Appendix. Target Phrases.

Phrase	Abbreviation	Statement context	Question context
She was here (?)	SWH	Bob asks if you have seen Jane. You say…	John tells you that the Queen of England visited your house. You ask…
It was hot (?)	IWH	You just got back from holidays in Florida. Jane asks if the weather was nice. You say…	Jane says her vacation to Alaska was too hot. You ask…
She loves dogs (?)	SLD	Molly asks if Sue likes dogs. You say…	Jane doesn't like animals. Someone tells you she loves dogs. You ask…
Play it loud (?)	PIL	You can't hear the message on your answering machine. You tell Jane to …	It's midnight and someone asks you to turn up the volume on the stereo. You ask…
Use some soap (?)	USS	You ask Jenny to get a out a stain from your shirt. She asks how. You say…	You run out of shampoo while washing your hair. Your attendant tells you to just use soap. You ask…
She said no (?)	SSN	Bill proposed to Sue but she refused. Someone asks what she said. You say…	John offered Beth a million dollars. She declined. You ask…
It's Thursday (?)	ITD	Someone asks what day it is. You say…	Someone tells you it's Thursday on the weekend. You ask…
He lives there (?)	HLT	Doug asks you where Chris lives. You point to the blue house and say…	Chris is a millionaire. Paul says he lives in a small run down apartment downtown. You ask…
Pass some salt (?)	PSS	Your food is a little bland. You say…	While eating desert Jim asks you for the salt. You ask…
Give it back (?)	GIB	Someone grabs your pen. You say…	Someone gives you a gift and then asks for it back. You ask…