Lexical boundary error analysis in hypokinetic and ataxic dysarthria

Julie M. Liss\textsuperscript{a)} and Stephanie M. Spitzer
Motor Speech Disorders Laboratory, Arizona State University, Box 871908, Tempe, Arizona 85281

John N. Caviness, Charles Adler, and Brian W. Edwards\textsuperscript{b)}
Department of Neurology, Mayo Clinic–Scottsdale, 13400 Shea Boulevard, Scottsdale, Arizona 85259

(Received 21 April 1999; revised 15 October 1999; accepted 1 March 2000)

This investigation is the second in a series to examine a potential source of reduced intelligibility in dysarthric speech, namely the mismatch between listeners’ perceptual strategies and the acoustic information available in the dysarthric speech signal. Lexical boundary error (LBE) analysis was conducted on listener transcripts from phrases produced by speakers with hypokinetic dysarthria, ataxic dysarthria, and normal controls. By design, the hypokinetic and ataxic dysarthric tapes elicited similar intelligibility (words-correct) scores. However, they elicited different numbers and patterns of lexical boundary errors. The nature of the error pattern differences can be traced to the listeners’ use of available syllabic strength information to segment the acoustic stream. Specifically, although both dysarthric speech samples elicited numerous lexical boundary errors, those for the hypokinetic speech generally conformed to predictions offered from studies of degraded normal speech. Those for the ataxic speech did not conform strongly to such predictions. It appears that the prosodic deficits of the ataxic speech (tendency toward syllabic isochrony, excessive loudness variation, and reduced vowel working space consequent to reductions in vowel strength) posed more of a problem for listeners than did the prosodic deficits of the hypokinetic speech (rapid rate, monotony, reduced vowel working space). © 2000 Acoustical Society of America.

PACS numbers: 43.70.Dn, 43.70.Fq, 43.71.Gv, 43.71.Bp [DOS]

INTRODUCTION

Though the construct of speech intelligibility is fundamental to both clinical and research endeavors in dysarthria, surprisingly little is known about the perceptual processes that underlie it. For example, it is not known which perceptual strategies listeners invoke to decipher the degraded acoustic signal of dysarthria, or the extent to which speech signal characteristics might influence the choice or the effectiveness of these perceptual strategies. Answers to these questions will bear on both theories of speech perception and the clinical use of speech intelligibility measures.

This paper is the second in a series that examines the interface between the dysarthric speech signal characteristics and the listener’s response to that signal in an effort to identify a source for reductions in intelligibility. To explore this interface, it was necessary to adopt a paradigm that accommodated both listener and speaker variables in a meaningful and interpretable way (see Lindblom, 1990; see also Connolly, 1986; Weismer and Martin, 1992). Toward this end, we chose to examine the perceptual task of segmenting the continuous acoustic stream into its component words. Lexical segmentation was selected for several reasons, including the primacy of the task for lexical access, the availability of a working model for perceptual strategies to accomplish the task, and the ability of the working model to accommodate dysarthric speech. Each of these is addressed in turn.

By most theoretical accounts, identification of word boundaries in connected speech triggers lexical search (e.g., Cole and Jakimik, 1980; Forster, 1989; Goldinger et al., 1989; Gow and Gordon, 1995; Marslen-Wilson, 1989). If speech intelligibility is defined as a listener’s success in deciphering the words spoken, word-boundary identification is an obvious and critical step. A large number of investigations have examined acoustic cues that may be involved in the prelexical processing of speech input (Cutler and Butterfield, 1990, 1991, 1992; Cutler and Norris, 1988; Gaygen and Luce, 1998; Grosjean and Gee, 1987; Gow and Gordon, 1995; Klatt, 1980; Lehiste, 1972; Mattys and Samuel, 1997; Nakatani and Schaffer, 1978; Quené, 1992; Vitevitch and Luce, 1998). This research has examined both segmental and suprasegmental cues produced by normal speakers that may be involved in the signaling of word boundaries to listeners.

One contemporary view is that listeners rely on utterance prosody—in particular, the juxtaposition of strong and weak syllables—to guide lexical segmentation decisions [metrical segmentation strategy hypothesis (MSS); Cutler and Norris, 1998; Cutler and Butterfield, 1992]. According to this model, strong syllables are those that contain full vowels and that may or may not receive prosodic stress. Weak syllables contain reduced vowels and do not receive prosodic stress (Cutler and Butterfield, 1990, 1991; Cutler and Carter, 1987; Fear et al., 1995; Smith et al., 1989). Central to the MSS hypothesis is the assumption that, in English, segmentation of the speech signal is activated by the occurrence of a strong syllable. This hypothesis is supported by research that has examined patterns of naturally occurring “slips of the
and by lexical boundary errors elicited by the presentation of speech at very low listening levels, just above threshold (Cutler and Butterfield, 1992). Specifically, when listeners recognize strong syllables as word onsets, they will be more likely to erroneously insert lexical boundaries before strong than before weak syllables. They also will be more likely to erroneously delete lexical boundaries before weak than before strong syllables. Although this model does not claim to completely explain the cognitive processes involved in lexical segmentation, it does provide a theoretical framework upon which to base hypotheses and interpret results.

Because the MSS hypothesis emphasizes prosodic contrastivity, it is an attractive model for the investigation of intelligibility in dysarthria. All varieties of dysarthria affect the prosody of connected speech in one way or another (Darley et al., 1969; Duffy, 1995). If prosodic cues play a crucial role in signaling word boundary location in normal speech perception, their disturbance in dysarthric speech likely has a negative impact on lexical segmentation. By examining the precise nature of the prosodic disturbances relative to the patterns of errors they elicit, we obtain information about the source of intelligibility decrement. This information is then dysarthria- (or, more accurately, speech pattern-) specific.

In our previous study, we examined perceptual strategies and hypokinetic dysarthric speech within the context of the MSS hypothesis (Liss et al., 1998). Hypokinetic dysarthria was chosen as an entry point for this line of research because it was believed that the cardinal speech features, by their very nature, serve to diminish syllabic strength contrastivity. Although the precise acoustic correlates of syllabic strength have yet to be determined, there is evidence that they lie in the domains of relative syllable duration, intersyllabic pitch and loudness differences, and vowel quality or vowel strength (Cutler and Butterfield, 1990, 1991, 1992; Fear et al., 1995; Halle and Keyser, 1971; Klatt, 1980; Lehiste, 1972; Nakatani and Schaffer, 1978; Quéné and Koster, 1998; van Ooijen et al., 1997). Thus, the hypokinetic dysarthric speech features of rapid speaking rate, a tendency toward monotony and monoloudness, and phoneme imprecision (Ackermann and Ziegler, 1991; Adams, 1991; Darley et al., 1969; Duffy, 1995; Forrest et al., 1989; Logemann and Fisher, 1981; Ludlow and Bassich, 1984; Ramig, 1992; Weismer, 1984, 1991) should reduce syllabic strength contrastivity.

We expected to see that, if listeners rely on syllabic strength contrasts to identify word boundaries in hypokinetic speech, the LBE patterns should reveal reduced effectiveness of this strategy. Instead, we found that the general patterns of LBEs were identical to those predicted from studies of degraded normal speech (Cutler, 1993; Cutler and Butterfield, 1992; Cutler and Norris, 1988). Even though the phrases produced by these speakers contained reduced syllabic strength contrasts according to both perceptual and acoustic evaluation, the information was apparently sufficiently robust to be utilized by the listeners. However, two important points were of note. First, the strength of adherence to the predicted pattern of LBEs across all speech samples was less than that reported in the literature for normal degraded speech. Second, speech samples with the most severely degraded strong–weak contrasts (according to perceptual and acoustic indices) elicited the lowest strength of adherence values. Thus, this investigation provided evidence that listeners rely on acoustic information about syllabic strength to locate word boundaries in hypokinetic dysarthric speech, but that listeners have difficulty applying the strategy as syllabic strength information is increasingly degraded.

Having established that the severity of syllabic strength degradation is a source of reduced intelligibility in hypokinetic dysarthria, the present study posed the following questions: Does the form of the dysarthria have a definable effect on listeners’ abilities to apply the Metrical Segmentation Strategy? For comparison, we selected ataxic dysarthria because it differs antithetically from hypokinetic dysarthria in its pattern of prosodic deficit. Ataxic dysarthria is characterized by a slow speaking rate with excess and equal stress, and excessive loudness variation, with irregular articulatory breakdown and articulatory impression (Ackermann and Hertrich, 1994; Darley et al., 1969; Duffy, 1995; Kent et al., 1979). Our goal was to select two groups of speakers, hypokinetic and ataxic dysarthrics, whose speech samples did not differ in terms of intelligibility scores, but whose speech samples exhibited very specific and different forms of prosodic deficit. In this way, when listeners mis-segment the acoustic stream, the perceptual errors could be interpreted relative to specific production characteristics rather than to fundamental differences in intelligibility.

Thus, both dysarthric groups exhibited reductions in syllabic strength contrastivity but for different reasons. The question was whether these two speech patterns would show evidence of posing different challenges to the listeners’ task of perceptual segmentation. If this is found to be the case, we will have identified different sources of intelligibility reduction for these two speech pattern types.

Based on our previous study (Liss et al., 1998), we expected that listeners would have difficulty identifying lexical boundaries in the hypokinetic speech, but that they would make use of the available syllabic strength information to venture guesses. This would be apparent in two ways: the presence of a large number of LBEs, and a LBE pattern that generally conformed to that predicted by the MSS hypothesis. We also expected that because syllabic strength contrastivity is reduced in these phrases, the strength of adherence to the predicted pattern would be less than that reported in the literature for degraded normal speech. If these results were found, they would replicate those of our previous investigation and strengthen the validity of those original findings.

With regard to the ataxic speech, we expected that listeners would have difficulty identifying lexical boundaries in the phrases—again, because of the presumed decrement in syllabic strength contrasts. This would manifest as a large number of LBEs. However, it was not known the extent to which the nature of the prosodic deficit would affect the systematic application of the metrical segmentation strategy, nor whether listeners would show evidence of using syllabic strength contrasts to parse the phrases. Within our sample of phrases, the opportunity to produce lexical boundary dele-
tions and insertions before strong and weak syllables was exactly opposite the MSS prediction pattern (see Liss et al., 1998; Table III). Thus, if the LBE pattern conformed to the predicted pattern and did so strongly, it could be surmised that the syllabic strength information was sufficient for the effective use of this perceptual strategy. If the LBE pattern conformed weakly, it would signify a mismatch between the acoustic information and the perceptual strategy. In other words, it would indicate that listeners are relying on inferior or insufficient acoustic information to conduct the task of lexical segmentation. If the LBE pattern matched the opportunity for certain errors to occur (i.e., more deletion than insertion errors; more insertions before weak than strong syllables; more deletions before strong than weak syllables), it could be interpreted either as the failure of the perceptual strategy and/or use of an alternate perceptual strategy that did not capitalize on syllabic strength information.

I. METHOD

A. Listeners

The 60 listeners were 30 men and 30 women who ranged in age from 18–44 years. Most were undergraduate students at Arizona State University who were compensated for their participation in this study. All listeners self-reported normal hearing, were native speakers of Standard American English, and reported having little or no experience listening to dysarthric speech. The listeners were quasi-randomly assigned to one of the listening groups (Control, Hypokinetic, and Ataxic) such that each group contained ten men and ten women.

B. Speech stimuli

Speech stimuli consisted of three audiotapes of 60 phrases produced by three groups of speakers: six speakers with hypokinetic dysarthria, six with ataxic dysarthria, and six neurologically normal control speakers. It was critical to the success of this investigation that two conditions were met in terms of the two dysarthria tapes. First, LBEs must be interpretable relative to syllabic strength. This required careful construction of the stimulus phrases themselves. Second, the two dysarthria tapes must be of equivalent intelligibility (as measured by words-correct scores), and all of the phrases on each tape must be representative of the operational definitions of the respective dysarthria. This allowed differences in the dependent variables to be interpreted relative to differences in speech production characteristics as they relate to syllabic strength contrasts. Toward this end, speakers and phrase tokens were selected through a series of steps.

The 60 phrases, modeled after Cutler and Butterfield (1992), were designed to permit the interpretation of LBEs. The phrases themselves were of low interword predictability to reduce the contribution of semantic information to word perception. They consisted of six syllables that alternated in phrasal stress patterns. Half of the phrases alternated strong–weak (SWSWSW), and the other half alternated weak–strong (WSWSWS). The majority of the strong and weak syllables contained full and reduced vowels, respectively. The phrases ranged in length from three to five words and no word contained more than two syllables. None of the words in the phrases was repeated except articles and auxiliary verbs; all English phonemes except /zh/ were represented.

To meet the requirements of similar within-group speech characteristics and equivalent between-group phrase intelligibility, more speakers were recruited and more phrases were recorded than ultimately were used. Forty-seven hypokinetic and 48 ataxic dysarthric speakers were identified by their neurologists and speech pathologists as potential subjects for this investigation. Following an initial telephone conversation with the first author, 13 hypokinetic and 12 ataxic speakers were invited to participate based on their moderate to severe levels of intelligibility deficit and predominant speech characteristics. Our operational definition of hypokinetic dysarthria was as follows: perceptually rapid speaking rate with monopitch and monoloudness; little use of variation in pitch or loudness to achieve differential syllabic stress; imprecise articulation that gives the impression of a blurring of phonemes and syllables; and a breathy and perhaps hoarse/harsh voice. Ataxic dysarthria was defined as perceptually slow speaking rate with a tendency toward equal and even syllable duration (scanning speech); excessive loudness variation; and irregular articulatory breakdown. Of the 25 speakers who provided samples, five hypokinetic and six ataxic speakers ultimately were not used because of additional speech characteristics either not present or not noted during the initial telephone screening. These characteristics included the presence of a distinctive regional accent, a pervasive vocal tremor and oral dyskinesia, insufficient intelligibility, and the absence of one or more components of our operational definitions of dysarthria subtype. Thus, speech samples from eight hypokinetic dysarthrics, six ataxic dysarthrics, and eight neurologically normal controls were regarded as potential candidates for construction of the listening tapes.

Speech samples were collected during a single hour-long session with each speaker. The details of the speech protocol are reported in Liss et al. (1998). Briefly, each speaker produced words and sentences for the Assessment of Intelligibility of Dysarthric Speech (AIDS: Yorkston and Beukelman, 1981), a reading passage, vowel productions, several minutes of spontaneous speech, and the set of stimulus phrases. Each speaker typically produced three or four iterations of each stimulus phrase during the course of speech sample collection. The first token, which contained no word omissions, substitutions, dysfluences, or interword pauses, and which most closely represented our operational definitions, was selected as the experimental token. The 60 phrases per speaker were low-pass filtered at 10 kHz, digitized at a 22-kHz sampling rate, and stored in a computer file using CSpeech Laboratory Automation System (Milenkovic and Read, 1992). Prior to downloading onto DAT audiocassettes for use in the perceptual experiment, the files were subjected to a customized gain program (MATLAB) to equilibrate mean rms energy across the phrases.

The next step was to identify the 18 speakers whose phrases would comprise each of the three listening tapes, and to select the 60 phrase productions (ten per speaker) for each tape. The first and second authors used information provided
by eight certified speech language pathologists,\textsuperscript{3} and the results of the previous study on hypokinetic dysarthria (Liss \textit{et al.}, 1998) to accomplish this task. First, the six hypokinetic speakers whose intelligibility levels most closely matched the six ataxic speech samples were identified. Second, ten phrases from each speaker were selected such that all 60 phrases were represented for each dysarthria tape. Although the goal of phrase selection was only to create hypokinetic and ataxic tapes with similar overall intelligibility distributions, it was possible to select phrase pairs with similar intelligibility levels for the vast majority of phrases. Thus, two dysarthria tapes were constructed—and there was every indication from the preliminary steps that the two would elicit similar words-correct scores from the listeners in the present study.\textsuperscript{4}

Finally, three men and three women whose age range was most similar to that of the 12 dysarthric speakers were selected from among the eight control speakers. Because the intelligibility of their phrases ranged from 98\%–100\%, phrase selection for the listening tape was straightforward. Ten phrases were selected from each of the six speakers in the quasi-random fashion, such that no speaker provided consecutive phrases.

The three audiotapes each contained one production of the 60 phrases, ten phrases for each of the six speakers in each group. Phrases were preceded by the phrase number (1–60) spoken by a neurologically normal female, and followed with a 12-s interstimulus silent interval. The subjective recording quality was judged to be high and signal intensity consistent across speakers and phrases.

### 1. Perceptual-acoustic characterization of phrases

As noted earlier, the precise acoustic correlates of syllabic strength contrastivity have yet to be determined but are thought to lie in the domains of vowel quality, relative syllable duration, and pitch and loudness contours. Perceptually, all of the phrases produced by the dysarthric speakers were characterized by reductions in syllabic contrastivity, as compared with those produced by the control group. However, the source of this reduced contrastivity differed between the two dysarthria subtypes, as per the operational definitions herein. Syllabic contrastivity reductions in the hypokinetic phrases were thought to arise from the perceptually rapid rate, blurring of syllabic boundaries, and monotony. The ataxic phrases were characterized by a tendency toward equal and even syllable duration, and a labored, slow rate of speech. The acoustic correlates of these perceived reductions were examined for all of the stimulus phrases. Intraphrase measures included duration (phrase, syllable, and vowel), and fundamental frequency and amplitude variation. Intragroup measures of strong vowel formant frequencies ($F_1$–$F_2$) were plotted to determine the “outer limits” of the vowel working space for each set of phrases. By inference, smaller working spaces may reflect reduced capacity for strong–weak distinctions because the strong vowels are relatively reduced.\textsuperscript{5}

### 2. Duration

Phrase duration was obtained during the initial editing of the phrases by placing cursors on the first and last acoustic evidence of phonemes on the spectrographic display. This included the first or last glottal pulse in the case of initial or final voiced phonemes, respectively; the beginning or end of noise energy in the case of initial or final fricatives; and the beginning or end of the burst release in the case of initial or final stop consonants. One-hundred milliseconds of silence was then appended to the beginning and end of each phrase to reduce onset–offset effects and the entire screen was saved as a digital file for all subsequent acoustic analysis.

Syllable and vowel durations were measured using the cursor function of CSpeech on a spectrographic display. To promote uniform measurement criteria and high measurement reliability, a hardcopy of each phrase as produced by the control speaker was segmented according to traditional phoneme landmarks (Crystal and House, 1998a, b; Weismer, 1984), and operational definitions of each segment were created by the first author. The hardcopy and definitions were used as a guide for the on-screen segmentation of the other control and dysarthria phrases. The first author made all measurements, and approximately 20\% of the dysarthric phrases were resampled in a quasi-random fashion by the first author and one other judge for reliability purposes. Reliability measures for the control phrases were not undertaken because of the use of the hardcopy reference. Data for reliability of vowel and syllable measures were collapsed. The mean interjudge difference was 44 ms (range, 0–123 ms), and intrajudge differences ranged from 0–112 ms (mean, 31 ms).

<table>
<thead>
<tr>
<th>Speaker group</th>
<th>Median phrase duration\textsuperscript{a}</th>
<th>Median strong syllable duration\textsuperscript{b}</th>
<th>Median weak syllable duration\textsuperscript{c}</th>
<th>Median strong-weak vowel duration ratio\textsuperscript{d}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>1681.5</td>
<td>339.0</td>
<td>171.5</td>
<td>1.88</td>
</tr>
<tr>
<td>Hypokinetic</td>
<td>1192.5</td>
<td>231.5</td>
<td>121.0</td>
<td>1.86</td>
</tr>
<tr>
<td>Ataxic</td>
<td>2253.5</td>
<td>462.0</td>
<td>272.0</td>
<td>1.48</td>
</tr>
</tbody>
</table>

\textsuperscript{a}$H(2) = 135.9, P < 0.0001$, all comparisons significant $P < 0.05$.

\textsuperscript{b}$H(2) = 193.4, P < 0.0001$, all comparisons significant $P < 0.05$.

\textsuperscript{c}$H(2) = 141.2, P = 0.0001$, all comparisons significant $P < 0.05$.

\textsuperscript{d}$H(2) = 7.98, P = 0.0185$, control and hypokinetic versus ataxic significant $P < 0.05$. The second and third are based on syllable duration measures. The ratios presented in the final column are based on intraphrase strong-weak vowel duration contrasts, not on values in the previous data columns.
The first three data columns of Table I contain the median phrase, strong syllable, and weak syllable durations for each of the speaker groups. In all cases, the duration measures of the hypokinetic phrases were significantly shorter and the ataxic phrases significantly longer than those of the other two groups.

An index of local durational contrastivity between vowels in strong and weak syllables was computed. For each phrase, the duration of each strong vowel was divided by the duration of its adjacent weak vowel. To avoid the artifacts of final syllable lengthening (Oller, 1973; Klatz, 1975), vowels located in final syllables were not included. Therefore, each phrase yielded four S/W ratios and a mean. The median values for each group are presented in the final data column of Table I. A Kruskal–Wallis one way ANOVA on ranks and post hoc Student–Newman–Keuls revealed that the ratio of the ataxic phrases was significantly lower than that of the other two groups. Although the ratio was slightly lower for the hypokinetic than control phrases, this did not reach significance.

To summarize, the phrase, syllable, and vowel duration measures support the perception of rapid hypokinetic and slow ataxic speaking rate relative to the controls. The perception of equal and even syllable duration in the ataxic phrases as compared to those of the control and hypokinetic phrases was supported by the strong-to-weak vowel duration ratios.

### 3. Vowel quality

First and second formant frequencies were measured at the temporal midpoints of seven occurrences of the vowels /i/, /æ/, /a/, and /u/ embedded in strong syllables (see Liss et al., 1998 for word contexts). Measurements were made using both broadband spectrograms and LPC displays. Hard-copy visuals and operational definitions also were created on copy visuals and operational definitions also were created on final syllables were not included. Therefore, each speaker was located in final syllables were not included. Therefore, each phrase yielded four S/W ratios and a mean. The median values for each group are presented in the final data column of Table I. A Kruskal–Wallis one way ANOVA on ranks and post hoc Student–Newman–Keuls revealed that the ratio of the ataxic phrases was significantly lower than that of the other two groups. Although the ratio was slightly lower for the hypokinetic than control phrases, this did not reach significance.

To summarize, the phrase, syllable, and vowel duration measures support the perception of rapid hypokinetic and slow ataxic speaking rate relative to the controls. The perception of equal and even syllable duration in the ataxic phrases as compared to those of the control and hypokinetic phrases was supported by the strong-to-weak vowel duration ratios.

### 4. F0 and amplitude variation

Fundamental frequency (F0) and its variation within each digitized phrase was computed automatically using the CSpeech short-term autocorrelation function with center clipping. All pitch traces were inspected visually to identify and edit tracking errors, which occurred occasionally in the dysarthric phrases. The rms amplitude envelope of each phrase was converted automatically to mean decibels and variation around the mean was calculated. To compare mean intraphrase F0 and amplitude variations across phrases and speaker groups, coefficients of variation were calculated by dividing each standard deviation by its mean. These results, along with F0 ranges, are provided in Table II.

In summary, the values support the perceptual impression that the hypokinetic phrases tended toward monotonicity. Ataxic and control phrases had higher coefficients of variation for F0 than the hypokinetic group [$H(2) = 65.1, P < 0.0001$; comparisons significant at $P < 0.05$]. The values also support the perceptual impression that the ataxic phrases contained excess loudness variation. With regard to loudness variation, ataxic dB coefficients of variation were significantly greater than those of the control and hypokinetic groups [$H(2) = 32.6, P < 0.0001$; comparisons significant at $P < 0.05$].

### C. Procedures

The three groups of listeners were instructed to listen to each phrase and to write down exactly what they heard. They were told that all phrases consisted of real words in the English language produced by several different male and female speakers. They were told that some of the phrases may be difficult to understand, but that they should guess if they did not know what the speaker was saying. They were told that if they could not venture a guess, they were to use a slash to indicate that part of the phrase they could not understand.
D. Analysis

The listeners were seated in individual cubicles. The audio tapes were presented via the Tandberg Educational sound system in the ASU Language Laboratory over high-quality Tandberg supra-aural headphones. Equivalent sound pressure levels across headphones were verified with a head- phone coupler sound level meter (Quest 215 Sound Level Meter). Listeners were instructed to adjust the volume to a comfortable listening level (in 4-dB increments up or down) during the preliminary instructions. They were directed not to alter the volume once the stimulus phrases had begun. The listeners transcribed three practice phrases, which were read by a neurologically normal female speaker. Listeners who made more than one word-transcription error in the practice phrases were not eligible for the study. No listeners were excluded by this criterion.

TABLE III. Examples of coding lexical boundary errors from the listeners’ transcriptions.

<table>
<thead>
<tr>
<th>Target phrase</th>
<th>Listener response</th>
<th>Error type(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>to rest and not appear</td>
<td>to wish another beer</td>
<td>DW, IS</td>
</tr>
<tr>
<td>alive and eager smile</td>
<td>a liar when he goes well</td>
<td>IS, IW</td>
</tr>
<tr>
<td>their jury is below</td>
<td>there’s serious below</td>
<td>DW</td>
</tr>
<tr>
<td>always lobby water</td>
<td>always love your water</td>
<td>IW</td>
</tr>
<tr>
<td>will travel after court</td>
<td>will travel if you’re caught</td>
<td>IW</td>
</tr>
<tr>
<td>a term arranged inside</td>
<td>I turn her engine fine</td>
<td>IS, DW, IS</td>
</tr>
<tr>
<td>confuse the very back</td>
<td>confusing with the bank</td>
<td>DW, IW</td>
</tr>
<tr>
<td>amend the slower page</td>
<td>I meant to sort the page</td>
<td>IS, IW</td>
</tr>
<tr>
<td>the rally found some light</td>
<td>the raw impounds some light</td>
<td>IW, DS</td>
</tr>
<tr>
<td>answer dying temper</td>
<td>entertaining temper</td>
<td>DS</td>
</tr>
</tbody>
</table>

Note: IS refers to insertion of a lexical boundary before a strong syllable; IW refers to insertion before a weak syllable. DS and DW refer to deletions of lexical boundaries before strong and weak syllables, respectively. The first five examples are from transcripts from the ataxic tape and the second five are from hypokinetic tape transcripts.

The total data set consisted of 60 transcripts of the 60 phrases. Three trained judges independently coded the listener transcriptions for the presence and type of LBEs. Lexical boundary violations were defined as erroneous insertions or deletions of lexical boundaries. These insertions or deletions were coded as occurring either before strong or before weak syllables (as determined by the target phrasal stress pattern of the phrase, SWSWSW or WSWSWS). Thus, four error types were possible and each phrase had the possibility of containing more than one LBE. Examples from the actual transcripts are provided in Table III. Other tabulations for each transcript included words-correct scores for intelligibility calculations, the number of word substitutions (defined as incorrect words that did not violate lexical boundaries, and that may or may not bear phonetic similarities to the target), and the number of words for which no response was attempted.

The 60 phrases consisted of 360 syllables, 60 of which were phrase-initial syllables and were therefore not subject to LBEs. Of those 300 non-phrase-initial syllables, 102 were word-initial strong syllables; 80 were word-initial weak syllables; 48 were non-word-initial strong syllables; and 70 were non-word-initial weak syllables. The occurrence of each word-initial strong syllable in the target corresponded to the opportunity for the deletion of a lexical boundary before a strong syllable. Similarly, word-initial weak syllables corresponded to the opportunity for the deletion of a lexical boundary before a weak syllable; non-word-initial strong syllables to the opportunity for an insertion of a lexical boundary before a strong syllable; and non-word-initial weak syllables to the opportunity for an insertion of a lexical boundary before a weak syllable. Thus, the opportunities for producing the different types of LBEs were not equal, but are representative of the opportunities generally available in the English language (Cutler and Carter, 1987).

The codes generated by each of the three judges were merged into one composite data set that reflected instances in which there was 100% agreement among the judges. Twenty potential LBEs were discarded due to interjudge disagreement. The number, type (insertion or deletion), and location (before strong or before weak syllables) were then tallied for each listener group. Words-correct scores were calculated for each tape. Chi-squares, t-tests, and analyses of variance were conducted to identify significant differences between category and group means. When the data violated the assumption of a normal distribution, the Kruskal–Wallis analysis of variance by ranks was performed to test for differences within coding categories and between listener groups. Post hoc comparisons were evaluated by the Student–Newman–Keuls procedure.

II. RESULTS

A. Transcript tabulations

Table IV contains the results of the coding and tabulation procedures for each of the three listening groups. Intelligibility values are listed in the first column of this table. Recall that every effort was made during phrase selection and tape construction to ensure that the mean intelligibility scores elicited by the dysarthric tapes would not differ. The
mean words-correct scores of 43.2% for the ataxic phrases and 41.8% for the hypokinetic phrases were found not to be significantly different, so LBE analysis proceeded without modification of the phrase composition.

Of the words that were not correctly transcribed, some consisted of word substitutions, some were not attempted by the listeners (no response), and some consisted of words that violated lexical boundaries. These are listed in the next data columns of Table IV.

The transcripts of the hypokinetic tape contained more instances of word substitution errors than did the transcripts of the ataxic tape (1059 vs 925), although the magnitude of this difference is not great. Actual examples of word substitution errors include, ‘‘excess’’ for ‘‘expect,’’ ‘‘quacked’’ for ‘‘stopped,’’ ‘‘little’’ for ‘‘legal,’’ and ‘‘carcass’’ for ‘‘caucus.’’ In contrast, the ataxic tape transcripts contained many more instances in which no attempt was made for given words than the hypokinetic tape transcripts (1035 vs 669 words). Thus, of the responses that did not obviously violate lexical boundaries, the hypokinetic and ataxic tapes elicited approximately the same number of words-correct and a similar number of word substitution errors, but the transcripts of the ataxic tape contained more instances of unattempted word transcriptions.

Both dysarthric samples elicited many LBEs from the listeners. In all, 1430 LBEs were identified unanimously by the three independent judges: 610 for the ataxic phrases and 820 for the hypokinetic phrases. The control group elicited only a negligible number of LBEs (22) and their data will not be discussed further. No predictions were made regarding the absolute differences in the incidence of LBEs elicited by the two dysarthric tapes. However, the hypokinetic tape elicited significantly more LBEs than the ataxic tape \[ t(38) = -3.227; \ p = 0.003. \]

### C. Liss et al. (1998) replication

Our previous investigation (Liss et al., 1998) provided a large-scale detailed examination of LBEs elicited by hypokinetic speech. Seventy listeners transcribed 420 phrases produced by seven different speakers with hypokinetic dysarthria. The present investigation differed in several important ways. First, only six speakers provided ten phrases each to produce one listening tape of 60 phrases, but each of the 20 listeners heard ten phrases from all six speakers in their transcription task. This is in contrast to the 20-phrase blocks of three voices heard by the listeners in the first study. Second, although there was some overlap in speakers between the first and second studies, the range of dysarthria severity in the present investigation was more narrow. Also, the phrases selected for the listening tapes were selected on the basis of intelligibility and representativeness of our definition of hypokinetic dysarthria. This served to further increase the homogeneity of the phrases on the hypokinetic listening tape.

Despite these important differences, the results of the present investigation mirrored those of our previous report. In both studies, the hypokinetic speech samples elicited large numbers of LBEs; insertion errors exceeded deletion errors by nearly three times; insertion errors before strong syllables outnumbered those before weak syllables; and the opposite pattern occurred for deletion errors. Strength of adherence patterns were virtually identical in both studies. The IS–IW ratio of the present and previous studies was 1.7; the DW–DS ratio was 1.8 in the present and 1.5 in the previous study. This consistency of findings across the two studies suggests the results are robust and valid for hypokinetic speech.

### B. Lexical boundary error pattern

As seen in Table V, LBE insertions outnumbered deletions by approximately three times. This was the case for both the hypokinetic and ataxic phrases (2.92 and 2.98 times, respectively). The first four data columns of this table contain the numbers of insertion and deletion errors that occurred before strong and weak syllables for each group. Both groups elicited more LBE insertions before strong than before weak syllables, and more LBE deletions before weak than before strong syllables. However, this difference for error type and location for the ataxic data was very small. A chi-square analysis indicated a significant interaction between the variables of insert/delete and strong/weak for the data generated by the hypokinetic phrases \[ \chi^2(1, N = 4) = 47.637, \ p < 0.001, \] but not for the data generated by the ataxic phrases.

In our previous paper (Liss et al., 1998), we used two ratio calculations to express strength of adherence to the expected patterns of errors relative to syllabic strength. The first was the number of insertions before strong syllables relative to those before weak syllables (IS–IW). The second was the number of definitions before weak syllables relative to those before strong syllables (DW–DS). Ratio values of ‘‘1’’ indicate that insertions and deletions occur gradually as often before strong and weak syllables; the greater the positive distance from ‘‘1,’’ the greater the strength of adherence to the predicted pattern. For the present hypokinetic data, insertion errors occurred 1.7 times more often before strong than before weak syllables, and deletion errors occurred 1.8 times more often before weak than before strong syllables. The values for the ataxic data signify weak or lack of adherence to the expected pattern: IS–IW was 1.2 and DW–DS was 1.0.

### TABLE V. Lexical boundary error distributions for the dysarthric groups.

<table>
<thead>
<tr>
<th>Speaker group</th>
<th>LB insertions</th>
<th>LB deletions</th>
<th>IS–IW ratio</th>
<th>DW–DS ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypokinetic</td>
<td>386</td>
<td>225</td>
<td>74</td>
<td>135</td>
</tr>
<tr>
<td>Ataxic</td>
<td>247</td>
<td>210</td>
<td>75</td>
<td>78</td>
</tr>
</tbody>
</table>

Note: The values in the first four data columns are raw scores.
III. DISCUSSION

The present study provides compelling evidence that, as with the perception of normal speech, listeners attend to syllabic strength contrasts to segment the acoustic stream in the cases of hypokinetic and ataxic dysarthria. Further, the patterns of LBEs reveal that form of the prosodic deficit in the two dysarthria subtypes differentially influences the listeners' ability to apply the metrical segmentation strategy. This finding is critical for our understanding of speech intelligibility in dysarthria because it demonstrates the reciprocal relationship between listeners' perceptual strategies and the integrity of specific aspects of the acoustic signal. We therefore conclude that reductions in intelligibility for these two speech samples may be thought to arise from different perceptual challenges.

Several authors have suggested that listeners modify their perceptual strategies to decipher degraded speech (Forster, 1989; Marslen-Wilson, 1989; McQueen, 1991; Pisoni and Luce, 1986). This may take the form of hypothesis testing and postperceptual processing. In our previous paper, we hypothesized that when articulatory integrity is most impaired, syllabic contrastivity "matters more." Listeners may rely then on the relatively robust cues associated with syllabic strength contrasts in the face of phonetic uncertainty (Cutler and Butterfield, 1992).

The speakers, phrases, and phrase tokens used in this investigation were carefully selected to ensure that any differences in LBE patterns elicited by the two dysarthric tapes would be interpretable relative to syllabic strength contrasts. Like the hypokinetic speech, the ataxic tape elicited a large number of LBEs in which insertion errors outnumbered deletion errors by nearly three times. Although this finding alone is not interpretable relative to syllabic strength contrasts, it does speak to the listeners' use of syllable onset information to parse the acoustic stream. Recall that our corpus of phrases provided many times more opportunities for lexical boundary deletions than insertions. The high insertion–deletion ratio for both dysarthric tapes provides compelling evidence that the listeners exploited the statistical probabilities of the English language that favors the occurrence of single syllable words (Cutler and Carter, 1987).

In addition to syllable onset information, it is clear that listeners of both dysarthric tapes utilized available syllabic strength information to hypothesize about word-ontsets. If they had not, the proportion of insertion and deletion errors that occurred before strong and weak syllables would have mirrored the opportunities for such errors to occur. As stated above, the pattern of LBEs for the hypokinetic tape adhered to the pattern predicted by the MSS, though less strongly than that reported for normal speech. The ataxic tape elicited a pattern of even weaker adherence to the predicted MSS pattern: LBE insertions and deletions occurred with equal frequency before strong and weak syllables.

Given these findings, it can be surmised that the prosodic deficits contributed differentially to the systematic application of the metrical segmentation strategy. Specifically, the characteristics of the ataxic dysarthric speech appeared to pose more of an impediment to the strategy than did the hypokinetic speech. It is not possible from this investigation to know precisely which acoustic correlates of syllabic strength account for our pattern of results. However, we speculate that the relative syllabic isochrony of the ataxic phrases (see Table II) may have contributed to the decreased effectiveness of the perceptual strategy of attending to syllabic strength contrasts.

Hertrich and Ackermann (1998) published a preliminary report that examined the effects of synthetically altered segment durations on intelligibility and naturalness of ataxic and normal speech. Their goal was to determine whether "normalizing" segment durations in ataxic dysarthria would result in improved speech, and whether distorting normal temporal profiles to match ataxic templates would result in ataxic-sounding speech. The synthetic alterations of the ataxic speech were judged more favorably than the original utterances on a number of perceptual dimensions. However, intelligibility judgments were not affected by the temporal alterations.

The relevance of the Hertrich and Ackermann study is not so much its overall findings. Their speech sample (five sentences presented in seven resynthesized variants each) and listener pool (six speech-language pathologists) were quite limited. The perceptual measures were global ratings that were not sufficiently sensitive to quantify subtle changes. Instead, the relevance to the present report lies in the notion that temporal parameters in ataxic dysarthria may indeed affect listener performance in quantifiable ways. Temporal manipulations of the ataxic phrases produced for our present investigation would provide a perfect test for our hypothesis about syllabic isochrony and perceptual segmentation.

Although all of our hypotheses were based on LBE patterns and proportions, it is of note that the ataxic tape elicited overall fewer LBEs than the hypokinetic tape (see Table IV). If ataxic speech truly poses more of a challenge to listeners' application of the metrical segmentation strategy than hypokinetic speech, why did it appear to elicit fewer lexical boundary errors overall? We speculate that the difference in raw numbers is a byproduct of the transcription/compositions. Recall that we assembled phrases for both the hypokinetic and ataxic tapes from larger pools of phrases. Our goal was to select phrases such that the tapes would be of equivalent intelligibility. For this study, the transcribers of the ataxic tape obtained slightly higher words-correct scores than did the transcribers of the hypokinetic tape. This difference was not statistically significant so we met our goal of equivalent intelligibility. However, it does mean that, functionally, there were fewer opportunities for LBEs in the transcripts of the ataxic phrases. If we had modified the ataxic tape by replacing more intelligible phrases with less intelligible ones (which was our plan had equivalent intelligibility not been attained), our LBE count may have increased. In addition, the ataxic phrases elicited a greater number of "no response" from the listeners. It is not immediately obvious why the transcribers of the ataxic tape chose not to venture guesses on many phrases. However, this also serves to reduce the possibility for the identification of LBEs. Taken together, the slight edge in intelligibility and the greater "no responses" may account for the difference in raw numbers.
The data could have been corrected to express the absolute number of LBEs as a function of opportunities available to determine the actual significance of the observed difference. However, none of our predictions about LBE patterns called for such a comparison. If we find significant differences in LBE raw numbers in our subsequent studies when all other things are equal (i.e., intelligibility and number of “no responses”), an alternative explanation will be warranted.

Finally, the findings of our study must be viewed within the limitations of the methodology and the design. First, despite equivalent intelligibility of the dysarthric tapes, and the measures taken to ensure the presence of cardinal dysarthria features for all phrases included on the tapes, individual differences did exist among speakers. Because speakers only provided ten phrases each, it was not possible to assess the full impact of individual variation on listener performance. Second, the results were interpreted relative to one presumed cognitive strategy of perceptual segmentation. Although the results strongly suggest that listeners indeed used this strategy, other strategies certainly were operative as well. The LBE analysis is not designed to reveal alternative strategies. A qualitative assessment of listener errors would be required to speculate about such alternative strategies. On a related issue, our study did not explicitly examine the influence of segmental integrity or goodness on lexical segmentation decisions. Listeners surely used hypotheses about phoneme identity to guide lexical searches. There is no doubt that this interacted with their hypotheses about which syllables constituted word-onsets. Again, a qualitative assessment of listener responses may shed light on the nature of this interaction and yield testable hypotheses that serve further to elucidate the nature of intelligibility deficits in these two dysarthria groups.

Despite the limitations of this investigation, the results offer compelling evidence that a normal model of language processing may be used successfully to examine the perception of dysarthric speech. This constitutes a significant paradigm shift away from traditional approaches to intelligibility that focus exclusively or primarily on output characteristics. Our findings suggest that the classical notions of motor speech disorders be revisited and revised to take into account relevant listener variables.

### IV. CONCLUSION

The examination of LBEs revealed that listeners used the perceptual strategy of attending to syllabic strength to segment the acoustic stream in both ataxic and hypokinetic dysarthria. The pattern of errors elicited by the hypokinetic speech more closely adhered to predictions offered by the MSS hypothesis than did the errors elicited by the ataxic speech. It is hypothesized that the tendency toward syllabic isochrony in ataxic speech is especially detrimental to the application of the metrical segmentation strategy.

### ACKNOWLEDGMENTS

This research was supported by Research Grant No. 5 R29 DC 02672 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health. Gratitude is extended to the patients of the Mayo Clinic—Scottsdale who participated in this investigation, and to Diana Acevedo, Shannon Beatty, Kellie Filter Sciortino, Michele Marshall, and Melissa Miller for their assistance on the project.

---

1. No listeners in the present investigation participated in the Liss et al. (1998) study.
2. A list of the phrases is available electronically from the first author.
3. Eight speech-language pathologists with experience in dysarthria independently transcribed the 60 phrases produced by a quasi-random sampling of the 14 dysarthric speakers (8 hypokinetic and 6 ataxic).
4. If words-correct scores would have been different for the two listener groups, a phrase analysis would have been conducted to identify and replace those phrases responsible for the discrepancy.
5. Although vowel space calculations have been used to capture reductions in articulatory excursions in dysarthria (Turner et al., 1995), vowel spaces are notoriously variable in even the normal population (Tsao and Weismer, 1999). Their use in the present study is an attempt to quantify the perceptual impression of vowel reduction or centralization among our dysarthric speakers (see also Bradlow et al., 1996).
6. It is recognized that the dB ratio values represent a nonlinear function. Our purpose for these data was simply to make ordinal observations about amplitude variation, and not to extrapolate about magnitude of differences.


Liss et al.: Lexical boundary error analysis