The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria

Julie M. Liss and Stephanie M. Spitzer
Motor Speech Disorders Laboratory, Arizona State University, Box 871908, Tempe, Arizona 85281
John N. Caviness and Charles Adler
Department of Neurology, Mayo Clinic–Scottsdale, Scottsdale, Arizona 85259

(Received 2 May 2002; accepted for publication 15 August 2002)

This study is the third in a series that has explored the source of intelligibility decrement in dysarthria by jointly considering signal characteristics and the cognitive–perceptual processes employed by listeners. A paradigm of lexical boundary error analysis was used to examine this interface by manipulating listener constraints with a brief familiarization procedure. If familiarization allows listeners to extract relevant segmental and suprasegmental information from dysarthric speech, they should obtain higher intelligibility scores than nonfamiliarized listeners, and their lexical boundary error patterns should approximate those obtained in misperceptions of normal speech. Listeners transcribed phrases produced by speakers with either hypokinetic or ataxic dysarthria after being familiarized with other phrases produced by these speakers. Data were compared to those of nonfamiliarized listeners [Liss et al., J. Acoust. Soc. Am. 107, 3415–3424 (2000)]. The familiarized groups obtained higher intelligibility scores than nonfamiliarized groups, and the effects were greater when the dysarthria type of the familiarization procedure matched the dysarthria type of the transcription task. Remarkably, no differences in lexical boundary error patterns were discovered between the familiarized and nonfamiliarized groups. Transcribers of the ataxic speech appeared to have difficulty distinguishing strong and weak syllables in spite of the familiarization. Results suggest that intelligibility decrements arise from the perceptual challenges posed by the degraded segmental and suprasegmental aspects of the signal, but that this type of familiarization process may differentially facilitate mapping segmental information onto existing phonological categories. © 2002 Acoustical Society of America. [DOI: 10.1121/1.1515793]

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

I. INTRODUCTION

With great facility, we are able to extract spoken words from a continuous acoustic stream that contains virtually no reliable and consistent word boundary cues (Lehiste, 1972; Nakatani and Schaffer, 1978). We easily negotiate the spoken messages of friends and strangers; men, women, and children; fast and slow talkers; synthetic speech; and the messages of speakers whose first language or dialect is not our own. Even in suboptimal listening conditions, such as noisy rooms or poor phone connections, we execute the task of lexical segmentation with surprising accuracy. This ability to be successful in the face of variable acoustic manifestations of words has proven problematic for theories of speech perception (see McQueen and Cutler, 2001). Learning, adaptation, and normalization appear to be highly operative processes that underlie the plasticity of our speech perception capabilities.

But these perceptual capabilities have limits, and cognitive effort increases and accuracy decreases as the acoustic information becomes degraded or unreliable (Munro and Derwing, 1995; Pisoni et al., 1987). This is precisely the case with the perception of dysarthric speech. Although reductions in speech intelligibility secondary to dysarthria are well documented, we know very little about the cognitive–perceptual source of these decrements. That is to say, we do not understand how the nature of the degraded speech signal affects our ability to process it. The study of the perception of dysarthric speech has not only clinical implications, but offers a test case for theories and models of normal speech perception processes.

The current report is the third in a series of studies that has attempted to identify part of the source of intelligibility decrements in connected dysarthric speech by examining the mistakes listeners make in lexical segmentation. The Metric Segmentation Strategy (MSS) proposed by Cutler and Norris (1988) was selected as a framework for these investigations because of its ability to jointly represent signal characteristics (i.e., the dysarthric speech patterns) and listener constraints (as evidenced by lexical boundary error patterns). Furthermore, its predictions based on prosodic patterns can be applied directly to the prosodic disruptions that are common in dysarthria. The MSS posits that listeners capitalize on the rhythmic structures of language to identify the location of word boundaries. In English, listeners will be highly successful in their lexical segmentation if they attend to strong syllables as potential word-onsets (Cutler and Carter, 1987). Dysarthric speech, particularly the types targeted in the current series of studies, is characterized in part by disruptions in prosody or rhythm. It was assumed that if listeners rely on syllabic stress information for identification of word boundaries, then their patterns of lexical boundary
errors (LBE) should reveal difficulties in applying this strategy when the prosodic structure is disturbed. As reported previously, this was found to be the case (Liss et al., 1998; Liss et al., 2000). The data were consistent with the idea that the degraded prosodic pattern made it more difficult to identify word boundaries from stress for some forms of dysarthria than others (Liss et al., 2000). Hypokinetic dysarthria, characterized by a perceptually rapid rate and monotonicity, generated a large number of lexical boundary errors whose patterns generally conformed to the predictions based on normal degraded speech, particularly at higher levels of intelligibility. However, ataxic dysarthria characterized by a slow rate of speech with equal and even syllabic stress, elicited LBE patterns that did not conform to predictions. Erroroneous insertions and deletions of lexical boundaries occurred equally often before strong and weak syllables. It appeared that the type of prosodic degradation associated with ataxic dysarthria made it particularly difficult for listeners to use stress cues for lexical segmentation. This distinction between LBE patterns elicited by these hypokinetic and ataxic dysarthric speech samples is of particular note because the two samples were of equivalent intelligibility. These studies shed light on how the nature of the speech signal contributes to the efficiency with which normal cognitive–perceptual processes can be applied.

In contrast to our two previous reports on the perception of dysarthric speech, the current study focused on the manipulation of listener constraints. We wondered whether listeners could learn something about the degraded signal that would assist them in applying the metrical segmentation strategy, and whether this would be evident in their pattern of LBEs. Our previous studies suggested a reduced ability to distinguish strong and weak syllables in dysarthric speech, particularly in ataxic dysarthric speech. If listeners can acquire information that facilitates the distinction of strong and weak syllables, their LBE patterns should align more closely with predicted patterns. A process of brief familiarization was selected to examine this issue.

The construct of familiarization encompasses a broad array of methodological and conceptual categories, including exposure, training, adaptation, and experience. Each of these categories can vary along a number of continua, including amount of exposure, duration of exposure, type of information presented, and the quality and quantity of performance feedback provided. Irrespective of these methodological variations, experience with a degraded speech signal has been shown to facilitate subsequent processing of that signal in many studies. Benefits have been demonstrated as improved intelligibility of synthetic or electronically modified speech signals (Dupoux and Green, 1997; Greenspan et al., 1988; Rosen et al., 1999; Sebastián-Gallés et al., 2000; Schwab et al., 1985), disordered speech (Dagenais et al., 1999; DePaul and Kent, 2000; Fliipsen, 1995; Robinson and Summerfield, 1996; Tjaden and Liss, 1995; although see Yorkston and Beukelman, 1983, for evidence to the contrary), and non-native speech (e.g., Logan et al., 1991; Tremblay et al., 1997). Although the precise cognitive–perceptual mechanisms underlying the benefits of familiarization have not been discovered, it is hypothesized that it promotes the normalization process and the mapping of speech stimulus features onto existing phonological representations (Dupoux and Green, 1997; Guenther et al., 1999; Schwab et al., 1985).

The purpose of the present study was to examine the effects of modifying listener constraints through the process of a brief prior exposure to either hypokinetic or ataxic dysarthric speech. The following questions were addressed: (1) Do listeners who are familiarized with dysarthric speech in general obtain higher intelligibility scores than those who have had no prior exposure; (2) Is there a dysarthria-specific effect of familiarization on intelligibility, in which exposure to hypokinetic or ataxic speech improves intelligibility of hypokinetic or ataxic speech, respectively; (3) Do lexical boundary error patterns provide evidence that listeners learn about prosodic form from the familiarization procedure? Evidence for the final question would be found in the distribution of lexical boundary errors relative to strong and weak syllables. Specifically, do the patterns of LBEs produced by the listeners who were familiarized with the dysarthric speech adhere more strongly to the predicted patterns than those elicited from nonfamiliarized listeners?

II. METHOD

A. Study overview

A between-group design was selected for the present study to more closely control the familiarization effects than a within-group design would permit. By its nature, the phenomenon of familiarization is cumulative and irreversible within the context of a circumscribed investigation. Although a within-subjects design is preferable for a variety of reasons, we believed it was necessary first to establish the magnitude of any effect in separate groups.

Two groups of listeners were familiarized with either hypokinetic or ataxic speech and then they transcribed a series of phrases produced by speakers with the corresponding type of dysarthria. Appended to the 60 phrase series was a subset of 20 low-intelligibility phrases produced by speakers with the other type of dysarthria. Thus, one listener group was familiarized with hypokinetic speech. They transcribed 60 hypokinetic phrases (hypokinetic familiarized), then they transcribed 20 low intelligibility ataxic phrases (familiarized with other). The second listener group was familiarized with ataxic speech. They transcribed 60 ataxic phrases (ataxic familiarized), then they transcribed 20 low intelligibility hypokinetic phrases (familiarized with other). Results for the 60-phrase and 20-phrases were compared with the corresponding data from a third and fourth group (reported previously in Liss et al., 2000) who transcribed the phrases without prior exposure to any dysarthric speech, and were therefore considered the control groups (hypokinetic control and ataxic control) for this investigation. The comparisons of interest are summarized in Table I.

B. Listeners

Data from two groups of 40 listeners were collected for this investigation. These data were compared with those of two control groups of 20 listeners that did not receive expo-
TABLE I. Comparisons of interest.

<table>
<thead>
<tr>
<th>60 phrases (between-group):</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypokinetic control versus hypokinetic familiarized</td>
<td></td>
</tr>
<tr>
<td>Ataxic control versus ataxic familiarized</td>
<td></td>
</tr>
<tr>
<td>20 phrases (between-group):</td>
<td></td>
</tr>
<tr>
<td>Hypokinetic control versus hypokinetic familiarized versus familiarized with other</td>
<td></td>
</tr>
<tr>
<td>Ataxic control versus ataxic familiarized versus familiarized with other</td>
<td></td>
</tr>
</tbody>
</table>

sure to dysarthric speech prior to the transcription task (Liss et al., 2000). Each of the four listener groups contained equal numbers of men and women whose ages ranged from 18 to 50 years old. Most were undergraduate students at Arizona State University, and all 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.

C. Speech stimuli

Construction of the stimulus tapes has been described in detail in our previous reports (Liss et al., 1998, 2000). Briefly, three audiotapes of phrases were produced by three groups of speakers: six speakers with hypokinetic dysarthria, six with ataxic dysarthria, and six neurologically normal control speakers (whose tape was not used in the present investigation). All of the phrases on the hypokinetic and ataxic tapes complied with our operational definitions, which were derived from the Mayo Classification System (Darley et al., 1969; Duffy, 1995). All of the hypokinetic phrases were characterized by a 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. The ataxic phrases were characterized by a perceptually slow speaking rate with a tendency toward equal and even syllable duration (scanning speech); excessive loudness variation; and irregular articulatory breakdown.

These perceptual impressions of reduced syllabic strength (see Fear et al., 1995) were corroborated by acoustic measures of phrase duration, strong-to-weak vowel duration calculations, vowel formant frequencies and point-vowel quadrilateral areas, and fundamental frequency and amplitude variation (see Liss et al., 2000, Tables I and II). Briefly, the hypokinetic phrases were significantly shorter in duration than those of the ataxic or neurologically normal speakers; they had a significantly smaller range of fundamental frequency variation; and their vowel quadrilateral areas were 50% smaller than those of the normal speakers. This coincided with the perception of rapid and monotonous speech. The ataxic phrases were significantly longer in duration than those of the hypokinetic or neurologically normal speakers; their adjacent strong and weak vowel durations were of similar durations; and their vowel quadrilateral areas also were 50% smaller than those of the control speakers. This corroborated the perception of slow, equal, and even speech.

By design, the phrases on the two dysarthria tapes were of equivalent intelligibility. As reported in Liss et al. (2000), the mean words-correct score for the ataxic tape was 43.2% and the mean for the hypokinetic tape was 41.8%. This allowed differences in the dependent variables to be interpreted as arising from differences in speech production characteristics, specifically syllabic strength contrasts.

The phrases, modeled after Cutler and Butterfield (1992), were designed to permit the interpretation of LBE patterns. 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 3 to 5 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.

The stimulus audiotapes consisted of phrases produced by dysarthric speakers, each preceded by a neurologically normal female saying the phrase number, and followed by 12 seconds of silence in which to transcribe what had been heard. Each tape contained one production of the 60 phrases, 10 phrases per each of the six speakers in each group.

In addition to the data from the 60 core phrases, we wished to obtain information upon which to compare the effects of specific versus general familiarization with dysarthric speech. The 60 phrases provided data for specific familiarization effects, or benefits to the perception of hypokinetic or ataxic speech after being familiarized with hypokinetic or ataxic speech, respectively. To obtain a measure of more general effects, 20 phrases from the other dysarthria type were appended to the end of each 60-phrase tape. The 60 phrases of the hypokinetic speakers were followed by 20 phrases from the ataxic speakers; the 60 ataxic phrases were followed by 20 phrases from the hypokinetic speakers. Because the 20 phrases were a subset of the 60 phrases, it was necessary to minimize the possibility that listeners would recognize the phrases, thereby improving transcription performance. Thus, 20 low-intelligibility phrases were selected from the two dysarthria tapes. The 20 phrases were selected based on phrase-level intelligibility data acquired from Liss et al. (2000). The mean words-correct intelligibility scores of the 20 phrases were 18.0% and 18.2% for the hypokinetic and ataxic phrases, respectively.

In addition to the stimulus tapes, two familiarization tapes were constructed. Each of the hypokinetic and ataxic familiarization tapes contained 18 novel phrases, three from each of the six speakers. The familiarization phrases were identical to the test phrases in their syllabic strength alternation, low interword predictability, phoneme representation, and general composition. However, there was no phrase overlap between the stimulus and familiarization phrases.
TABLE II. 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>Younger rusty viewers</td>
<td>Younger rest if you are</td>
<td>IW, IW</td>
</tr>
<tr>
<td>The rally found some light</td>
<td>The real effects of light</td>
<td>IW, DS</td>
</tr>
<tr>
<td>Soon the men were asking</td>
<td>Cinnamon were asking</td>
<td>DW, DS</td>
</tr>
<tr>
<td>His display collects it</td>
<td>Hands did spray collected</td>
<td>IS, DW</td>
</tr>
<tr>
<td>Friendly slogans catch it</td>
<td>Fred eats slow with ketchup</td>
<td>IW, IW, DW</td>
</tr>
<tr>
<td>Govern proper landings</td>
<td>Car and prop for landings</td>
<td>IW, IW</td>
</tr>
<tr>
<td>Call a random voter</td>
<td>Coming from the motor</td>
<td>DW, IW</td>
</tr>
<tr>
<td>A term arranged inside</td>
<td>A turmoil raged inside</td>
<td>DW, IS</td>
</tr>
<tr>
<td>Convince the council here</td>
<td>In winds the cows will here</td>
<td>IS, IW</td>
</tr>
<tr>
<td>She describes a nuisance</td>
<td>Strangers cause a nuisance</td>
<td>DW, IS</td>
</tr>
</tbody>
</table>

*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 familiarized transcripts, and the second five are from hypokinetic familiarized transcripts.

D. Procedures

The listeners were seated in individual cubicles. The audiotapes 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 headphone 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 would not be eligible for the study. No listeners were excluded by this criterion.

Prior to the transcription task, listeners in the two familiarization groups were given a list of the 18 familiarization phrases. They were asked to follow along carefully as the various speakers read the phrases. The listeners then received instructions for the transcription task. They were asked 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.

E. Analysis

The familiarization corpus consisted of 6400 phrase transcriptions (80 listeners×80 phrases). A words-correct score was calculated for each phrase as an index of intelligibility. A word was counted as correct when it exactly matched the target, or when it differed only by tense (-ed) or plural (-s) and did not add another syllable. Substitutions between “a” and “the” were also regarded as correct. The percentages of words-correct were calculated for each listener and averaged within listening groups as indices of intelligibility. Intelligibility scores for the 60 phrases and the subset of 20 phrases were calculated separately.

Two trained judges (the first two authors) independently coded the listener transcripts for the presence and type of LBEs to obtain a corpus of errors identified and/or agreed upon by both judges. 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: Insert boundary before a strong syllable (IS); insert boundary before a weak syllable (IW); delete boundary before a strong syllable (DS); and delete boundary before a weak syllable (DW). Each phrase had the possibility of containing more than one LBE. Examples from the actual transcripts are provided in Table II.

Analyses of variance and post-hoc pairwise multiple comparisons were conducted to detect differences in mean intelligibility scores between and among groups (see Table I for comparisons of interest). Because 40 listeners were included in the familiarization groups as compared to the twenty listeners in the control groups, a conservative alpha of 0.001 was selected. This was intended to minimize the potential for an overpowered study, in which small between-group differences may attain statistical significance without being clinically (or perceptually) relevant. Chi-square procedures were conducted to determine the relationship between the variables of insert/delete and strong/weak for 10 sets of lexical boundary error data: 60-phrase control and familiarized; 20-phrase control, familiarized, and familiarized with other, for both sets of dysarthric speech.

In addition to the LBE proportion comparisons in the contingency tables, IS/IW and DW/DS ratios were calculated for all 10 sets of LBE data. These ratios permitted a comparison with previously published data regarding strength of adherence to predicted error patterns. Specifically, if listeners use the strategy of attending to syllabic strength to mark
word boundaries in the connected acoustic stream, they will most likely erroneously insert lexical boundaries before strong syllables, and delete them most often before weak syllables. Ratio values of 1 indicate that insertions and deletions occur equally as often before strong and weak syllables. Therefore, the greater the positive distance from “1,” the greater the strength of adherence to the predicted pattern. Between-group differences in median values among conditions were assessed by nonparametric Kruskal–Wallis one-way analysis of variance on ranks procedures.

III. RESULTS

A. Intelligibility

Figure 1 shows the mean percent words-correct intelligibility scores for each of the groups on the 60 phrases. A one-way analysis of variance showed significant differences among the means of the familiarized and nonfamiliarized listener groups \( F(3,116) = 22.1, \ P < 0.0001 \). The student–Newman–Keuls procedure indicated that both familiarized groups significantly outperformed the corresponding nonfamiliarized groups (\( P < 0.05 \)). The mean of the ataxic familiarized group was 51.6% (s.d. = 4.48), which was significantly greater than that of the ataxic control group (\( M = 43.2, \ s.d. = 6.41 \)). The hypokinetic familiarized group mean was 45.3% (s.d. = 4.91), which was significantly greater than that of the hypokinetic Control group (\( M = 41.8, \ s.d. = 5.58 \)).

The 20-phrase subset intelligibility means and standard deviations are shown Fig. 2. Three different values are presented for each of the two dysarthria tapes. The first bars for each dysarthria type correspond with the intelligibility score achieved by the control subjects from Liss \textit{et al.} (2000). The second values correspond with the scores derived from the familiarization condition (familiarization). This is the percentage of words-correct for these 20 phrases taken from the set of 60 when the listener had been familiarized with that particular dysarthria type. The third bars represent the percentage of words-correct on these 20 phrases when the listener had been familiarized with the \textit{other} dysarthria type (other). Analyses of variance were significant at \( P < 0.001 \) for both groups \( F(2,97) = 45.899 \) for hypokinetic; \( F(2,97) = 76.795 \) for ataxic. Post-hoc analysis showed significant differences (\( p < 0.05 \)) for all pairwise comparisons within both hypokinetic and ataxic groups. Thus, for both types of dysarthric speech, the specific familiarization condition resulted in the highest score, followed by the general familiarization condition, followed by the no-familiarization condition.

B. Lexical boundary error pattern

Table III contains the total LBEs, the LBE category proportions thereof,\(^5\) and the median IS/IW and DW/DS ratios

\[ \begin{array}{cccccc}
\text{Total LBEs} & \% \text{IS} & \% \text{IW} & \% \text{DS} & \% \text{DW} & \text{Median IS/IW} & \text{Median DW/DS} \\
\hline
\text{Hypokinetic control} (N = 20) & 820 & 47.1 & 27.4 & 9.0 & 16.5 & 1.75 & 1.83 \\
\text{Hypokinetic familiarized} (N = 40) & 1633 & 44.9 & 27.1 & 10.3 & 17.6 & 1.56 & 1.75 \\
\text{Ataxic control} (N = 20) & 610 & 40.5 & 34.4 & 12.3 & 12.8 & 1.18 & 0.90 \\
\text{Ataxic familiarized} (N = 40) & 1266 & 36.0 & 34.0 & 14.5 & 15.4 & 1.04 & 1.00 \\
\end{array} \]

TABLE III. LBE totals, category proportions expressed in percentages, and median ratio values for each of the four listener groups for the 60 phrases.
for each of the four groups for the 60 phrases. Contingency tables were constructed for the LBEs for each of the familiarized listener groups to determine whether the variables of lexical boundary error type (i.e., insertion/deletion) and lexical boundary error location (i.e., before strong/weak syllable) were significantly related. Consistent with our previous report on the hypnotic control (Liss et al., 2000), chi-square results were significant for the data derived from the hypnotic speech samples [hypokinetic familiarized, $\chi^2(1, N = 4) = 83.5, P < 0.0001$]. That is, there was a significant relationship between LBE type and the locations in which the errors occurred. Erroneous lexical boundary insertions occurred more often before strong than before weak syllables, and erroneous lexical boundary deletions occurred more often before weak than before strong syllables. This was not the case for LBE pattern for the ataxic data. Consistent with our previous report on ataxic control, there was no statistically significant relationship between LBE type and location among the ataxic familiarization data [$\chi^2(1, N = 4) = 0.759, P = 0.3838$].

The IS/IW and DW/DS ratios echo the chi-square findings, and offer a source of examining the strength of adherence to the predicted pattern of results both between and within dysarthria type. Because the ratio data were not normally distributed, Kruskal–Wallis ANOVA on Ranks procedures were performed on the two sets of ratio data. The median IS/IW ratios for the hypokinetic control and hypokinetic familiarization groups were significantly greater than those for the ataxic control and ataxic familiarization groups [$H(3) = 40.4, P < 0.0001$]. Similarly the median DW/DS ratios for the two hypokinetic groups were significantly greater than those for the two ataxic groups [$H(3) = 19.9, P = 0.0002$]. However, there were no significant differences between the median values of either the hypokinetic control and familiarization groups, or of the ataxic control and familiarization groups, according to Dunn’s method of all pairwise multiple comparisons ($P < 0.05$). Thus, both hypokinetic tapes elicited significantly higher strength of adherence values than did either of the two ataxic tapes.

Tables IV and V contain the LBE information for the 20-phrase subset taken from the control condition, the familiarized condition, and the familiarized with other condition. As with the data from the 60 phrases, significant relationships between LBE type and location were found for all data derived from the hypokinetic tapes as shown in Table IV [hypokinetic control, $\chi^2(1, N = 4) = 55.3, P < 0.0001$; hypokinetic familiarized, $\chi^2(1, N = 4) = 14.0, P = 0.0002$; familiarized with other, $\chi^2(1, N = 4) = 75.2, P < 0.0001$], but not for the LBE proportions derived from the ataxic tapes as shown in Table V [ataxic control, $\chi^2(1, N = 4) = 0.678, P = 0.4104$; ataxic familiarized, $\chi^2(1, N = 4) = 0.15, P = 0.6986$; familiarized with other, $\chi^2(1, N = 4) = 1.85, P = 0.174$].

Kruskal–Wallis ANOVAs on ranks were conducted on ratio values within each of the dysarthria categories to identify specific versus general effects of familiarization in the subset of 20 phrases. No significant differences were found among control, familiarized, and familiarized with other for any of the ataxic data [IS/IW, $H(2) = 2.45, P = 0.0294$; DW/DS, $H(2) = 0.20, P = 0.6386$], nor for the hypnotic data [$H(2) = 6.33, P = 0.0422$]. Thus, IS/IW and DW/DS ratios were similar and consistent within dysarthria subtype across all three conditions.

### IV. DISCUSSION

The results of the present study demonstrate perceptual benefits of familiarization with dysarthric speech. However, these benefits were apparent only in intelligibility scores, and not in lexical boundary error patterns. In fact, the LBE results were remarkably consistent with the data from nonfamiliarized listeners, in which it appears that different patterns of dysarthria may differentially affect listeners’ abilities to apply the MSS. If listeners had gleaned information about distinguishing strong and weak syllables in the two dysarthric speech patterns, LBE results should have reflected this. Although knowledge about prosodic patterns and syllabic stress manifestations may have benefited intelligibility, the absence of LBE differences suggests the familiarization procedure did not improve strong/weak syllable identification, but rather improved performance at some other (perhaps segmental) level.

The question of general and specific familiarization effects on intelligibility scores in dysarthria has significant clinical and theoretical implications. The familiarization procedure of the present investigation was very brief, just three phrases each from the six speakers with dysarthria who provided the stimulus phrases. It also required little active participation by the listeners. They simply were instructed to follow along with a written transcript of the 18 familiarization phrases as they listened. They were not instructed to listen for certain features or characteristics, nor were they even told the purpose of the procedure. Nonetheless, this brief and relatively passive experience resulted in significantly higher intelligibility scores for the familiarized groups.

<table>
<thead>
<tr>
<th>Total LBEs</th>
<th>%IS</th>
<th>%IW</th>
<th>%DS</th>
<th>%DW</th>
<th>Median IS/IW</th>
<th>Median DW/DS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control ($N=20$)</td>
<td>434</td>
<td>47.0</td>
<td>22.1</td>
<td>9.0</td>
<td>21.9</td>
<td>1.81</td>
</tr>
<tr>
<td>Familiarized ($N=40$)</td>
<td>710</td>
<td>39.1</td>
<td>30.0</td>
<td>12.7</td>
<td>18.2</td>
<td>1.56</td>
</tr>
<tr>
<td>Familiarized with other ($N=40$)</td>
<td>883</td>
<td>44.8</td>
<td>21.4</td>
<td>12.5</td>
<td>21.3</td>
<td>2.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total LBEs</th>
<th>%IS</th>
<th>%IW</th>
<th>%DS</th>
<th>%DW</th>
<th>Median IS/IW</th>
<th>Median DW/DS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control ($N=20$)</td>
<td>284</td>
<td>39.1</td>
<td>34.5</td>
<td>12.3</td>
<td>14.1</td>
<td>1.05</td>
</tr>
<tr>
<td>Familiarized ($N=40$)</td>
<td>585</td>
<td>35.7</td>
<td>32.9</td>
<td>15.7</td>
<td>15.7</td>
<td>1.04</td>
</tr>
<tr>
<td>Familiarized with other ($N=40$)</td>
<td>744</td>
<td>35.9</td>
<td>35.6</td>
<td>12.6</td>
<td>15.9</td>
<td>1.00</td>
</tr>
</tbody>
</table>
than those obtained by nonfamiliarized listeners. When the
Dysarthria of the familiarization procedure matched the dys-
arthria of the transcription task, the benefits were even
greater.

Although the data cannot provide a definitive source for
the intelligibility benefit, they do offer a number of clues.
First, familiarization effects were not of the same magnitude
for the two Dysarthria subtypes. Transcriptions of both the
hypokinetische and ataxic phrases evidenced familiarization ef-
effects, however the effects were greater for the ataxic speech.
This is particularly compelling because the two Dysarthria
Tapes were constructed to ensure equivalent intelligibility,
and to ensure that each phrase was perceptually representa-
tive of the operational definitions of the respective Dysarthria
(Liss et al., 2000). Thus, the 8.4% advantage for the ataxic
tape can be compared directly to the clinically less impres-
sive 3.5% advantage for the hypokinetische tape. Listeners who
heard then transcribed ataxic speech benefited more from
their exposure than did listeners who heard and then tran-
scribed hypokinetische speech.

This pattern is even more robust in the subset of data
from the 20 low-intelligibility phrases for which the familiar-
ization procedure produced greater gains. The ataxic fa-
miliarized listeners who transcribed the ataxic phrases en-
joyed a 21% advantage over those who were not
familiarized, while the hypokinetische familiarized realized a
15.6% gain over those who transcribed the hypokinetische
phrases without prior exposure to that particular Dysarthria.

In both of these cases, the benefit of being familiarized with
the specific Dysarthria was roughly double the intelligibility

This possibility has some support in the psycholinguistic
literature. The construct of perceptual adaptation is related to
the process of familiarization in the present study, and has
been the subject of a number of investigations with time-
compressed speech. As listeners are exposed to a time-
compressed speech signal, their ability to understand that
signal improves to a point. This improvement is taken as
evidence of perceptual adaptation to the distorted signal.
Sebastián-Gallés and colleagues (2000) examined language-
specific rhythmical, lexical, and phonological factors in per-
ceptual adaptation to time-compressed speech. Their
Spanish-speaking listeners transcribed time-compressed
Spanish sentences after exposure to time-compressed speech
of different languages. These listeners showed perceptual
benefits only from prior exposure to time-compressed Span-
ish, and from Italian and French, which are rhythmically
(syllable-timed) and phonologically similar to Spanish. The
listeners did not demonstrate significant gains in intelligibil-
ity with exposure to the dissimilar languages of English (a
stress-timed language) or Japanese (a mora-timed language).
Moreover, a second experiment with time-compressed Greek
(a syllable-timed but not a Romance language) revealed the
effect was not simply one of lexical overlap or similarity
between Spanish and Italian. They suggested that perceptual
adaptation benefits derive not only from the similarity of
language rhythm, but also from phonetic and phonologic
similarity. This conclusion coincides with the work of Dup-
oux and Green (1997) who found robust and relatively long-
term perceptual adaptation effects for time-compressed
speech. They also concluded that the adaptation mechanism

This points toward the possibility that the familiarization
procedure allowed the listeners to map aspects of the de-
graded acoustic signal onto pre-existing phonological tem-
plates. This hypothesis has been explicated in the psycholin-
guistic literature (e.g., Dupoux and Green, 1997; Dupoux
et al., 2001; Francis et al., 2000; Greenspan et al., 1988) but
empirical data from motor speech disorders are limited.
Tjaden and Liss (1995) attempted to tease out the relative
contributions of segmental and suprasegmental information
in the familiarization process. Using the speech of a Korean
woman with moderate–severe spastic–ataxic Dysarthria sec-
ondary to cerebral palsy, they created two familiarization
tapes. The first contained a paragraph, consisting of 12 six-
word sentences, read by the woman. The second consisted of
a list of all 72 words contained in the paragraph, but ar-
 ranged in a random order and read as a word list. Thus the
content of the two tapes was identical, but the paragraph tape
provided listeners with sentence level prosodic and interword
coarticulatory information not found in the word list. The
task consisted of transcribing 48 six-word sentences pro-
duced by the woman. It was expected that the Paragraph
group would obtain higher words-correct intelligibility
scores than the other two groups because they would have
received the most information relevant to the transcription
task (see Greenspan et al., 1988). However, the results dem-
onstrated that although the two familiarized groups outper-
formed the control group, there was no significant perfor-
ance difference between the Paragraph and Word groups.
Although the high degree of variability in listener scores may
have masked actual differences, perhaps the segmental infor-
mation gleaned from both familiarization tapes accounted for
the majority of familiarization benefit.

This hypothesis has been explicated in the psycholin-
guistic literature. The construct of perceptual adaptation is related to
the process of familiarization in the present study, and has
been the subject of a number of investigations with time-
compressed speech. As listeners are exposed to a time-
compressed speech signal, their ability to understand that
signal improves to a point. This improvement is taken as
evidence of perceptual adaptation to the distorted signal.
Sebastián-Gallés and colleagues (2000) examined language-
specific rhythmical, lexical, and phonological factors in per-
ceptual adaptation to time-compressed speech. Their
Spanish-speaking listeners transcribed time-compressed
Spanish sentences after exposure to time-compressed speech
of different languages. These listeners showed perceptual
benefits only from prior exposure to time-compressed Span-
ish, and from Italian and French, which are rhythmically
(syllable-timed) and phonologically similar to Spanish. The
listeners did not demonstrate significant gains in intelligibil-
ity with exposure to the dissimilar languages of English (a
stress-timed language) or Japanese (a mora-timed language).
Moreover, a second experiment with time-compressed Greek
(a syllable-timed but not a Romance language) revealed the
effect was not simply one of lexical overlap or similarity
between Spanish and Italian. They suggested that perceptual
adaptation benefits derive not only from the similarity of
language rhythm, but also from phonetic and phonologic
similarity. This conclusion coincides with the work of Dup-
oux and Green (1997) who found robust and relatively long-
term perceptual adaptation effects for time-compressed
speech. They also concluded that the adaptation mechanism

must contain a component of phonetic-to-phonological mapping.

Consistent with Sebastián-Gallés et al. (2000), the current data support the notion that the intelligibility decrement derives from both difficulty in decoding the rhythmic structure of the signal, and mapping the degraded acoustic phonetic information onto existing phonological templates. In the current study, intelligibility was most improved when the prosodic and segmental features of the familiarization stimulus matched those of the transcription task. However, the LBE analysis indicates that the improvement cannot be attributed predominantly to knowledge about prosodic form, so a segmental explanation must be entertained. The data of the present report were not collected in a way to confirm a phonetic-to-phonologic adaptation mechanism, but some preliminary support is offered in a methodological paper from our laboratory that evaluated word substitution errors in a subset of these transcribed phrases. Spitzer et al. (2000) proposed an analysis method by which to capture evidence of the segmental benefits of the familiarization procedure. Twenty phrases transcribed by 34 listeners were analyzed for the presence of word substitution errors. Word substitutions were defined as whole words that did not violate the lexical boundaries or the syllabic constitution of the intended target words. The goal was to develop a classification scheme that would quantify phonemic preservation of either vowels or consonants within these word substitutions. It was hypothesized that if the familiarization procedure facilitates mapping of the degraded segmental information onto stored representations for phonemes, word substitutions should bear phonemic resemblance to the target word. The results from this preliminary analysis were encouraging. In this subset of phrases, more word substitutions from the familiarized groups bore phonemic resemblance to the target than did those word substitutions from the nonfamiliarized groups. Additionally, a dysarthria-specific difference was discovered. The word substitutions taken from transcriptions of the ataxic speech had significantly more phonetically related than nonphonetically related word substitutions. This difference was not found in the word substitutions elicited by the hypokinetic speech. Because such a small proportion of the full data set was examined in this methodological paper, the inconclusive findings for the hypokinetic speech are not particularly interpretable. However, the patterns of phoneme preservation elicited by the ataxic speech are consistent with the idea that familiarization improves the perceptual process of phonetic–phonologic mapping.

V. CONCLUSION

The present report offers additional evidence for the bidirectional relationship between the nature of the degraded signal and the cognitive–perceptual processes brought to bear on that signal (Lindblom, 1990). The LBE data suggest three important conclusions. First, consistent with our previous reports, it appears that syllabic strength contrasts are a source of information to listeners as they attempt to segment these types of dysarthric speech. Were this not the case, the LBE patterns would have been more closely aligned with chance, based on the number of opportunities to commit certain errors (in the case of this phrase set, deletions > insertions, IW > IS, and DS > DW). Second, the data are consistent with the notion that the form of the prosodic disturbance offers differential challenge to the application of the MSS. As in our previous report, there was no difference between error location for either erroneous insertions or deletions for the ataxic speech (IS = IW, DW = DS). Third, the level and type of familiarization provided herein did not convey sufficient knowledge about syllabic strength contrastivity to facilitate the application of a metrical segmentation type of strategy. Nonetheless, the familiarized listeners obtained higher intelligibility scores than those who did not receive familiarization, and a dysarthria-specific effect was evident. Thus, it can be hypothesized that intelligibility decrements may arise from inefficient processing of the degraded segmental and suprasegmental aspects of the signal, but this familiarization process differentially facilitates the cognitive–perceptual process of mapping the degraded segmental information onto existing phonological categories.

The study of naturally degraded speech offers ecological validity, but is replete with its own set of limitations. Although the two dysarthria tapes were of equivalent intelligibility and each phrase was representative of the corresponding dysarthria, the specific articulatory deficits varied across speakers both between and within groups. Some speakers may have had more consistent articulatory errors than others, making it conceivably easier to “break the code” during the brief familiarization procedure. It is not possible with these natural speech samples to control for critical acoustic features that would allow us to draw firm conclusions about the cognitive mechanisms underlying the perception of dysarthric speech. However, the data are a springboard for constructing controlled experiments with (re)synthesized samples, perhaps emulating the various syllabic contrastivity differences between the two dysarthria subtypes.

There are several aspects of the study design that call for caution in the interpretation of the data. The 20 phrases of the familiarization with other condition were not novel, as the listeners had transcribed them as part of the 60 phrase familiarization list. Because listeners understood very little of the phrases in the original exposure (18% of words), we expected that perceptual benefits would be minimal. Our design did not permit us to ascertain the relative contributions of this prior exposure to the phrases and exposure to the dysarthric speech. However, the pattern of data does support our interpretation of a dysarthria-specific benefit to perception. Rather than a uniform benefit that might be predicted from prior exposure to phrases that were 18% intelligible, a differential benefit to the ataxic phrases was discovered. Future experiments may control for such possible confounds by ensuring that phrases across conditions do not overlap, or overlap in controlled and systematic ways.

Interpretive power would be additionally strengthened by including a more extensive nondysarthric familiarization procedure for the control group. These listeners did hear a three-phrase practice series recorded by a neurologically normal speaker, which consisted of low-predictability phrases similar to the stimulus phrases. However, they did not hear 18 phrases, as did the familiarization groups. This step would
have permitted us to identify any perceptual improvements afforded by simply learning more about the nature of the phrases, separate from the influences of exposure to dysarthric speech.

Finally, the familiarization procedure of the present study was brief and relatively passive. It is likely that more impressive gains in intelligibility, and perhaps evidence of more efficient application of the MSS, can be realized with more extensive familiarization procedures (Francis et al., 2000). A paradigm in which different types of information are offered may shed light on the dysarthria-specific findings of the present (and previous) studies. In addition, the 15–20% intelligibility advantage evidenced in the low intelligibility phrases strongly supports the integration of familiarization procedures in motor speech disorders clinical practice.

ACKNOWLEDGMENTS

This research was supported by research Grant No. 5 R29 DC 02672 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. Gratitude is extended to the patients and families of the Mayo Clinic-Scottsdale who participated in this investigation.

1To ensure a sufficient number of errors for the LBE analyses of the 20-phrase subset, 40 listeners were recruited for the familiarization groups (as compared to 20 for the groups from the previous investigations).

2A list of the phrases is available electronically from the first author.

3The opportunities for producing the different types of LBEs were not equal, however are representative of the opportunities generally available in the English language (Cutler and Carter, 1987). Please refer to Liss et al. (2000) for a breakdown of the possible error sites in this set of phrases.

4The data from the two control conditions were reported in Liss et al. (2000) and are provided here for ease of comparison.

5The same criteria for words-correct were applied in all published studies related to the larger investigation (Liss et al., 1998; Liss et al., 2000; Spitzer et al., 2000).

6To provide a sufficient number of errors for the LBE analyses of the 20-phrase subset, 40 listeners were recruited for the familiarization groups (as compared to 20 for the groups from the previous investigations).


