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

Julie M. Liss[a],

Motor Speech Disorders Laboratory, Arizona State University, Box 871908, Tempe, Arizona 85281

Stephanie M. Spitzer,

Motor Speech Disorders Laboratory, Arizona State University, Box 871908, Tempe, Arizona 85281

John N. Caviness, and

Department of Neurology, Mayo Clinic–Scottsdale, Scottsdale, Arizona 85259

Charles Adler

Department of Neurology, Mayo Clinic–Scottsdale, Scottsdale, Arizona 85259

Abstract

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.

[a]julie.liss@asu.edu.

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 Metrical 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 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. Erroneous 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; Flipsen, 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-phrases 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 exposure to dysarthric speech prior to the transcription task[1] (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

---

[1]To 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).

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 inLiss *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.[2]

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

---

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

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.

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 $\times$ 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.[3] 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.[4] 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 (IS), and delete them most often before weak syllables (DW). 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

---

[3]The 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).
[4]The 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 toLiss *et al.* (2000) for a breakdown of the possible error sites in this set of phrases.

values among conditions were assessed by nonparametric Kruskal–Wallis oneway 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 fromLiss *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 *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 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 hypokinetic control (Liss *et al.*, 2000), chi-square results were significant for the data derived from the hypokinetic 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

---

[5]The data from the two control conditions were reported inLiss *et al.* (2000) and are provided here for ease of comparison.

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,$H(2)=2.45$,$P =0.029\ 43$;DW/DS,$H(2)=2.05$,$P=0.3586$], nor for the hypokinetic 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 than those obtained by nonfamiliarized listeners. When the dysarthria of the familiarization procedure matched the dysarthria 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 hypokinetic and ataxic phrases evidenced familiarization 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 representative 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 impressive 3.5% advantage for the hypokinetic tape. Listeners who heard then transcribed ataxic speech benefited more from their exposure than did listeners who heard and then transcribed hypokinetic speech.

This pattern is even more robust in the subset of data from the 20 low-intelligibility phrases for which the familiarization procedure produced greater gains. The ataxic familiarized listeners who transcribed the ataxic phrases enjoyed a 21% advantage over those who were not familiarized, while the hypokinetic familiarized realized a 15.6% gain over those who transcribed the hypokinetic 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 gain for the familiarized with other condition. Thus, the *general* benefit of exposure to dysarthric speech was of the same magnitude for transcribers of both dysarthric tapes, but the *specific* benefits were substantially higher for the ataxic familiarization group than for the hypokinetic familiarization group.

If only the intelligibility data were considered, it could only be concluded that *something* about exposure to the ataxic speech was disproportionately beneficial to listeners who transcribed the ataxic phrases. However, the second clue of LBE patterns allows us to speculate that the cognitive–perceptual source of benefit may not lie in the mapping of suprasegmental patterning. The LBE patterns elicited by the hypokinetic familiarized and ataxic familiarized conditions were nearly identical to those of the hypokinetic control and

ataxic control, respectively. This finding is quite unexpected because it suggests that the phenomenon underlying the LBE patterns must be extraordinarily robust to persist across different listener groups and listening circumstances (see Cutler *et al.*, 1997). If the assumption is correct that LBE analysis offers a window to lexical segmentation strategies relative to syllabic strength, it must be concluded that this particular familiarization procedure did not facilitate recognition of syllabic strength contrasts. As in the previous study, listeners in the current study appeared to have less success applying this strategy to the ataxic speech than to the hypokinetic speech.

This points toward the possibility that the familiarization procedure allowed the listeners to map aspects of the degraded acoustic signal onto pre-existing phonological templates. This hypothesis has been explicated in the psycholinguistic 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 secondary 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 arranged 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 produced 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 demonstrated that although the two familiarized groups outperformed the control group, there was no significant performance difference between the Paragraph and Word groups. Although the high degree of variability in listener scores may have masked actual differences, perhaps the segmental information gleaned from both familiarization tapes accounted for the majority of familiarization benefit.

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 perceptual 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 Spanish, and from Italian and French, which are rhythmically (syllable-timed) and phonologically similar to Spanish. The listeners did not demonstrate significant gains in intelligibility 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 Dupoux 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 phonologic 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

References

Cutler A, Butterfield S. Rhythmic cues to speech segmentation: Evidence from juncture misperception. J. Mem. Lang. 1992; 31:218–236.

Cutler A, Carter DM. The predominance of strong syllables in the English vocabulary. Comput. Speech Lang. 1987; 2:133–142.

Cutler A, Dahan D, van Donselaar W. Prosody in the comprehension of spoken language: A literature review. Lang. Speech. 1997; 40:141–201. [PubMed: 9509577]

Cutler A, Norris D. The role of strong syllables in segmentation for lexical access. J. Exp. Psychol. Hum. Percept. Perform. 1988; 14:113–121.

Dagenais PA, Watts CR, Tarnage LM, Kennedy S. Intelligibility—acceptability of moderately dysarthric speech by three types of listeners. J. Med. Speech-Lang. Path. 1999; 7:91–96.

Darley F, Aronson A, Brown J. Differential diagnostic patterns of dysarthria. J. Speech Hear. Res. 1969; 12:246–269. [PubMed: 5808852]

DePaul R, Kent RD. A longitudinal case study of ALS: Effects of listener familiarity and proficiency on intelligibility judgements. Am. J. Speech-Lang. Path. 2000; 9:230–240.

Duffy, JR. Motor Speech Disorders. St. Louis: Mosby; 1995.

Dupoux E, Green K. Perceptual adjustments to highly compressed speech: Effects of talker and rate changes. J. Exp. Psychol. Hum. Percept. Perform. 1997; 23:914–927. [PubMed: 9180050]

Dupoux E, Pallier C, Kakehi K, Mehler J. New evidence for prelexical phonological processing in word recognition. Language Cognitive Processes. 2001; 16:491–505.

Fear BD, Cutler A, Butterfield S. The strong/weak syllable distinction in English. J. Acoust. Soc. Am. 1995; 97:1893–1904. [PubMed: 7699171]

Flipsen PJ. Speaker-listener familiarity: Parents as judges of delayed speech intelligibility. J. Commun. Dis. 1995; 28:3–19.

Francis AL, Baldwin K, Nusbaum HC. Effects of training on attention to acoustic cues. Percept. Psychophys. 2000; 62:1668–1680. [PubMed: 11140187]

Greenspan SL, Nusbaum HC, Pisoni DB. Perceptual learning of synthetic speech produced by rule. J. Exp. Psychol. Learn. Mem. Cogn. 1988; 14:421–433. [PubMed: 2969941]

Guenther FH, Husain FT, Cohen MA, Shinn-Cunningham BG. Effects of categorization and discrimination training on auditory perceptual space. J. Acoust. Soc. Am. 1999; 106:2900–2912. [PubMed: 10573904]

Lehiste I. The timing of utterances and linguistic boundaries. J. Acoust. Soc. Am. 1972; 51:2018–2024.

Lindblom, B. Explaining phonetic variation: A sketch of the H and H theory. In: Hardcastle, WJ.; Marchal, A., editors. Speech Production and Speech Modeling. The Netherlands: Kluwer Academic; 1990. p. 403-439.

Liss JM, Spitzer SM, Caviness JN, Adler C, Edwards B. Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. J. Acoust. Soc. Am. 1998; 104:2457–2466. [PubMed: 10491707]

Liss JM, Spitzer SM, Caviness JN, Adler CA. Lexical boundary decisions in the perception of hypokinetic and ataxic dysarthric speech. J. Acoust. Soc. Am. 2000; 107:3415–3424. [PubMed: 10875386]

Logan JS, Lively SE, Pisoni DB. Training Japanese listeners to identify English /r/ and /l/: A first report. J. Acoust. Soc. Am. 1991; 82:874–885. [PubMed: 2016438]

McQueen JM, Cutler A. Spoken word access processes: An introduction. Language Cognitive Processes. 2001; 16:469–490.

Munro MJ, Derwing TM. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. Lang. Speech. 1995; 38:289–306. [PubMed: 8816082]

Nakatani LH, Schaffer JA. Hearing "words" without words: Prosodic cues for word perception. J. Acoust. Soc. Am. 1978; 63:234–245. [PubMed: 632416]

Pisoni DB, Manous LM, Dedina MJ. Comprehension of natural and synthetic speech: effects of predictability on the verification of sentences controlled for intelligibility. Comput. Speech Lang. 1987; 2:303–320. [PubMed: 23226919]

Robinson K, Summerfield AQ. Adult auditory learning and training. Ear Hear. 1996; 17:51S–65S. [PubMed: 8807276]

Rosen S, Faulkner A, Wilkinson L. Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. J. Acoust. Soc. Am. 1999; 106:3629–3636. [PubMed: 10615701]

Schwab EC, Nusbaum HC, Pisoni DB. Some effects of training on the perception of synthetic speech. Hum. Factors. 1985; 27:395–4081. [PubMed: 2936671]

Seastián-Gallés N, Dupoux E, Costa A, Mehler J. Adaptation to time-compressed speech: Phonological determinants. Percept. Psychophys. 2000; 62:834–842. [PubMed: 10883588]

Spitzer SM, Liss JM, Caviness JN, Adler C. An exploration of familiarization effects in the perception of hypokinetic and ataxic dysarthric speech. J. Med. Speech-Lang. Path. 2000; 8:285–293.

Tjaden KK, Liss JM. The role of listener familiarity in the perception of dysarthric speech. Clin. Ling. Phon. 1995; 9:139–154.

Tremblay K, Kraus N, Carrell TD, McGee T. Central auditory system plasticity: Generalization to novel stimuli following listening training. J. Acoust. Soc. Am. 1997; 102:3762–3773. [PubMed: 9407668]

Yorkston, KM.; Beukelman, DR. The influence of judge familiarization with the speaker on dysarthric speech intelligibility. In: Berry, W., editor. Clinical Dysarthria. San Diego, CA: College-Hill Press; 1983. p. 155-164.
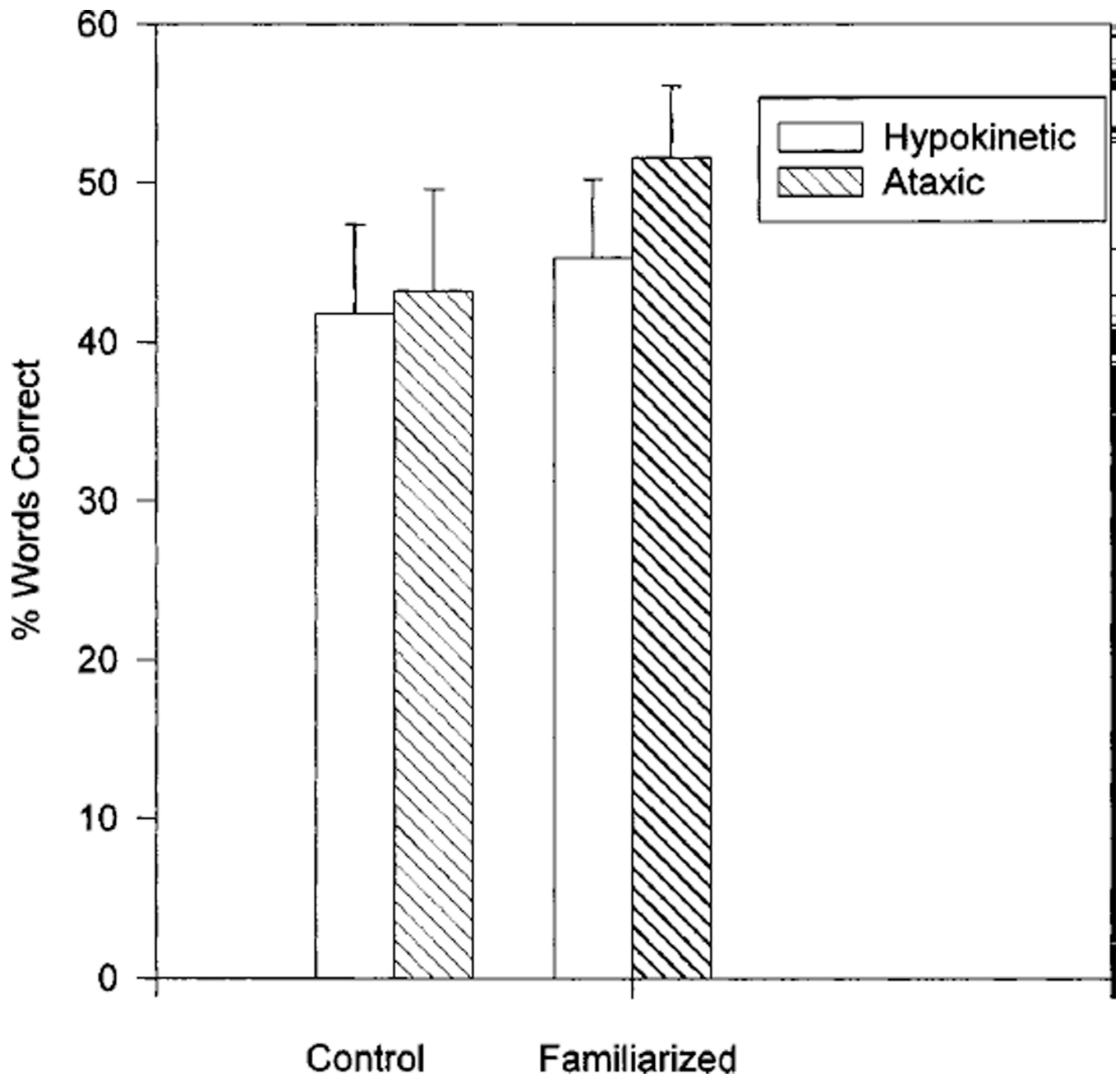
FIG. 1.
Intelligibility scores for the 60 phrases. The first two bars show the mean (+1 standard deviation) for the control groups, and the second bars show the same values for the familiarized groups. Means of the hypokinetic familiarized and ataxic familiarized were significantly higher than their respective control group ($P<0.05$).
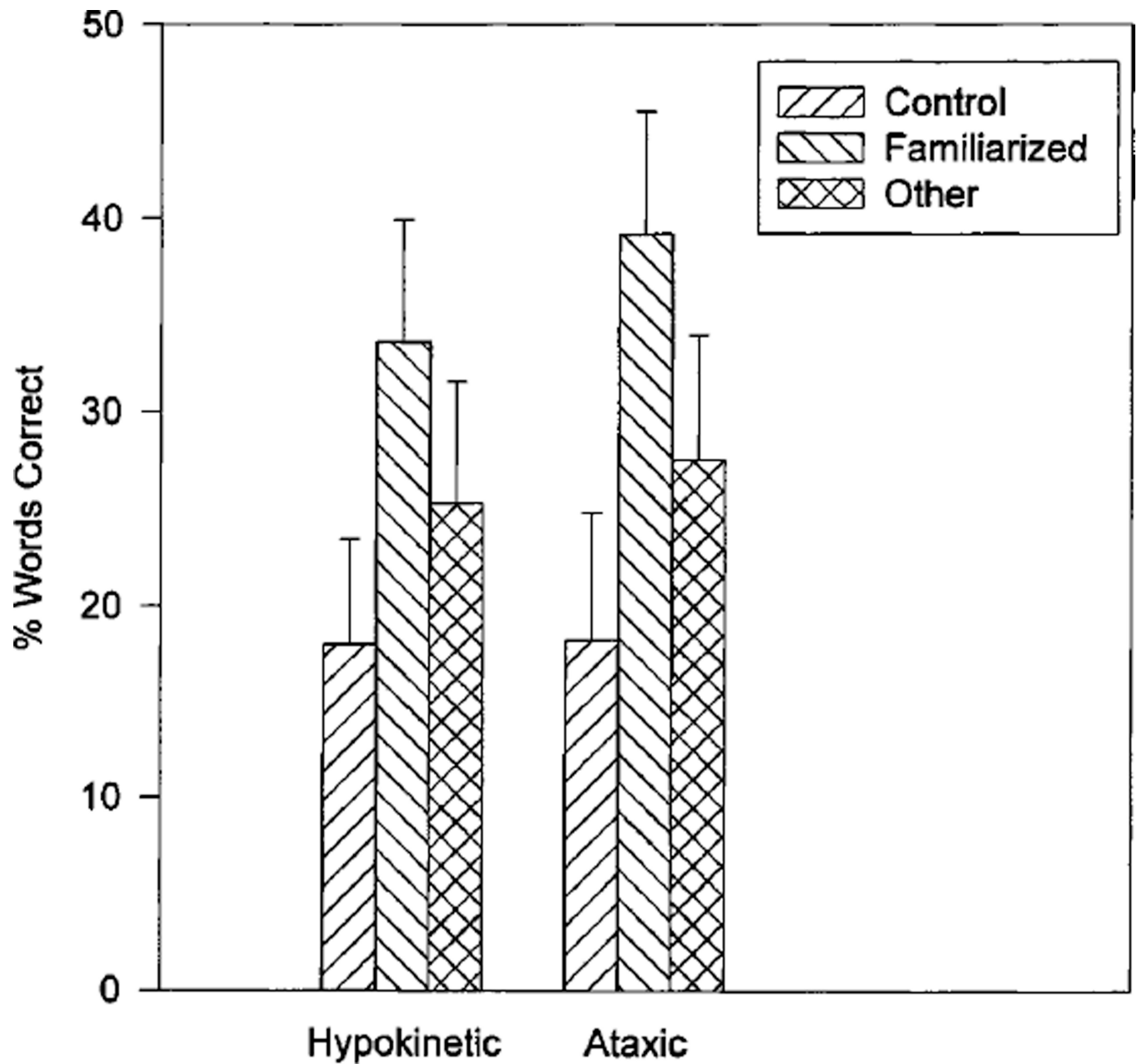
FIG. 2.
Intelligibility scores for the 20-phrase subset. The first three bars show the mean (+1 standard deviation) for the control, familiarized, and familiarized with other conditions for the hypokinetic tape. The second three bars show the same information for the ataxic tape. For each tape, significant differences were found among all three conditions (*P*<0.05).

TABLE I

Comparisons of interest.

60 phrases (between-group):
Hypokinetic control versus hypokinetic familiarized
Ataxic control versus ataxic familiarized
20 phrases (between-group):
Hypokinetic control versus hypokinetic familiarized versus familiarized with other
Ataxic control versus ataxic familiarized versus familiarized with other

TABLE II

Examples of coding lexical boundary errors from the listeners' transcriptions.[a]

Target phase	Listener response	Error type(s)
Younger rusty viewers	Younger rest if you are	IW, IW
The rally found some light	The real effects of light	IW, DS
Soon the men were asking	Cinnamon were asking	DW, DS
His display collects it	Hands did spray collected	IS, DW
Friendly slogans catch it	Fred eats slow with ketchup	IW, IW, DW
Govern proper landings	Car and prop for landings	IW, IW
Call a random voter	Coming from the motor	DW, IW
A term arranged inside	A turmoil raged inside	DW, IS
Convince the council here	In winds the cows will here	IS, IW
She describes a nuisance	Strangers cause a nuisance	DW, IS

[a]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.

TABLE III

LBE totals, category proportions expressed in percentages, and median ratio values for each of the four listener groups for the 60 phrases.

	Total LBEs	%IS	%IW	%DS	%DW	Median IS/IW	Median DW/DS
Hypokinetic control (*N* = 20)	820	47.1	27.4	9.0	16.5	1.75	1.83
Hypokinetic familiarized (*N* = 40)	1633	44.9	27.1	10.3	17.6	1.56	1.75
Ataxic control (*N* = 20)	610	40.5	34.4	12.3	12.8	1.18	0.90
Ataxic familiarized (*N* = 40)	1266	36.0	34.0	14.5	15.4	1.04	1.00

TABLE IV

LBE totals, category proportions expressed in percentages, and median ratio values for the 20-phrase subset for the hypokinetic tapes.

	Total LBEs	%IS	%IW	%DS	%DW	Median IS/IW	Median DW/DS
Control (*N* = 20)	434	47.0	22.1	9.0	21.9	1.81	1.72
Familiarized (*N* = 40)	710	39.1	30.0	12.7	18.2	1.56	1.75
Familiarized with other (*N*= 40)	883	44.8	21.4	12.5	21.3	2.00	1.63

TABLE V

LBE totals, category proportions expressed in percentages, and median ratio values for the 20-phrase subset for the ataxic tapes.

	Total LBEs	%IS	%IW	%DW	%DS	Median IS/IW	Median DW/DS
Control (N = 20)	284	39.1	34.5	12.3	14.1	1.05	1.08
Familiarized (N = 40)	585	35.7	32.9	15.7	15.7	1.04	1.00
Familiarized with other (N = 40)	744	35.9	35.6	12.6	15.9	1.00	1.00