Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech

Stephanie A. Borrie[1,2], **Megan J. McAuliffe**[1,2], **Julie M. Liss**[3], **Cecilia Kirk**[4], **Gregory A. O'Beirne**[1,2], and **Tim Anderson**[5]

[1]Department of Communication Disorders, University of Canterbury, Christchurch, New Zealand [2]New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, Christchurch, New Zealand [3]Department of Speech and Hearing Science, Arizona State University, Tempe, AZ, USA [4]Department of Special Education and Clinical Services, University of Oregon, Eugene, OR, USA [5]Van der Veer Institute for Parkinson's and Brain Research, Christchurch, New Zealand

Abstract

This investigation evaluated the familiarisation conditions required to promote subsequent and more long-term improvements in perceptual processing of dysarthric speech and examined the cognitive-perceptual processes that may underlie the experience-evoked learning response. Sixty listeners were randomly allocated to one of three experimental groups and were familiarised under the following conditions: (1) neurologically intact speech (control), (2) dysarthric speech (passive familiarisation), and (3) dysarthric speech coupled with written information (explicit familiarisation). All listeners completed an identical phrase transcription task immediately following familiarisation, and listeners familiarised with dysarthric speech also completed a follow-up phrase transcription task 7 days later. Listener transcripts were analysed for a measure of intelligibility (percent words correct), as well as error patterns at a segmental (percent syllable resemblance) and suprasegmental (lexical boundary errors) level of perceptual processing. The study found that intelligibility scores for listeners familiarised with dysarthric speech were significantly greater than those of the control group, with the greatest and most robust gains afforded by the explicit familiarisation condition. Relative perceptual gains in detecting phonetic and prosodic aspects of the signal varied dependent upon the familiarisation conditions, suggesting that passive familiarisation may recruit a different learning mechanism to that of a more explicit familiarisation experience involving supplementary written information. It appears that decisions regarding resource allocation during subsequent processing of dysarthric speech may be informed by the information afforded by the conditions of familiarisation.

Keywords

Dysarthria; Perceptual learning; Speech perception

Perceptual performance can improve with experience and listeners can become better at perceiving a speech signal that is initially difficult to understand (e.g., Davis, Johnsrude, Herrvais-Adelman, Taylor, & McGettigan, 2005; Francis, Nusbaum, & Fenn, 2007). This experience-evoked capacity to retune or adapt the speech perception system, known as

Correspondence should be addressed to: Stephanie A. Borrie, Department of Communication Disorders, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. steph.borrie@gmail.com.

perceptual learning, is defined as "relatively long-lasting changes to an organisms perceptual system that improves its ability to respond to its environment and are caused by this environment" (Goldstone, 1998, p. 585). According to interactive models of speech perception, an individual's perceptual system is flexible and dynamically adjusts to successfully navigate the incoming acoustic information (e.g., McClelland, Mirman, & Holt, 2006; Norris, McQueen, & Cutler, 2003).

Evidence for an adaptable speech perception system has been demonstrated in numerous studies. These have investigated perceptual learning with a variety of speech signals that vary significantly along multiple acoustic dimensions to that of typically encountered speech, including foreign-accented (e.g., Bradlow & Bent, 2008; Weill, 2001) and hearing-impaired speech (e.g., McGarr, 1983), as well as artificially manipulated speech signals such as those that have been noise-vocoded (e.g., Davis & Johnsrude, 2007; Davis et al., 2005), computer-synthesised (e.g., Francis, Nusbaum, & Fenn, 2007; Greenspan, Nusbaum, & Pisoni, 1988; Hoover, Reichle, Van Tasell, & Cole, 1987), and time-compressed (e.g., Golomb, Peelle, & Wingfield, 2007; Pallier, Sebastian-Galles, Dupoux, & Christophe, 1998). Taken together, this body of research provides substantial evidence that experience with atypical speech can facilitate improved recognition of the signal during subsequent encounters.

Although debate continues regarding the source of perceptual benefit (see Borrie, McAuliffe, & Liss, in press), it is commonly assumed that listeners extract regularities in the atypical acoustic pattern that facilitates or accommodates subsequent processing. Research using healthy speech variants or laboratory modified speech provide excellent examples of this regularity, wherein segmental and/or suprasegmental aspects of these speech signal vary in consistent ways. However, the acoustic degradation that characterises the speech of those with neurological disease or injury varies in both systematic and nonsystematic ways.

Neurological conditions may manifest in a variety of atypical speech patterns, termed dysarthrias (Duffy, 2005). Produced upon a platform of impaired muscle tone, inadequate respiratory support, phonatory instability, and deficient articulatory movement, breakdowns in the speech of individuals with dysarthria frequently occur in irregular and unpredictable ways. Phonemes produced adequately in one context may be distorted or omitted in the next word, speech may deteriorate in a mumbled rush of speech at the end of a sentence, and voicing may break or cease intermittently. Despite this nonsystematic acoustic variation, a small number of studies have demonstrated improved word recognition for listeners familiarised with dysarthric speech (e.g., Liss, Spitzer, Caviness, & Adler, 2002). Such findings suggest that at least some aspect of the dysarthric speech signal may be *learnable*.

The clinical significance of improved recognition of dysarthric speech should not be underestimated. Dysarthria very rarely occurs in isolation. Physical, cognitive, and memory deficits commonly co-occur, all of which can greatly reduce the individual's capacity to learn and maintain benefits from the more traditionally employed speaker-oriented interventions (Duffy, 2005). An opportunity exists to develop treatments that bypass these speaker limitations; instead, focusing on the neurologically intact listener (e.g., family members, friends, or carers). Investigation into perceptual learning of dysarthric speech may prove key to the developments of such treatments and hence, the optimisation of communication success for this speaker population (see Borrie et al., in press).

While research with other forms of atypical speech has afforded a strong consensus that experience can facilitate improved signal processing, current experimental evidence regarding perceptual learning of dysarthric speech is limited and findings have been equivocal (see Borrie et al., in press, for a full review of the literature). A significant

methodological variation across the existing research is found in the type of familiarisation conditions employed. Some studies have utilised a relatively passive familiarisation approach, whereby listeners are presented with auditory productions of the degraded speech only (e.g., Garcia & Cannito, 1996; Hustad & Cahill, 2003). In contrast, other studies have employed a more explicit familiarisation experience, in which the degraded auditory productions are supplemented with written feedback of the spoken targets (e.g., Liss et al., 2002; Spitzer, Liss, Caviness, & Adler, 2000). Mixed findings have been reported. There is some evidence that passive familiarisation may facilitate intelligibility improvements (Hustad & Cahill, 2003). However, other studies have failed to observe a perceptual benefit for listeners familiarised with dysarthric speech under passive conditions (Garcia & Cannito, 1996; Yorkston & Beukelman, 1983). Similarly with explicit familiarisation, some studies have identified significant performance gains (D'Innocenzo, Tjaden, & Greenman, 2006; Liss et al., 2002; Spitzer et al., 2000; Tjaden & Liss, 1995), whereas others have not (Yorkston & Beukelman, 1983). To date, only one study has directly compared intelligibility scores following passive versus explicit familiarisation (Yorkston & Beukelman, 1983). This study, limited by small participant numbers, found no significant difference when word recognition scores were compared across the experimental conditions: passive familiarisation (n=3), explicit familiarisation ($n = 3$), and no familiarisation (n=3). Currently, knowledge of the conditions required to induce perceptual learning of dysarthric speech is largely undefined.

Taking a traditional view of speech perception, we can hypothesise that the learnable and useful regularities that characterise dysarthric speech will facilitate the perceptual processes of lexical segmentation, lexical activation, and lexical competition (see Jusczyk & Luce, 2002). One could imagine, for example, that prior exposure to the rapid articulation rate may allow listeners to modify their expectations or internal representations of phoneme duration which in turn, may reduce ambiguity and facilitate lexical activation and competition. Or perhaps experience with the reduced variation in fundamental frequency facilitates lexical segmentation by encouraging increased attention towards alternative syllabic strength cues. While both segmental and suprasegmental learning have been postulated, few studies have attempted to document the cognitive-perceptual processes that underpin improved processing of dysarthric speech and existing findings have not led to clear answers.

Liss et al. (2002) hypothesised that a brief familiarisation procedure with either hypokinetic or ataxic dysarthria—two distinctly different forms of dysarthric speech—would improve intelligibility, as measured by words correct, and that these gains may be traced to improved lexical segmentation (and hence enhanced lexical activation and competition). This work borrowed predictions from the Metrical Segmentation Strategy (MSS), which claims that when segmental information affords insufficient cues, listeners will exploit prosodic properties of the signal to predict the onset of a new word (Cutler & Butterfield, 1992; Cutler & Norris, 1988; see also Mattys, White, & Melhorn, 2005). Based on the statistical probabilities of the English language, speech segmentation will be largely successful if listeners treat strong syllables (those receiving relative stress through longer duration, fundamental frequency change, increased loudness, and relatively full vowel) as word onsets (Cutler & Carter, 1987). Evidence of this perceptual strategy can be found in a listener's lexical boundary error (LBE) patterns—manifested in the tendency to mistakenly insert lexical boundaries before strong syllables and mistakenly delete boundaries before weak syllables. Because the two dysarthrias targeted in Liss' work had distinctly different types of prosodic degradation, it was anticipated that the dysarthria type would pose different perceptual challenges to the application of the MSS, and that familiarisation would mitigate these challenges by facilitating lexical segmentation strategies. However, in spite of significant intelligibility improvements for both dysarthria types, the LBE patterns for familiarised listeners did not differ from those of nonfamiliarised control groups. This

implies that the intelligibility gains were not sourced from an improved ability to detect syllabic stress as a cue to promote lexical segmentation. Alternatively, the 18 familiarisation phrases employed may have been insufficient to elicit perceptual changes in processing of suprasegmental information. A post-hoc segmental exploration of their data revealed that word substitution errors (in which no lexical boundaries were violated) contained a higher portion of target consonants for the familiarised listeners than those unfamiliarised, but that this finding held only for one form of dysarthria (ataxic) (Spitzer et al., 2000). This suggests that the source of benefit may vary dependent upon the type of signal to be learned, however the link at this point is unclear.

Current study

The purpose of the present study was to examine the familiarisation conditions required to promote subsequent and more long-term improvements in perceptual processing of dysarthric speech and to examine the source of these intelligibility benefits. The following four questions were addressed: (1) Do listeners familiarised with dysarthric speech achieve higher intelligibility scores relative to listeners familiarised with neurologically intact speech; (2) Is there an effect of familiarisation condition, in which the magnitude of perceptual gain is regulated by the type of familiarisation experience (passive versus explicit); (3) Do perceptual gains remain stable after a period of 7 days in which no further dysarthric speech input is received; and (4) Are perceptual gains accompanied by changes at the segmental and/or the suprasegmental level of cognitive-perceptual processing? Hypokinetic dysarthria—the speech disorder common to Parkinson's disease (PD)—was targeted for this investigation as it presents an acoustic signal in which both segmental (imprecise articulation) and suprasegmental (monopitch, reduced stress, monoloudness, and short rushes of speech) properties are significantly compromised. An audio example of hypokinetic dysarthria of PD in American English can be found at http://www.asu.edu/clas/shs/liss/Motor_Speech_DIsorders_Lab/Sound_Files.html.

METHOD

Study overview

A between-group design was used to investigate perceptual learning effects associated with different familiarisation conditions. Three groups of listeners were familiarised with passage readings under one of three experimental conditions: (1) neurologically intact speech (control), (2) dysarthric speech (passive familiarisation), or (3) dysarthric speech coupled with written information (explicit familiarisation). Following familiarisation, all listeners completed an identical phrase transcription task. Listeners familiarised with dysarthric speech returned 7 days later and completed a second transcription task involving novel phrases.

Listeners

Data were collected from 60 young healthy individuals (47 females and 13 males) aged 19 to 40 years ($M = 25.53$; $SD = 5.2$). All listener participants were native speakers of New Zealand English (NZE), passed a pure tone hearing screen at 20 dB HL for 1000, 2000, and 4000 Hz and at 30 dB HL for 500 Hz bilaterally, reported no significant history of contact with persons having motor speech disorders, and reported no identified language, learning, or cognitive disabilities. Listener participants were recruited from first year undergraduate classes and the local community.

Speech stimuli

Three male native speakers of NZE, with moderate hypokinetic dysarthria and a primary diagnosis of PD, and three male native speakers of NZE with neurologically intact speech (controls) provided the speech stimuli for the present study. The speakers ranged in age from 70 to 77 years, with a mean age of 72 years. Further details of the speakers are provided in Table 1.

Speakers with neurologically degraded speech were selected for the current study based on speech features characteristic of hypokinetic dysarthria. The operational definition of hypokinetic dysarthria was similar to that of Liss, Spitzer, Caviness, Adler, and Edwards (1998) and derived from the Mayo Classification System (Darley, Aronson, & Brown, 1969; Duffy, 2005). Under this definition, speakers must exhibit a perceptually rapid speaking rate, monopitch, monoloudness, reduced syllable stress, imprecise consonants, and perhaps a weak and breathy voice. The presence of these perceptual impressions were judged independently by three speech-language pathologists associated with the study (SB, MM, JL) and were verified objectively by relevant acoustic measures. The present study also required the impaired speech signal to fall within a tightly constrained operational definition of a moderate intelligibility impairment—defined as a score between 65% and 75% words correct on the Sentence Intelligibility Test[1] (SIT; Yorkston, Beukelman, & Hakel, 1996).

An initial pool of 43 individuals with hypokinetic dysarthria was identified from neurologist recommendations and local speech-language therapy clinics as potential speakers. Speech screening was conducted via telephone, and a total of nine individuals were identified as broadly fitting the selection criteria and subsequently completed a full speech evaluation. Of the nine speakers assessed, three individuals exhibited the speech characteristics and degree of impairment as described by the operational definition of a moderate hypokinetic dysarthria. Thus, the three speakers who provided speech samples for the current study presented with highly similar segmental and suprasegmental acoustic degradation. Control speakers were selected according to the following criteria: (1) speakers of NZE, (2) male, and (3) age-matched (within 2 months) to one of the three speakers with dysarthria. The three control speakers used in the current study reported no history of neurological injury or disease, or any speech, language, hearing, or voice disorder.

Speech samples were collected in a sound-attenuated booth with a head-mounted microphone at a 5 cm mouth-to-microphone distance. Speech output elicited during the speech tasks was recorded digitally to a laptop computer using *Sony Sound Forge* (v 9.0, Madison Media Software, Madison, WI) at 48 kHz (16 bit sampling rate) and stored as individual.wav files on a laptop. Samples included: (1) 15 sentences that comprised the SIT (Yorkston et al., 1996), (2) a standard passage reading, the Rainbow Passage (Fairbanks, 1960), and (3) 72 experimental phrases. Speech stimuli for the three speech tasks were presented to the speakers via a PowerPoint presentation displayed on a second laptop positioned directly in front of the speakers. During the production of the passage reading and experimental phrases, speakers were encouraged to use their "normal conversational" voice.

The experimental phrases were modelled on the work of Cutler and Butterfield (1992), which hypothesised that listeners rely on syllable strength to determine lexical boundaries during perception of connected speech. Each phrase consisted of six syllables and alternated phrasal stress patterns to enable LBEs to be interpreted relative to syllabic strength. Half the phrases were trochaic and alternated strong-weak (SWSWSW), and the other half were iambic and alternated weak-strong (WSWSWS). The majority of the strong syllables

---

[1]Only speakers with dysarthria completed the SIT.

contained full vowels and the majority of the weak syllables contained reduced vowels. The length of the phrases ranged from three to five words and all words were either mono- or bi-syllabic. Phrases contained correct grammatical structure but no sentence level meaning (semantically anomalous) to reduce the effects of semantic and contextual knowledge on speech perception.

A single set of 72 experimental phrases was created by selecting 24 novel experimental phrases from each of the three speakers with dysarthria. Phrasal stress patterns were balanced, so that of the 24 phrases from each speaker, 12 were trochaic and 12 were iambic in nature. Perceptual ratings of each phrase were used to ensure that each phrase included in the single speech set meet the operational definition of a moderate hypokinetic dysarthria. A second set of 72 experimental phrases was created using the corresponding control phrases produced by the neurologically intact speakers. Acoustic analysis was performed on the two sets of experimental phrases. Using *Time-Frequency Analysis* Software (TF32; Milenkovic, 2001), measures of phrase duration, fundamental frequency variation, amplitude variation, and vowel space were calculated using standard operational definitions and procedures (Peterson & Lehiste, 1960; Weismer, 1984). These metrics were chosen to validate the presence of fast rate of speech, monotone, monoloudness, and reduced syllable strength contrastivity, respectively. Table 2 summarises the phrase duration, fundamental frequency variation, and amplitude variation for the phrases produced by the speakers with dysarthria and the control speakers.

To examine vowel quality, the first (F1) and second (F2) formant frequencies were measured at the temporal midpoints of six occurrences (two productions from each of the three speakers) of the vowels /i/, /a/, and /ɔ/, using both broadband spectrograms and Linear predictive coding (LPC) displays. Mean formant values for each of the three vowels were used to calculate the vowel triangle area as an overall measure of vowel space for the speakers with dysarthria and matched controls. The vowel triangle area generated by the speakers with dysarthria was approximately 25% smaller than the area generated by the identical vowels produced by the control speakers. The perceptual impression of reduced vowel strength contrasts in the dysarthric phrases was therefore supported by the indirect measure of reduced vowel working space and the geometric area occupied by the vowel triangle derived from point vowels in strong syllables. Twenty percent of the phrases were re-measured by the first author (intra-judge) and by a second trained judge (inter-judge) to obtain reliability estimates for the acoustic metrics. Discrepancies between the re-measured data and the original data revealed that agreement was high (all r>.95), with only minor absolute differences.

Passage readings from both the speakers with hypokinetic dysarthria and matched controls were used as familiarisation material. The 72 phrases produced by the three speakers with hypokinetic dysarthria, which had been verified perceptually and acoustically, were used as test material. Two speech sets were created: initial test speech set and follow-up test speech set. The speech sets were balanced on a number of variables, including: (1) number of phrases (36 phrases); (2) number of phrases produced by each speaker (12 phrases per speaker); (3) syllable stress pattern of the phrases (six trochaic and six iambic phrases per speaker); (4) number of words and syllables; and (5) number and type of LBEs. Note that no phrase was repeated either within or across the two speech sets. Using a Brüel & Kjær Head and Torso Simulator Type 4128-C (Brüel & Kjær, Nærum, Denmark), all individual experimental stimuli.wav files and recordings of the Rainbow Passage (used for familiarisation material) were loudness calibrated to levels within ±0.1 dB. Audio presentation of all speech stimuli was set to 65 dB (A).

Procedure

The 60 listener participants were randomly assigned to one of three experimental groups, so that each group consisted of 20 participants. The three experimental groups were as follows: (1) control, (2) passive-passages, and (3) explicit-passages. The experiment was conducted in two primary phases: (1) familiarisation phase and (2) initial test phase, and the passive-passages and explicit-passages groups participated in a third (3) follow-up test phase.

The experiment was conducted in a quiet room using sound-attenuating headphones (Sennheiser HD 280 pro). Listeners were tested either individually or in pairs, located to eliminate visual distractions. The experiment was presented via a laptop computer, preloaded with the experimental procedure programmed in LabVIEW 8.20 (National Instruments, TX, USA) by one of the authors (G.O'B). Participants were told that they would undertake a listening task followed by a transcription task, and that task-specific instructions would be delivered via the computer programme. This process was employed to ensure identical stimulus presentation methods across participants.

During the *familiarisation* phase, listeners in the control group were presented with three readings of the rainbow passage, each produced by a different speaker with neurologically intact speech. To ensure each speaker was heard in each position a similar number of times, the order in which each of the 20 participants heard the three speakers was counterbalanced. For example, two of the speakers were heard in the first position seven times and one speaker six times, with similar ratios for the second and third positions. The order was then randomised using the Knuth implementation of the Fisher-Yates shuffling algorithm (Knuth, 1998). Participants were instructed to simply listen to the three speech samples. Listeners in the passive-passages group were also given the same instruction but were presented with three readings of the rainbow passage; each produced by a different speaker with dysarthria. Listeners in the explicit-passages group were presented with the same dysarthric stimuli as the passive group; however, they were provided with a written transcript of the intended targets on the computer screen and were instructed to carefully read along as they listened. The order of familiarisation material was controlled using identical procedures to that described for the control group.

Immediately following the familiarisation task, all three experimental groups participated in an identical *initial test* phase in which they transcribed the initial test speech set. Phrases were presented one at a time, and listeners were asked to listen carefully to each phrase and to type exactly what they heard. Listeners were told that all phrases contained real English words but that the phrases themselves would not make sense. They were told that some of the phrases would be difficult to understand, and that they should guess any words they did not recognise. Listeners were told to place an "X" to represent part of the phrase, if they were unable to make a guess. They were given 12 seconds to type each response. Listeners in the passive-passages and explicit-passages groups were asked to return in 7 days to participate in the *follow-up test* phase, in which they transcribed the follow-up test speech set. The 36 phrases in both the initial and follow-up test speech sets were presented randomly to each of the 60 listener participants.

Transcription analysis

The total data set consisted of 100 transcripts of 36 experimental phrases: 60 transcripts of the initial test speech set and 40 of the follow-up test speech set. The first author independently analysed the listener transcripts for three primary measures: (1) percent words correct (PWC), (2) percent syllable resemblance (PSR), and (3) the presence and type of LBEs. A PWC score, out of a total of 141 words, was tabulated for each listener transcript. From this, the mean PWC for the 20 participants in each listener group was determined. This

score reflects a measure of intelligibility for each of the experimental conditions. Words correct were defined as those that matched the intended target exactly, as well as those that differed only by the tense "ed" or the plural "s". In addition, substitutions between "a" and "the" were regarded as correct.[2]

Transcripts were also analysed using a measure of PSR in incorrectly transcribed words. This was defined as the number of syllables that contained at least 50% phonemic accuracy to the syllable target, divided by the total number of syllable errors made. Thus, to be scored as a syllable that resembled the target, syllables with two phonemes required one correct phoneme, syllables with three phonemes required two correct phonemes, syllables with four phonemes required at least two correct phonemes, and syllables with five phonemes required at least three correct phonemes. The number of syllables that resembled the target were tallied for each transcript and divided by the total number of syllables in error for that transcript, so that the final PSR score for each transcript reflected the percentage of syllable errors that resembled the correct syllable target. Mean PSR scores for each condition were calculated. In addition, transcripts were analysed for percent syllable correct (PSC) in order to examine PSR within the overall context of intelligibility. Syllables correct were defined as those that matched the intended target exactly, as well as substitutions between "a" and "the". Each 36 phrase speech set contained a total of 216 syllables.

Finally, transcripts were analysed with regards to LBEs, defined as incorrect insertions or deletions of lexical boundaries. Insertion and deletion errors were further coded for location, occurring either before a strong or before a weak syllable. Accordingly, four types of errors could be coded: (1) insert boundary before a strong syllable (IS); (2) insert boundary before a weak syllable (IW); (3) delete boundary before a strong syllable (DS); and (4) delete boundary before a weak syllable (DW) (see Liss et al., 1998, for error coding examples). LBE proportions for each error type were calculated as a percent score for each condition group at both initial and follow-up testing. In addition to the LBE proportion comparisons, IS/IW and DW/DS ratios based on the sum of group errors were calculated, again for each condition group at both initial and follow-up testing. According to Cutler and Butterfield (1992), these ratios are considered to reflect the strength of adherence to predicted error patterns: it is postulated that if listeners rely on syllabic strength to determine word boundaries, they will most likely make IS and DW errors. Thus, a ratio value of "1" reflects an equal occurrence of insertion and deletion errors before strong and weak syllables, and as the distance from "1" positively increases, so too does the strength of adherence to the predicted patterns of error.

Reliability of transcription coding

Twenty-five randomly selected transcripts were re-analysed by the first author (intra-judge) and by a second trained judge (inter-judge) to obtain reliability estimates for the coding of the three primary dependent variables. Discrepancies between the re-analysed data and the original data revealed that agreement was high (all $r>.95$), with only minor absolute differences.

RESULTS

Percent words correct

Figure 1 reflects the mean PWC scores for the three experimental groups at initial and follow-up tests. A one-way analysis of variance (ANOVA) showed a significant group effect

---

[2]The criteria for words correct were based on other published studies which have also examined listener transcripts following familiarisation with dysarthric speech (Liss et al., 2002, 1998; Liss, Spitzer, Caviness, Adler, & Edwards, 2000).

for PWC scores immediately following familiarisation, $F(2, 57) =89.15$, $p<.001$, $\eta^2=.76$. *Post-hoc* tests, using Bonferroni correction, revealed that PWC scores achieved by the explicit-passages group were significantly higher than those evident in the passive-passages group, $t(38) =5.30$, $p<.001$, $d=1.84$, and the control group, $t(38) =13.24$, $p<.001$, $d=3.76$, and that PWC scores achieved by the passive-passages group were significantly higher than those evident in the control group, $t(38) =8.09$, $p<.001$, $d=2.66$. Thus, immediate intelligibility improvements were realised for both groups familiarised with dysarthric passages, with the greatest gains observed for the listeners familiarised under explicit conditions.

Paired *t*-tests were used to examine the within-group stability of intelligibility gains over time by comparing PWC scores from the initial and follow-up tests. Comparisons revealed that the PWC scores for both the passive-passages group, $t(19) =13.94$, $p<.001$, $d=3.72$, and the explicit-passages group, $t(19) =12.48$, $p<.001$, $d=2.47$, declined significantly over the 7 day interval. When PWC scores from the passive-passages and explicit-passages groups at follow-up were compared with the control group, a one-way ANOVA revealed a significant group effect, $F(2, 57) =11.99$, $p<.001$, $\eta^2= .30$. *Post-hoc* tests, using Bonferroni correction, indicated that while the PWC scores for the passive-passages group at follow-up were similar to those evident in the control group, $t(38) =0.53$, $p =1.0$, $d=0.19$, the PWC scores for the explicit-passages group at follow-up were significantly higher than both the control group, $t(38) =4.48$, $p<.001$, $d=1.22$, and the passive-passages group, $t(38) =3.94$, $p<.001$, $d=1.37$. Thus, while intelligibility declined over 7 days for both groups familiarised with dysarthric passages, some intelligibility carry-over was observed for the listeners familiarised under explicit conditions.

Syllable resemblance

Figure 2 reflects the mean PSR scores, in addition to the mean PSC scores, for the three experimental groups at initial and follow-up tests. Pearson product-moment correlation coefficients demonstrated a strong relationship between the variables of PSC and PWC for all conditions. Accordingly, statistical analysis was performed on the PSR data only, as PSC findings are reflected in the analysis of PWC.

A one-way ANOVA on the PSR scores revealed a significant group effect immediately following familiarisation, $F(2, 57) =11.17$, $p<.001$, $\eta^2=.28$. *Post-hoc* tests, using Bonferroni correction, demonstrated that PSR scores achieved by both the passive-passages group, $t(38) =2.98$, $p =.01$, $d=1.05$, and the explicit-passages group, $t(38) =4.67$, $p<.001$, $d=1.44$, were significantly higher than the control group. There was no significant difference in PSR scores achieved by the passive-passages and explicit-passages groups, $t(38) =1.69$, $p =.29$, $d=0.50$. Thus, passive familiarisation with dysarthric passages facilitated similar benefits to a segmental measure of perceptual processing as explicit familiarisation with dysarthric passages.

Paired *t*-tests were used to examine the within-group stability of segmental gains over time by comparing PSR scores from the initial and follow-up tests. Comparisons revealed that while a small increase in the PSR scores was observed at follow-up for both groups, these differences were not significant for the passive-passages group, $t(19) =1.3$, $p =.20$, $d=0.40$, and the explicit-passages group, $t(19) =1.6$, $p =.11$, $d=.40$. When PSR scores from the passive-passages and explicit-passages groups at follow-up were compared with the control group, a one-way ANOVA revealed a significant group effect, $F(2, 57) =20.69$, $p<.001$, $\eta^2=.42$. *Post-hoc* tests, using Bonferroni correction, demonstrated that PSR scores achieved by both the passive-passages group, $t(38) =4.49$, $p<.001$, $d=1.37$, and the explicit-passages group, $t(38) =6.24$, $p<.001$, $d=2.18$, were significantly higher than the control group. There was no significant difference in PRS scores achieved by the passive-passages and explicit-

passages groups at follow-up, $t(38) = 1.75$, $p = .26$, $d = 0.50$. Taken together, the within- and between-group comparisons on the PSR data show that the benefits to a measure of segmental processing for both groups familiarised with dysarthric passages remained robust over 7 days.

LBE patterns

Table 3 contains the LBE category proportions and the sum IS/IW and DW/DS ratios for the three experimental groups at the initial and follow-up tests. Contingency tables were constructed for the total number of LBEs by error type (i.e., insertion/deletion) and error location (i.e., before strong/weak syllable) for the groups to determine whether the variables were significantly related. A within-group chi-square analysis revealed a significant interaction between the variables of type (insert/delete) and location (strong/weak) for the data generated by the control group, $\chi^2(1, N=20) = 33.15$, $p < .001$, and the explicit-passages group—both immediately following familiarisation, $\chi^2(1, N=20) = 76.95$, $p < .001$, and at follow-up, $\chi^2(1, N=20) = 128.27$, $p < .001$. In both the control and the explicit-passage groups, 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. Such LBE error patterns are predicted (Cutler & Butterfield, 1992). Ratio figures reflect the strength of adherence to these predicted error patterns. Relative to the control group, the magnitude of the IS/IW ratio was substantially greater for explicit-passages group. This indicated that listeners familiarised with dysarthric passages under explicit conditions learnt to utilise syllabic stress contrast cues to inform speech segmentation. This finding was not evidenced in the data of the passive-passages group, at either the initial or the follow-up testing. While there was a small increase in the number of erroneous lexical boundary insertions that occurred before a strong syllable relative to a weak syllable, there was a small decrease in the number of erroneous lexical boundary deletions that occurred before a weak syllable relative to a strong syllable. Differences, however, were not significant both immediately following familiarisation, $\chi^2(1, N=20) = 0.22$, $p = .71$, and at follow-up, $\chi^2(1, N=20) = 2.25$, $p = .14$. The lack of relationship between the type and location of LBEs for the passive-passages group indicated that the listeners familiarised with dysarthric passages under passive conditions did not learn to utilise syllabic stress contrast cues to inform speech segmentation.

A between-group chi-square analysis was used to examine differences in error distribution between the three experimental groups. Results identified significant differences in error distribution between the control and passive-passages groups, $\chi^2(3, N=40) = 38.98$, $p < .001$, and the passive-passages and explicit-passages groups, $\chi^2(3, N=40) = 109.19$, $p < .001$. No significant difference was found between the control and explicit-passages groups, $\chi^2(3, N=40) = 6.34$, $p = .10$. Thus, the relative distribution of errors observed for the control group were similar to those observed for the listeners familiarised with dysarthric passages under explicit conditions, but this error pattern was significantly different to that observed for the listeners familiarised with dysarthric passages under passive conditions.

DISCUSSION

The present study provides evidence of perceptual learning for listeners familiarised with dysarthric speech and enables a number of conclusions to be drawn. First, intelligibility improved substantially following a relatively brief familiarisation experience with dysarthric stimuli. Second, the magnitude and robustness of the intelligibility benefits were influenced by the familiarisation conditions. Finally, performance gains were associated with changes in the processing of both segmental and suprasegmental aspects of the degraded signal. However, the perceptual changes at these processing levels appeared to vary as a function of familiarisation condition. Such findings support a dynamic and adaptable speech perception

system, which is further discussed with regards to speech intelligibility and cognitive-perceptual processing.

Significantly higher intelligibility scores were observed for listeners familiarised with dysarthric speech compared with those familiarised with control speech. Improved processing of the dysarthric signal demonstrates that listeners can learn to better understand neurologically degraded speech. This provides evidence for a dynamic model of perceptual processing that enables online adjustments to acoustic features of dysarthric speech. Key, however, is that explicit familiarisation offered superior performance gains than those afforded by passive familiarisation, as has been previously reported with perceptual learning of noise-vocoded speech (Davis et al., 2005; Loebach, Pisoni, & Svirsky, 2010). In addition to significantly larger intelligibility benefits, explicit familiarisation also facilitated some intelligibility carry-over (at 7 days postfamiliarisation). Listeners who received passive familiarisation did not exhibit any performance gains at follow-up. From the intelligibility data, it would appear that passive familiarisation with the degraded signal alone is not sufficient to facilitate any long-term changes in perceptual processing. This likely reflects the fact that there was less learning in the passive condition because, based on the performance of the control group, only approximately 25% of the words in the phrases were recognisable. Even if limited, it has been proposed that the ability to recognise some words enables listeners to use acoustic-phonetic information to modify phonemic representations (e.g., Eisner & McQueen, 2005; Norris et al., 2003). Thus, it can be speculated that the addition of the passive-passage familiarisation allowed listeners to better exploit the 25% understandable words for an additional 13% gain. Less robust learning would lead to faster decay if, as in modular theories, learning is viewed as a temporary perceptual adjustment, allowing representations to return to preperceptual learning parameters over time (Kraljic & Samuel, 2005).

If intelligibility scores were considered in isolation, the explanation that the performance benefit associated with passive familiarisation was simply enhanced when familiarisation was more explicit could be assumed. However, error patterns at segmental and suprasegmental levels of perceptual processing revealed that intelligibility differences between experimental groups were not simply a case of the magnitude of learning. Listeners familiarised with dysarthric speech achieved a significantly higher percentage of syllables that bore phonemic resemblance to the targets (not including correctly transcribed syllables) relative to the control group. Thus, it appears that experience with dysarthric speech enabled listeners to better map acoustic-phonetic aspects of the disordered signal onto existing mental representations of speech sounds. This finding extends support for previous studies which have postulated that improved recognition of dysarthric speech is sourced from segmental level benefits (Liss et al., 2002; Spitzer et al., 2000). However, it is difficult to account for the superior intelligibility benefits observed in listeners who received explicit familiarisation, given that the PSR scores were similar for both passive and explicit familiarisation conditions. Furthermore, there appeared to be relative maintenance of the segmental benefits afforded by both passive and explicit familiarisation conditions at follow-up. The PSR scores did not diminish on day 7 for either of the familiarised groups. Thus, despite poorer words-correct intelligibility performance in the passive-passages group, the perceptual benefits to segmental processing appeared to remain. Given that word recognition scores returned to levels similar to that of the controls for passively familiarised listeners, robust improvements in phoneme perception at follow-up for this group are unexpected. Stable PSR scores in the face of a substantial intelligibility decline would serve to demonstrate that passive familiarisation to dysarthric speech does improve subsequent acoustic-phonetic mapping at 7 days following the exposure experience. If the measure of syllabic resemblance is a valid index of phoneme perception accuracy, these findings raise the possibility that learning decay may occur at different rates across different levels of

analysis. However, it is also possible that the decay in word recognition scores, to some degree, may be influenced by the amount of familiarisation listeners receive. While the quantity of familiarisation material was substantially more than the amount that is generally employed to study this phenomenon (e.g., D'Innocenzo et al., 2006; Liss et al., 2002; Tjaden & Liss, 1995), whether increased periods of familiarisation would facilitate more robust intelligibility benefits provides a valuable direction for future investigations.

Another unexpected finding calls into question the conclusion that the difference between passive and explicit familiarisation simply reflects how much the listener has learnt. Comparison of the LBE error patterns of the control and explicit-passages groups reveal expected results. Both groups made significantly more predicted (IS and DW) errors than nonpredicted (IW and DS) errors, a pattern which conforms to the MSS hypothesis (Cutler & Butterfield, 1992; Cutler & Norris, 1988). Furthermore, this pattern was stronger for the group that received explicit familiarisation than for the control group. While reduced syllable stress contrasts are a cardinal feature of hypokinetic dysarthria (Darley et al., 1969), the presence of written information during experience with the degraded signal presumably enabled listeners to learn something about the reduced and aberrant syllabic stress contrast cues by drawing attention to relevant acoustic information (e.g., Goldstone, 1998; Nosofsky, 1986). Such findings are supported by evidence that listeners relied on syllabic stress information to facilitate lexical segmentation of speech produced by individuals with hypokinetic dysarthria (Liss et al., 1998), although a relatively small familiarisation procedure in a subsequent study did not elicit significant changes in LBE error patterns (Liss et al., 2002).

The unexpected finding, then, comes with the analysis of the passive familiarisation LBE data. This group appeared to largely ignore syllabic strength contrast cues to inform speech segmentation. In contrast to listeners in the control and explicit-passages groups, listeners who received passive familiarisation were just as likely to make unpredicted errors (IW and DS) as they were to make predicted errors (IS and DW). This is a remarkable finding given that the sole difference between the passive and explicit groups was the addition of written information for listeners familiarised with dysarthric speech under explicit conditions. Furthermore, similar LBE patterns were observed for both passive and explicit groups at follow-up suggesting, perhaps, the persistence of cognitive-perceptual strategies that were engendered by each familiarisation procedure. Thus, LBE data reveals that familiarisation conditions may differentially influence learning of suprasegmental properties. The presence of written information regarding the lexical targets appeared to promote syllabic stress contrasts as an informative acoustic cue, whereas experience with degraded signal alone essentially eliminated any cognitive attention toward this prosodic information. Interestingly, this conclusion appears to be at odds with some of the perceptual learning literature that has speculated on conditions required to achieve learning. Research has identified that perceptual learning of a signal in which segmental properties have been artificially manipulated (e.g., noise-vocoded speech) may depend on knowledge of the lexical targets (e.g., Davis et al., 2005), whereas improved recognition of a signal in which the suprasegmental information has been modified (e.g., time-compressed speech) has been reported in the absence of any supplementary information regarding the degraded productions (e.g., Pallier et al., 1998; Sebastian-Galles, Dupoux, Costa, & Mehler, 2000). Future studies are needed to investigate why, with the neurologically degraded signal, segmental properties appear to be learned relatively automatically and yet attention towards suprasegmental information may necessitate more explicit learning conditions. In addition, research with other types and severities of dysarthric speech will enable a more comprehensive picture of perceptual learning processes to be established.

CONCLUSION

The current study yields empirical support for perceptual learning of dysarthric speech. There is evidence to suggest that greater and more robust performance gains are achieved when the degraded signal is supplemented with written information under explicit learning conditions. However, there is also evidence to suggest that, for this particular pattern and level of speech degradation, the learning afforded by passive familiarisation may be qualitatively different to that afforded by explicit familiarisation. Thus, the current study has revealed a possible relationship between familiarisation conditions (passive verses explicit) and subsequent processing of dysarthric speech. Further research is, however, required to validate such a speculation.

Acknowledgments

References

Borrie SA, McAuliffe MJ, Liss JM. Perceptual learning of dysarthric speech: A review of experimental studies. Journal of Speech, Language, and Hearing Research. (in press).

Bradlow AR, Bent T. Perceptual adaption to non-native speech. Cognition. 2008; 106:707–729. [PubMed: 17532315]

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

Cutler A, Carter DM. The predominance of strong initial syllables in the English vocabulary. Computer Speech and Language. 1987; 2:133–142.

Cutler A, Norris DG. The role of strong syllables in segmentation for lexical access. Journal of Experimental Psychology: Human Perception and Performance. 1988; 14:113–121.

Darley FL, Aronson AE, Brown JR. Differential diagnosis patterns of dysarthria. Journal of Speech and Hearing Research. 1969; 12:246–269. [PubMed: 5808852]

Davis MH, Johnsrude IS. Hearing speech sounds: Top-down influences on the interface between audition and speech perception. Hearing Research. 2007; 229:132–147. [PubMed: 17317056]

Davis MH, Johnsrude IS, Herrvais-Adelman A, Taylor K, McGettigan C. Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. Journal of Experimental Psychology: General. 2005; 134(2):222–241. [PubMed: 15869347]

D'Innocenzo J, Tjaden K, Greenman G. Intelligibility in dysarthria: Effects of listener familiarity and speaking condition. Clinical Linguistics and Phonetics. 2006; 20(9):659–675. [PubMed: 17342875]

Duffy, JR. Motor speech disorders: Substrates, differential diagnosis, and management. 2. St. Louis, MS: Elsevier Mosby; 2005.

Eisner F, McQueen JM. The specificity of perceptual learning in speech processing. Perception and Psychophysics. 2005; 67(2):224–238. [PubMed: 15971687]

Fairbanks, G. Voice and articulation drillbook. 2. New York, NY: Harper & Row; 1960.

Francis AL, Nusbaum HC, Fenn K. Effects of training on the acoustic-phonetic representation of synthetic speech. Journal of Speech, Language, and Hearing Research. 2007; 50:1445–1465.

Garcia, JM.; Cannito, MP. Top down influences on the intelligibility of a dysarthric speaker: Addition of natural gestures and situational context. In: Robin, D.; Yorkston, KM.; Beukelman, DR., editors. Disorders of motor speech. Baltimore, MD: Paul H. Brookes; 1996. p. 67-87.

Goldstone R. Perceptual learning. Annual Review of Psychology. 1998; 49:585–612.

Golomb JD, Peelle JE, Wingfield A. Effects of stimulus variability and adult aging on adaption to time-compressed speech. Journal of the Acoustical Society of America. 2007; 121(3):1701–1708. [PubMed: 17407906]

Greenspan SL, Nusbaum HC, Pisoni DB. Perceptual learning of synthetic speech. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1988; 14(3):421–433.

Hoover JR, Reichle J, Van Tasell D, Cole D. The intelligibility of synthesized speech: Echo versus votrax. Journal of Speech and Hearing Research. 1987; 30:425–431. [PubMed: 2959817]

Hustad KC, Cahill MA. Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. American Journal of Speech-Language Pathology. 2003; 12:198–206. [PubMed: 12828533]

Jusczyk PW, Luce PA. Speech perception and spoken word recognition: Past and present. Ear and Hearing. 2002; 23:2–40. [PubMed: 11881915]

Knuth, DE. The art of computer programming. 3. Vol. 2. Boston, MA: Addison-Wesley; 1998.

Kraljic T, Samuel AG. Perceptual learning for speech: Is there a return to normal? Cognitive Psychology. 2005; 51:141–178. [PubMed: 16095588]

Liss JM, Spitzer SM, Caviness JN, Adler C. The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. Journal of the Acoustical Society of America. 2002; 112(6):3022–3030. [PubMed: 12509024]

Liss JM, Spitzer SM, Caviness JN, Adler C, Edwards B. Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. Journal of the Acoustical Society of America. 1998; 104(4):2457–2466. [PubMed: 10491707]

Liss JM, Spitzer SM, Caviness JN, Adler CA, Edwards B. Lexical boundary error analysis in hypokinetic dysarthria. Journal of the Acoustical Society of America. 2000; 107(6):3415–3424. [PubMed: 10875386]

Loebach JL, Pisoni DB, Svirsky MA. Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant. Journal of Experimental Psychology. 2010; 36(1):224–234. [PubMed: 20121306]

Mattys SL, White L, Melhorn JF. Integration of multiple speech segmentation cues: A hierarchical framework. Journal of Experimental Psychology: General. 2005; 134(4):477–500. [PubMed: 16316287]

McClelland JL, Mirman D, Holt LL. Are there interactive processes in speech perception? Trends in Cognitive Sciences. 2006; 10:363–369. [PubMed: 16843037]

McGarr NS. The intelligibility of deaf speech to experienced and inexperienced listeners. Journal of Speech and Hearing Research. 1983; 26:451–458. [PubMed: 6645470]

Milenkovic, P. TF32: Time-frequency analysis for 32-bit windows [Computer software]. Madison: Wisconsin; 2001.

Norris D, McQueen JM, Cutler A. Perceptual learning in speech. Cognitive Psychology. 2003; 47:204–238. [PubMed: 12948518]

Nosofsky RM. Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General. 1986; 115(1):39–57. [PubMed: 2937873]

Pallier C, Sebastian-Galles N, Dupoux E, Christophe A. Perceptual adjustment to time-compressed speech: A cross-linguistic study. Memory and Cognition. 1998; 26:844–851. [PubMed: 9701975]

Peterson GE, Lehiste I. Duration of syllable nuclei in English. Journal of the Acoustical Society of America. 1960; 32:693–703.

Sebastian-Galles N, Dupoux E, Costa A, Mehler J. Adaptation to time-compressed speech: Phonological determinants. Perception and Psychophysics. 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. Journal of Medical Speech-Language Pathology. 2000; 8:285–293.

Tjaden KK, Liss JM. The role of listener familiarity in the perception of dysarthric speech. Clinical Linguistics and Phonetics. 1995; 9(2):139–154.

Weill SA. Foreign-accented speech: Encoding and generalization. Journal of the Acoustical Society of America. 2001; 109:2473 (A).

Weismer G. Acoustic descriptions of dysarthric speech: Perceptual correlates and physiological inferences. Seminars in Speech and Language. 1984; 5:293–314.

Yorkston, KM.; Beukelman, DR. The influence of judge familiarization with the speaker on dysarthric speech intelligibility. In: Berry, W., editor. Clinical dysarthria. Austin, TX: Pro-Ed; 1983. p. 155-164.

Yorkston, KM.; Beukelman, DR.; Hakel, M. Speech intelligibility test for windows. Lincoln, NE: Communication Disorders Software; 1996.
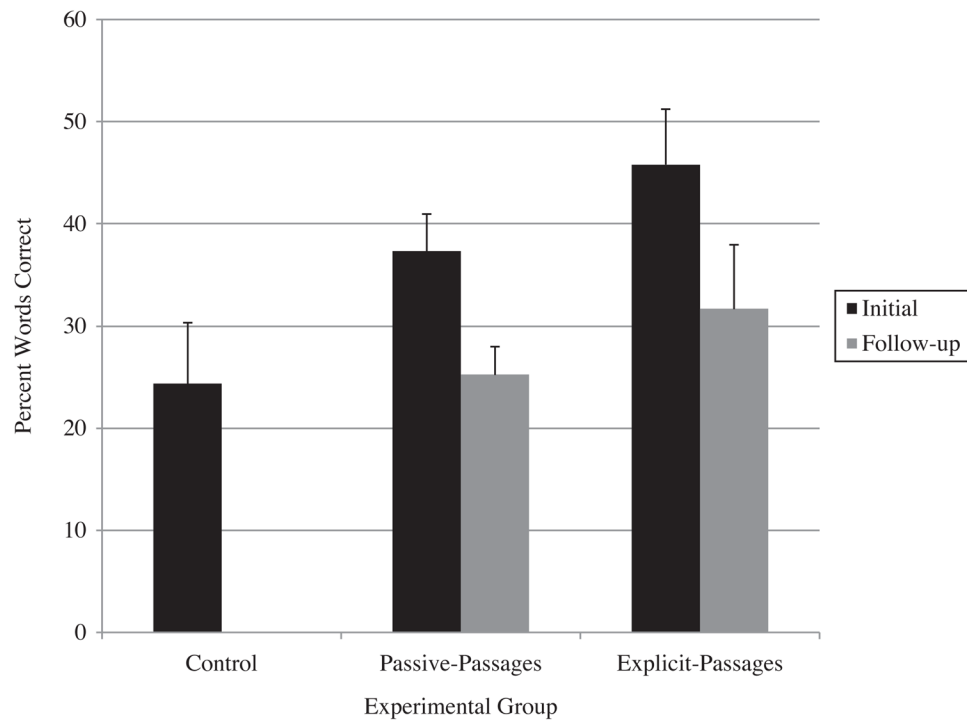
Figure 1.
Mean PWC for listeners by experimental group at the initial and follow-up tests. Bars
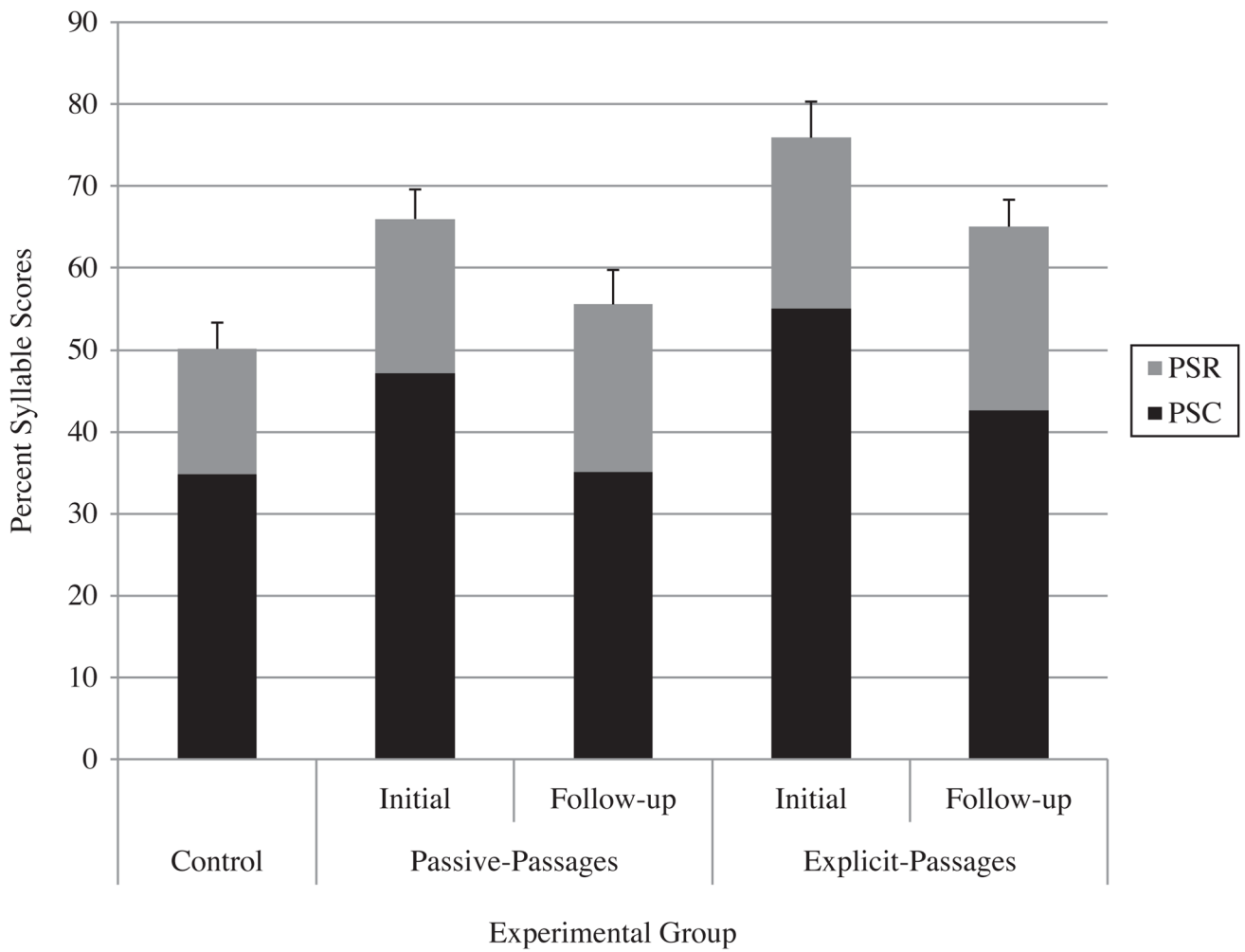delineate +1 standard deviation of the mean.

Figure 2.
Mean PSC and mean PSR for listeners by experimental group at the initial and follow-up tests. Bars delineate +1 standard deviation of the mean PSR data.

TABLE 1

Characteristics of the speakers with PD and neurologically intact controls

Speakers with dysarthria	Age	Years post-Dx	SIT score (%)	Control speakers	Age
HD1	77	12	65	CO1	77
HD2	70	11	70	CO2	71
HD3	70	13	75	CO3	70

Note: "HD" and "CO" refer to hypokinetic dysarthric and control speakers, respectively. The age of the HD speakers and the number of years that have elapsed since their diagnosis of PD (years post-Dx) are presented in the first two data columns. The third data column contains the HD speaker's scores on the SIT (Yorkston et al., 1996) as rated by one naïve listener.

TABLE 2

Mean values and independent *t*-test results of the acoustic variables of interest across the experimental phrases

	Mean values (*SD*)		
	Dysarthric speakers	**Control speakers**	t_{142}
Phase duration (ms)	1,020.09 (116.46)	2,031.37 (349.60)	22.93[*]
Pitch variation (Hz)	17.67 (4.05)	25.96 (7.23)	12.36[*]
Amplitude variation (dB)	6.73 (1.32)	10.92 (2.55)	8.34[*]

[*]
p <.001.

TABLE 3

Category proportions of LBEs expressed in percentages and sum error ratio values for listeners by experimental group

Group[a]	% IS	% IW	% DS	% DW	IS-IW Ratio	DW-DS Ratio
Control	37.15	15.84	19.55	28.21	2.4	1.4
Passive-passages	27.31	22.69	28.41	21.59	1.2	0.8
Passive-passages: FU	29.48	28.87	23.92	17.73	1.0	0.7
Explicit-passages	42.42	12.31	16.70	28.57	3.5	1.7
Explicit-passages: FU	42.12	14.95	12.06	30.87	2.8	2.6

Note: "IS", "DS", "IW" and "DW" refer to LBEs defined as insert boundary before strong syllable, delete boundary before strong syllable, insert boundary before weak syllable, and delete boundary before weak syllable, respectively. FU, Follow-up.

[a] *n*=20.