A follow-up investigation into the mechanisms that underlie improved recognition of dysarthric speech

Stephanie A. Borrie[a)] and Megan J. McAuliffe
Department of Communication Disorders and New Zealand Institute of Language,
Brain and Behaviour, University of Canterbury, Private Bag 4800, Christchurch 8140,
New Zealand
steph.borrie@gmail.com, megan.mcauliffe@canterbury.ac.nz

Julie M. Liss
Department of Speech and Hearing Sciences, Arizona State University,
P.O. Box 870102, Tempe, Arizona 85287-0102
julie.liss@asu.edu

Greg A. O'Beirne
Department of Communication Disorders and New Zealand Institute of Language,
Brain and Behaviour, University of Canterbury, Private Bag 4800, Christchurch 8140,
New Zealand
gregory.obeirne@canterbury.ac.nz

Tim J. Anderson
New Zealand Brain Research Institute, 66 Stewart St, Christchurch 8011, New Zealand
Tim.Anderson@cdhb.govt.nz

Abstract:　　Differences in perceptual strategies for lexical segmentation of moderate hypokinetic dysarthric speech, apparently related to the conditions of the familiarization procedure, have been previously reported [Borrie *et al.*, Language and Cognitive Processes (2012)]. The current follow-up investigation examined whether this difference was also observed when familiarization stimuli highlighted syllabic strength contrast cues. Forty listeners completed an identical transcription task following familiarization with dysarthric phrases presented under either passive or explicit learning conditions. Lexical boundary error patterns revealed that syllabic strength cues were exploited in both familiarization conditions. Comparisons with data previously reported afford further insight into perceptual learning of dysarthric speech.

1. Introduction

Fundamental to recognizing spoken language is the ability to perform *lexical segmentation*, the perceptual process that enables a continuous stream of acoustic energy to be parsed into its individual word components (Jusczyk and Luce, 2002). Most recent accounts of lexical segmentation assume an integrative framework in which listeners can employ both lexically-driven (e.g., semantic plausibility, syntactic rules) and sublexically-driven (e.g., phonotactic probabilities, syllabic stress) strategies to segment speech (McQueen and Cutler, 2001). Based on the assumption that listeners will utilize the most economical means to achieve lexical segmentation, it is postulated that perceptual strategy use is dependent upon the quality of the acoustic signal and the richness of the

---

[a)]Author to whom correspondence should be addressed. Current address: Department of Speech and Hearing Science, Arizona State University, Tempe, AZ 85287-0102.

contextual information (Cutler and Butterfield, 1992; Cutler and Norris, 1988; Mattys *et al.*, 2005). The Metrical Segmentation Strategy (MSS) asserts that when segmental information is ambiguous (i.e., affords insufficient perceptual cues), listeners will exploit prosodic properties of the acoustic signal to inform word boundary decisions (Cutler and Butterfield, 1992; Cutler and Norris, 1988). Specifically, the presence of strong syllables will dictate the onset of a new word. Evidence in support of the MSS is observed in the lexical boundary error (LBE) patterns of listener transcriptions of degraded speech—manifested in the tendency to mistakenly insert word boundaries before strong syllables and delete word boundaries before weak syllables (Cutler and Butterfield, 1992).

Recent research investigating perceptual learning of dysarthric speech found that not only did magnitude of intelligibility gain depend on whether familiarization was passive (dysarthric speech) or explicit (dysarthric speech coupled with written information), but condition discrepancies were also observed in speech segmentation strategies (Borrie *et al.*, 2011). Listeners familiarized with reading passages produced by individuals with dysarthria under explicit conditions conformed strongly to MSS predictions. That is, syllabic stress contrast cues were exploited to inform decisions regarding the location of word boundaries. In contrast, listeners familiarized with reading passages produced by individuals with dysarthria under passive conditions did not adhere to the predicted error patterns, revealing a perceptual shift away from these prosodic segmentation cues. Borrie *et al.* speculated that perceptual learning afforded by passive familiarization may be qualitatively different from that which occurs with explicit familiarization.

The current study investigated the empirical basis for this premise and examined whether performance differences associated with passive and explicit familiarization are more than simply a magnitude of benefit. If perceptual strategy use is indeed dependent upon the presence or absence of written information at the time of learning, new listeners familiarized with dysarthric speech under passive or explicit conditions should exhibit the same error patterns observed by Borrie and colleagues. However, if listeners are familiarized with unique speech stimuli designed to specifically draw attention to the syllabic stress contrast cues in the transcription phrases, this LBE condition discrepancy may disappear. This finding would provide evidence that the locus of learning is dependent on the information supplied, and its density, within the familiarization procedure. Thus, the following question was addressed: Do listeners familiarized with dysarthric stimuli that highlight syllabic strength cues exploit this information regardless of learning condition? In addition, the study examined whether there was an effect of familiarization procedure, in which the magnitude of intelligibility gain is regulated by the structure of familiarization stimuli (phrases versus passage readings) and learning condition (passive versus explicit).

2. Method

2.1 Study overview

Two groups of listeners, novel to the current investigation, were familiarized with speech stimuli produced by speakers with dysarthria under one of two experimental conditions: (1) Auditory presentation of experimental phrases (passive-phrases) or (2) concurrent auditory and written presentation of experimental phrases (explicit-phrases). Following familiarization, all listeners completed an identical phrase transcription task. Data from the current study were compared with the corresponding data from experimental groups reported by Borrie and colleagues (2011): (1) Auditory presentation of passage readings (passive-passages), and (2) concurrent auditory and written presentation of passage readings (explicit-passages).

2.2 Participants

Data were collected from 40 young healthy individuals aged 19–40 years [$M = 24.4$ years; standard deviation (SD) $= 6.3$] using identical inclusion criteria to that used in the initial study (Borrie *et al.*, 2011).

2.3 Stimuli and procedure

Speech stimuli consisted of two speech sets, familiarization speech set and transcription speech set, established by Borrie and colleagues (2011). Stimuli were produced by three individuals with a primary diagnosis of Parkinson's disease and a perceptual and acoustically verified moderate hypokinetic dysarthria (the reader is directed to Borrie *et al.*, 2011 for comprehensive information regarding the nature of the speakers). In brief, each speech set consisted of 36 semantically anomalous novel phrases. All phrases contained six syllables with alternating phrasal stress which enabled LBEs to be interpreted relative to syllabic strength. Speech sets were balanced so that each set included the same number of phrases produced by each speaker (12 phrases per speaker), syllable stress pattern of the phrases (6 trochaic and 6 iambic phrases per speaker), and number and type of LBE opportunities. No phrase was repeated either within or across the two speech sets. See Borrie *et al.* (2011) for further details regarding the nature of the experimental stimuli and speech set development.

The 40 listener participants were randomly assigned to one of two listener groups, passive-phrases or explicit-phrases, so that each group consisted of 20 participants. The experiment was conducted in two distinct phases: (1) *Familiarization phase* and (2) *transcription phase*. As per the perceptual learning procedure performed in Borrie *et al.* (2011), the experiment was conducted in a quiet room using sound-attenuating headphones (Sennheiser HD 280 pro) and a laptop computer pre-loaded with the experiment. Participants were told that they would undertake a listening task followed by an orthographic transcription task, and that task-specific instructions would be delivered via the computer program. During the familiarization phase, listeners in the passive-phrases group were presented with auditory productions of the familiarization speech set and were instructed to simply listen to the phrases. Listeners in the explicit-phrases group were also presented with auditory productions of the familiarization speech set, in addition to written transcripts of the intended phrase targets, and were instructed to read these alongside the auditory productions. Immediately following the familiarization task, both of the listener groups participated in an identical transcription phase in which they transcribed the transcription speech set.

2.4 Analysis

The total data set consisted of 40 transcripts of the 36 experimental phrases that made up the transcription speech set. The first author independently analyzed the listener transcripts for type (i.e., insertion/deletion) and location (i.e., before strong/weak syllable) of LBEs and a measure of speech intelligibility, percent words correct (PWC), using identical coding rules to that reported previously (Borrie *et al.*, 2011; Liss *et al.*, 1998). Twenty-five randomly selected transcripts were reanalyzed by the first author (intra-judge) and by a second trained judge (inter-judge) to obtain reliability estimates for the coding of LBE and PWC data. Discrepancies between the reanalyzed data and the original data revealed that agreement was high (all $r > 0.95$), with only minor absolute differences. Chi-square analyses were performed on LBE data to determine the relationship between error type and error location and PWC data were analyzed using a two-way analysis of variance (ANOVA).

3. Results

3.1 Lexical boundary errors

Table 1 contains the LBE category proportions and the sum IS/IW and DW/DS ratios for the two listener groups familiarized with experimental phrases under passive (passive-phrases) and explicit (explicit-phrases) conditions. Corresponding data from

Table 1. 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
Passive-phrases	51.60	16.80	11.40	20.20	3.1	1.8
Explicit-phrases	51.95	14.29	11.69	22.08	3.6	1.9
[*]Passive-passages	27.31	22.69	28.41	21.59	1.2	0.8
[*]Explicit-passages	42.42	12.31	16.70	28.57	3.5	1.7
[*]Control	37.15	15.84	19.55	28.21	2.4	1.4

Note: "IS," "IW," "DS," and "DW" refer to LBEs defined as insert boundary before strong syllable, insert boundary before weak syllable, delete boundary before strong syllable, and delete boundary before weak syllable, respectively.
[*]= LBE data previously reported [Borrie *et al.* (2011)] and included for visual comparison only.
[a]$n = 20$.

Borrie *et al.* (2011) for listeners familiarized with passage readings under passive (passive-passages) and explicit (explicit-passages) conditions, as well as a control group, are included for visual comparisons only. Contingency tables were constructed for the total number of LBEs by error type and error location for the passive-phrases and explicit-phrases 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 passive-phrases group, $X^2(1, N = 20) = 71.84$, $p < 0.001$, and the explicit-phrases group, $X^2(1, N = 20) = 89.06$, $p < 0.001$. In both of the listener groups, erroneous lexical boundary insertions occurred more often before strong (IS) than before weak syllables (IW), and erroneous lexical boundary deletions occurred more often before weak (DW) than before strong syllables (DS). Such LBE error patterns conform to MSS predictions (Cutler and Butterfield, 1992). A between-group chi-square analysis was also used to examine differences in the distribution of errors exhibited by the passive- and explicit-phrases groups. The comparison revealed no significant difference in error distribution between the two groups, $X^2(3, N = 40) = 3.9$, $p = 0.27$. Thus, the relative distribution of errors observed for the passive-phrases group were similar to those observed for listeners in the explicit-phrases group. Ratio figures signal the strength of adherence to predicted error patterns, the greater the positive distance from "1" reflecting increased adherence. The ratio values for both the passive- and explicit-phrases groups support exploitation of syllabic stress contrast cues to inform speech segmentation, similar to that observed in the corresponding data from the explicit-passages group (Borrie *et al.*, 2011).

3.2 PWC

Figure 1 reflects the mean PWC scores for the two listener groups familiarized with phrases under passive (passive-phrases) or explicit (explicit-phrases) conditions, as well as the corresponding data for listeners familiarized with passage readings under passive (passive-passages) and explicit (explicit-passages) conditions (from Borrie *et al.*, 2011). Control data from the prior report is included for visual comparisons only. A two-way ANOVA was conducted on PWC scores of the listeners familiarized with dysarthric speech, with condition (passive or explicit) and stimuli (passages or phrases) as between subject variables. The ANOVA revealed a significant main effect of condition, $F(1, 76) = 122.51$, $p < 0.001$, $\eta^2 = 0.27$. Thus, explicit familiarization afforded significantly greater intelligibility gains than passive familiarization. The main effect of stimuli was also significant, $F(1, 76) = 251.90$, $p < 0.001$, $\eta^2 = 0.55$. Thus, familiarization with the passage stimuli afforded significantly greater intelligibility gains than familiarization
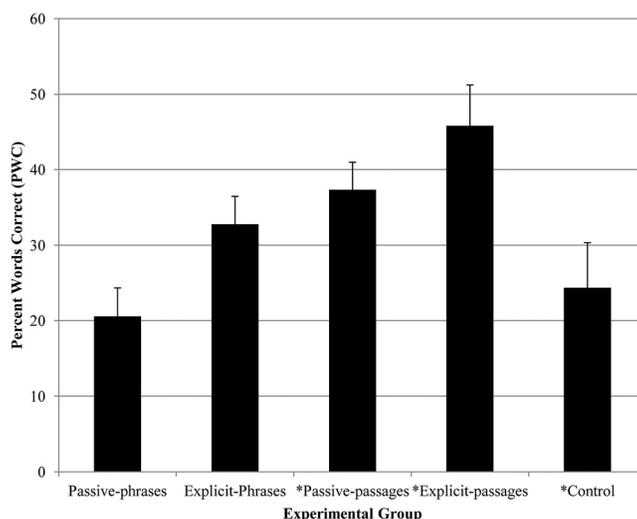
Fig. 1. Mean PWC for listeners by experimental group. Bars delineate +1 standard deviation of the mean. * Data previously reported [Borrie *et al.* (2011)].

with the phrase stimuli. The interaction between condition and stimuli was not significant, $F(1, 76) = 70.58$, $p = 0.05$, $\eta^2 = 0.01$.

4. Discussion

Consistent with the intelligibility data from an earlier report (Borrie *et al.*, 2011), performance was superior when the familiarization phase included written transcripts of the spoken targets (explicit-phrases) relative to a more passive experience involving only the auditory information (passive-phases). This finding is in line with a number of studies that have evidenced the utility of knowledge-of-target in perceptual learning of degraded speech (e.g., Davis *et al.*, 2005; Loebach *et al.*, 2010). However, key is that the speech segmentation condition discrepancy reported in the initial study, wherein syllabic stress cues were exploited only when familiarization was explicit, was not observed in the current data. Listeners familiarized with stimuli designed to emphasize syllabic stress contrast cues readily applied this segmentation strategy, regardless of whether learning conditions were passive or explicit. This was evidenced in that the LBE patterns for both listener groups adhered strongly to the MSS hypothesis—a greater number of predicted (IS and DW) versus non-predicted (IW and DS) errors (Cutler and Norris, 1988). Thus, the LBE analysis reveals that the alternating stress structure of the familiarization stimuli used in the current study (i.e., trochaic and iambic phrases versus the variable stress patterning that characterizes the previously used story passage) served to direct the listeners' attention toward syllabic stress information.

Intelligibility benefits for listeners familiarized with dysarthric speech were only realized under explicit conditions (explicit-passages; explicit-phrases) or when passage stimuli were utilized (passive-passages). PWC scores exhibited by listeners familiarized with phrases under passive conditions (passive-phrases) did not exceed those of the control group reported in Borrie *et al.* (2011). This raises the question as to why the passive-phrases group did not experience the moderate intelligibility benefit achieved by the passive-passages group in the initial study. If attendance to syllabic stress information is a functional perceptual strategy for lexical segmentation in instances of signal degradation, the passive-phrases group could be expected to outperform the passive-passage group (who displayed no tendency to exploit syllabic stress cues). Even considering the transfer-appropriate processing theory, which postulates that

improvements may be magnified when learning conditions are reinstated at testing (e.g., Lockhart, 2002; Rajaram *et al.*, 1998), performance gains would be anticipated. Unlike the experimental phrases, the reading passage did not afford similarities to the test stimuli. However, word recognition benefits associated with a familiarization experience were not evident for the passive-phrases group.

　　　The most obvious explanation is that the two passive familiarization groups experienced different levels of knowledge-of-target during their passive listening task. While the reading passage consisted of semantically and syntactically predictable sentences within a rich story context, the experimental phrases were semantically anomalous. Thus, the ability to recognize one word in a phrase did not aid in priming surrounding words. As such, listeners familiarized with the phrases were disadvantaged in their capacity to deploy top-down, predictive processes to decipher the spoken word targets. It was only when information regarding the intended productions was provided with more explicit instruction (written transcripts), that listeners were able to extract information required for performance gains during subsequent encounters with the degraded speech. Thus, in the absence of robust knowledge-of-target, listeners in the passive-passages group were only able to reliably extract, and subsequently apply, suprasegmental information.

　　　It should be noted that the current study was not designed to serve as a test of efficacy regarding different familiarization stimuli—the cross-experiment PWC comparisons were performed to supplement LBE findings. Accordingly, a number of factors, including word familiarity, frequency, and quantity of stimuli were not controlled for. Future studies should address the lack of empirical evidence regarding the role of signal-independent information in perceptual learning of dysarthric speech.

　　　In summary, the present investigation adds interpretability to the findings of an earlier report (Borrie *et al.*, 2011). Specifically, results suggest that the locus of learning is not solely determined by the learning conditions—the presence or absence of written information during familiarization. Rather, it appears that perceptual learning of dysarthric speech is differentially influenced by both the structure of the familiarization material and the extent of knowledge-of-target afforded during the familiarization procedure. Research going forward will aim to explicate this relationship further.

Acknowledgments

References and links

Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., and Anderson, T. (**2012**). "Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech," in *Language and Cognitive Processes*, Lang. Cognit. Processes, http://dx.doi.org/10.1080/01690965.01692011.01610596.

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

Cutler, A., and Norris, D. G. (**1988**). "The role of strong syllables in segmentation for lexical access," J. Exp. Psychol. Hum. Percept. Perform. **14**, 113–121.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (**2005**). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**(2), 222–241.

Jusczyk, P. W., and Luce, P. A. (**2002**). "Speech perception and spoken word recognition: Past and present," Ear Hear. **23**, 2–40.

Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., and Edwards, B. (**1998**). "Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech," J. Acoust. Soc. Am. **104**(4), 2457–2466.

Lockhart, R. S. (**2002**). "Levels of processing, transfer-appropriate processing, and the concept of robust encoding," Memory **10**, 397–403.

Loebach, J. L., Pisoni, D. B., and Svirsky, M. A. (**2010**). "Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant," J. Exp. Psychol. **36**(1), 224–234.

Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple speech segmentation cues: A hierarchical framework," J. Exp. Psychol. Gen. **134**(4), 477–500.

McQueen, J. M., and Cutler, A. (**2001**). "Spoken word access processes: An introduction," Lang. Cognit. Processes **10**(3), 309–331.

Rajaram, S., Srinivas, K., and Roediger, H. (**1998**). "A transfer-appropriate processing account of context effects in word-fragment completion," J. Exp. Psychol. [Hum. Learn.] **24**, 993–1004.