A relationship between processing speech in noise and dysarthric speech

Stephanie A. Borrie[a)]
Department of Communicative Disorders and Deaf Education, Utah State University, Logan, Utah 84322, USA

Melissa Baese-Berk
Department of Linguistics, University of Oregon, Eugene, Oregon 97403, USA

Kristin Van Engen
Department of Psychological and Brain Sciences, Washington University, St. Louis, Missouri 63130, USA

Tessa Bent
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405, USA

There is substantial individual variability in understanding speech in adverse listening conditions. This study examined whether a relationship exists between processing speech in noise (environmental degradation) and dysarthric speech (source degradation), with regard to intelligibility performance and the use of metrical stress to segment the degraded speech signals. Ninety native speakers of American English transcribed speech in noise and dysarthric speech. For each type of listening adversity, transcriptions were analyzed for proportion of words correct and lexical segmentation errors indicative of stress cue utilization. Consistent with the hypotheses, intelligibility performance for speech in noise was correlated with intelligibility performance for dysarthric speech, suggesting similar cognitive-perceptual processing mechanisms may support both. The segmentation results also support this postulation. While stress-based segmentation was stronger for speech in noise relative to dysarthric speech, listeners utilized metrical stress to parse both types of listening adversity. In addition, reliance on stress cues for parsing speech in noise was correlated with reliance on stress cues for parsing dysarthric speech. Taken together, the findings demonstrate a preference to deploy the same cognitive-perceptual strategy in conditions where metrical stress offers a route to segmenting degraded speech. © 2017 Acoustical Society of America.
[http://dx.doi.org/10.1121/1.4986746]

I. INTRODUCTION

It is now widely acknowledged that cognitive, perceptual, and linguistic skills underlie the variability in an individual's success, or lack thereof, in recognizing speech in adverse listening conditions (e.g., Watson *et al.*, 1996; Adank and Janse, 2010; Borrie *et al.*, 2017). Emerging from this area of enquiry is the question of whether *successful* listeners are universally successful, irrespective of the context (e.g., speech in noise, accented speech, and disordered speech), or whether success is modulated by the type of listening adversity. In examining correlations between intelligibility performance on different types of signal adversity, we recently observed that native English listeners who were good at understanding a non-native accent (Spanish accent) were also good at understanding a regional dialect (Irish English) and disordered speech (ataxic dysarthia), although no correlation between understanding the regional dialect and the dysarthric speech was noted (Bent *et al.*, 2016). We surmised that, rather than a global skill for recognizing speech in all types of adverse listening conditions, listeners may possess specific cue sensitivities and/or favor perceptual strategies that allow them to be successful with particular types of listening adversity. This speculation warrants further study.

Central to recognizing spoken language is lexical segmentation, the task of parsing the continuous speech stream into individual words (Cutler *et al.*, 1986). In favorable listening conditions, this task is a relatively easy one as the speech stream carries an abundance of segmentation cues. Such cues have been broadly characterized as lexical and sublexical, with sublexical cues further categorized as segmental and suprasegmental. Lexical cues refer to higher-level information arising from knowledge of individual words and syntactic, semantic, and pragmatic relations between words. Segmental cues include phonemic, phonotactic, and co-articulatory features, and suprasegmental cues refer to speech rhythm properties, including metrical stress (see White *et al.*, 2012, for more detail of the segmentation cue categories).

A hierarchy of cues to lexical segmentation, introduced by Mattys *et al.* (2005), postulates that cue utilization is modulated by the quality and quantity of lexical and sublexical information afforded by the speech signal. According to this framework, when all cues are optimally available, speech segmentation is lexically driven. However, when lexical information is unavailable, ambiguous, or impoverished,

a)Electronic mail: stephanie.borrie@usu.edu

listeners will resort to lower level segmental cues to parse the speech stream. In cases where the segmental cues are themselves degraded (e.g., speech in noise, disordered speech), listeners will turn their attention to suprasegmental information, namely, metrical stress. This work, therefore, builds a strong case for the notion that stress cues are especially important when processing speech in adverse listening conditions.

One perceptual strategy that exploits metrical stress to parse connected speech is the metrical segmentation strategy (MSS; Cutler and Norris, 1988; Cutler and Butterfield, 1992). According to the MSS, listeners exploit the statistics of English syllable stress to guide lexical segmentation. Strong syllables—those receiving relative stress through longer duration, fundamental frequency change, increased loudness, and unreduced vowels—are treated as the onset of a new word (Cutler and Butterfield, 1992). Evidence of MSS application can be found in listener's lexical missegmentations, or lexical boundary errors, including when they mistakenly insert or delete word boundaries in contexts that reflect stress-initial segmentation (i.e., inserting a boundary before a strong syllable and/or deleting a boundary before a weak syllable). When a listener makes more predicted lexical boundary errors (those that treat strong syllables as word initial) than unpredicted lexical boundary errors (those that treat weak syllables as word initial), adherence to the MSS can be assumed (see Table I for examples of predicted and unpredicted errors). Evidence supporting adherence to the MSS in adverse listening conditions has been observed in listener attempts to segment speech presented in noise (e.g., Smith *et al.*, 1989; Woodfield and Akeroyd, 2010) or at reduced intensity (e.g., Cutler and Butterfield, 1992). Even when speech rhythm is disrupted, as is the case of dysarthria, a neurogenic motor speech disorder characterized by both segmental and suprasegmental degradation, listeners' lexical boundary errors adhere to the MSS, reflecting stress cue utilization (e.g., Liss *et al.*, 1998; Borrie *et al.*, 2012).

Given that listeners have been shown to exploit speech rhythm cues to segment speech in adverse listening conditions, specifically those contexts where lexical and segmental cues are unavailable, ambiguous, or impoverished, it follows that listeners with enhanced rhythm sensitivity will be better equipped to navigate such signals. Indeed, empirical evidence has shown that individual expertise in rhythm perception, specifically sensitivity to temporal differences in musical signals, provides a significant processing advantage

TABLE I. Lexical boundary error examples. Classifications based on the type (insertion or deletion) and location (before a strong or weak syllable) of the error. The syllables surrounding the boundary errors in each response and corresponding stimulus are underlined. Note: Listeners adhering to the MSS are predicted to make more erroneous insertions before strong (IS) syllables and deletions before weak (DW) syllables than the other two error classes.

Target	Response	Lexical boundary errors
Unseen machines agree	I see machines agree	Insertion before strong (IS)
Push her equal culture	Picture equal culture	Delete before weak (DW)
Music useful rising	Music used for rising	Insertion before weak (IW)
Define respect instead	Define respect he said	Delete before strong (DS)

for both speech in noise (Slater and Kraus, 2015) and dysarthric speech (Borrie *et al.*, 2017). These findings suggest that some listeners may be better equipped to exploit metrical stress to segment speech in noise and dysarthric speech, rendering them successful as listeners in both types of listening adversity. Whether a relationship between processing speech in noise and dysarthric speech exists, with regard to intelligibility performance and the use of metrical stress to segment connected speech, has yet to be examined.

In the present study, we examined individual differences in perceptual processing of two types of listening adversity: speech in noise and dysarthric speech. Speech in noise, particularly speech-shaped noise, is characterized by disruption to a number of segmental features, but relatively intact suprasegmental cues (Parikh and Loizu, 2005; Jiang *et al.*, 2006). Conversely, dysarthria, with its imprecise articulation and rhythmic disturbances, is characterized by both segmental and suprasegmental degradation. While differentially degraded, both signals have been independently identified to induce stress-based segmentation. The following three key research questions were addressed in the current study: (a) is there a relationship between intelligibility performance for speech in noise and dysarthric speech, (b) is degree of reliance on stress cues to segment speech in adverse listening conditions differentially affected by the type of listening adversity, and (c) is there a relationship between the use of stress cues to segment speech in noise and dysarthric speech? Our hypotheses were rooted in a body of work showing that metrical stress is important for processing speech in adverse listening conditions (Mattys *et al.*, 2005), that listeners exploit such cues for segmenting speech in noise (e.g., Woodfield and Akeroyd, 2010) and dysarthric speech (e.g., Borrie *et al.*, 2012), and that individual listeners with expertise in rhythm perception are more successful understanding speech in noise (Slater and Kraus, 2015) and dysarthric speech (Borrie *et al.*, 2017). Thus, we postulated that intelligibility scores for perceiving speech in noise would be positively correlated with intelligibility scores for perceiving dysarthric speech, implying similar cognitive-perceptual processing mechanisms for different types of signal degradation in which segmental cues are compromised. In support of this implication, we hypothesized that both types of listening adversity would induce stress-based segmentation, but given the greater integrity of stress cues afforded by speech in noise relative to dysarthric speech, we posited that listeners would exploit stress cues to a greater degree when parsing speech in noise. Finally, we postulated that individual reliance on stress cues to segment speech in noise would be positively correlated with reliance on stress cues to segment dysarthric speech, demonstrating a preference to deploy the same perceptual strategy, regardless of the origin of degradation.

II. METHOD

A. Listener participants

Participating in the experiment were 90 adults (41 females and 49 males) aged 23–71 years old [mean (M) = 34.7; standard deviation (SD) = 11.35]. All participants were native speakers

J. Acoust. Soc. Am. **141** (6), June 2017

Borrie *et al.*   4661

of American English and currently residing in the United States (U.S.). Participants reported no history of speech, language, or hearing problems and no significant prior contact with persons having neurogenic speech disorders.

Participants were recruited using the crowdsourcing website, Amazon Mechanical Turk (MTurk).[1] Termed MTurk workers, all participants were considered voluntary workers, protected through MTurk's participation agreement and privacy notice, and received remuneration ($5) in exchange for their participation. We utilized the MTurk setup option regarding worker requirements, limiting participation to high-performing workers with an earned "Masters" status, an option recommended for researchers wishing to obtain high-quality data (Peer *et al.*, 2014). We also utilized a location restriction, permitting only workers confirmed as current residents in the U.S to participant in this study. The use of human subjects recruited via online crowdsourcing was approved by the Utah State University Institutional Review Board (IRB).

B. Speech stimuli

The speech stimuli set consisted of 160 experimental phrases. These phrases, taken from Liss *et al.* (1998) and modeled on those of Cutler and Butterfield (1992), were developed specifically for the purposes of examining listener reliance on syllable stress cues to guide speech segmentation decisions. They have been used in a number of studies examining the MSS hypothesis in processing of dysarthric speech (e.g., Liss *et al.*, 2000; Borrie *et al.*, 2012; McAuliffe *et al.*, 2014). The phrases ranged in length from three to five words, and were syntactically plausible but semantically anomalous to reduce the influence of lexical cues on word recognition and speech segmentation (e.g., *amend estate approach*). Each phrase contained six syllables, alternating strong (S) and weak (W) syllables, such that half of the phrases contained a SWSWSW phrasal stress pattern and the other half contained a WSWSWS phrasal stress pattern. These alternating syllabic stress patterns enable errors in speech segmentation (lexical boundary errors) to be interpreted relative to a perceptual strategy hypothesis, the MSS.

The 160 phrases were divided into 2 80-phrase speech sets, 1 to be used for speech in noise stimuli and the other for dysarthric speech stimuli. The speech sets were balanced for number of phrases, number of words, alternating stress patterns, and number of lexical boundary error opportunities. This stimuli design facilitates interpretation and comparison of MSS ratios within and between speech sets.

For the creation of the dysarthric speech stimuli, the 80 phrases that made up the dysarthric speech set were elicited from a 23-year-old male native speaker of American English with dysarthric speech secondary to traumatic brain injury. Audio recordings of the phrases produced by the speaker with dysarthria were independently rated by three speech-language pathologists with expertise in assessment and diagnosis of motor speech disorders. The dysarthria was diagnosed as a classic *spastic* type, characterized perceptually by reduced stress, monotony, monoloudness, slow speaking rate, imprecise articulation, and a strained–strangled vocal quality. A pilot experiment with four naive listeners

indicated that listeners could identify ∼40% of the words in the dysarthric speech stimuli.

For the speech in noise stimuli, the 80 phrases that made up the speech in noise speech set were elicited from a 24-year-old male native speaker of American English with no history of neurological disease or injury. Using the scale intensity function in Praat (Boersma and Weenink, 2016), the audio recordings were leveled to equalize root mean square (RMS) amplitude. The files were then mixed with steady speech-shaped noise that had been created using the long-term average spectrum of the target sentences. A pilot experiment with four listeners indicated that a −6 dB signal-to-noise ratio (SNR) was required for listeners to identify ∼40% of the words in the speech in noise stimuli. This SNR resulted in similar overall intelligibility for the dysarthria and speech in noise conditions.

The speech stimuli files were programed in a web-based listener-perception application hosted on a secure university-based web server. The basic function of the web application was to present the listener participants with the experimental phrases (see Procedure, Sec. II C, for additional details) and have them type out what they thought the person was saying. Phrases were blocked by type of listening adversity and counterbalanced across the listeners. Thus, half the listeners heard and transcribed speech in noise phrases followed by dysarthric speech phrases, and the other half heard and transcribed dysarthric speech phrases followed by speech in noise phrases. Presentation order of the phrases within each speech set was randomized across all listeners to eliminate potential order effects.

C. Procedure

A brief description of the study task (i.e., listening to speech that is difficult to understand and typing out what you think the person is saying), requirements (use of headphones), time commitment (∼45 min), and remuneration amount ($5) was posted on MTurk. Interested workers were directed to a web page loaded with the study. Before beginning the study, MTurk workers/listeners were required to read through the IRB approved consent form. By clicking "Agree," listeners indicated that they had read and understood the information provided in the consent from and still voluntarily agreed to participate. Listeners were then required to complete a brief questionnaire regarding age, gender, and basic background questions related to inclusion/exclusion criteria (see Sec. II A for details). Progression to the experimental procedure commenced when responses to all questions were provided.

Upon completion of the questionnaire, the experimental portion of the procedure was loaded. Listeners were told that they would be presented with 160 phrases that were difficult to understand. They were informed that half the phrases would be challenging because there would be a lot of background noise. The other half would be challenging because the phrases would be produced by someone with a severe speech disorder. They were told that the phrases all contained real English words but would not make sense (e.g., "*throwing down the houses*"). Listeners were then given a non-test speech audio file (male speaker not used in the study) and

instructed to adjust the volume setting on their processor to a comfortable listening level. This file could be played as many times as needed to establish a comfortable volume. Listeners were prompted to click the return key when they were ready to begin the experiment.

Phrases were presented one at a time, and following each presentation, listeners were instructed to use the keyboard to type out exactly what they thought was being said. Listeners were strongly encouraged to make a guess at any words they did not recognize and use an "X" to represent any part of a provided phrase where a guess could not be made. Once they had finished typing their response, listeners were prompted to press the return key to move on to the next phrase. The self-paced experimental procedure took, on average, 45 min to complete.

D. Transcript analysis

The total data set consisted of 90 listener transcripts, each containing 80 speech in noise phrases and 80 dysarthric speech phrases. Transcripts were analyzed for words correct, using previously established scoring criteria (Liss *et al.*, 1998; Borrie *et al.*, 2012). Words were counted correct if they matched the intended target or differed only by tense (*-ed*) or plurality (*-s*). Homophones and obvious spelling errors were also counted as correct. Percent words correct (PWC) scores, one for speech in noise and one for dysarthric speech, were tabulated for each listener to reflect intelligibility performance for each type of listening adversity. Twenty percent of the transcripts were randomly selected and reanalyzed to examine inter-judge reliability for coding words correct. Discrepancies revealed high agreement between two judges with a Pearson correlation *r* score above 0.981.

Transcripts were also analyzed for lexical boundary errors to facilitate the calculation of a measure considered to reflect reliance on stress cues to segment speech. Boundary errors were classified according to type (insertion or deletion) and location (before a weak or strong syllable). Thus, four classes of errors can be coded: insertion before a strong syllable (IS), insertion before a weak syllable (IW), deletion before a strong syllable (DS), and deletion before a weak syllable (DW). Examples of the four classes of boundary errors can be viewed in Table I.

The MSS posits that if listeners are using stress cues to guide segmentation, the large majority of their errors will treat strong syllables as word initial (Cutler and Norris, 1988). Thus, adherence to the MSS should result in more IS and DW errors (predicted errors) than IW and DS errors (unpredicted errors). A simple ratio, the MSS ratio, quantifies strength of adherence to the stress-based segmentation strategy. This ratio is calculated by taking the sum of predicted errors over the total number of errors. Two MSS ratio scores, one for speech in noise and one for dysarthric speech, were tabulated for each listener. MSS ratios less than 0.5 signify no adherence to the stress-based segmentation strategy, whereas ratios greater than 0.5 signify adherence—the greater the ratio, the greater degree of adherence (Spitzer *et al.*, 2007). Twenty percent of the transcripts were randomly selected and reanalyzed to examine inter-judge

reliability for coding words correct. Discrepancies revealed high agreement between two judges with a Spearman correlation ρ score above 0.939.

III. RESULTS

A. Intelligibility performance

Intelligibility performance, expressed by the mean PWC scores, for each listener across the two types of listening adversity is illustrated in Fig. 1. The top panel of Fig. 1 shows that as a group, intelligibility performance on speech in noise ($M = 42.08$, SD $= 7.46$) was comparable to dysarthric speech ($M = 42.81$, SD $= 8.12$), $t(89) = 0.883$, $p = 0.38$. This finding is not surprising given that the SNR used to create the speech in noise stimuli was specifically selected, through piloting, to approximate the intelligibility level of the dysarthric speech stimuli.

The top panel of Fig. 1 also illustrates the large individual variability in PWC scores for both types of listening adversity. As listener age and gender were not controlled for during the online crowdsourcing recruitment strategy, a linear regression with age, gender, and type of listening adversity
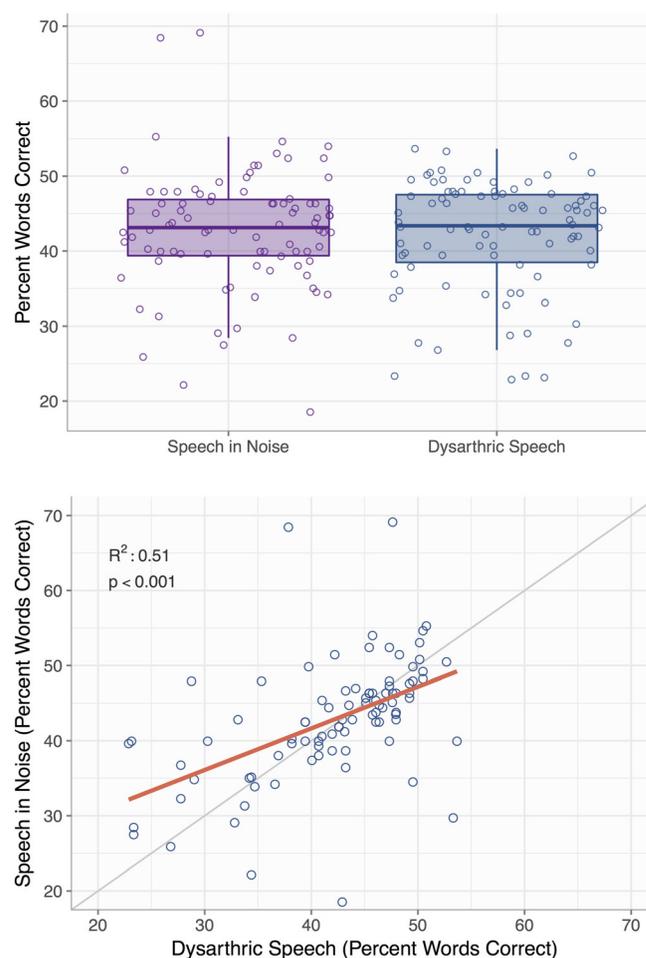

FIG. 1. (Color online) Intelligibility performance for speech in noise and dysarthric speech. The top panel reflects individual performance in each type of listening adversity with whiskers representing the range of the data (excluding outliers). The bottom panel reflects the relationship between intelligibility performance in both types of listening adversity.

J. Acoust. Soc. Am. **141** (6), June 2017

Borrie *et al.*   4663

entered as predictors was conducted. This analysis revealed that none of the variables emerged as significant predictors of intelligibility performance (age: $\beta = 0.04$, $p = 0.47$; gender: $\beta = -0.38$, $p = 0.75$; type: $\beta = 0.72$, $p = 0.54$).

To address the first research question regarding whether there is a relationship between intelligibility performance on the two types of listening adversity, a correlation between listeners' PWC scores on speech in noise and dysarthric speech was conducted. As shown in the bottom panel of Fig. 1, this analysis revealed a significant positive association between the two PWC scores, $r^2 = 0.51$, $p < 0.001$. Thus, the listeners who are successful at understanding speech in noise are the same listeners who are successful at understanding dysarthric speech.

B. Lexical boundary errors and MSS ratios

Table II provides a summary of lexical boundary errors for each type of listening adversity. A chi-square analysis on the errors confirmed that the error patterns were not uniform across the conditions, $\chi^2 (1) = 42.04$, $p < 0.001$. However, the error distribution for both types of listening adversity conformed to a stress-based segmentation strategy, with significantly more predicted than unpredicted errors for speech in noise, $\chi^2 (1) = 383.46$, $p < 0.001$, and dysarthric speech, $\chi^2 (1) = 181.56$, $p < 0.001$.

Degree of adherence to the MSS, expressed as ratios of predicted errors over total errors, for each listener across the two types of listening adversity is illustrated in Fig. 2. The top panel of Fig. 2 shows that while the large majority of listeners demonstrated MSS adherence for both types of listening adversity (i.e., data points above 0.5), large individual variability exists. It is also evident that some listeners did not adopt the MSS (i.e., data points below 0.5). A linear regression with age, gender, and type of listening adversity entered as predictors was conducted. Results of this analysis revealed that only type of listening adversity emerged as a significant predictor of the MSS ratios (age: $\beta = -0.0004$, $p = 0.47$; gender: $\beta = -0.023$, $p = 0.054$; type: $\beta = 0.059$, $p < 0.001$).

The results of the linear regression, in combination with the chi-square analyses on predicted and unpredicted errors, address the second research question of the study, demonstrating that degree of reliance on stress cues to segment
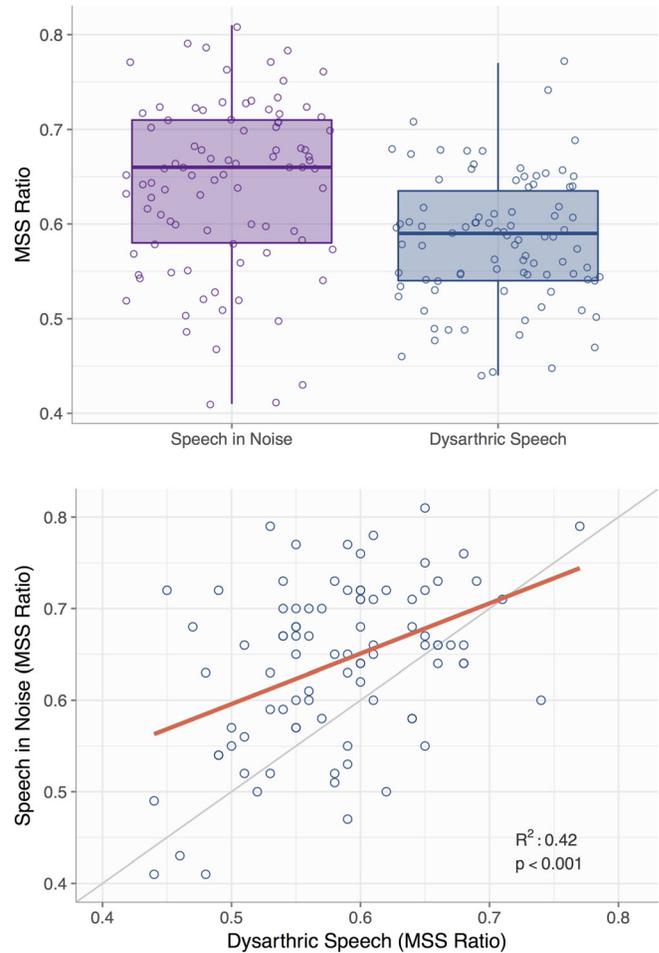

FIG. 2. (Color online) Adherence to the MSS for speech in noise and dysarthric speech. The top panel reflects individual MSS ratios in each type of listening adversity with whiskers representing the range of the data (excluding outliers). The bottom panel reflects the relationship between MSS ratios in both types of listening adversity.

speech in adverse listening conditions is differentially affected by the type of listening adversity. Specifically, listeners adhered to the MSS when segmenting both speech in noise and dysarthric speech, but they relied more heavily on this stress-based segmentation strategy with speech in noise.

To address the final research question regarding whether a relationship between the use of stress cues to segment speech in noise and dysarthric speech exists, a correlation between participants' MSS ratios with speech in noise and dysarthric speech was conducted. As shown in the bottom panel of Fig. 2, this analysis revealed a significant positive association between the two variables, $r^2 = 0.42$, $p < 0.001$. Thus, the listeners who relied heavily on stress cues to segment speech in noise are the same listeners who relied heavily on stress cues to segment dysarthric speech.

Listeners made a large number of lexical boundary errors with both speech in noise and dysarthric speech. Speech in noise elicited a total of 4952 errors across the 90 listeners, whereas dysarthric speech elicited a total 7058 errors. While no predictions were made regarding the absolute differences in the incidence of total number of lexical boundary errors elicited by the two types of listening

TABLE II. Pattern of lexical boundary errors by adverse listening condition. Note: MSS, metrical segmentation strategy, taking the sum of predicted errors over the total number of errors.

Lexical boundary errors	Speech in noise	Dysarthric speech
Total number of errors	4952	7058
Total number of predicted errors	3165	4095
Total number of unpredicted errors	1787	2963
Mean MSS ratio (SD)	0.64 (0.09)	0.58 (0.07)
Breakdown of errors (total number)		
Insert before strong (IS)	2429 (49%)	2491 (35%)
Insert before weak (IW)	610 (12%)	1767 (25%)
Delete before strong (DS)	1177 (24%)	1196 (17%)
Delete before weak (DW)	736 (15%)	1604 (23%)

adversity, a paired *t*-test revealed that listeners made significantly more lexical boundary errors with dysarthric speech, $t(89) = -16.95$, $p < 0.001$.

IV. DISCUSSION

The purpose of this study was to examine whether a relationship exists between an individual's ability to process speech in noise and dysarthric speech, with regard to intelligibility performance and the use of metrical stress to segment the degraded speech signals. The results of the study, consistent with our hypotheses, demonstrated that listeners who were successful at understanding speech in noise were also successful at understanding dysarthric speech. While the origin of degradation differs, with speech in noise arising from environmental degradation and dysarthric speech considered a source degradation (see Mattys *et al.*, 2012), similar cognitive-perceptual processing mechanisms may support both. The current lexical segmentation findings substantiate this speculation, with listeners utilizing metrical stress cues to segment speech in both types of listening adversity. Indeed, independent studies have previously reported that listeners exploit metrical stress cues to decipher speech in noise (e.g., Smith *et al.*, 1989) and dysarthric speech (e.g., Liss *et al.*, 1998). Here, we replicate these findings with both types of listening adversity, and extend them by demonstrating (i) an effect of type of listening adversity on the degree to which stress cues are utilized, and (ii) a relationship between the reliance on stress cues to segment speech in noise and dysarthric speech. These two findings are expanded upon in the ensuing discussion.

Examination of degree of adherence to the MSS revealed that, while metrical stress cues were used to inform speech segmentation decisions in both types of listening adversity, these cues were used more heavily to segment speech in noise. This pattern of results likely arises from the fact that speech in noise affords more reliable suprasegmental cues than dysarthric speech. While noise imposes substantial segmental degradation, prosodic integrity, including strong–weak syllabic stress cues, remains relatively intact (e.g., Mattys, 2004; Smith *et al.*, 1989). In contrast, dysarthric speech is characterized by prosodic disturbances (Darley *et al.*, 1969; Duffy, 2013). More specifically, the cardinal suprasegmental deficits of spastic dysarthria, including reduced stress, monotony, monoloudness, and slow rate of speech, collectively serve to diminish the strong–weak syllabic stress cues. Here, while reduced, the syllabic stress contrasts in spastic dysarthria are sufficiently robust to induce stress-based segmentation, consistent with findings of an earlier report of lexical segmentation of spastic speech (Borrie, 2015). However, the diminished syllabic stress contrasts weaken the effectiveness of the stress-based segmentation strategy when compared with speech in noise. The idea that severity of prosodic degradation could modulate adherence to the MSS has been previously raised by Liss *et al.* (1998). In an examination of stress-based segmentation in hypokinetic dysarthria, the authors observed that the speech samples with the most diminished syllabic strength contrasts induced the lowest degree of adherence to the MSS. A similar pattern of results was reported in a follow-up study comparing stress-based segmentation in two types of dysarthria with unique prosodic deficits (Liss *et al.*, 2000). Taken together, the findings suggest that listeners, as a group at least, rely on metrical stress cues to segment degraded speech, but application of this segmentation strategy is reduced when the very cues that signal syllabic stress are impoverished.

At the level of the individual, examination of degree of adherence to the MSS in both types of listening adversity in this study revealed that reliance on stress cues to segment speech in noise was correlated with reliance on stress cues to segment dysarthric speech. This finding suggests that individuals deploy similar perceptual strategies across different types of listening adversity, despite differences in the origin of signal degradation. The data also suggest that some listeners are better able and/or prefer to exploit metrical stress cues for parsing connected speech in adverse listening conditions. This finding is consistent with evidence of large individual variation in the ability to perceive rhythm cues (e.g., Slater and Kraus, 2015) and previous speculations that listeners with expertise in rhythm perception may be better placed to leverage rhythm cues for segmenting degraded speech (Borrie *et al.*, 2017). Indeed, even when the signal affords sufficient syllabic stress cues for stress-based segmentation, as evidenced by group adherence to the MSS for both speech in noise and spastic dysarthria, not all individuals adopted this segmentation strategy (see Fig. 2, MSS ratios below 0.5). Thus, it appears that the salience of rhythm cues for the task of lexical segmentation is influenced by the information afforded in the speech signal *and* the individual listeners' ability/preference to detect and utilize rhythm cues in degraded speech streams. Of course, individual factors such as conditions in which the listener carried out the task (e.g., fatigued, anxious, etc.) may also contribute to an explanation of the results.

Although not part of our research questions and hypotheses, it is noteworthy that the dysarthric stimuli elicited significantly more lexical boundary errors than the speech in noise stimuli in this study. This finding supports the idea that dysarthric speech, relative to speech in noise, poses a greater challenge to lexical segmentation, specifically application of the MSS. Further, given that intelligibility levels for both types of listening adversity were comparable, we postulate that speech in noise, relative to dysarthric speech, may pose a greater challenge to processing segmental information. That is, listeners may be relatively good at exploiting metrical stress to identify word boundaries with speech in noise (reflected in fewer boundary errors and stronger adherence to the MSS), but relatively poor at navigating the disrupted segmental cues to identify the correct target (e.g., "*about*" for "*account*"). A detailed analysis of segmental errors for both types of listening adversity would provide further insight into this speculation.

The semantically anomalous stimuli used in the current study were explicitly designed to examine use of metrical stress cues to inform word boundary decisions, or put another way, use of the MSS (Liss *et al.*, 1998, 2000). Lexical cues were reduced (semantically anomalous phrases) and opportunities for predicted and unpredicted lexical boundary errors were balanced within (and across) speech

J. Acoust. Soc. Am. **141** (6), June 2017

Borrie *et al.*   4665

sets to enable adherence to the MSS to be quantified. Thus, the distribution of error opportunities in the experimental stimuli do not represent the error distribution encountered in everyday English (i.e., 90% of content words with stress on the first syllable; Cutler and Carter, 1987). Accordingly, regressing MSS ratios with PWC scores offers little value. However, under everyday listening circumstances where the majority of content words in English have a metrically stressed syllable at their onset, strength of adherence to the MSS should be a robust predictor of intelligibility success. Future studies in this line of work should include an additional set of stimuli that conform more closely to the distribution of lexical stress patterns in English. Had such additional stimuli been included in the current study, we would expect to observe a positive relationship between strength of adherence to the MSS and intelligibility performance on the stimuli set reflecting everyday error opportunities.

While adherence to the MSS does not afford a complete account of the cognitive-perceptual processes involved in lexical segmentation in adverse listening conditions, it does offer insight into individual differences in the use of metrical stress to understand different types of listening adversity. It is, however, likely that listeners use multiple cues to segment speech (e.g., Mattys *et al.*, 2005; Sanders and Neville, 2000). Thus, in addition to examining the use of other suprasegmental cues for speech segmentation in different types of listening adversity, a systematic exploration of individual variance in the use of segmental cues is warranted. Previous research has suggested that even when segmental cues are degraded, listeners can adapt to the novel acoustic-phonetic information to support perceptual processing of a degraded speech signal (e.g., Francis *et al.*, 2007; Spitzer *et al.*, 2000).

V. CONCLUSION

Consistent with our hypotheses, intelligibility scores for speech in noise were positively correlated with intelligibility scores for spastic dysarthria, suggesting similar cognitive-perceptual processing mechanisms may support both. The speech segmentation findings corroborated this finding. While stress-based segmentation was stronger with speech in noise relative to dysarthric speech, listeners utilized stress cues to parse both types of listening adversity. Further, we observed that reliance on stress cues with speech in noise was correlated with reliance on stress cues with dysarthric speech. Taken together, the speech segmentation data demonstrate a preference to deploy the same perceptual strategy in conditions where metrical stress offers a route to parsing ambiguous or impoverished speech streams. Additionally, the data shows that the utility of metrical stress for lexical segmentation in listening adversity is jointly influenced by the integrity of stress cues available in the speech signal and the individual listeners' ability to exploit such cues. More broadly, the current findings add to the body of literature demonstrating the role of rhythm in recognizing speech in adverse listening conditions.

ACKNOWLEDGMENTS

[1]Participants recruited via MTurk (http://www.mturk.com; last viewed November 2016) have been shown to be more representative of the U.S. population than in-person convenience samples (Berinsky *et al.*, 2012). In addition, empirical evidence suggests that speech perception data collected from MTurk are comparable to the same data collected in the laboratory (Lansford *et al.*, 2016; McAllister Byun *et al.*, 2015).

Adank, P., and Janse, E. (**2010**). "Comprehension of a novel accent by young and older listeners," Psychol. Aging **25**, 736–740.

Bent, T., Baese-Berk, M., Borrie, S. A., and McKee, M. (**2016**). "Individual differences in the perception of regional, nonnative, and disordered speech varieties," J. Acoust. Soc. Am. **140**, 3775–3786.

Berinsky, A. J., Huber, G. A., and Lenz, G. S. (**2012**). "Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk," Polit. Anal. **20**, 351–368.

Boersma, P., and Weenink, D. (**2016**). "Praat: Doing phonetics by computer (version 6.0.09) [computer program]," http://www.praat.org/ (Last viewed 2 August 2016).

Borrie, S. A. (**2015**). "Visual speech information: A help or hindrance in perceptual processing of dysarthric speech," J. Acoust. Soc. Am. **137**, 1473–1480.

Borrie, S. A., Lansford, K. L., and Barrett, T. S. (**2017**). "Rhythm perception and its role in perception and learning of dysrhythmic speech," J. Speech Lang. Hear. Res. **60**, 561–570.

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," Lang. Cogn. Process. **27**, 1039–1055.

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

Cutler, A., and Carter, D. M. (**1987**). "The predominance of strong initial syllables in the English vocabulary," Comput. Speech Lang. **2**, 133–142.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (**1986**). "The syllable's differing role in the segmentation of French and English," J. Mem. Lang. **25**, 385–400.

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.

Darley, F. L., Aronson, A. E., and Brown, J. R. (**1969**). "Differential diagnostic patterns of dysarthria," J. Speech Lang. Hear. Res. **12**, 246–269.

Duffy, J. R. (**2013**). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management* (Elsevier Mosby, St. Louis, MO), pp. 1–499.

Francis, A. L., Nusbaum, H. C., and Fenn, K. (**2007**). "Effects of training on the acoustic-phonetic representation of synthetic speech," J. Speech Lang. Hear. Res. **50**, 1445–1465.

Jiang, J., Chen, M., and Alwan, A. (**2006**). "On the perception of voicing in syllable-initial plosives in noise," J. Acoust. Soc. Am. **119**, 1092–1105.

Lansford, K. L., Borrie, S. A., and Bystricky, L. (**2016**). "Use of crowdsourcing to assess the ecological validity of perceptual-training paradigms in dysarthria," Am. J. Speech Lang. Pathol. **25**, 233–239.

Liss, J. M., Spitzer, S., 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**, 2457–2466.

Liss, J., Spitzer, S., Caviness, J., Adler, C., and Edwards, B. (**2000**). "Lexical boundary error analysis in hypokinetic and ataxic dysarthria," J. Acoust. Soc. Am. **107**, 3415–3424.

Mattys, S. L. (**2004**). "Stress versus coarticulation: Towards an integrated approach to explicit speech segmentation," J. Exp. Psychol. Hum. Percept. Perform. **30**, 397–408.

Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (**2012**). "Speech recognition in adverse conditions: A review," Lang. Cogn. Process. **27**, 953–978.

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

McAllister Byun, T., Halpin, P. F., and Szeredi, D. (**2015**). "Online crowd-sourcing for efficient rating of speech: A validation study," J. Commun. Disord. **53**, 70–83.

McAuliffe, M. J., Gibson, E. M. R., Kerr, S. E., Anderson, T., and LaShell, P. J. (**2014**). "Vocabulary influences older and younger listeners' processing of dysarthric speech," J. Acoust. Soc. Am. **134**, 1358–1368.

Parikh, G., and Loizou, P. (**2005**). "The influence of noise on vowel and consonant cues," J. Acoust. Soc. Am. **118**, 3874–3888.

Peer, E., Vosgerau, J., and Acquisti, A. (**2014**). "Reputation as a sufficient condition for data quality on Amazon Mechanical Turk," Behav. Res. Methods **46**, 1023–1031.

Sanders, L. D., and Neville, H. J. (**2000**). "Lexical, syntactic, and stress-pattern cues for speech segmentation," J. Speech Lang. Hear. Res. **43**, 1301–1321.

Slater, J., and Kraus, N. (**2015**). "The role of rhythm in perceiving speech in noise: A comparison of percussionists, vocalists and non-musicians," Cogn. Process. **17**, 79–87.

Smith, M. R., Cutler, A., Butterfield, S., and Nimmo-Smith, I. (**1989**). "The perception of rhythm and word boundaries in noise-masked speech," J. Speech Lang. Hear. Res. **32**, 912–920.

Spitzer, S. M., Liss, J. M., Cavines, J. N., and Adler, C. A. (**2000**). "An exploration of familiarization effects in the perception of hypokinetic and ataxic dysarthric speech," J. Med. Speech Lang. Pathol. **8**, 285–293.

Spitzer, S. M., Liss, J. M., and Mattys, S. L. (**2007**). "Acoustic cues to lexical segmentation: A study of resynthesized speech," J. Acoust. Soc. Am. **122**, 3678–3687.

Watson, C. S., Qiu, W. W., Chamberlain, M. M., and Li, X. F. (**1996**). "Auditory and visual speech perception: Confirmation of a modality independent source of individual differences in speech recognition," J. Acoust. Soc. Am. **100**, 1153–1162.

White, L., Mattys, S. L., and Widget, L. (**2012**). "Segmentation cues in conversational speech: Robust semantics and fragile phonotactics," Front. Psychol. **3**, 375.

Woodfield, A., and Akeroyd, M. A. (**2010**). "The role of segmentation difficulties in speech-in speech understanding in older and hearing-impaired adults," J. Acoust. Soc. Am. **128**, EL26–EL31.

J. Acoust. Soc. Am. **141** (6), June 2017

Borrie *et al.* 4667