The role of linguistic and indexical information in 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

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

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

Tim J. Anderson
New Zealand Brain Research Institute, 66 Stewart St, Christchurch 8011, New Zealand

(Received 12 January 2012; revised 8 November 2012; accepted 12 November 2012)

This investigation examined perceptual learning of dysarthric speech. Forty listeners were randomly assigned to one of two identification training tasks, aimed at highlighting either the linguistic (word identification task) or indexical (speaker identification task) properties of the neurologically degraded signal. Twenty additional listeners served as a control group, passively exposed to the training stimuli. Immediately following exposure to dysarthric speech, all three listener groups completed an identical phrase transcription task. Analysis of listener transcripts revealed remarkably similar intelligibility improvements for listeners trained to attend to either the linguistic or the indexical properties of the signal. Perceptual learning effects were also evaluated with regards to underlying error patterns indicative of segmental and suprasegmental processing. The findings of this study suggest that elements within both the linguistic and indexical properties of the dysarthric signal are learnable and interact to promote improved processing of this type and severity of speech degradation. Thus, the current study extends support for the development of a model of perceptual processing in which the learning of indexical properties is encoded and retained in conjunction with linguistic properties of the signal. © *2013 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4770239]

PACS number(s): 43.71.An, 43.71.Gv, 43.71.Bp [BRM] Pages: 474–482

I. INTRODUCTION

The speech signal carries both *linguistic* and *indexical* information. Linguistic information conveys the content of the utterance. This includes the phonological, morphological, syntactic, and semantic information provided within the word, phrase, and sentence structures of the acoustic signal (Levi and Pisoni, 2007). Indexical signal properties, on the other hand, are speaker-specific and reflect information pertaining to the talker's identity, including gender (see, e.g., Munson *et al.*, 2006), regional dialect (see, e.g., Hagiwara, 1997; Hillenbrand *et al.*, 1995), and emotional state (see, e.g., Costanzo *et al.*, 1989; Murry and Arnott, 1993). These properties manifest acoustically in measures such as fundamental frequency, formant spacing, relative segment durations, and overall speaking rate (Nygaard, 2008). Indexical information introduces substantial variability, both within and between speakers, and can profoundly influence the acoustic realizations of speech.

Founded on the premise that the perceptual system disregards any speaker-specific variation in an attempt to *normalize* the signal to a stable linguistic form (see, e.g., Brown and Carr, 1993; Halle, 1985; Joos, 1948; Ladefoged and Broadbent, 1957), conventional models of spoken language recognition have focused on the processing of linguistic information, largely ignoring the potential contributions of indexical information (see, e.g., Luce and Pisoni, 1998; Morton, 1969; Norris, 1994). Although acknowledging the existence of speaker-specific properties, traditional theoretical paradigms contend that such information is processed independently of linguistic information (see, e.g., Halle, 1985). According to the processes of normalization, the perceptual system removes any distinctive and variable features imposed by the speaker, reducing the acoustic signal to its canonical form. With an abstractly defined, stable representation of the linguistic information imprinted within a listener's memory, speech perception can continue to be successful in the face of substantial individual acoustic variability (for detailed reviews, see Goldinger, 1998; Tenpenny, 1995).

However, these conventional models have been challenged by research which demonstrates that, rather than being discarded in the process of recognizing spoken language, indexical properties may play a key role in speech perception (see, e.g., Johnson, 1997; Mullennix and Pisoni,

[a] Author to whom correspondence should be addressed. Electronic mail: steph.borrie@gmail.com

1990; Summerfield *et al.*, 1984). Evidence that linguistic processing is influenced by speaker-specific information is demonstrated in a number of studies that report a perceptual benefit with indexical consistency (see, e.g., Creelman, 1957; Goldinger *et al.*, 1991; Mattys and Liss, 2008). For example, Creelman (1957) correlated word recognition in noise with percent recognition accuracy when word lists were produced under single- versus multiple-speaker conditions. This study found an inverse relationship between intelligibility and number of speakers—word recognition scores increased as speaker numbers decreased. The same-speaker advantage in perceiving speech is robust with hearing-impaired adults (Kirk *et al.*, 1997) and with preschool children (Ryalls and Pisoni, 1997). This growing body of literature reveals that indexical properties of the speech signal can inform processing of spoken language. It appears that speech perception is in fact a highly integrated process, whereby indexical properties of lexical items are retained and encoded alongside linguistic information (see, e.g., Goldinger, 1996, 1998; Johnson, 2006; Pisoni, 1997).

In addition, a small collection of studies show that experience with speaker-specific properties may promote *improved* processing of linguistic information during subsequent encounters with the same talker. Nygaard and colleagues (1994) found that listeners trained to identify the names of ten unfamiliar speakers achieved higher recognition accuracy scores when presented with novel words produced by these now-familiar speakers, relative to listeners presented with the same novel words produced by unfamiliar speakers. Improved linguistic processing in noisy conditions for listeners familiarized with indexical properties of the speech signal was replicated in a follow-up study involving sentence-level stimuli (Nygaard and Pisoni, 1998) and a study involving both younger and older individuals (Yohan and Sommers, 2000). More recently, Loebach *et al.* (2008) reported on the perceptual benefit of training listeners to attend to indexical speech information within an artificially degraded speech signal—noise-vocoded speech. Significant intelligibility improvements were observed following a speaker identification task and further, the magnitude of performance gain was comparable to that achieved by a group of listeners who participated in linguistic-based training task. Thus, there is preliminary evidence to suggest that experience with indexical information may be as valuable as experience with linguistic information in facilitating improved recognition of degraded speech. Although research efforts have yet to document the learning mechanisms associated with exploiting indexical information for subsequent signal processing, it may be inferred that learned regularities within the indexical input assist the cognitive-perceptual processes involved in the perception of speech—lexical segmentation, lexical activation, and lexical competition (see Jusczyk and Luce, 2002).

If indexical properties provide a source of learning for processing of speech in noise or noise-vocoded speech, one may readily assume the same to be true for all forms of speech degradation. However, a significant challenge arises when attempting to adopt phenomena observed in experiments using highly constrained artificially degraded speech to that of the neurologically degraded speech (i.e., dysarthric speech). To illustrate, noise-vocoded speech is created by the systematic removal of specific spectral aspects of the acoustic signal (Shannon *et al.*, 1995). However, dysarthric speech is produced upon a platform of impaired muscle tone, inadequate respiratory drive, phonatory instability, and deficient articulatory movement. The implication for speech production is that although some acoustic degradation present in dysarthric speech may be relatively consistent, other breakdowns occur in nonsystematic and unpredictable ways (Borrie *et al.*, 2012a). To date, no study has addressed the role of indexical information in perceptual learning of dysarthric speech.

Recently, however, Mattys and Liss (2008) reported a perceptual advantage associated with indexical consistency of dysarthric speech. Listeners were more successful at recalling words if played in the same voice, as opposed to a different voice, between two consecutive blocks of speech stimuli. Further, the magnitude of perceptual benefit was significantly greater than that observed for listeners recalling neurologically healthy speech. It appears that speaker-specific detail may be especially informative when the perceptual system is challenged by the degradation that characterizes the speech of individuals with dysarthria. However, it is currently not known whether attention directed toward indexical properties of the signal, as with linguistic signal information (Borrie *et al.*, 2012b; Borrie *et al.*, 2012c), will enhance perceptual learning of dysarthric speech. Such knowledge is critical to inform the development of a model of perceptual processing that accounts for adaptation to neurologically degraded speech. Further, an understanding of the role that indexical information plays in perceptual learning of dysarthric speech is imperative to establish a theoretical framework that supports the development of listener-based treatment for the management of neurogenic speech disorders (Borrie *et al.*, 2012a).

The purpose of the current study was to investigate whether directing attention toward indexical information within the dysarthric signal could facilitate improved recognition of this type of speech and further, how this learning compares to that afforded by directing attention toward linguistic signal properties. The following key questions were addressed: (1) Do listeners trained to attend to the indexical properties of the dysarthric signal demonstrate similar intelligibility benefits as those achieved by listeners trained to attend to the linguistic information; and (2) Does training to attend to indexical versus linguistic properties differentially influence speech segmentation strategies? It was hypothesized that the magnitude of intelligibility benefit for listeners trained to attend to indexical signal information would be comparable to that achieved by listeners trained to attend to linguistic signal information—consistent with findings reported for perceptual learning of noise-vocoded speech (Loebach *et al.*, 2008). However, to the extent that the locus of learning is constrained by the different task requirements (focus on word identity versus focus on speaker identity), it was anticipated that differences would be observed in the ability to resolve phoneme ambiguity, with a lexical task advantage. In contrast, it was predicted that the more global focus associated with speaker identification learning may incur more success in exploiting prosodic cues to lexical segmentation. Thus, both groups were expected to exhibit intelligibility gains with training, but owing to different sources of learning. As per initial reports (Borrie *et al.*, 2012b; Borrie *et al.*, 2012c), perceptual learning of dysarthric speech is investigated by jointly considering intelligibility scores (percent words correct) and associated error patterns considered indicative of processing segmental (percent syllable resemblance) and suprasegmental (lexical boundary errors) level information.

II. METHOD

A. Overview

A between-group design was used to investigate perceptual learning effects for listeners familiarized with dysarthric speech via one of two types of training: (1) linguistic training (word identification task), or (2) indexical training (speaker identification task). A group of listeners who received no training formed a third comparison group, (3) control group (passive familiarization). Following familiarization, listeners in all three experimental groups engaged in an identical transcription task with 36 novel phrases produced by the speakers with dysarthria.

B. Listeners

Data were collected from 60 young healthy individuals (45 females and 15 males) aged 19–40 years ($M = 24.08$; SD = 6.25, where M is mean and SD is standard deviation). All listener participants were native speakers of New Zealand English (NZE), passed a pure tone hearing screen at 20 dB hearing level (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 undergraduate classes at the University of Canterbury.

C. Speech stimuli

Speech stimuli used in the current study were described in detail in an earlier report (Borrie *et al.*, 2012b). In brief, a standard reading passage (Rainbow Passage: Fairbanks, 1960) and a series of experimental phrases were elicited from three male native speakers of NZE with moderate hypokinetic dysarthria associated with Parkinson's disease. A moderate intelligibility impairment was defined as a score between 65% and 75% words correct on the Sentence Intelligibility Test[1] (SIT; Yorkston *et al.*, 1996). All of the selected experimental phrases were characterized perceptually by a rapid speaking rate, monopitch, monoloudness, reduced syllable stress, and imprecise consonants. Acoustic metrics—phrase duration, fundamental frequency variation, amplitude variation, and vowel space—were used to provide objective evidence of the presence of the perceived deviant speech features relative to speech produced by age- and gender-matched control speakers (see Borrie *et al.*, 2012b; Tables I and II).

TABLE I. Mean difference (MD) and Pearson product-moment correlation (R) coefficients for intra- and inter-judge reliability of the transcript analysis.

	Intra-judge		Inter-judge	
Analysis	MD (SD[a])	R[b]	MD (SD[a])	R[b]
Percent words correct	0.21 (0.34)	0.99*	0.52 (0.46)	0.99*
Percent syllable resemblance	0.40 (0.52)	0.97*	1.00 (0.47)	0.91*
Lexical boundary errors	0.30 (0.48)	0.99*	0.90 (0.57)	0.98*

[a] Standard deviation.

[b] An asterisk (*) indicates $p < 0.001$.

Experimental phrases were designed to enable speech segmentation errors to be interpreted relative to the Metrical Segmentation Strategy (MSS) predictions that listeners exploit strong syllables to determine word onsets in processing connected speech (Cutler and Butterfield, 1992; Cutler and Norris, 1988). All phrases were six syllables and employed either a strong–weak (SWSWSW) or weak–strong (WSWSWS) stress pattern (Liss *et al.*, 1998). Phrases ranged from three to five words in length and were semantically anomalous to eliminate known effects of semantic and contextual knowledge on speech perception. Using the phrases produced by the speakers with dysarthria, two speech sets were created for use in the perceptual learning paradigm—a *training speech set* and a *test speech set* (see the Appendix). These speech sets contained novel phrases but were balanced for: (a) number of phrases (36 phrases); (b) number of phrases produced by each speaker (12 phrases per speaker); (c) syllable stress pattern of the phrases (six trochaic and six iambic phrases per speaker); (d) number of words and syllables; and (e) number and type of lexical boundary error (LBE) opportunities. The 1 s average A-weighted sound pressure levels of all experimental stimuli (phrases and passage readings) were calibrated to within ±0.1 dB using a Brüel and Kjær Head and Torso Simulator Type 4128-C (Brüel and Kjær, Nærum, Denmark). Audio presentation of all speech stimuli was set to 65 dB (A). See Borrie *et al.* (2012b) for a comprehensive description on the construction and collection of speech stimuli.

D. Procedure

The first 40 listener participants were randomly assigned to one of two training conditions, linguistic (word identification) or indexical (speaker identification), so that each experimental group consisted of 20 participants. Control data from an additional 20 listener participants were collected to validate findings. The experiment was conducted in three distinct phases: (1) familiarization phase, (2) training phase, and (3) test phase. Figure 1 contains a diagrammatic representation of the perceptual learning paradigm employed.

The experiment was conducted in a quiet room using sound-attenuating headphones (Sennheiser HD 280 PRO). Listeners were tested individually. The experiment was presented via a laptop computer preloaded with the experimental procedure. Participants were told that they would undertake a listening task followed by a transcription task,

TABLE II. Category proportions of lexical boundary errors expressed in percentages and sum error ratio values for listeners by experimental group.[a]

Group[b]	%IS	%IW	%DS	%DW	IS/IW ratio	DW/DS ratio
Control	37.92	14.68	24.77	22.63	2.6	0.9
Linguistic	51.72	19.27	11.56	17.44	2.7	1.5
Indexical	54.53	16.21	10.53	18.74	3.4	1.8

[a] IS, IW, DS, and DW refer to lexical boundary errors defined as insert boundary before strong syllable, insert boundary before weak syllable, delete boundary before strong syllable, and delete boundary before weak syllable, respectively. Error ratio scores reflect strength of adherence to predicted error patterns, with the greater the positive distance from "1" indicative of increased adherence.

[b] $N = 20$.

and that task-specific instructions would be delivered via the computer program. This process was carried out to ensure identical stimulus presentation methods across participants.

During the *familiarization* phase, all listeners, regardless of group assignment, were presented with three readings of the rainbow passage—each produced by a different speaker with dysarthria. To ensure each speaker was heard in each position a similar number of times, the order in which each of the 20 participants in each experimental group 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 randomized using the Knuth implementation of the Fisher–Yates shuffling algorithm (Knuth, 1998). In addition to the readings, listeners in the indexical training group also received the name[2] of the speaker producing the passage (John, Bob, or Peter). Prior to passage presentation, all listeners were informed that they would hear some short passage readings. Additionally, listeners assigned to the training groups were notified of the nature of their subsequent task and given relevant instructions regarding attention allocation during familiarization with the passage readings—listeners in the linguistic training (word identification) group were instructed to listen carefully to any information that may help them learn to recognize what was being said[3] and listeners in the indexical training (speaker identification) group were instructed to listen carefully to any information that may help them learn to recognize the speaker.[4] Listeners in the control group received no additional instructions regarding attentional allocation during familiarization.

Immediately following the familiarization phase, listeners engaged in a *training* phase, which involved the 36 experimental phrases that made up the training speech set. Following the presentation of each individual phrase, listeners in the training groups participated in either a word (linguistic training) or speaker identification (indexical training) task. Listeners undertaking the word identification task were presented with three phonetically similar words and asked to use the mouse to select which word they thought they heard within the phrase (target word position varied with each phrase). They were told that they would have heard only one of the three words. Listeners were given as long as required to make their word selection. Upon selection of a word choice, regardless of accuracy, the correct response was highlighted as feedback regarding task performance. Listeners undertaking the speaker identification task were presented with the names of all three speakers and asked to use the mouse to select the speaker they thought they heard. As with the listeners in the word identification group, these listeners were given as long as required to make their name selection, and upon their selection of a name, the correct response was highlighted. Listeners in the control group were presented with the 36 training speech set phrases, however training (task and feedback) was not provided. The training phrases were presented randomly to each of the 60 listeners.

In order to ensure listeners trained with either the word or speaker identification task recognized the desired properties within the signal, linguistic or indexical respectively, a 70% criterion[5] across the 36 training items was selected. The software program that delivered the perceptual learning paradigm automatically identified whether a response was "correct" or "incorrect" on the word or speaker identification task. Responses were then tallied across the 36 items and converted into a single percent item correct score for each individual listener. All listeners performed above the 70% criterion on the training task and subsequently, the final analysis involved analysis of all 20 listener transcripts per training group. An independent *t*-test between percent correct identification for listeners who received the word identification training task ($M = 77.36$, SD $= 4.2$) and listeners who

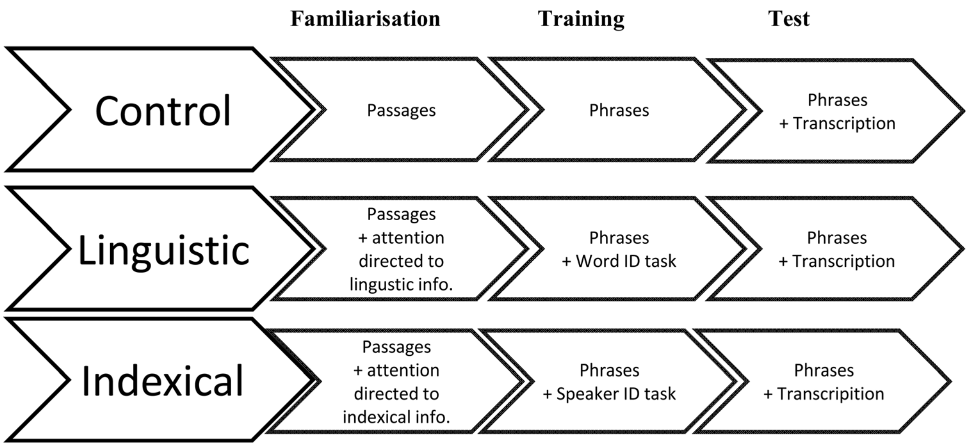

FIG. 1. Perceptual learning paradigm.

received the speaker identification training task ($M = 77.56$, SD $= 4.6$) revealed no statistically significant difference between the two training groups, $t(38) = 0.08$, $p = 0.97$, $d = 0.02$. This would suggest that similar levels of attention toward the intended training targets across the two training groups was achieved.

Immediately following the training task phase, the training groups and the control group participated in an identical *test* phase, in which they transcribed the 36 novel phrases that made up the test speech set. Transcription task instructions were identical to those of the previous two studies (Borrie *et al*., 2012b; Borrie *et al*., 2012c). 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 recognize. Listeners were told to place an "X" to represent part of a phrase, if they were unable to make a guess. They were given 12 s to type each response. The 36 phrases that made up the test speech set were presented randomly to each of the 60 listeners.

E. Transcription analysis

The total data set consisted of 60 transcripts of the 36 experimental phrases that made up the test speech set. The first author independently analyzed each of the listener transcripts for an index of intelligibility as well as measures considered indicative of segmental and suprasegmental processing. Measures were averaged across the 20 listener participants that comprised each of the experimental groups.

Percent words correct (PWC) was used to ascertain a standard measure of speech intelligibility for recognizing dysarthric speech. To be counted as correct, words were required to be an exact match to the intended target or differ only by the tense ''ed'' or the plural ''s.'' Word substitutions between "a" and "the" were also coded as correct. A PWC score, out of a total of 141 words, was tabulated for each individual listener transcript.

Percent syllable resemblance (PSR), a metric of segmental goodness developed by the first author and reported in Borrie *et al*. (2012b), was employed to afford insight into perceptual processing of segmental cues. This measure reflects whether syllables perceived in error, *resemble* their intended target, to at least some degree. A syllable is deemed to resemble its target, if it contains at least 50% phonemic accuracy—syllables with two phonemes required one correct phoneme, syllables with three phonemes required two correct phonemes, syllables with four phonemes required two or three correct phonemes, and syllables with five phonemes required three or four correct phonemes. The total number of syllables that resemble their 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. Percent syllable correct (PSC) scores were also collected for each transcript to enable PSR scores to be interpreted within the overall context of intelligibility. To be scored as correct, syllables had to match the intended target exactly. Each 36 phrase speech set contained a total of 216 syllables.

Finally, transcripts were also analyzed for LBEs, which offer information regarding processing of suprasegmental cues. Errors with lexical boundaries were coded for type (incorrect insertion or deletion of a boundary) and location (incorrect boundary occurring either before a strong or weak syllable). Thus, LBE errors could be coded into the following four categories: (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) (for error coding examples, see Liss *et al*., 1998). As per LBE analysis reported in Borrie *et al*. (2012b), error category proportions were calculated as a percent score for each experimental group and error ratios, IS/IW and DW/DS[6] were based on the sum of group errors for each group.

F. Reliability of transcription coding

Twenty-five percent of the listener transcripts were randomly selected according to a computer-generated random number list and were reanalyzed by one of the authors (intra-judge) and by a second trained judge (inter-judge) to obtain reliability estimates for the dependent variables PWC, PSR and number of LBEs. Discrepancies between the reanalyzed data and the original data analysis are reported in terms of absolute mean difference and Pearson's correlation coefficients reveal the degree of association between the data sets. Discrepancies between the reanalyzed data and the original data revealed that agreement was high (all $r > 0.90$), with only minor absolute differences. Table I summarizes the results.

III. RESULTS

A. Percent words correct

Figure 2 details the mean PWC scores for the three listener groups—those familiarized with dysarthric speech via a linguistic training task, an indexical training task or a control group who received no training. Prior to statistical analysis, the percentage correct values were transformed using rationalized arcsine transformation (Studebaker, 1985). One-way analysis of variance (ANOVA) was performed on the transformed values. Results showed a significant effect of group for PWC scores following familiarization with dysarthric speech, $F(2, 57) = 11.51$, $p < 0.001$, $\eta^2 = 0.32$. Post hoc tests, using Bonferroni correction, indicated that PWC scores of listeners in both the indexical, $t(38) = 3.68$, $p < 0.001$, $d = 1.16$, and linguistic, $t(38) = 4.88$, $p < 0.001$, $d = 1.54$, training groups were significantly higher than that of the control group. There was no significant difference in PWC scores between the linguistic or indexical training groups, $t(38) = 0.03$, $p = 0.871$, $d = 0.05$. Thus, similar intelligibility scores after training were observed for the listeners who received linguistic training and the listeners who received indexical training. Both groups' intelligibility

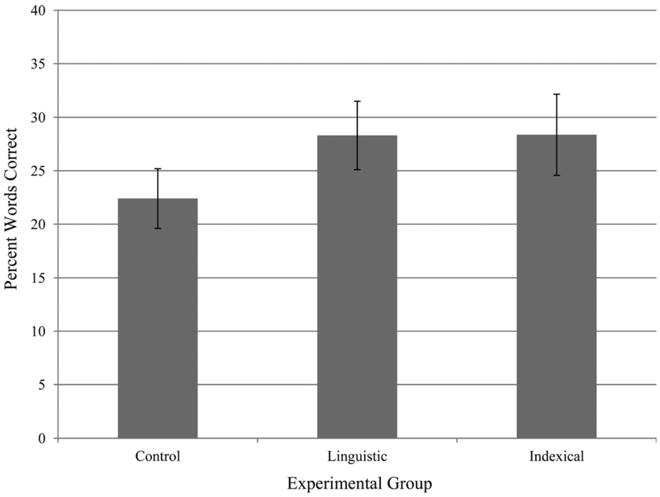

FIG. 2. Mean percent words correct (PWC) for listeners by training group. Bars delineate +1 standard deviation of the mean.

scores were significantly greater than those of listeners who did not receive training.

B. Percent syllable resemblance

Figure 3 details the mean PSR and PSC scores for the three listener groups—those familiarized with dysarthric speech via a linguistic training task, an indexical training task or a control group who received no training. Pearson product-moment correlation coefficients demonstrated a strong relationship between the variables of PSC and PWC for all three experimental groups ($r > 0.80$). Accordingly, statistical analysis was performed on the PSR data only, as PSC findings are reflected in the analysis of PWC. A one-way ANOVA showed no significant effect of group for PSR scores following familiarization with dysarthric speech, $F(2, 57) = 2.09$, $p = 0.132$. Thus, similar PSR scores were observed for all listeners familiarized with dysarthric speech, even when the familiarization did not involve training.

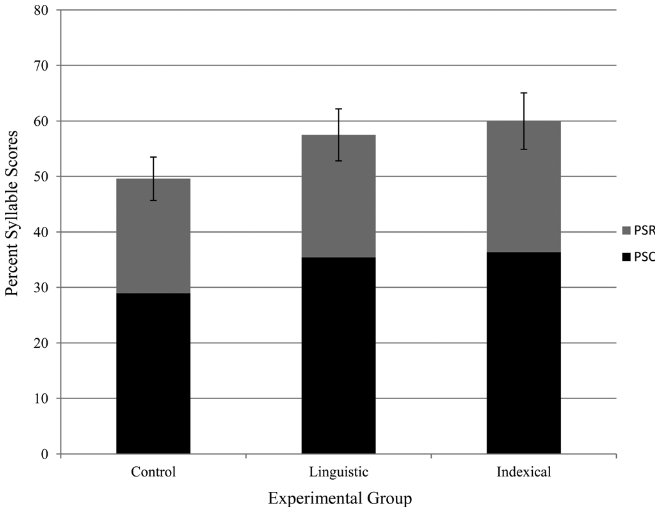

FIG. 3. Mean percent syllable correct (PSC) and mean percent syllable resemblance (PSR) for listeners by training group. Bars delineate +1 SD of the mean PSR data.

C. Lexical boundary errors

Table II details the LBE category proportions and the sum IS/IW and DW/DS ratios for the three listener groups—those familiarized with dysarthric speech via a linguistic training task, an indexical training task or a control group who received no training. Contingency tables, categorized by error type (i.e., insertion/deletion) and error location (i.e., before strong/weak syllable), were constructed using the total number of LBEs exhibited by each experimental group to determine whether the variables were significantly related. Within-groups χ^2 analyses revealed a significant interaction effect between the variables of type (insert/delete) and location (strong/weak) for the data generated by the group of listeners who received the linguistic training task, $X^2(1, N = 20) = 47.57$, $p < 0.001$, the group of listeners who received the indexical training task, $X^2(1, N = 20) = 73.10$, $p < 0.001$, and the control group, $X^2(1, N = 20) = 13.71$, $p < 0.001$. Listeners in all three groups made more predicted (IS and DW) than nonpredicted errors (IW and DS). This error pattern, according to the MSS hypothesis, indicates that all listeners utilized syllabic stress information to inform word boundary decisions.

Error ratios reflect the degree to which listeners utilize syllable stress to segment speech, with the greater the positive distance from "1" indicative of increased adherence to predicted error patterns. The ratio scores reveal that training groups conformed more strongly to predicted patterns relative to the control group. Further, error ratios exhibited by the indexical training group are greater than those observed in the linguistic training group, suggesting that attention toward speaker-specific acoustic information promotes increased reliance on syllabic stress contrast cues.

A between-group χ^2 analysis was used to examine differences in the error distribution between the experimental groups. Results identified significant differences in error distribution between the control and linguistic training group, $X^2(3, N = 40) = 70.01$, $p < 0.001$, and the control and indexical training group, $X^2(3, N = 40) = 82.69$, $p < 0.001$. No significant difference was found between the linguistic and indexical training groups, $X^2(3, N = 40) = 6.34$, $p = 0.10$. Thus, the relative distribution of errors observed for the control group was significantly different to that observed for the listeners trained to attend to either linguistic or indexical properties of the dysarthric signal. A between-group χ^2 analysis revealed no significant difference in error distribution between the linguistic and indexical training groups, $X^2(3, N = 40) = 4.50$, $p = 0.21$. Thus, the relative distribution of errors observed for the linguistic training group were similar to those observed for the indexical training group.

IV. DISCUSSION

The perceptual benefit of prior experience with a signal in which linguistic properties of the dysarthric stimuli are emphasized has been previously established (Borrie *et al.*, 2012b; Borrie *et al.*, 2012c; Liss *et al.*, 2002). The assumed mechanism for this benefit is that listeners learn to pair the degraded signal with corresponding mental representations of phonemes or words, thereby mapping (or remapping) for the characteristics of the speech (see, e.g., Eisner and McQueen, 2005; Francis *et al.*, 2007; Greenspan *et al.*, 1988). The present investigation sought to determine whether attending instead to indexical properties of the speaker would afford learning benefits as well. However, the hypothesized locus of learning would not be specifically at the linguistic level because, first, attention would be focused on speaker identity rather than the message, and second, the speech material is difficult to understand. This latter point is important because existing literature reports no perceptual benefit for listeners passively familiarized (no written feedback) to semantically anomalous phrases produced by speakers with a moderate hypokinetic dysarthria (Borrie *et al.*, 2012c). The current data suggest that a focus on either the linguistic aspects of the speech during training, or on indexical properties, both result in comparable benefits to speech intelligibility. In addition, there was evidence that both the word and speaker identification tasks facilitated increased use of suprasegmental information for lexical segmentation, with a trend of greater adherence to predicted error patterns for the indexical training group. Further, although the magnitude of syllable errors (PSR scores) were similar regardless of whether or not training was received, PSC scores suggest that listeners in the training groups learned something about mapping the degraded signals to phonemes. The present investigation provides additional support for the influence of linguistic information in improved recognition of dysarthric speech and offers preliminary insight into the role of indexical information in this learning process. The finding that perceptual learning afforded by an indexical training task was comparable to that achieved with a linguistic training task is discussed with regard to theoretical implications for processing of degraded speech.

Listeners who participated in a training task that emphasized indexical properties of the neurologically degraded signal achieved intelligibility scores that were significantly higher than that of a control group. Thus, it appeared that attention to the indexical elements of the dysarthric signal may provide a source of learning in perceptual adaption to this type and severity of speech degradation. Although different perceptual learning paradigms were employed, the current findings validate those reported by Nygaard and Pisoni (1998) and Nygaard *et al.* (1994), wherein improved linguistic processing of speech in noise was reported for listeners trained to identify the names of the speakers who produced the speech stimuli. The results also parallel recent findings in the sociophonetic perception literature, whereby experience with indexical information enhances bilingual speech processing—the translational priming effect (Szakay *et al.*, 2012). Further, the present study found that intelligibility improvements following an indexical training task paralleled those observed for listeners engaged in a training task in which linguistic properties were highlighted. Comparable intelligibility scores between the two training groups suggests that directing perceptual attention toward indexical elements of the signal offers similar performance gains to that achieved by directing attention toward the linguistic properties. This finding is consistent with studies examining perceptual learning of noise-vocoded speech—intelligibility scores for listeners familiarized with indexical elements of the signal (speaker identification task) were equivalent to those achieved by listeners familiarized with linguistic elements of the signal (transcription task) (Loebach *et al.*, 2008). Thus, the current findings reveal that the perceptual benefit of indexical information on processing of a speech signal that has been systematically degraded continues to be robust under the highly variable and frequently inconsistent acoustic degradation that characterizes dysarthric speech.

From the performance data alone, two conclusions can be drawn: that training to attend to indexical properties of the neurologically degraded signal does provide some perceptual benefit (relative to passive familiarization), and that this level of benefit is similar to that afforded by training with the linguistic aspects of the signal. Traditional views of perceptual processing do not account for the processing of speaker-specific detail, and thus the current findings extend support for the development of new theoretical paradigms in which indexical properties inform processing of spoken language (see, e.g., Goldinger, 1998; Palmeri *et al.*, 1993; Pisoni, 1997).

Examination of error patterns enables insight into the cognitive-perceptual mechanisms that underlie the performance benefits associated with focused training. Analysis of segmental-level errors revealed no significant difference in the number of syllables that resembled their phonetic target (PSR) between listeners trained to attend to linguistic information, listeners trained to attend to indexical information and listeners who received no focused training. Similar findings with processing of segmental information in perceptual learning of dysarthric speech have been previously reported, wherein the syllable errors following a passive familiarization task were similar in magnitude to the syllable errors following a more explicit familiarization task involving written feedback (Borrie *et al.*, 2012b). Syllable correct scores (PSC) aligned closely with the analysis of PWC data. Thus, regardless of which signal properties were emphasized, training to attend to specific aspects of the dysarthric signal enabled listeners to glean information about learnable acoustic–phonetic features.

Analysis of the LBE error patterns revealed that all three groups attended to syllabic stress cues in their attempts to decipher dysarthric speech, a conclusion evidenced in a greater proportion of predicted versus nonpredicted errors. Group differences, however, were evident in the degree to which syllabic stress information was exploited for the purpose of speech segmentation. Listeners in both training groups utilized these prosodic cues to a greater extent than the control (no training) group. Thus, it appears that a specific training task (word or speaker identification) involving iambic and tropic speech stimuli, serves to increase cognitive attention toward available stress information. In addition, error ratio discrepancies between training groups (higher ratios for the indexical training group) reveal that the speaker identification task promoted learning of this segmentation cue to a greater degree than the word identification task. This raises an interesting hypothesis for further testing—that stress patterns may be part of the indexical representation of the acoustic properties of dysarthric speech.

Taken together, findings from both training groups performance gains as well as error patterns observed with segmental and suprasegmental processing are remarkably similar, regardless of which signal properties are highlighted during training. These findings suggest that talker and phonetic information are integrally related properties, fundamentally linked in perceptual processing (see Nygaard, 2008). Although it is certainly possible that a longer training period would have facilitated more detectable group differences in the learning mechanisms that underlie enhanced speech processing following indexical or linguistic training, significant performance gains relative to the control group would suggest the current training paradigm was sufficient to promote substantial learning.

Future work in this area will aim to further examine the potential mechanisms involved with encoding and retaining indexical properties in conjunction with linguistic properties of the signal. To this end, the need to validate the assumption that the speaker and word identification tasks adequately encouraged attention toward desired information—indexical and linguistic properties, respectively—is required. Although a 70% test item criterion was employed to confirm task attention, tasking the linguistic group with identifying the speakers and the indexical group, the lexical items, would yield increased evidence for such a claim. Additionally, increasing speaker numbers may serve to increase attentional requirements of indexical training and facilitate more robust processing of speaker-specific detail. Such studies will build on knowledge regarding learning source associated with improved recognition of degraded speech (Borrie *et al.*, 2012b; Borrie *et al.*, 2012c; Liss *et al.*, 2002).

V. CONCLUSION

The present study provides preliminary evidence that both linguistic and indexical information can promote perceptual learning of dysarthric speech. Thus, there is empirical validation to support the development of a theoretical model that accounts for the interaction—or in fact, relationship—between linguistic and indexical properties as a source of learning in improved recognition of the neurologically degraded speech signal. These findings add to the growing body of literature that challenges long-standing theoretical paradigms that postulate independent processing of such information. Indeed, functional processing of linguistic and indexical information appears to be fundamentally linked.

ACKNOWLEDGMENTS

We thank the participants with Parkinson's disease, and their families, for their participation in this study. We also gratefully acknowledge the contribution of the listener participants. Primary support for this study was provided by a University of Canterbury Doctoral Scholarship (S.A.B.). Support from the New Zealand Neurological Foundation (Grant No. 0827-PG) and Health Research Council of New Zealand (Grant No. HRC09/251) (M.J.M.) and National Institute on Deafness and Other Communicative Disorders Grant (R01 DC 6859) (J.M.L.) is also gratefully acknowledged.

APPENDIX

See Table III for experimental phrases.

TABLE III. Experimental phrases.

Training speech set[a]	Test speech set
Account for who could *knock*	Address her meeting time
Admit the *gear* beyond	Amend estate approach
Afraid beneath *demand*	Assume to catch control
And *spoke* behind her sin	Attend the trend success
Attack *became* concerned	Award his drain away
Avoid or *beat* command	Beside a sunken bat
Balance *clamp* and bottle	Bolder ground from justice
Bush is chosen after	Cheap control in paper
Career *despite* research	Confused but roared again
Commit such *used* advice	Darker painted baskets
Connect the *beer* device	Define respect instead
Constant willing *walker*	Distant leaking basement
Cool the jar in private	Had eaten junk and train
Divide across retreat	Embark or take her sheet
Done with finest handle	For coke a great defeat
Frame her *seed* to answer	Forget the joke below
It's harmful *note* abounds	Functions aim his acid
Increase a *grade* sedate	Hold a page of fortune
Indeed a *tax* ascent	Mate denotes a judgment
Listen final station	Mistake delight for heat
Mark a single ladder	Mode campaign for budget
Measure *fame* with legal	Pick a chain for action
Model *sad* and local	Pooling pill or cattle
Narrow *seated* member	Push her equal culture
Her owners *arm* the phone	Remove and name for stake
Perceive sustained *supplies*	Rowing father matters
Rampant *boasting* captain	Seat for locking runners
Resting older earring	Secure but lease apart
Rocking modern *poster*	Signal breakfast pilot
Rode the lamp for testing	Sinking rather tundra
Round and *bad* for carpet	Stable wrist and load it
Spackle *enter* broken	Target keeping season
Submit *his* cash report	Transcend almost betrayed
Support with *dock* and cheer	Unless escape can learn
Technique *but* sent result	Unseen machines agree
To *sort* but fear inside	Vital seats with wonder

[a] Trained words (word identification task) identified in italics.

[1] The SIT has 15 semantically predictable sentences and aims to give a gross overall measure of speech performance.

[2] Names changed to comply with participant confidentiality agreement.

[3] Specific instructions: "You are going to hear some short passage readings. You will hear the same short passage produced by three different speakers. Following this, you will participate in a 'word identification task' so please listen carefully to any information that may help you learn to recognize what is being said."

[4] Specific instructions: "You are going to hear some short passage readings. You will hear the same short passage produced by three different speakers. Following this, you will participate in a 'speaker identification task' so please listen carefully to any information that may help you learn to recognize who is speaking."

[5] Based on the study by Nygaard and Pisoni (1998) in which the authors employed a 70% criterion to separate "good" from "poor" learners.

[6] According to predicted error patterns in which strong syllables cue word onsets (Cutler and Butterfield, 1992), errors should be largely of IS and DW in nature. The stimuli were constructed such that the opportunities to commit insertion and deletion errors before strong and weak syllables was

slightly higher for unpredicted errors (IW and DS).Thus, error ratios reflect strength of adherence to predicted error patterns. An error ratio of "1" indicates equal occurrence of insertion and deletion errors before strong and weak syllables. Greater ratio scores depict reliance on syllabic stress contrast cues for speech segmentation.

Borrie, S. A., McAuliffe, M. J., and Liss, J. M. (**2012a**). "Perceptual learning of dysarthric speech: A review of experimental studies," J. Speech Lang. Hear. Res. **55**, 290–305.

Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., and Anderson, T. (**2012b**). "Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthic speech," Lang. Cogn. Process. **27**, 1039–1055.

Borrie, S. A., McAuliffe, M. J., Liss, J. M., O'Beirne, G. A., and Anderson, T. (**2012c**). "A follow-up investigation into the mechanisms that underlie improved recognition of dysarthric speech," J. Acoust. Soc. Am. **132**, EL102–EL108.

Brown, J., and Carr, T. (**1993**). "Limits on perceptual abstraction in reading: Asymmetric transfer between surface forms differing in typicality," J. Exp. Psychol. Learn. Mem. Cogn. **19**, 1277–1296.

Costanzo, F. S., Markel, N. N., and Costanzo, P. R. (**1989**). "Voice quality profile and perceived emotion," J. Counsel. Psychol. **16**, 267–270.

Creelman, C. D. (**1957**). "The case of the unknown talker," J. Acoust. Soc. Am. **29**, 655.

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.

Eisner, F., and McQueen, J. M. (**2005**). "The specificity of perceptual learning in speech processing," Percept. Psychophys. **67**(2), 224–238.

Fairbanks, G. (**1960**). *Voice and Articulation Drillbook*, 2nd ed. (Harper and Row, New York), p. 234.

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.

Goldinger, S. D. (**1996**). "Words and voices: Episodic traces in spoken word identification and recognition memory," J. Exp. Psychol. Learn. Mem. Cogn. **22**, 1166–1183.

Goldinger, S. D. (**1998**). "Echoes of echoes?: An episodic theory of lexical access," Psychol. Rev. **105**, 251–279.

Goldinger, S. D., Pisoni, D. B., and Logan, J. S. (**1991**). "On the nature of talker variability effects of spoken word lists," J. Exp. Psychol. Learn. Mem. Cogn. **17**, 152–162.

Greenspan, S. L., Nusbaum, H. C., and Pisoni, D. B. (**1988**). "Perceptual learning of synthetic speech," J. Exp. Psychol.: Learn. Mem. Cogn. **14**(3), 421–433.

Hagiwara, R. (**1997**). "Dialect variation and formant frequency: The American English vowels revisited," J. Acoust. Soc. Am. **102**, 655–658.

Halle, M. (**1985**). "Speculations about the representation of words in memory," in *Phonetic Linguistics*, edited by V. A. Fromkin (Academic, New York), pp. 101–104.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Johnson, K. (**1997**). "Speech perception without speaker normalization: An exemplar mode," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic Press, San Diego, CA), pp. 145–165.

Johnson, K. (**2006**). "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology," J Phonetics **34**, 485–499.

Joos, M. A. (**1948**). "Acoustic phonetics," Lang. Monogr. **23**(23), 136.

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

Kirk, K. I., Pisoni, D. B., and Miyamoto, R. C. (**1997**). "Effects of stimulus variability on speech perception in listeners with hearing impairment," J. Speech Lang. Hear. Res. **40**, 1395–1405.

Knuth, D. E. (**1998**). *The Art of Computer Programming*, 3rd ed. (Addison–Wesley, Boston, MA), Vol. 2, pp. 145–146.

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

Levi, S. V., and Pisoni, D. B. (**2007**). "Indexical and linguistic channels in speech perception: Some effects of voiceovers on advertising outcomes," in *Psycholinguistics Phenomena in Marketing Communications*, edited by T. M. Lowrey (Lawrence Erlbaum, Mahwah, NJ), pp. 203–219.

Liss, J. M., Spitzer, S. M., Caviness, J. N., and Adler, C. (**2002**). "The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria," J. Acoust. Soc. Am. **112**(6), 3022–3030.

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.

Loebach, J. L., Bent, T., and Pisoni, D. B. (**2008**). "Multiple routes to the perceptual learning of speech," J. Acoust. Soc. Am. **124**(1), 552–561.

Luce, P. A., and Pisoni, D. B. (**1998**). "Recognizing spoken words: The neighbourhood activation model," Ear Hear. **19**, 1–36.

Mattys, S. L., and Liss, J. M. (**2008**). "On building models of spoken-word recognition: When there is as much to learn from natural 'oddities' as artificial normality," Percept. Psychophys. **70**(7), 1235–1242.

Morton, J. (**1969**). "Interaction of information in word recognition," Psychol. Rev. **76**, 165–178.

Mullennix, J. W., and Pisoni, D. B. (**1990**). "Stimulus variability and processing dependencies in speech perception," Percept. Psychophys. **47**, 379–390.

Munson, B., McDonald, E. C., DeBoe, N. L., and White, A. R. (**2006**). "The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech," J. Phonet. **34**, 202–240.

Murry, I. R., and Arnott, J. L. (**1993**). "Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion," J. Acoust. Soc. Am. **93**, 1097–1108.

Norris, D. (**1994**). "Shortlist: A connectionist model of continuous speech recognition," Cognition **52**, 189–234.

Nygaard, L. C. (**2008**). "Perceptual integration of linguistic and nonlinguistic properties of speech," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Malden, MA), pp. 390–413.

Nygaard, L. C., and Pisoni, D. B. (**1998**). "Talker-specific learning in speech perception," Percept. Psychophys. **60**, 355–376.

Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (**1994**). "Speech perception as a talker-contingent process," Psychol. Sci. **5**, 42–46.

Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (**1993**). "Episodic encoding of voice attributes and recognition memory for spoken words and voices," J. Exp. Psychol. Learn. Mem. Cogn. **18**, 915–930.

Pisoni, D. B. (**1997**). "Some thoughts on 'normalization' in speech perception," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego, CA), pp. 9–32.

Ryalls, B. O., and Pisoni, D. B. (**1997**). "The effect of talker variability on word recognition in preschool children," Dev. Psychol. **33**(3), 441–452.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primary temporal cues," Science **62**(4), 834–842.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Summerfield, Q., Haggard, M., Foster, J., and Gray, S. (**1984**). "Perceiving vowels from uniform spectra: Phonetic exploration of an auditory after effect," Percept. Psychophys. **35**, 203–213.

Szakay, A., Babel, M., and King, J. (**2012**). "Sociophonetic markers facilitate translation priming: Maori English GOAT – A different kind of animal," University of Pennsylvania Working Papers in Linguistics **18**(2), Article 16, http://repository/upenn.edu/pwpl/vol18/iss2/16.

Tenpenny, P. L. (**1995**). "Abstractionist versus episodic theories of repetition, priming and word identification," Psychonom. Bull. Rev. **2**, 339–363.

Yohan, C. A., and Sommers, M. S. (**2000**). "The effects of talker familiarity on spoken word identification in younger and older adults," Psychol. Aging **15**, 88–99.

Yorkston, K. M., Beukelman, D. R., and Hakel, M. (**1996**). "Speech intelligibility test for windows," Institute for Rehabilitation Science and Engineering at Madonna Rehabilitation Hospital, Lincoln, NE.