Research Note

The Effect of Background Noise on Intelligibility of Dysphonic Speech

Keiko Ishikawa,^a Suzanne Boyce,^a Lisa Kelchner,^a Maria Golla Powell,^a
Heidi Schieve,^a Alessandro de Alarcon,^b and Sid Khosla^c

Purpose: The aim of this study is to determine the effect of background noise on the intelligibility of dysphonic speech and to examine the relationship between intelligibility in noise and an acoustic measure of dysphonia: cepstral peak prominence (CPP).

Method: A study of speech perception was conducted using speech samples from 6 adult speakers with typical voice and 6 adult speakers with dysphonia. Speech samples were presented to 30 listeners with typical hearing in 3 noise conditions: quiet, signal-to-noise ratio (SNR)+5, and SNR+0. Intelligibility scores were obtained via orthographic transcription as the percentage of correctly identified words. Speech samples were acoustically analyzed using CPP, and the correlation

between the CPP measurements and intelligibility scores was examined.

Results: The intelligibility of both typical and dysphonic speech was reduced as the level of background noise increased. The reduction was significantly greater in dysphonic speech. A strong correlation was noted between CPP and intelligibility score at SNR+0.

Conclusions: Dysphonic speech is relatively harder to understand in the presence of background noise as compared with typical speech. CPP may be a useful predictor of this intelligibility deficit. Future work is needed to confirm these findings with a larger number of speakers and speech materials with known predictability.

Individuals with dysphonia frequently report difficulty with intelligibility, especially in a noisy communication environment (Jacobson et al., 1997; Smith et al., 1998). To some extent, this situation is not surprising, as dysphonia affects a number of acoustic cues that listeners use for understanding speech. A brief review of typical speech acoustics and some examples of dysphonic speech acoustics may make this clearer. In a typical voice, vocal fold vibration provides a harmonically rich sound source, which is interrupted at intervals by devoiced consonants, vocal tract closures for stop consonants, and pauses. When this sound source is operational, the changing shapes of the vocal tract during speech production selectively attenuate and amplify clusters of harmonics at frequencies specific to speech sounds (Fant, 1960). These patterns of

vocal tract resonance, called *formants*, supply most of the information in the speech signal for vowels and sonorant consonants (Dalston, 1975; Peterson & Barney, 1952). They also supply most of the information for place of articulation with regard to obstruent consonants as well as a host of additional segmental and suprasegmental information involved in conveying meaning (Delattre, Berman, & Cooper, 1962; Delattre, Liberman, & Cooper, 1955). The greatest fraction of the total intensity is contributed by the fundamental frequency of vocal fold vibration and the low frequencies associated with it; for adults, these frequencies are mostly below 400 Hz (Baken & Orlikoff, 2000). Thus, for instance, the vowels of “seen” and “soon” have similar first formant frequencies in the loudest region of the signal, below 400 Hz, but are distinguished from each other by differences in the softer regions of the signal: the second formant between 1000 and 3000 Hz (Peterson & Barney, 1952). In contrast, the stop consonants of “bit” and “pit” are distinguished from one another more by the timing patterns of harmonic information versus vocal tract information than by frequency information. The interplay between vocal tract changes and the sound source also provides the major source for other linguistic (voice onset time, pausing, vowel-intrinsic fundamental frequency, amplitude, and duration) and paralinguistic (intonation, emphasis,

^aDepartment of Communication Sciences and Disorders, University of Cincinnati, OH

^bDepartment of Otolaryngology, Cincinnati Children’s Hospital Medical Center, OH

^cDepartment of Otolaryngology, University of Cincinnati, OH

Correspondence to Keiko Ishikawa: ishikak@mail.uc.edu

Editor: Julie Liss

Associate Editor: Bharath Chandrasekaran

Received January 10, 2016

Revision received July 19, 2016

Accepted January 20, 2017

https://doi.org/10.1044/2017_JSLHR-S-16-0012

Disclosure: The authors have declared that no competing interests existed at the time of publication.

speaker identity, etc.) information in speech. If the accessibility of information in the signal is reduced, it is likely that intelligibility is affected.

Speakers with dysphonia typically have typical functioning supraglottic articulators, with the exception of cases secondary to central neurological disorders. Therefore, acoustic abnormality in dysphonic speech is most likely the result of a problem with sound source production. Dysphonic voices frequently show a loss of total intensity (Casper & Leonard, 2006), so that, for any distance between the speaker and listener, the total amount of information available to the listener is reduced. This in itself may be sufficient to account for patients' complaints of an intelligibility deficit due to dysphonia. However, other aspects of dysphonic speech should also be considered. The incomplete glottic closure often seen in dysphonic speech significantly reduces the intensity of harmonics in the frequency ranges of the higher formants beyond what is seen in typical speech (Bickley, 1982; Hillenbrand, Cleveland, & Erickson, 1994). This significant reduction may make it harder for listeners to decide which speech sound was produced. Under ideal conditions, a listener who fails to hear information from higher formants may be able to guess what was said on the basis of louder low-frequency information; however, the information may be unrecoverable when external environmental noise is present (Parikh & Loizou, 2005). Incomplete glottic closure also introduces turbulent noise in the higher-frequency range (Isshiki, Kitajima, Kojima, & Harita, 1978). This noise may mask information from formants and secondary noise sources that distinguish obstruent consonants such as fricatives, affricates, and stops. Some cases of dysphonia are secondary to a mass lesion on the vocal folds. The mass can vibrate at a different frequency from the rest of the vocal fold tissue or create unexpected onsets and offsets of phonation. These spectral and temporal abnormalities in dysphonic speech tend to make information contained in the speech signal more ambiguous. When the "bottom-up" information is ambiguous, listeners will need to use their linguistic knowledge and consider more alternative interpretations of what is said. The need to keep up with the speaking rate of the speaker will force them to make hurried perceptual decisions, which may result in more communication breakdowns. Therefore, even when loudness is held constant, listeners are likely to need to expend more cognitive resources to understand dysphonic speech.

It is well known that background noise affects intelligibility. At 1.5 m, which is a typical distance between communication partners, the average intensity of connected speech is 56 dB (Hall, 1966). However, modern society is a noisy place. Background noise levels in homes and classrooms without speech by communication partners range from 45 to 55 dB sound pressure level (SPL) (Pearsons, Bennett, & Fidell, 1977). The level is substantially higher in nonresidential environments, averaging 71 dBA SPL in restaurants (Lebo et al., 1994), and could be as high as 95 dBA SPL on urban streets (McAlexander, Gershon, & Neitzel, 2015). This means that for most communication situations, the intensity level of the communication—the

signal—is frequently equal to or less than the level of the background noise. In signal-to-noise ratio (SNR) terms, when the SNR is +0 dB, the intensity of the speech (i.e., the signal) and noise is equal. SNR+10 dB means that the intensity of speech is 10 times greater than the background noise. In a daily communication environment, a minimum of SNR+10 dB is required for speech to be 100% intelligible at 1 m away from the origin of the speech signal (Pearsons et al., 1977). Accordingly, even when speakers have typical voices, listeners may be receiving speech at approximately SNR+0 dB during conversation even in relatively quiet environments. The situation for dysphonic voices may be more extreme.

The effect of background noise on the intelligibility of typical speech has been extensively explored in the hearing science literature (Bradlow, Torretta, & Pisoni, 1996; Draegert, 1951; Miller & Nicely, 1955; Payton, Uchanski, & Braida, 1994). On the other hand, the effect of background noise on dysphonic speech has received very limited attention in the literature. Two specific types of dysphonic speech have been examined, namely, that of alaryngeal speakers (Holley, Lerman, & Randolph, 1983; McColl, Fucci, Petrosino, Martin, & McCaffrey, 1998) and that of speakers with Parkinson's disease (Leszcz, 2012). Studies that examined the effect of background noise on intelligibility of speakers with Parkinson's disease reported that the background noise further reduces intelligibility of these speakers in comparison with speakers with typical speech (Adams, Dykstra, Jenkins, & Jog, 2008; Leszcz, 2012). McColl et al. (1998) examined the effect of background noise on tracheoesophageal speech at nine SNR levels, ranging from -15 dB to +65 dB. Their findings indicated that listeners had more difficulty with understanding tracheoesophageal speech in comparison with typical speech, and the difference began to emerge at SNR-5 dB. Holley et al. (1983) reported that esophageal, electrolarynx, and typical speech showed no difference in intelligibility at SNR+3 dB but that a difference did become significant at SNR-1 dB. Interestingly, the greatest intelligibility was found with esophageal speech, followed by electrolarynx speech, and typical speech. The authors attributed this unexpected finding to possible acoustic similarity between typical speech and multispeaker babble that was used as the background noise. These studies confirm the hypothesis that background noise affects intelligibility for all speech but may affect the intelligibility of dysphonic speech differently from that of typical speech. However, findings on alaryngeal speech and Parkinsonian speech are not necessarily generalizable to speakers with dysphonia with other types of etiologies. The acoustic characteristics of phonation in typical speech are significantly different from those of alaryngeal speech (Qi & Weinberg, 1995). In the case of speakers with Parkinson's disease, a role for the involvement of supraglottic articulators in intelligibility cannot be ruled out.

Currently, a typical dysphonia evaluation protocol does not include a direct intelligibility assessment, and the relationship between any particular acoustic measure and intelligibility of dysphonic speech is unknown. A practical reason for the absence may be that the effect of background

noise on dysphonic speech is difficult to measure in the quiet office, which is typical of a voice clinic. Another practical reason may stem from an underlying clinical assumption that intelligibility and the harmonic power of speech are tightly correlated; thus, the intelligibility deficit will automatically improve as the laryngeal condition is addressed and associated harmonic power improves. Although this is an extremely reasonable assumption, the absence of measurement is problematic as it prevents clinicians from testing the validity of the assumption and tracking the correlation of the patient's improvement in harmonic power with improvement in intelligibility. Furthermore, not all laryngeal pathologies are equally responsive to the currently available treatment (Benninger et al., 1996; MacArthur, Kearns, & Healy, 1994), leaving the intelligibility deficit unresolved for individuals with those conditions.

Acoustic analysis is frequently used in clinical voice evaluation. For the purposes of this article, however, it should be noted that the majority of the acoustic parameters used in voice evaluation are measured with a sustained vowel. Although this task provides a reasonable context for illustrating the relationship between laryngeal behavior and output sound, its ecological validity as a predictor of real-life challenges is poor. The use of sentence material would likely reveal intelligibility deficit better than sustained vowels. Cepstral peak prominence (CPP) is a periodicity measure based on the cepstrum, which is a Fourier transformation of the speech spectrum (Awan, Roy, Jetté, Meltzner, & Hillman, 2010). Unlike F_0 -based measures or inverse filtering, the measure includes information about supra-glottal articulatory movements. CPP from connected speech seems to correlate with judgment of dysphonic voice quality better than other acoustic measures (Awan et al., 2010; Halberstam, 2003; Heman-Ackah, Michael, & Goding, 2002; Hillenbrand & Houde, 1996; Lowell, Colton, Kelley, & Mizia, 2013). The periodicity of the signal is expressed as a distance from the cepstral peak to a linear regression line that represents the average sound energy. A highly periodic voice signal that is typically associated with typical voice has high-amplitude CPP, whereas a weakly periodic, or aperiodic, voice signal that is often associated with dysphonic voice has low-amplitude CPP. It has been reported that a CPP of less than 5 dB indicates dysphonic voice quality (Heman-Ackah et al., 2003). CPP is an attractive measure for examining the relationship between dysphonia and intelligibility because (a) it can be calculated on continuous speech and (b) it is sensitive to the human perception of dysphonia both with sustained vowels and continuous speech.

The first step toward understanding the interaction between dysphonia, background noise, and intelligibility is to quantify the effect of background noise at realistic levels on the intelligibility of dysphonic speech. Therefore, the first aim of this study was to compare the intelligibility of dysphonic and typical speech in different levels of background noise. We hypothesized that the intelligibility reduction would be greater in dysphonic speech in comparison with typical speech and that the difference between typical

and dysphonic speech would manifest at a particular SNR. Because the effect of reduced speech intensity on intelligibility (at the same communication distance) is well established, we aimed to separate this effect from other potential effects from the acoustic properties of dysphonic speech. Therefore, total intensity of speech samples was standardized for this experiment. Time and labor are at a premium in the clinical process. If an acoustic measure that is already obtained in voice evaluation could indicate the level of intelligibility deficit, we could negate the need for a manual intelligibility measurement. Thus, we also aimed to determine whether a particular acoustic measure might be a useful predictor of intelligibility. Specifically, we focused on CPP at the phrase or sentence level. Accordingly, the second aim of this study was to examine the relationship between intelligibility of dysphonic speech in noise and a clinically used acoustic measure of dysphonia, namely, CPP. It was hypothesized that CPP would predict intelligibility of dysphonic speech in noise.

Method

Speakers and Speech Recording

Speech samples were collected from six adults with typical voice and six adults with varying degrees of dysphonia. The average age of speakers with typical speech was 55.85 years ($SD = 10.75$ years). The average age of speakers with dysphonia was 68.12 years ($SD = 12.02$ years). Each group consisted of three women and three men. Alaryngeal speakers and speakers with dysphonia secondary to central neurological disorders were excluded from the study. The clinical diagnosis of dysphonia was confirmed by a laryngologist and speech-language pathologists at the University of Cincinnati Health Performance and Professional Voice Center. Each speaker recorded six sentences from the Consensus of Auditory Perceptual Evaluation of Voice (CAPE-V), 10 sentences of 20 semantically anomalous sentences (Vaughan, Storzbach, & Furukawa, 2006), and five spontaneous utterances elicited by the Cookie Theft picture (Goodglass & Kaplan, 1983). Samples were recorded in a quiet room with a unidirectional dynamic microphone (Shure SM48; Shure Incorporated, Niles, IL) held at a constant 5 cm off-axis position from the mouth. The recordings were digitized at the sampling frequency of 50 kHz using the Computerized Speech Lab (Model 4500; KayPENTAX, Lincoln Park, NJ). Three speech-language pathologists specializing in the care of voice disorders individually rated the dysphonia severity for each speaker and sentence using the standard perceptual CAPE-V protocol.

Table 1 shows auditory-perceptual ratings of all speakers with dysphonia on CAPE-V by three speech-language pathologists. The CAPE-V measures severity of dysphonia with one main perceptual category, overall severity, and five subcategories: roughness, breathiness, strain, pitch, and loudness. The degree of abnormality is measured on a visual analog scale, with 0 given to typical voice quality and 100 given to the most severe dysphonic

Table 1. Consensus of Auditory Perceptual Evaluation of Voice ratings.

Speaker	Overall	Rough	Breathy	Strain
D1	5.00 (1.53)	4.67 (1.45)	0.33 (0.33)	0.67 (0.67)
D2	37.33 (13.86)	36.33 (11.05)	2.00 (2.00)	2.00 (2.00)
D3	53.67 (10.04)	48.00 (7.21)	36.67 (5.93)	5.33 (2.33)
D4	46.67 (7.86)	31.67 (10.48)	31.00 (9.29)	20.00 (8.02)
D5	35.67 (7.31)	11.67 (11.67)	32.67 (5.61)	8.33 (1.20)
D6	9.33 (3.71)	9.33 (3.71)	4.00 (3.06)	0.00 (0.00)

Note. The values are the average of ratings from three raters. Note that the ratings were shown only on the following categories: "Overall Severity," "Roughness," "Breathiness," and "Strain." Values shown in the parentheses are standard errors. A letter D indicates "dysphonic" status of the speakers.

voice quality. All speakers with typical speech scored 0 on all categories. On average, the overall severity of our speakers with dysphonia ranged from 5 to 53.67 out of 100. The severity rating was lowest for Speaker D1 (i.e., the mildest dysphonia), followed by D6, D5, D2, D4, and D3. The primary feature of dysphonia was "roughness" for D1, D2, D3, and D6, indicated by the higher values in this category compared with "breathiness" and "strain." The interrater reliability of the raters was moderate (intraclass correlation coefficient = .64; 95% confidence interval [0.09, 0.9]).

Preparation of Stimuli

All samples were manually edited using WaveSurfer acoustic analysis software (Sjölander & Beskow, 2011) to extract the speech portion of the recording. The edited samples were down sampled to 22500 Hz, and the intensity of the signal was standardized at 71.2–71.5 dB SPL using the transform function of WaveSurfer. Five hundred ms of silence was added before and after the speech signal following the intensity standardization. Two levels of noise were then added to the stimuli, resulting in three different versions of the same stimulus: quiet, SNR+5 dB, and SNR+0 dB. These SNRs were chosen to reflect daily communication scenarios in which background noise is either at the same level or louder than the level of the speaker. The noise was added using a custom designed MATLAB program, which selected a random noise segment of the same length as the speech from a recording of cafeteria noise (Auditec, St. Louis, MO). The intensity of the speech signal was held constant, and noise was added to achieve the targeted SNR. Scaling was done as necessary to avoid clipping of the combined signal.

A total of three semantically anomalous sentences and two sentences from the Cookie Theft picture (Goodglass & Kaplan, 1972) were selected for each speaker. To avoid exposing the listeners to the same sentence multiple times, sentences that were selected for multiple speakers were distributed to two separate master stimuli lists. A schematic representation of the stimulus list design is shown in Table 2. Each master list contained recordings from three

speakers with typical speech and three speakers with disordered speech, for a total of 30 stimuli per list. Three sublists were then created for each master list, differing in the noise condition for each stimulus. The sublists were balanced for noise condition and sentence type (i.e., semantically anomalous sentences and the Cookie Theft sentences). Consequently, no listener heard the same stimuli twice, yet every stimulus was presented in every condition over the course of the experiment. The stimuli were loaded onto Superlab 4 (Cedrus Corporation, San Pedro, CA), which presented them in a random order.

Listeners

A total of 30 naïve adult listeners between 20 and 30 years of age participated in the study. The average age of the listeners was 26.05 years, with a standard deviation of 2.31 years. There were 16 women and 14 men in the group. All listeners were native speakers of American English. Prior to the experiment, the listeners' speech recognition ability was screened by the Word Intelligibility by Picture Identification Test (Ross & Lerman, 1971). The test was designed for a pediatric population; however, it has also been used with adults for experimental purposes (Bradley & Sato, 2004; Lenhardt, Skellett, Wang, & Clarke, 1991; Papso & Blood, 1989). It has been reported that adults with typical hearing have no difficulty with word identification in this test (Papso & Blood, 1989). Therefore, it was assumed that listeners' performance on this test would confirm their basic speech recognition ability. All listeners scored at least 95% on the Word Intelligibility by Picture Identification Test.

Listening Procedures

The listening experiment was conducted in a quiet room in the University of Cincinnati Communication Sciences and Disorders Speech laboratory. Following acquisition of informed consent, listeners were asked to sit on a desk chair in front of a computer monitor and keyboard. One of six stimuli lists was assigned to a listener. There were five listeners per list. The stimuli were presented binaurally through a passive noise cancellation headphone (Sennheiser HD 201; Sennheiser electronic GmbH & Co. KG, Wedemark, Germany) at a comfortable listening level of 68–72 dB SPL. Listeners were asked to listen to the stimuli, repeat the stimuli into a microphone, and type what they heard. Experimenters referred to the audio recordings in a review of the listeners' responses if clarification of the typed response was necessary.

Acoustic Analysis of Speech Samples

CPP was obtained from all samples used in the listening experiment using Analysis of Dysphonia in Speech and Voice (KayPENTAX). The analysis was conducted after the down sampling and intensity normalization. The beginning and end of speech were visually confirmed on the basis of a waveform, and the Analysis of Dysphonia in

Table 2. Schematic display of the stimulus list design.

List 1					List 2				
Speaker ID	Sentence ID	List 1a	List 1b	List 1c	Speaker ID	Sentence ID	List 2a	List 2b	List 2c
N1	SAS 1	Quiet	SNR+5	SNR+0	N4	SAS 1	Quiet	SNR+5	SNR+0
N1	SAS 2	SNR+0	Quiet	SNR+5	N4	SAS 2	SNR+0	Quiet	SNR+5
N1	SAS 3	SNR+5	SNR+0	Quiet	N4	SAS 3	SNR+5	SNR+0	Quiet
N1	CT 1	Quiet	SNR+5	SNR+0	N4	CT 13	Quiet	SNR+5	SNR+0
N1	CT 2	SNR+0	Quiet	SNR+5	N4	CT 14	SNR+0	Quiet	SNR+5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
D1	SAS 10	Quiet	SNR+5	SNR+0	D4	SAS 13	Quiet	SNR+5	SNR+0
D1	SAS 11	SNR+0	Quiet	SNR+5	D4	SAS 14	SNR+0	Quiet	SNR+5
D1	SAS 12	SNR+5	SNR+0	Quiet	D4	SAS 15	SNR+5	SNR+0	Quiet
D1	CT 7	Quiet	SNR+5	SNR+0	D4	CT 19	Quiet	SNR+5	SNR+0
D1	CT 8	SNR+0	Quiet	SNR+5	D4	CT 20	SNR+0	Quiet	SNR+5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
D3	SAS 16	SNR+0	Quiet	SNR+5	D6	SAS 18	SNR+0	Quiet	SNR+5
D3	SAS 17	SNR+5	SNR+0	Quiet	D6	SAS 19	SNR+5	SNR+0	Quiet
D3	SAS 18	Quiet	SNR+5	SNR+0	D6	SAS 20	Quiet	SNR+5	SNR+0
D3	CT 11	SNR+0	Quiet	SNR+5	D6	CT 23	SNR+0	Quiet	SNR+5
D3	CT 12	SNR+5	SNR+0	Quiet	D6	CT 24	SNR+5	SNR+0	Quiet

Note. SNR = signal-to-noise ratio. Table displays only a part of the participants, sentence ID, and noise conditions for the lists. N and D for speaker ID indicate normal and dysphonic, respectively. SAS and CT for the Sentence ID indicate type of speech material, namely, “Semantically Anomalous Sentences” and “Cookie Theft Sentences.” Numbers after these identifiers indicate different sentences. Note that some of the SAS were used in both List 1 and List 2, and they were distributed to the separate main lists to avoid exposing a listener to the same sentence twice.

Speech and Voice algorithm automatically calculated CPP for the selected portion of the signal.

Statistical Analyses

Interrater reliability among the expert raters who rated dysphonia severity was examined with intraclass correlation coefficients. Mixed-effect logistic regression analysis was performed to examine the effect of diagnosis (i.e., typical vs. dysphonic) and noise conditions (quiet, SNR+5 dB, and SNR+0 dB) on intelligibility. The model’s response variable was set in a binary form with a number of words that were correctly identified and total number of words in a sentence. Diagnosis and noise conditions were set as fixed effects and listener was set as random effects. For examining a relationship between intelligibility scores and CPP, the intelligibility scores were calculated as the percentage of correctly identified words in a sentence. First, the relationship between the intelligibility scores at three noise conditions and CPP was examined at the sentence level using a multiple linear regression model. The mean of intelligibility scores was calculated across the listeners for this analysis. Because the model diagnostics indicated the presence of influential outliers, a robust regression model was used to evaluate the relationship. The robust regression model is a form of a weighted least square method, which diminishes the influence of outliers by assigning weights to observations on the basis of the residuals, thus allowing inclusion of all data in the analysis (Huber, 1964). The correlation between intelligibility scores and CPP was also examined at a speaker level. The means of CPP across sentences and intelligibility scores across sentences and listeners were calculated for each speaker, and their correlation was examined with the

regression models. The significance level for these tests was set at $p < .05$. All statistical tests were carried out with statistical software R version 3.1.0 (R Core Team, 2014).

Results

CPP of Speech Samples

CPP values averaged across speech samples used in the listening experiment are shown in Table 3. The group composed of speakers with dysphonia had CPP values ranging from 2.49 to 7.759 dB. Speaker D1 had the highest mean CPP value, followed by D6, D2, D5, D3, and D4. The group of speakers with typical speech had CPP values ranging from 5.226 to 7.322 dB. Speaker N6 had the highest mean CPP value, followed by N3, N2, N1, N5, and N4. CPP values of all speakers with typical speech were above the reported threshold value of dysphonia, 5 dB (Heman-Ackah et al., 2003). The average CPP values across speakers were 4.26 dB ($SD \pm 1.9$) for the group of speakers with dysphonia and 6.09 dB ($SD \pm 0.79$) for the speakers with typical speech. A Wilcoxon rank sum test indicated that the difference in CPP between the groups was not statistically significant ($W = 30$, $p = .06$). As D1’s CPP value was considerably higher than that of the other speakers in the dysphonic group, it was suspected that the lack of group difference was driven by this speaker. A Wilcoxon rank sum test was repeated after removal of this speaker from the group. The test indicated a statistically significant group difference ($W = 0$, $p = .004$).

Effect of Background Noise on Intelligibility

The presence of background noise reduced intelligibility in both the typical and dysphonic groups (see Figure 1).

Table 3. Mean intelligibility scores in all noise conditions and cepstral peak prominence (CPP) of each speaker.

Speaker	SNR+0	SNR+5	Quiet	CPP (dB)
D1	79.352 (4.812)	85.765 (5.249)	89.289 (5.132)	7.759 (0.051)
D2	9.238 (3.985)	41.814 (7.048)	76.114 (6.397)	4.136 (0.078)
D3	21.667 (5.865)	71.810 (5.420)	84.643 (5.831)	2.830 (0.094)
D4	26.786 (4.911)	89.500 (3.168)	96.000 (1.754)	2.490 (0.063)
D5	41.752 (8.064)	71.410 (6.534)	86.857 (5.012)	3.825 (0.114)
D6	28.825 (4.262)	90.952 (2.837)	93.460 (2.924)	4.544 (0.080)
N1	57.198 (6.565)	87.698 (2.792)	95.913 (1.221)	5.700 (0.063)
N2	68.818 (6.678)	93.000 (3.106)	94.000 (3.237)	6.048 (0.140)
N3	78.629 (4.682)	93.214 (2.575)	95.571 (2.075)	6.703 (0.040)
N4	45.267 (5.949)	86.567 (4.685)	96.600 (1.973)	5.226 (0.065)
N5	61.333 (6.354)	89.621 (2.694)	92.222 (3.585)	5.547 (0.060)
N6	60.429 (6.621)	86.429 (3.603)	90.500 (3.250)	7.322 (0.144)

Note. The values are the mean across all sentences. The values in the parentheses are standard errors. Letters D and N indicate “dysphonic” and “normal” status of the speakers. SNR = signal-to-noise ratio.

For both speaker groups, intelligibility scores were the lowest in the noisiest condition (i.e., SNR+0 dB), followed by SNR+5 dB and in quiet. The scores of the group with speakers with typical speech were 94.13% for quiet, 89.42% for SNR+5 dB, and 61.95% for SNR+0 dB. The scores of the group of speakers with dysphonia were 87.73% for quiet, 75.21% for SNR+5 dB, and 34.6% for SNR+0 dB. The logistic regression model indicated that the effect of listener was minimal ($\sigma^2 = 1e^{-14}$), and the overall effect of dysphonia and noise level was statistically significant ($\chi^2 = 79$,

$df = 1$, $p < .01$ for dysphonia; $\chi^2 = 419.3$, $df = 2$, $p < .01$ for noise level; see Table 4). A post hoc test (Wilcoxon rank sum test with Bonferroni correction) was carried out to evaluate the difference in intelligibility scores between typical and dysphonic groups in all noise conditions. The intelligibility scores were calculated as the percentage of correctly identified words for this test. The test indicated that the intelligibility scores are significantly different between the groups at SNR+5 dB ($p < .001$) and SNR+0 dB ($p < .001$). There was no statistically significant difference between the groups in the quiet condition ($p = .018$).

Figure 1. Mean intelligibility scores of normal and dysphonic groups in all noise conditions. The values are in percentages. Error bars indicate standard errors. SNR = signal-to-noise ratio.

Correlation Between Intelligibility Scores and CPP

A multiple linear regression analysis at the sentence level (i.e., intelligibility scores were averaged across listeners) indicated that the association between intelligibility scores at the three noise conditions and CPP was significant, but only a small amount of variance in CPP was explained by the model, $F(3, 296)$, 8.93, $p < .01$, $R^2 = .08$, $R^2_{adjusted} = .07$. The analysis showed that CPP was significantly associated with intelligibility score at SNR+0 dB and SNR+5 dB but not in the quiet condition, $\beta = 1.124$, $t(299) = 3.996$, $p < .01$ for SNR+0 dB; $\beta = 0.864$, $t(299) = 2.19$, $p = .03$ for SNR+5 dB; $\beta = 0.528$, $t(299) = 1.055$, $p = .29$ for quiet. The robust regression model confirmed these results: $\beta = 1.272$, $t(299) = 4.208$, $p < .01$ for SNR+0 dB; $\beta = 0.9$, $t(299) = 2.121$, $p = .035$ for SNR+5 dB; $\beta = 0.697$, $t(299) = 1.294$, $p = .196$, for quiet. Results of these models are shown in Table 5 and Figure 2.

A multiple linear regression analysis at the speaker level (i.e., intelligibility scores were averaged across listeners and sentences) indicated that the association between intelligibility scores at the three noise conditions and CPP was significant, and a large amount of variance in CPP was explained by the model, $F(3, 8)$, 9.616, $p < .01$, $R^2 = .78$, $R^2_{adjusted} = .70$. The analysis showed that CPP was significantly associated with intelligibility score at SNR+0 dB but

Table 4. Results of mixed-effect logistic regression model with the effects of listener, diagnosis of dysphonia, and noise level on intelligibility score.

Effect and variable		Value		
Random effect		Variance	SD	
Listener		10^{-14}	10^{-07}	
Fixed effects		β	SE	p
Intercept		-1.063	0.06	<2e-16
Diagnosis (base = dysphonic)				
Dysphonic		0.58	0.077	6.96E-14
Noise (base = quiet)				
SNR+5		0.764	0.076	<2e-16
SNR+0		0.923	0.075	<2e-16
Interaction: Diagnosis $\times$ Noise				
Normal $\times$ SNR+5		-0.405	0.101	6.00E-05
Normal $\times$ Quiet		-0.508	0.1	3.54E-07

Note. SNR = signal-to-noise ratio.
*** $p < .001$.

not at SNR+5 dB or in the quiet condition, $\beta = 0.075$, $t(11) = 4.436$, $p < .01$ for SNR+0 dB; $\beta = -0.015$, $t(11) = -0.244$, $p = .813$ for SNR+5 dB; $\beta = -0.045$, $t(11) = -0.34$, $p = .743$ for quiet. The robust regression model confirmed these results: $\beta = 0.075$, $t(11) = 3.594$, $p < .01$ for SNR+0 dB; $\beta = -0.018$, $t(11) = -0.236$, $p = .818$ for SNR+5 dB; $\beta = -0.039$, $t(11) = -0.238$, $p = .816$ for quiet. Results of these models are shown in Table 6 and Figure 3.

Spearman rank correlation tests were conducted to examine the strength of correlation between CPP and intelligibility score. At the sentence level, the test indicated that the correlation was statistically insignificant at quiet ($r_s = .058$, $p = .314$). The correlation was statistically significant but very weak at SNR+5 dB ($r_s = .18$, $p = .002$) and moderate at SNR+0 dB ($r_s = .516$, $p < .001$). The correlations at the speaker level were statistically insignificant at quiet and SNR+5 dB ($r_s = .105$, $p = .750$ for quiet; $r_s = .308$, $p = .331$ for SNR+5 dB). The correlation was statistically significant and strong at SNR+0 dB ($r_s = .881$, $p < .001$).

Table 5. Comparison between linear regression model and robust regression model, which examined the relationship between cepstral peak prominence and intelligibility scores at the sentence level in three noise conditions.

Model	β	SE	t value	Pr(> t)	Significance
Linear regression					
(Intercept)	3.443	0.549	6.268	1.29×10^{-9}	***
SNR+0	1.124	0.281	3.996	8.12×10^{-5}	***
SNR+5	0.864	0.395	2.188	0.030	*
Quiet	0.528	0.501	1.055	0.292	
Robust regression					
(Intercept)	3.187	0.591	5.396	1.39×10^{-7}	***
SNR+0	1.272	0.302	4.208	3.42×10^{-5}	***
SNR+5	0.900	0.425	2.121	0.035	*
Quiet	0.697	0.538	1.294	0.196	

Note. SNR = signal-to-noise ratio.

* $p < .05$. *** $p < .001$.

Discussion

This study was designed as the first step in examining the relationship between background noise, dysphonia, and intelligibility to determine the extent to which this clinical problem could be experimentally characterized. The overarching hypothesis of this study was that this relationship could be illustrated via a perception experiment in a controlled laboratory setting. A perception experiment was conducted to determine a noise level at which dysphonic speech becomes less intelligible than typical speech. Furthermore, it was hypothesized that the greater reduction of intelligibility in dysphonic speech would be attributed to spectral and temporal abnormalities in dysphonic speech; thus, the reduction would be observed even when intensity was held constant. The second aim of the study was to explore utility of an acoustic marker for predicting the intelligibility deficit. Specifically, suitability of CPP for prediction of intelligibility was explored.

Intelligibility of Speakers With Dysphonia Is More Negatively Affected by Background Noise

The results of this study confirm that intelligibility of dysphonic speech is more vulnerable than typical speech when produced in the presence of background noise. The difference between the dysphonic and typical speech was not significant in the quiet condition; however, it was significant in the presence of background noise. This finding supports the commonly expressed concern by speakers with dysphonia. Of the two SNRs tested in the study, the difference was most pronounced when the intensities of speech and background noise were equal (i.e., SNR+0 dB). Our finding was in line with a report by McColl et al. (1998), who studied the effect of background noise on intelligibility of tracheoesophageal speech. SNR used in their study ranged from SNR+65 dB to -15 dB. Intelligibility of their speaker with dysphonia was inferior to their speaker

Figure 2. Scatter plots of cepstral peak prominence (CPP) and intelligibility scores in quiet, SNR+5, and SNR+0 conditions. Intelligibility scores are averaged across listeners. The shaded area shows the 95% confidence interval for the mean value of the intelligibility score for a given value of CPP for all values of CPP. SNR = signal-to-noise ratio.

with typical speech even at SNR+65 dB, where the influence of the noise would have been minimal. On the other hand, the difference between the speakers was undetectable at the lowest SNRs, SNR−10 dB and −15 dB, which demonstrates that both types of speech were equally masked at these noise levels. The difference between the speakers began to show at SNR−5 dB and was most pronounced at SNR+5 dB. Although our speakers are considerably different from their speakers, it is interesting that the difference between our speaker groups was observed in the range of SNRs where McColl et al. observed the difference between the speakers. Although we did not examine the effect of noise at lower SNRs, people often need to speak in situations where background noise is as noisy as, or noisier than, speech. It is possible that dysphonic speech is even more challenged at the greater degree of background noise, and the intelligibility deficit is better illustrated at a different

SNR. This aspect of the relationship between dysphonia and intelligibility remains to be explored.

CPP Predicts Intelligibility of Dysphonia in Noise

Measuring the effect of background noise on intelligibility directly in a clinical setting may not be an easy task. Such assessment would require simulating daily communication environments in an office, where a speaker's intelligibility is assessed in the presence of background noise at various levels. This method would likely be impractical for the limited time allowed for an evaluation. Automatic, computerized measurement would help quantify intelligibility deficit quickly; however, no algorithm that measures the deficit is readily available today. This lack of suitable measurement motivated us to explore the potential of an acoustic marker. Specifically, a measure that is routinely used for characterizing dysphonic severity was selected as a candidate for the intelligibility marker.

Results of this study suggested that CPP may be a good predictor for the effect of background noise on intelligibility at SNR+0. Since the landmark article by Hillenbrand and Houde (1996), which showed a strong correlation between CPP and breathiness, there has been a steady growth of literature that supports the clinical and research utility of this measure. The measure is especially known to be a strong predictor for perception of overall dysphonic severity and breathiness; however, its predictive power is limited for roughness. Speakers in our study varied in their severity, and the primary perceptual category was roughness for most of the speakers (Table 1). It is possible that the association between CPP and intelligibility score would be different for breathy voices. Whether a particular type of dysphonic speech is more vulnerable to background noise than other types is an intriguing question and should be pursued in future studies.

Table 6. Comparison between linear regression model and robust regression model, which examined the relationship between cepstral peak prominence and intelligibility scores at the speaker level in three noise conditions.

Model	β	SE	t value	Pr(> t)	Significance
Linear regression					
(Intercept)	6.921	7.863	0.880	0.404	
SNR+0	0.075	0.017	4.436	0.002	**
SNR+5	−0.015	0.062	−0.244	0.813	
Quiet	−0.045	0.133	−0.340	0.743	
Robust regression					
(Intercept)	6.603	9.701	0.681	0.510	
SNR+0	0.075	0.021	3.594	0.816	**
SNR+5	−0.018	0.076	−0.236	0.818	
Quiet	−0.039	0.164	−0.238	0.004	

Note. SNR = signal-to-noise ratio.

** $p < .01$.

Figure 3. Scatter plots of cepstral peak prominence (CPP) and intelligibility scores in quiet, SNR+5, and SNR+0 conditions. Intelligibility scores are averaged across sentences per speaker and listeners. The shaded area shows the 95% confidence interval for the mean value of intelligibility score for a given value of CPP for all values of CPP. SNR = signal-to-noise ratio.

The correlation between CPP and intelligibility score improved when intelligibility scores were pooled across the sentences. This observation suggests that CPP varied across the sentences, likely because of their phonetic context. It has been shown that the phonetic context of speech material affects CPP values (Watts, 2015). As the sentences used for this study were rather short (i.e., seven words on average), it is possible that the effect of phonetic context on CPP was greater than the effect of dysphonia for some sentences. The improvement also suggests that variation in intelligibility score across the sentences for a given speaker weakened the association between the score and CPP. The speech material used for this study was chosen to partially replicate a clinical scenario, and it is possible that some sentences were inherently more predictable than others. Intelligibility can be influenced by multiple linguistic factors including semantic, syntactic, and phonetic context (Bradlow et al., 1996; Kalikow, Stevens, & Elliott, 1977; Picheny, Durlach, & Braida, 1986). Although semantically anomalous sentences were used in a portion of our speech samples, these linguistic aspects of the speech samples were neither controlled nor examined in this study. The effect of semantic and phonetic context of speech materials can be controlled by using materials whose predictability has been established. Future studies may consider using such materials.

Limitations of the Study

There are several aspects of this study that limit generalization of its findings to clinical practice in dysphonia. First, this is a small study, which included only six speakers with dysphonia and six speakers with typical speech. There is a wide range of intelligibility even among speakers with typical speech (Bradlow et al., 1996). It is possible that some of our speakers were inherently more or less intelligible than others regardless of the presence of dysphonia. A study with a larger sample would help reduce the effect of this within-group variability. In addition, this study did not

include speakers with severe dysphonia. It is likely that intelligibility would be more negatively affected as the severity of dysphonia increases; however, this assumption may not hold on the basis of the past study with speakers with alaryngeal speech (Holley et al., 1983).

Another limitation of the study can be attributed to a speech style used by our speakers for recording their samples and the possible effect of the intensity normalization on the results. In our study, the speakers were instructed to record their speech in a conversational style. Speakers adjust the way they speak according to the level of ambient noise, a phenomenon known as the *Lombard effect*. Acoustically, greater intensity and fundamental frequency and flatter spectral tilt are associated with Lombard speech in comparison with conversational speech (Lu & Cooke, 2009). It has been shown that Lombard speech is more intelligible than conversational speech even when excluding the effect of increased intensity, and the gain in intelligibility has been associated with the spectral change. It is likely that our speakers would also engage in Lombard speech when challenged by background noise, and when they do so, their intelligibility may or may not improve. Furthermore, although total intensity of speech was normalized in this study, the intensity of dysphonic speech could be reduced in real clinical cases, which could negatively affect intelligibility. Lastly, the noise levels at which the intelligibility was examined were rather specific, and they are held constant; however, the noise level fluctuates in daily communication environments. The experimental approach of our study does not replicate these situations, thus limiting the generalizability of the findings to real-life communication scenarios.

Conclusions

To our knowledge, this study is the first attempt to describe the intelligibility deficit in dysphonic speech that is experienced by speakers with benign, structural laryngeal

disorders. Findings of our study exemplify the importance of considering the effect of background noise on intelligibility when assessing this population. CPP may be a potential acoustic marker for describing the intelligibility deficit in the presence of background noise. These findings should be confirmed in future studies using larger number of speakers and linguistically controlled speech materials.

Acknowledgments

The authors would like to thank Joel MacAuslan at STAR Analytical Services for sharing the noise generation software, Sarah Hamilton for her technical assistance, and Lindsey Adams, Elizabeth Hary, Jenna LeFevre, Lindsay Mullins, and Meredith Meyer for proofreading this work.

References

- Adams, S., Dykstra, A., Jenkins, M., & Jog, M. (2008). Speech-to-noise levels and conversational intelligibility in hypophonia and Parkinson's disease. *Journal of Medical Speech-Language Pathology, 16*, 165–172.
- Awan, S. N., Roy, N., Jetté, M. E., Meltzner, G. S., & Hillman, R. E. (2010). Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V. *Clinical Linguistics & Phonetics, 24*, 742–758.
- Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. Boston, MA: Cengage Learning.
- Benninger, M. S., Alessi, D., Archer, S., Bastian, R., Ford, C., Koufman, J., ... Woo, P. (1996). Vocal fold scarring: Current concepts and management. *Otolaryngology—Head and Neck Surgery, 115*, 474–482.
- Bickley, C. (1982). Acoustic analysis and perception of breathy vowels. *Speech Communication Group Working Papers, 1*, 71–81.
- Bradley, J. S., & Sato, H. (2004). Speech intelligibility test results for grades 1, 3 and 6 children in real classrooms. Retrieved from https://www.acousticalsurfaces.com/soundproofing_tips/html/speech-intelligibility-in-classrooms.html
- Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*, 255–272.
- Casper, J. K., & Leonard, R. (2006). *Understanding voice problems: A physiological perspective for diagnosis and treatment*. Baltimore, MD: Lippincott, Williams & Wilkins.
- Dalston, R. M. (1975). Acoustic characteristics of English/w, r, l/spoken correctly by young children and adults. *Journal of the Acoustical Society of America, 57*, 462–469.
- Delattre, P. C., Berman, A., & Cooper, F. S. (1962). Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. *Studia Linguistica, 16*, 104–122.
- Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America, 27*, 769–773.
- Draeger, G. L. (1951). Relationships between voice variables and speech intelligibility in high level noise. *Communications Monographs, 18*, 272–278.
- Fant, G. (1960). *Acoustic theory of speech production*. The Hague, the Netherlands: Mouton.
- Goodglass, H., & Kaplan, E. (1972). *The assessment of aphasia and related disorders*. Philadelphia, PA: Lea and Febiger.
- Goodglass, H., & Kaplan, E. (1983). *Boston diagnostic aphasia examination booklet*. Philadelphia, PA: Lea & Febiger.
- Halberstam, B. (2003). Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. *Journal for Oto-Rhino-Laryngology and its Related Specialties, 66*(2), 70–73.
- Hall, E. T. (1966). *Hall, the hidden dimension*. Garden City, NY: Doubleday.
- Heman-Ackah, Y. D., Michael, D. D., Baroody, M. M., Ostrowski, R., Hillenbrand, J., Heuer, R. J., ... Sataloff, R. T. (2003). Cepstral peak prominence: A more reliable measure of dysphonia. *Annals of Otolaryngology, Rhinology & Laryngology, 112*, 324–333.
- Heman-Ackah, Y. D., Michael, D. D., & Goding, G. S. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice, 16*, 20–27.
- Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research, 37*, 769–778.
- Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research, 39*, 311–321.
- Holley, S. C., Lerman, J., & Randolph, K. (1983). A comparison of the intelligibility of esophageal, electrolaryngeal, and normal speech in quiet and in noise. *Journal of Communication Disorders, 16*, 143–155.
- Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics, 35*, 73–101.
- Isshiki, N., Kitajima, K., Kojima, H., & Harita, Y. (1978). Turbulent noise in dysphonia. *Folia Phoniatrica et Logopaedica, 30*, 214–224.
- Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The voice handicap index (VHI): Development and validation. *American Journal of Speech-Language Pathology, 6*(3), 66.
- Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*, 1337–1351.
- Lebo, C. P., Smith, M., Mosher, E., Jelonek, S., Schwind, D., Decker, K., ... Kurz, P. (1994). Restaurant noise, hearing loss, and hearing aids. *Western Journal of Medicine, 161*, 45.
- Lenhardt, M. L., Skellett, R., Wang, P., & Clarke, A. M. (1991). Human ultrasonic speech perception. *Science, 253*, 82–85.
- Leszcz, T. M. (2012). *The effect of multitalker background noise on speech intelligibility in Parkinson's disease and controls* (Master's thesis). The University of Western Ontario, London, ON, Canada.
- Lowell, S. Y., Colton, R. H., Kelley, R. T., & Mizia, S. A. (2013). Predictive value and discriminant capacity of cepstral- and spectral-based measures during continuous speech. *Journal of Voice, 27*, 393–400.
- Lu, Y., & Cooke, M. (2009). The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. *Speech Communication, 51*, 1253–1262.
- MacArthur, C. J., Kearns, G. H., & Healy, G. B. (1994). Voice quality after laryngotracheal reconstruction. *Archives of Otolaryngology—Head & Neck Surgery, 120*, 641–647.
- McAlexander, T. P., Gershon, R. R., & Neitzel, R. L. (2015). Street-level noise in an urban setting: Assessment and contribution to personal exposure. *Environmental Health, 14*, 1.
- McColl, D., Fucci, D., Petrosino, L., Martin, D. E., & McCaffrey, P. (1998). Listener ratings of the intelligibility of tracheo-esophageal speech in noise. *Journal of Communication Disorders, 31*, 279–289.

-
- Miller, G. A., & Nicely, P. E.** (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27, 338–352.
- Papso, C. F., & Blood, I. M.** (1989). Word recognition skills of children and adults in background noise. *Ear and Hearing*, 10, 235–236.
- Parikh, G., & Loizou, P. C.** (2005). The influence of noise on vowel and consonant cues. *Journal of the Acoustical Society of America*, 118, 3874–3888.
- Payton, K. L., Uchanski, R. M., & Braida, L. D.** (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of the Acoustical Society of America*, 95, 1581–1592.
- Pearsons, K., Bennett, R., & Fidell, S.** (1977). *Speech levels in various noise environments*. (Technical Report EPA-600/1-77-025). Washington, DC: U.S. Environmental Protection Agency.
- Peterson, G. E., & Barney, H. L.** (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175–184.
- Picheny, M. A., Durlach, N. I., & Braida, L. D.** (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, 29, 434–446.
- Qi, Y., & Weinberg, B.** (1995). Characteristics of voicing source waveforms produced by esophageal and tracheoesophageal speakers. *Journal of Speech, Language, and Hearing Research*, 38, 536–548.
- R Core Team.** (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2012.
- Ross, M., & Lerman, J.** (1971). *Word Intelligibility by Picture Identification*. Pittsburgh, PA: Stanwix House.
- Sjölander, K., & Beskow, J.** (2011). WaveSurfer (Version 1.8. 8) [Computer software].
- Smith, E., Taylor, M., Mendoza, M., Barkmeier, J., Lemke, J., & Hoffman, H.** (1998). Spasmodic dysphonia and vocal fold paralysis: Outcomes of voice problems on work-related functioning. *Journal of Voice*, 12, 223–232.
- Vaughan, N., Storzbach, D., & Furukawa, I.** (2006). Sequencing versus nonsequencing working memory in understanding of rapid speech by older listeners. *Journal of the American Academy of Audiology*, 17, 506–518.
- Watts, C. R.** (2015). The effect of CAPE-V sentences on cepstral/spectral acoustic measures in dysphonic speakers. *Folia Phoniatrica et Logopaedica*, 67, 15–20.