Influence of Phonatory Break Duration and Pause Time on Auditory-Perceptual Ratings of Speech Aceptability and Listener Comfort in Adductor-Type Laryngeal Dystonia

*Philip C. Doyle, †Rachel Woldmo, ‡Kathleen F. Nagle, ‖Natasha Crews, and ¶Nedeljko Jovanovic, *Stanford, ‡South Orange, and ¶London, Canada

Abstract: Introduction. This study empirically evaluated the influence of phonatory break duration and pause time on auditory-perceptual measures of speech produced by 26 adult speakers diagnosed with adductor-type laryngeal dystonia (AdLD).

Type of Study. Experimental.

Methods. Fifteen inexperienced, young adult normal-hearing listeners provided ratings of speech acceptability and listener comfort for samples of running speech. Four phonatory break and pause time conditions were assessed using visual analog scaling methods. All stimuli were randomized for presentation and listeners were presented with experimental stimuli in a counterbalanced manner.

Results. Results indicate that the duration of phonatory breaks directly influenced listener ratings of speech acceptability ($P < 0.001$) and listener comfort ($P < 0.001$), with significant differences between original and modified recordings for both. Speech acceptability and listener comfort ratings were strongly correlated across all timing conditions ($r = 0.85-0.97$).

Conclusions. The duration of phonatory breaks and pauses have significantly influence judgments of speech acceptability and listener comfort for AdLD. This suggests that temporal factors such as phonatory break duration and pause time in AdLD may carry substantial negative impact on listeners' perception relative to other auditory-perceptual features that co-exist in the signal.

Key Words: Adductor-type laryngeal dystonia−Voice disorders−Focal dystonia−Auditory-perception.

INTRODUCTION

Adductor-type laryngeal dystonia (AdLD) is a well-recognized neurological disorder that has been diagnostically classified as one form of dystonia.[1,2,1] As a distinct category of voice disorders, AdLD is characterized by abnormal and variable adductory control of the vocal folds during voice and speech production.[2] By clinical definition, AdLD is most commonly characterized by a vocal quality that is described by listeners as strained and strangled, as well as having variability that is influenced by vocal task.[3−6] Further, this abnormality in voice quality may be highly variable both within and across speakers and is perceptible to the listener across a range of severity.[1−6] Although researchers have investigated the auditory-perceptual characteristics exhibited by AdLD speakers, the most frequently reported auditory-perceptual characteristics have centered around aspects of laryngeal (adductory) strain. Accordingly, the most direct influence of AdLD is observed during vocal tasks that require continuous voicing.[1,3,5] In fact, one early auditory-perceptual dimension was termed laryngeal "overpressure" to reflect the increased degree of effort during voicing and the inconsistent spasmodic effects of the disorder on vocal fold adduction.[7] Because of this physiologic abnormality, as well as its variability and task-related intermittency, numerous concerns may be raised about the speaker's composite voice quality.[8−15]

Because of the underlying adductory laryngeal spasms associated with AdLD, one widely recognized auditory-perceptual characteristics of AdLD is the presence of intermittent and often inconsistent phonatory breaks. That is, during an adductory spasm, voicing may cease briefly or in its entirety, creating a noticeable disruption in the flow of speech.[1,5,8−10] Depending on the level of severity, laryngeal spasms that lead to voice breaks may vary in both their location within a spoken stimulus, as well as in the duration of the voicing break.[10−15] However, at present it is unknown how the temporal period of a voice break might influence listeners' more composite perceptual judgments of speakers with AdLD.

As a result of the types of intermittent changes that occur with AdLD, psychological and social impacts have also been reported.[16−18] In fact, previous research has shown strong relationships between the severity of AdLD speech and listener perceptions of speaker confidence and "tearfulness"[19]; inexperienced listeners have also reported that listening to AdLD speech is annoying, emotionally painful, and that they "must prepare to listen" to it [20, p.6]. Given

From the *Department of Otolaryngology Head and Neck Surgery, Division of Laryngology, Stanford University School of Medicine, Stanford University, Stanford, CA; †School of Communication Sciences and Disorders, Western University, London, Ontario, Canada; ‡Department of Speech-Language Pathology, Seton Hall University, South Orange, NJ; ‖Western University, London, Ontario, Canada; and the ¶Rehabilitation Sciences - Voice Production and Perception Laboratory, Western University, London, Ontario, Canada.

Address correspondence and reprint requests to Philip C. Doyle, Stanford University School of Medicine, Otolaryngology Head and Neck Surgery, 801 Welch Road, Stanford, CA 94305. E-mail: pdoyle2@stanford.edu

Journal of Voice, Vol. ■■, No. ■■, pp. ■■−■■
0892-1997

https://doi.org/10.1016/j.jvoice.2021.10.025

[1]Over the past 30 years, the widely used term for this disorder was adductor spasmodic dysphonia. Given the recent publication by Simonyan and colleagues (2021), we have adopted the use of AdLD for the purposes of this paper.

the paramount role of auditory-perceptual outcomes in the assessment of voice and voice treatment outcomes, listener perceptions are a critical component of communicative success for individuals with AdLD.

The question raised in this context is, if all other aspects of the vocal signal remain unchanged, does the independent, relative duration of phonatory breaks and pauses significantly influence listener perceptions? Therefore, within the study to follow, we directly sought to evaluate the influence of phonatory break and pause duration on listener ratings of two auditory-perceptual dimensions, speech acceptability (SA) and listener comfort (LC). Our selection of these two auditory-perceptual features was based on the following reasoning. First, by definition, both dimensions seek to describe the collective impact of a voice or speech disorder at a global level rather than focusing on a single, unitary feature (eg, pitch, strain, etc). Secondly, however, these dimensions also may reflect the listener's assessment of the composite nature of AdLD and its potential impact on the communication dyad between the speaker and the listener.[21,22] Consequently, both SA and LC were anticipated to provide ideal surrogates for understanding how the independent feature of phonatory breaks and pauses might influence listener judgments of AdLD speaker samples.

METHODS

Participant speakers

Twenty-four adults (18 women, 6 men) served as speakers for this study (mean age = 61.3, SD = 8.13, range = 43−75 years). All participant speakers were diagnosed by the same board-certified laryngologist as having AdLD. For all speakers, the disorder had been present for at least six months at the time of speech sample recording. Although all speakers were diagnosed with AdLD, none of them exhibited any coexisting vocal tremor, nor did they present with other sites of dystonia. Aside from their diagnosis of AdLD, all participants were self-reported to be in good general health. All methods were approved by our institution's Research Ethics Board, with all speakers consenting to recording. Neither participant speakers nor participant listeners were compensated for their time.

All 24 speakers had been receiving ongoing botulinum toxin (Botox®) injections as treatment for their AdLD. Speakers also ranged in their general level of severity, not only as a consequence of their voice disorder, but also in relation to where they fell in the regular Botox® treatment cycle. Thus, in order to obtain the most representative range of voice samples, all participant speaker samples were collected during a clinic visit, but prior to Botox® injection. Our decision to obtain recordings prior to treatment was based on the assumption that AdLD symptoms would be more apparent at this phase in the Botox® treatment cycle with voice abnormalities being more prominent at this time regardless of the base level of any given speaker's severity; however, at time of recording, listeners varied in the range of severity from mild to severe as a consequence of both the disorder itself and the period of time that had elapsed for each speaker since their last Botox® injection.

Prior to speaker selection, an assessment of speaker severity was made independently by one of the authors who is an experienced clinician. In seeking to identify potential participant speakers, speech samples (the Rainbow Passage) from a larger group of potential participants were evaluated by the same author and categorically assigned to one of five global, although arbitrary severity categories − mild, mild-to-moderate, moderate, moderate-to severe, and severe. This categorization judgment was based on a composite assessment of the speaker's sample that was based on the frequency of phonatory breaks (both within and between spoken words) and a range of increased vocal strain at varied intervals and loci in the speech sample. The sole purpose of this assessment and categorization was to identify a range of speaker severity prior to the modification of samples. Once this was categorization was completed, our series of speaker samples were then randomly selected to represent an approximately equal, albeit subjective, distribution across the categories of severity.

Speech stimuli and recording procedure

Each speaker recorded a stimulus sentence comprising only voiced sounds: "Early one morning a man and a woman were ambling along a one-mile lane running near Rainy Island Avenue".[23] This sentence was obtained as part of a larger, standardized speech recording protocol that was gathered in a consistent manner across all speakers. All recordings were made in a professional recording environment using Kay-Pentax Sona Speech II software (Pine Brook, NJ), a cardioid microphone and preamplifier, with all samples digitized at a 44.1 kHz sampling rate. All recordings were obtained by the same clinician while maintaining a fixed 15 cm microphone-to-mouth distance. All speakers were asked to produce the target sentence three times with a pause in between each production.

As part of the explicit instructions prior to audio recording, we asked participant speakers to "please read this sentence aloud using your typical voice and please use your typical pitch and loudness level. Do not worry if you have any disruption in your speech due to your voice disorder. We are seeking to obtain the most representative sample of your speech and we will only stop recording if you misread the sentence." After all recordings were obtained, all samples were reviewed to ensure that no errors in production occurred (ie, omitted or misspoken words). Following confirmation that no spoken errors existed in the samples, one of each speaker's three samples was randomly extracted for use in the current study.

Temporal editing of speech stimuli

Before any signal editing was initiated, all speaker samples were monitored and all vocal spasms within any spoken sentence that resulted in a phonatory break were perceptually identified. This process was performed independently by two of the authors through both auditory and visual identification and confirmation. Following this assessment, identifications were compared, resulting in 100% reliability specific to the occurrence and location of phonatory breaks. Once AdLD breaks/spasms were confirmed, the duration (in msecs) of all voice breaks and spasms was measured. Following confirmation of phonatory breaks, digital signal editing of speaker samples was initiated in order to create four sets of experimental stimuli.

All sound file editing was done using Audacity (Version 3.0.0), a digital recording and audio editing application. Using this software, all samples were manually edited to conform to a prescribed and standardized editing protocol in order to generate the stimuli used during the auditory-perceptual phase of the project. Four distinct signal modifications were undertaken resulting in four unique sets of stimuli. These four edited stimulus sets were identified as: 1) Normalized Original (NO), 2) Standardized Original (SO), 3) Modified Short (MS), and 4) Modified Long (ML). Specific audio file editing details for each set of experimental samples are outlined individually in the subsequent sections.

Normalized original (NO). Any phonatory break or spasm that exceeded 500 msecs in duration was edited to not exceed 500 msecs. In instances where a break/spasm needed to be shortened in duration, the appropriate temporal segment removed was always extracted from the midpoint of that break/spasm. Any break or spasm less than 500 msecs remained unchanged. Given that a range of severity was exhibited in our group of speakers, multiple edits were performed in this manner for some speakers; that is, the overall frequency of breaks remained, however, only those that exceeded 500 msecs were altered. Similarly, not all of the speaker samples required modification according to this predetermined, temporal editing requirement.

Standardized original (SO). For this editing condition, any break or spasm that was greater than 250 msecs in duration was edited to 250 msecs. That is, regardless of the duration of such a break, all but 250 msecs of that break/spasm was marked with cursors and deleted from the sample. This was again achieved by removing the necessary temporal segment from the mid-portion of any given break. Similar to the NO condition outlined previously, not all speakers demonstrated breaks or spasms of this duration; hence, not all speaker samples required modification. However, due to the range of severity of the participant speakers, differences in the number of edits that occurred varied. That is, for any break/speaker that met the *a priori* temporal duration, and regardless of the number of breaks that met this criterion, modifications were made.

Modified short (MS). For this condition, an artificial pause of 750 msecs was inserted into the modified samples at eight pre-established, between word boundaries. These temporal insertions occurred at points ranging from three to four syllables in length. The specific points where insertions were made were based on the following arbitrary truncation of the spoken sentence with segmentation of the experimental sentence stimuli occurring at points noted between the end brackets and start brackets of the stimulus as follows: *[Early one] [morning a] [man and a] [woman were] [ambling along] [a one-mile] [lane running near] [Rainy Island] [Avenue]*.[23] If a vocal abnormality existed in these intervals, that abnormality was removed in its entirety and replaced with the 750 msecs period of silence. The use of 3−4 syllable intervals was based on an arbitrary segmenting of each sample into nine segments of relatively short syllabic structure; however, we actively sought to avoid segmenting individual words as part of this study in an effort to avoid adding another perceptual variable into the study.[24,25] This resulted in eight inter-segment intervals wherein the 750 msecs silent (ie, phonatory break) durations were inserted for each sample. Because of the nature of this MS editing protocol, all speaker samples were modified using this method.

Modified long (ML). The same editing process utilized in the MS condition was followed to create this stimulus set. However, in this condition, a longer pause of 1250 msecs was inserted into the eight inter-segment sections of the sample. MS and ML were the only conditions where the 750 and 1250 msecs, respectively, were made. The extended duration of this "foil" was selected arbitrarily but it was our expectation that such a lengthy pause would almost certainly result in substantially negative auditory-perceptual ratings of samples for both of the auditory-perceptual features under study (ie, SA and LC). This assumption was based not only on the substantial temporal period of the pause (silence), but its interaction and interference with the overall rate and cadence of the running speech sample. In this regard, we believed that judgments of samples in this condition could serve as a "point of reference" for a substantial abnormality in the modified signal and listener evaluation relative to the other three editing conditions (NO, SO, and MS).

Participant listeners

Fifteen normally-hearing adults (11 women, 4 men) between the ages of 18;11−23;7 years served as listeners (*M* age 20;2 *SD* = 1;3). None reported prior experience listening to or evaluating voice disorders; thus, these listeners were considered inexperienced. Additionally, all listeners were native speakers of North American English and reported that they had no history of voice, speech, language, or hearing disorders. All listeners provided informed consent prior to their participation.

Procedure: Auditory-perceptual scaling of voice samples

Three master randomized lists of all speaker stimuli and duplicate samples (n = 24 speakers, 4 conditions) were created for the auditory-perceptual phase of the experiment. Each list also included 4 pseudo-randomly selected duplicate samples (one from each condition), inserted into the master randomizations at fixed intervals for purposes of assessing agreement. Thus, each stimulus list comprised 112 total samples [24 speakers x 4 conditions + 16 agreement duplicates (4 per condition). Listening sessions were conducted individually during a single listening session that lasted approximately 90-minutes; each session was conducted in a quiet laboratory room using a desktop computer and circumferential headphones (Sennheiser, HD 205). Listeners were randomly assigned to auditory-perceptual sessions in a counterbalanced manner; that is, they rated SA followed by LC, or vice versa. Following the rating of stimuli for any given feature, listeners were provided with a short break of approximately 10 minutes prior to initiating the second rating phase of the study in which the remaining auditory-perceptual feature was evaluated.

Prior to the onset of the auditory-perceptual rating session, listeners were informed that the speech samples presented were produced by both men and women and that samples may sound unusual. All listeners were instructed to rate each sample independently from others they would hear. These samples were productions of the same sentence as that used as experimental stimuli; however, they were not edited in any fashion. This task provided each listener with a representative sample of speech and voice characteristics of those with AdLD and was meant to reduce any unusual initial reactions to the abnormalities they would be exposed to during the experimental phase.

Listener assessments of SA were made based on the definition that indicates that this feature is a composite auditory-perceptual feature. More specifically, when rating of SA, listeners were asked to "*give careful consideration to the attributes of pitch rate, understandability, and voice quality. In other words, is the voice pleasing to listen to, or does it cause you some discomfort as a listener?*".[26] In regard to LC, this dimension was defined as "*how comfortable you would feel listening to the person's speech in a social situation. Your response should reflect your feelings about the way the person was speaking (ie, how comfortable you would feel listening to them), not what the person was saying or how their personality affected you.*".[27] Thus, while SA assesses a collective of perceptually salient factors inherent to the speaker's voice and speech, LC is defined in a manner that may more closely reflect an assessment relative to a face-to-face, social type of interaction. Thus, while the auditory-perceptual dimensions under evaluation (SA and LC) are not mutually exclusive constructs, they are distinguished by an aspect that asks the listener to make a judgment as if they were in fact interacting with the individual directly (ie, the LC condition). Further, judgments of LC require listeners to reflect on their own emotional state, as compared to rating the composite quality of the voice and speech signal that represented the SA perceptual feature. All listeners had access to a written definition of the feature of interest (SA and LC) throughout the session.

All listeners provided their independent judgment of each sample by marking an individual, undifferentiated 100 mm visual analog scale on a paper rating sheet for each dimension. Each stimulus item was listed with an individual rating scale (112 scales per dimension). Thus, listeners could move sequentially through randomized series of the experimental sentences. Listeners were instructed that following presentation of any stimulus sentence they were to identify their judgment of SA or LC at any point along the scale range that they believed best represented the sample for each dimension. Text anchors were placed at endpoints only; no numerals were included on any scale. For SA, the left side of the scale was identified as "extremely acceptable" and the right side of the scale as "extremely unacceptable"; for LC, the left side of the scale was identified as "extremely comfortable" and the right as "extremely uncomfortable". Thus, as numerical ratings increased from left to right (ie, with respective increases of one mm from 1 mm to 100 mm once measured), listener perceptions also represented increasing levels of *unacceptability* for SA and *less comfort* for LC. Again, written definitions of SA and LC were available throughout the perceptual task dependent upon which feature was being evaluated at any given time. As part of their instructions, listeners were directly informed that they could listen to any sample as many times as they desired prior to making a judgment. However, once the scaled rating was marked by bisecting the scaled line, listeners were asked to not return to any sample or modify any prior rating. However, each score sheet did have scales for 6 samples; so, listeners did have the potential capacity to see other ratings that they provided on a given score sheet regardless of whether SA or LC was being assessed. Yet it is important to note that all listeners were explicitly instructed to provide their rating of any sample independent from others and listener sessions were carefully monitored.

Analysis of auditory-perceptual data

All data for both judgments of SA and LC were recorded individually (for each listener) and then unrandomized for analysis. Data were then collated for each speaker/condition and measures of central tendency were calculated by condition. Statistical analyses were conducted in SPSS (version 26.0, Minneapolis, MN) using two one-way analyses-of-variance with timing condition (NO, SO, MS, and ML) as the independent variable and mean auditory-perceptual ratings (SA, LC) as the dependent variables. Pearson correlation coefficients were computed to examine the potential relationship between each of the four listening conditions across the LC and SA perceptual dimensions.

Measures of agreement between a listener's first evaluation of a given sample and a second rating of the same sample were calculated to assess intrarater consistency. The determination of agreement was based on whether second assessments of the same stimulus sample fell within one of four numerical categories. Specifically, we calculated whether the second rating was: 1) within +/- 5 mm of the original rating, 2) within +/- 10 mm of the original rating, or 3) within +/- 15 mm of the original rating. Determination of agreement for ratings was pooled across the 15 listeners for each condition (ie, SO, NO, MS, and ML). Thus, agreement was based on 60 comparisons (15 listeners X 4 duplicated samples) per temporal condition.

Intrarater reliability was calculated using a 2-way analysis of variance (ANOVA) with random effects (ie, Model 2).[28] Intraclass correlation (ICC) coefficients were found for LC and SA for all conditions and are reported as "single rater reliability" (ie, consistency between individual raters) and "average measures" reliability, or consistency of ratings with the group mean.

RESULTS

Data revealed that the changes in the duration of phonatory breaks and pauses significantly influenced judgments of both SA and LC. LC results indicated significant differences across the four conditions [F (1,31) =124.127, P =<.001]. Post-hoc pairwise comparisons using the Bonferroni adjustment revealed that ratings of LC were higher in the ML condition (M 60.23) compared to the NO (M 35.05, P =< 0.001), SO (M 34.48, P =< 0.001), and MS (M 57.19, P = 0.003) conditions. Listeners also rated LC significantly higher in the MS condition compared to the NO (P =< 0.001) and SO (P =< 0.001) conditions. There was no statistically significant difference between the NO and SO edited conditions (P = 0.485).

The one-way, within-subjects ANOVA also revealed a significant difference between the four timing conditions in the SA dimension [F(1,33) = 144.737, P =< 0.001]. Post-hoc analyses revealed that listeners' responses to stimuli were significantly less favorable at ML (M 60.04) compared to the NO (M 36.24, P =< 0.001) and SO (M 32.03, P =< 0.001) conditions. SA scores were significantly greater in the MS condition (M 59.69) compared to the NO (P =< 0.001) and SO (P =< 0.001) conditions, as well as in the NO versus SO edited conditions (P =< 0.001). The difference in SA ratings between the MS and ML conditions was not found to be statistically significant (P = 0.682). Summaries of these data for both LC and SA ratings across condition are provided in Tables 1 and 2, respectively.

Although auditory-perceptual ratings were similar between SO and NO conditions for both SA and LC, significantly increased scaled scores (less favorable) were observed for both dimensions in the MS and ML conditions. Graphic representation of each speaker's data reveal that despite individual speaker differences (likely due to differing individual levels of AdLD severity), the pattern of change across the conditions was remarkably consistent (see Figures 1 & 2).

TABLE 1.
Measures of Central Tendency For Scaled Ratings of LC Across Timing Conditions (NO, SO, MS, ML); Scores Provided are Represented in mm

	Mean	SD	Minimum	Maximum
NO	34.48	17.54	4.0	70.4
SO	35.04	17.80	5.1	69.0
MS	57.19	9.0	44.4	79.1
ML	60.23	9.26	46.3	78.2

Note: Higher scores represent less favorable auditory-perceptual judgments.
Abbreviations: ML, modified long stimuli; MS, modified short stimuli; NO, normalized original stimuli; SO, standardized original stimuli.

TABLE 2.
Measures of Central Tendency For Scaled Ratings of SA Across Timing Conditions (NO, SO, MS, ML); Scores Provided are Represented in mm

	Mean	SD	Minimum	Maximum
NO	32.03	16.48	5.7	67.3
SO	36.23	18.01	6.8	76.1
MS	59.68	8.75	45.1	78.8
ML	60.03	8.25	44.9	75.9

Note: Higher scores represent less favorable auditory-perceptual judgments.
Abbreviations: ML, modified long stimuli; MS, modified short stimuli; NO, normalized original stimuli; SO, standardized original stimuli.

In regard to correlational analyses, overall, there was a strong positive and significant correlation between the auditory-perceptual dimensions of LC and SA across all four timing conditions: NO (r = 0.97; P =< 0.001), SO (r = 0.95; P = < 0.001), MS (r = 0.85; P =< 0.001), and ML (r = 0.89; P =< 0.001).

Agreement and reliability

Intrarater agreement was calculated at three specific levels for SA and LC (ie, +/-5mm; +/-10mm; +/-15mm). Table 3 displays the gross number of agreements within each level for 60 pairs (15 listeners x 4 pairs) presented in each condition, and the relative percentage of those pairs in agreement at each level of agreement. The cumulative percentage across rows equals 100% in Table 3: the cumulative percent agreement within +/- 10mm was of particular interest for all timing conditions. At that level, overall intrarater agreement was 55.8% for SA and 56.1% for LC.

Single rater reliability was poor-moderate for both LC and SA in all conditions, based on 95% confidence intervals (Table 4).[29] Consistency of individual ratings with the group mean, as measured by average-measures ICCs, was strong for both LC and SA in all conditions, based on 95% confidence intervals (Table 5).

DISCUSSION

This study was designed to systematically evaluate the influence of phonatory break duration in speech samples produced by speakers with AdLD on auditory-perceptual judgments made by inexperienced listeners. In addition, we explored pause times that were systematically inserted into the samples at predetermined intervals. The perceptual judgments focused on the evaluation of two specific auditory-perceptual dimensions, speech acceptability (SA) and listener comfort (LC). The collective results of this work suggest that the duration of phonatory breaks does influence listener assessments of both SA and LC. Further, regardless

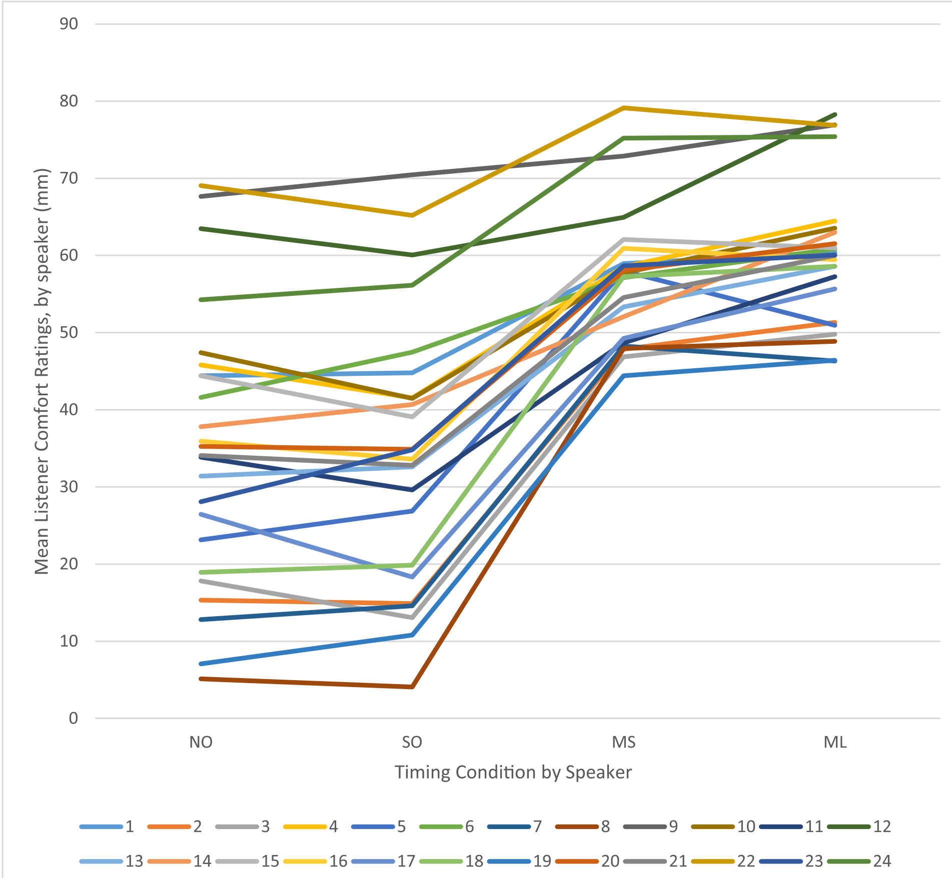

FIGURE 1. Mean listener ratings of LC for each of the 24 speakers across timing conditions NO, SO, MS, and ML (higher scores reflect less favorable ratings by listeners). Note: NO, normalized original stimuli; SO, standardized original stimuli; MS, modified short stimuli; ML, modified long stimuli.

of the duration of a phonatory break, the presence of pauses also negatively influences listener assessments of such samples. Regardless of the dimension being evaluated by listeners, as break duration increased, less favorable ratings of both SA and LC were observed. Further, the findings of this work for both dimensions (ie, the combined findings for both SA and LC) were highly consistent across both the timing modification conditions and the speakers assessed. Thus, our work supports the assumption that the duration of phonatory breaks (ie, the NO and SO conditions) carries clear perceptual salience to listener judgments. However, our data also indicate that other "breaks" that occur within an utterance in a systematic fashion (ie, the MS and ML conditions) will result in additional negative influences on auditory-perceptual judgments of listeners.

While the broader psychophysical implications of this work are complex, the current data provide evidence that break duration is one component that influences a listener's assessment of speakers with AdLD. In that regard, it would seem that the interactive relationships among a number of factors, and certainly those that are acoustic in nature, are multifaceted in relation to how AdLD is described clinically. Both the extent of variation and the interactions among multiple acoustic properties that exist in speakers with AdLD appear to be important areas for further study. Yet, the present data confirm that phonatory break duration alone significantly influences how listeners judge these samples.

Of particular note is the remarkable consistency observed from our data across speakers for both auditory-perceptual dimensions. This finding was observed within any given temporal condition across speakers in this study. Although the present study was designed in a manner that provided control over both the duration of the break and its location within the target sentence stimuli, we would suggest that both factors are critical to listener judgments. However, we might also note that if phonatory breaks occurred in an inconsistent manner, as is found in speakers with AdLD, this might have a differential effect on listener ratings of LC. This suggestion would appear to be of particular importance relative to the duration of such breaks and associated pauses, as well as the loci within any given utterance. Thus, we carefully considered concerns related to breaks associated with speech task, linguistic demand, and lexical boundaries.[4,25,30] This suggestion is based on the fact that by definition, LC does carry with it some inferential suggestion of a face-to-face interaction (ie, "...listening to the person's speech in a social situation"). Specifically, we might

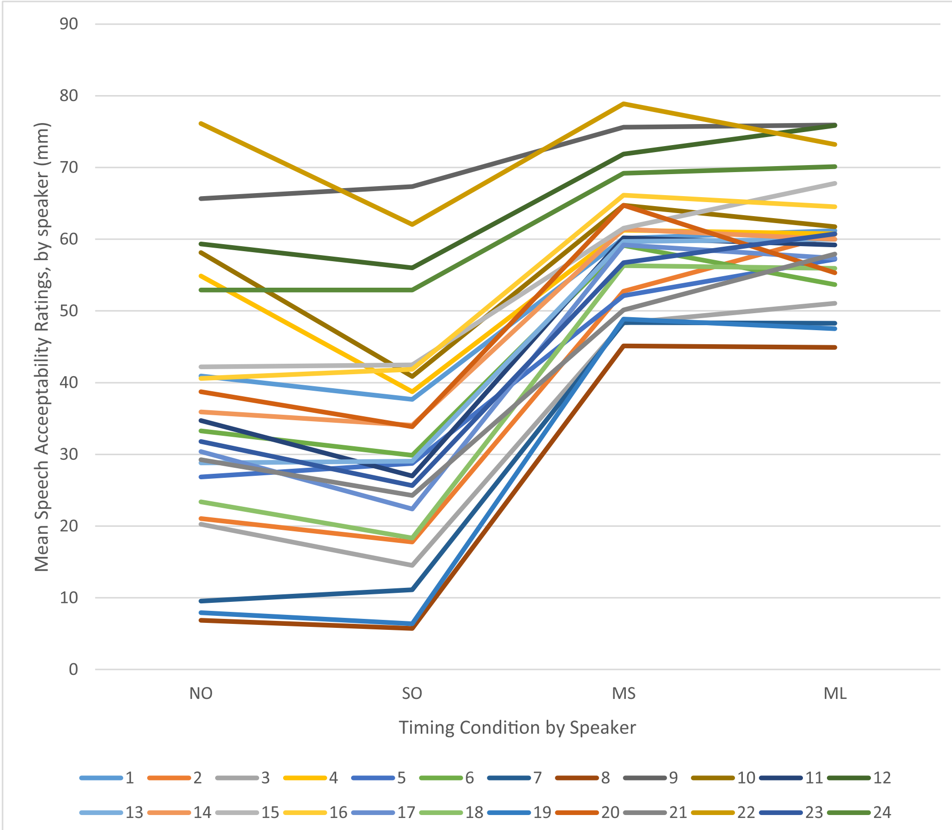

FIGURE 2. Mean listener ratings of SA for each of the 24 speakers across timing conditions NO, SO, MS, and ML (higher scores reflect less favorable ratings by listeners). Note: NO, normalized original stimuli; SO, standardized original stimuli; MS, modified short stimuli; ML, modified long stimuli.

find reduced LC reported due to the inconsistent nature of the spoken signal and the listener's active consideration of interacting with the speaker directly (ie, the level of effort required by the listener to monitor the signal). Likewise, the irregular phonatory breaks common to AdLD speech might affect judgments of SA, given its strong correlation with LC

TABLE 3.
Raw Number of Overall Intrarater Agreements Across ll Listeners For Auditory-Perceptual Ratings of SA And LC by Timing Condition, With Relative Percentage in Parentheses

Speech Acceptability	+/- 5	+/-10	+/-15	>15
NO	17 (28.3)	14 (23.3)	12 (20.0)	17 (28.3)
SO	25 (41.6)	11 (18.3)	9 (15.0)	15 (25.0)
MS	24 (40.0)	11 (18.3)	7 (11.7)	18 (30.0)
ML	20 (33.3)	12 (20.0)	11 (18.3)	17 (28.3)
Totals	*86 (35.8)*	*48 (20.0)*	*39 (16.3)*	*67 (27.9)*
Listener Comfort	**+/- 5**	**+/-10**	**+/-15**	**>15**
NO	20 (33.3)	14 (23.3)	4 (6.0)	22 (36.6)
SO	30 (50.0)	7 (11.7)	11 (18.3)	12 (20.0)
MS	21 (35.0)	8 (13.3)	13 (21.6)	18 (30.0)
ML	24 (40.0)	13 (21.6)	10 (16.7)	13 (21.6)
Totals	*95 (39.6)*	*42 (17.5)*	*38 (15.8)*	*65 (27.1)*

Abbreviations: ML, modified long stimuli; MS, modified short stimuli; NO, normalized original stimuli; SO, standardized original stimuli.

TABLE 4.
Single Rater Reliability For LC and SA For 15 Raters

	LC ICC (2, 1) [95% CI]	SA ICC (2, 1) [95% CI]
NO	.26 [.14 - .48]	.31 [.18 - .54]
SO	.25 [.14 - .48]	.29 [.16 - .52]
MS	.46 [.30 - .69]	.26 [.14 - .48]
ML	.46 [.30 - .69]	.34 [.20 - .57]

Abbreviations: ML, modified long stimuli; MS, modified short stimuli; NO, normalized original stimuli; SO, standardized original stimuli.

TABLE 5.
Average-Measures Reliability For LC and SA For 15 Raters

	LC ICC (2, *k*) [95% CI]	SA ICC (2, *k*) [95% CI]
NO	0.89 [0.80–0.96]	0.92 [0.84–0.97]
SO	0.89 [0.79–0.96]	0.91 [0.82–0.96]
MS	0.95 [0.91–0.98]	0.89 [0.79–0.96]
ML	0.95 [0.91–0.98]	0.92 [0.86–0.97]

Abbreviations: ML, modified long stimuli; MS, modified short stimuli; NO, normalized original stimuli; SO, standardized original stimuli.

across timing conditions in this study, and given the particular reference to pitch, rate and voice quality in the instructions to raters.

The significant difference in ratings of SA in the NO compared to the SO condition is interesting in the absence of such a difference for ratings of LC. These stimuli were relatively unmodified, in that we inserted no artificial pauses; however, all phonatory breaks were reduced to either 500 msecs (NO) or 250 msecs (SO). The finding of a disparity in SA between these conditions indicates that our inexperienced listeners perceived temporal differences among stimuli that most closely resembled "natural" AdLD speech. The lack of such a distinction in ratings of LC suggests that although listeners noticed a change, they did not appear to sense the degree of distress as a listener that they provided in their ratings of modified stimuli (ie, MS, ML). This difference provides evidence that there may be benefit in at least considering both signal-based (SA) and listener-based (LC) perceptual measures for disrupted speech, particularly as functional outcomes.

The artificial pauses in our samples were set at fixed word-based intervals that ranged in length from 3–4 syllables. However, although our determination of where a phonatory break occurred was selected arbitrarily, we actively modified our samples to ensure that no break existed within the structure of a single word. The breaks that did exist in our samples were inserted between words. We suspect that if breaks occurred within a multisyllabic word, both SA and LC would be further negatively affected.[24] For that reason, the temporal characteristics of the voice breaks associated with AdLD must be more fully evaluated. This raises the related concern of how phonatory breaks will influence the overall speech rate of any given sample. As pauses increase in either frequency of occurrence or duration, the total length of any utterance is altered. Past research into the influence of pause time on auditory-perceptual ratings of disordered speech samples has been reported most commonly in relation to stuttering and dysarthria subgroups.[25,31]

It is clear from this project that both linguistic pauses and articulatory speed will influence a speaker's rate of speech. As reported in prior studies with other clinical populations, temporal editing or disordered speech may provide a rich area for continued study.[31,32] As expectations of normal speech timing are violated, we would assume that listeners would provide less favorable ratings of any given speech sample.[33] We believe that this would also be observed in assessments of the speech of those with AdLD. For that reason, questions specific to a variety of temporal changes in the speech of AdLD speakers, and the effect of such changes on listener perceptions, may provide a valuable area for future inquiry.

The present data suggest that listener judgments of both SA and LC may offer a viable and valuable index of deficits common to those with AdLD. While further research needs to be pursued to provide external validity to our findings, the present correlation data are quite robust and the patterns of auditory-perceptual judgments of our listeners were consistent across the timing modifications assessed for both auditory-perceptual dimensions. In the context of treatment of AdLD, and more specifically the use of Botox® as a preferred and widely used method of treatment, it would be of interest to explore the acoustic characteristics of phonatory breaks relative to the disruption of the spoken signal. That is, as the flow of speech is altered due to spasms associated with AdLD and their associated phonatory breaks, we might expect that more negative judgments would emerge as both the number of breaks and their relative durations increase.[1,2,5,8,11,16] The location of these breaks within an utterance would also likely influence listener judgments. It is not unreasonable to expect that as breaks occur within a word, or in those instances where normal sentential pauses would occur as points of linguistic juncture, less favorable listener judgments would occur.

Accordingly, as these types of changes occur as a consequence of AdLD, we would logically assume that listener judgments of the overall severity of the disorder would also increase. This assumption requires confirmation through additional research; however, the present data do appear to offer some translational value in our efforts to better understand the relationship between auditory-perceptual dimensions, listener assessments of these dimensions, and the composite acoustic (frequency, intensity, and temporal) characteristics of the AdLD speech signal. The influence of less controlled or systematic interruptions of voice, as may be observed in spontaneous AdLD speech, may result in more complicated clinical outcomes, particularly given the unpredictable ways in which conversational partners may react to it.[34]

Limitations of the present study

Although the present study did reveal remarkable consistency across the speakers studied, several limitations exist. Most prominently, the arbitrary modification of break durations, as well as the pre-established loci of timing modifications in the conditions studied present a concern. Despite the fact that none of these changes truncated any given word, it is possible that these breaks and pauses fell at unusual points in what might be considered the normal prosodic flow of the sentence.[4,8,10,20] Hence, changes that occurred in the signal secondary to editing at these pre-established points in the MS and ML conditions may have been perceived as unusual or awkward by listeners. As a result, poorer ratings of both SA and LC in two conditions may have occurred simply due to the imposition of eight pauses within an utterance. Thus, it is possible that the existence and duration of a pause at these atypical intervals based on normal juncture, as well as the presence of other abnormal features within any given sample, carried perceptual weight relative to other points in the sample where adductory spasms occurred. However, when data from all speakers are viewed together, the pattern of change due to the timing modifications remains consistent.

Additionally, it is well documented that the location of phonatory breaks or other disruptions in speech flow for those with AdLD may often occur within a word or may be exacerbated within the context of particular phonetic entities. This was certainly the case with our speakers who ranged in their AdLD severity. Consequently, the location as well as the duration of phonatory breaks must be considered in future evaluation of listener judgments. As such, the collective influence of break duration and other features characteristic of AdLD relative to listener assessments may be raised. In this regard, we would also note that the motor speech behaviors that underlie both normal and disordered voice and speech production must be considered. More specifically, it is possible that a speaker with AdLD may be more likely to exhibit phonatory breaks in association with particular sounds or relative to a given articulatory or coarticulatory gesture associated with the demands of a given word environment rather than at the arbitrary points represented within our NO and SO conditions.

This finding supports the notion that although the arbitrary temporal modifications for some conditions and the points of pre-established durational change in the MS and ML conditions may have considerable primary perceptual salience, other characteristics inherent to the sample may be considered as additive to the listener. We sought to minimize "snap" judgments based on partial samples by explicitly requesting that listener judgments be made only following a listener's auditing of the entire sample. However, it is clear that considerable challenges remain relative to the auditory-perceptual assessment of voice quality regardless of the disorder being studied.[35−37]

Because listeners heard the same sentence multiple times, it is possible that individual judgments were affected by the eventual predictability of the pauses within the sentence, or frankly by fatigue from hearing the sentence itself; however, the use of three randomly presented sets of stimuli limited the likelihood of sequence effects on mean ratings of SA or LC. Relatedly, intrarater agreement was quite good overall, with listeners' second ratings of repeated samples tending to fall within 10mm of their first ratings for both SA and LC a majority of the time. ICC measures were also excellent. While listeners did have potential access to five other ratings on any given score sheet (i.e., there were six scales for samples on each page), previous ratings could have served as a personalized external visual standard against which a given listener could base other judgments on that page. In this way, listeners may have been guided less by their contemporaneous assessment of LC or SA for a particular stimulus than by an attempt to be consistent. However, listeners were instructed to make each judgment independently and to not return to any prior judgment, a process that was monitored by one of the experimenters. Visual evidence of prior judgments may, therefore, decrease the ecological validity of this study, given that such an external standard could artificially increase reliability of perceptual judgments[38]. Any effects of access to prior ratings would not have extended to other raters, however, and interrater reliability in this study was comparable with results from the literature on auditory-perceptual ratings of voice quality[39] Accordingly, we do acknowledge that listeners may exhibit increased sensitivity to particular changes in speech flow such as the frequency of occurrence and duration of breaks, as well as and the larger composite presentation of any given sample regardless of disorder.[40−42] When the dynamic nature of AdLD is considered, future efforts that seek to identify interactions between elements within a disordered speech sample and their relative impact on listener judgments would be of value.

CONCLUSIONS

This study sought to identify changes in the auditory-perceptual evaluation of temporally modified speech samples from speakers who had been diagnosed for AdLD. Two auditory-perceptual features, SA and LC were assessed. Our data indicate that the duration of phonatory breaks and pauses have a significant influence on judgments of SA and LC for AdLD. This suggests that temporal factors affecting AdLD speech may carry an additional and substantial negative impact on listener perceptions relative to other acoustic factors that co-exist in the signal. Although ratings of SA and LC were highly correlated across conditions, there were indications that listeners perceived relatively minor timing differences among "original" stimuli, based on differences in SA ratings. That LC ratings did not differ in these conditions suggests that they may reflect a unique aspect of communicative interactions with individuals with AdLD. The present data provide new information that may be further explored to determine the influence on temporal changes on listener judgments of AdLD speech.

REFERENCES

1. Simonyan K, Barkmeier-Kraemer J, Blitzer A, et al. Spasmodic dysphonia/laryngeal dystonia. laryngeal dystonia: multidisciplinary update on terminology, pathophysiology, and research priorities. *Neurology*. 2021;96(21):989–1001. https://doi.org/10.1212/WNL.0000000000011922.
2. Brin MFS, Fahn S, Blitzer ALO, et al. Movement disorders of the larynx. In: Blitzer A, Brin MF, Sasaki CT, eds. *Neurologic Disorders of the Larynx*. New York: Thieme Medical Publishers; 1992:248–278.
3. Cannito MP, Woodson G. The spasmodic dysphonias. In: Kent RD, Ball M, eds. *Voice Quality Measurement*. San Diego: Singular Publishing Group; 2000:411–430.
4. Froeschke LL. The influence of linguistic demand on symptom expression in adductor spasmodic dysphonia. *J Voice*. 2020;34:807–e11.
5. Ludlow CL, Domangue R, Sharma D, et al. Consensus-based attributes for identifying patients with spasmodic dysphonia and other voice disorders. *JAMA Otolaryngol Head Neck Surg*. 2018;144:657–665. https://doi.org/10.1001/jamaoto.2018.0644.
6. Fisher KV, Scherer RC, Guo CG, et al. Longitudinal phonatory characteristics after Botulinum Toxin Type A injection. *J Speech Lang Hear Res*. 1996;39:968–980. https://doi.org/10.1044/jshr.3905.968.
7. Izdebski K. Overpressure and breathiness in spastic dysphonia: An acoustic (LTAS) and perceptual study. *Acta Oto-Laryngol*. 1984;97(3-4):373–378.. Retrieved from; http://informahealthcare.com/doi/abs/10.3109/00016488409131003.
8. Sapienza CM, Murry T, Brown Jr. WS. Variations in adductor spasmodic dysphonia: Acoustic evidence. *J Voice*. 1998;12:214–222.
9. Nash EA, Ludlow CL. Laryngeal muscle activity during speech breaks in adductor spasmodic dysphonia. *Laryngoscope*. 1996;106:484–489.

10. Stewart CF, Allen EL, Tureen P, et al. Adductor spasmodic dysphonia: standard evaluation of symptoms and severity. *J Voice*. 1997;11:95–103. https://doi.org/10.1016/S0892-1997(97)80029-X.
11. Blitzer A, Brin MF, Stewart CF. Botulinum toxin management of spasmodic dysphonia (laryngeal dystonia): a 12-year experience in more than 900 patients. *Laryngoscope*. 1998;108:1435–1441.
12. Roy N, Whitchurch M, Merrill RM, et al. Differential diagnosis of adductor spasmodic dysphonia and muscle tension dysphonia using phonatory break analysis. *Laryngoscope*. 2008;118:2245–2253.
13. Rubin AD, Wodchis WP, Spak C, et al. Longitudinal effects of Botox injections on voice-related quality of life (V-RQOL) for patients with adductory spasmodic dysphonia: part II. *Arch Otolaryngol-Head Neck Surg*. 2004;130:415–420. https://doi.org/10.1001/archotol.130.4.415.
14. Chen Z, Li J, Ren Q, et al. Acoustic and perceptual analyses of adductor spasmodic dysphonia in mandarin-speaking Chinese. *J Voice*. 2019;33:333–339. https://doi.org/10.1016/j.jvoice.2017.12.007.
15. Zwirner P, Murry T, Swenson M, et al. Effects of botulinum toxin therapy in patients with adductor spasmodic dysphonia: acoustic, aerodynamic, and videoendoscopic findings. *Laryngoscope*. 1992;102:400–406. https://doi.org/10.1288/00005537-199204000-00006.
16. Yeung JC, Fung K, Bornbaum CC, et al. Clinical approach to monitoring variability associated with adductor spasmodic dysphonia. *J Otolaryngol Head Neck Surg*. 2011;40:343–349.
17. Kaptein AA, Hughes BM, Scharloo M, et al. Psychological aspects of adductor spasmodic dysphonia: A prospective population controlled questionnaire study. *Clin Otolaryngology*. 2010;35:31–38. https://doi.org/10.1111/j.1749-4486.2009.02070.x.
18. Rojas GVE, Ricz H, Tumas V, et al. Vocal parameters and self-perception in individuals with adductor spasmodic dysphonia. *J Voice*. 2017;31:391–399. e7-391.e18. https://doi.org/10.1016/j.jvoice.2016.09.029.
19. Isetti D, Xuereb L, Eadie T. Inferring speaker attributes in adductor spasmodic dysphonia: Ratings from unfamiliar listeners. *Amer JSpeech-Lang Path*. 2014;23:134–145.
20. Nagle KF, Eadie TL, Yorkston KM. Everyday listeners' impressions of speech produced by individuals with adductor spasmodic dysphonia. *J Comm Dis*. 2015;58:1–13. https://doi.org/10.1016/j.jcomdis.2015.07.001.
21. Eadie T, Nicolici C, Baylor C, et al. Effect of experience on judgments of adductor spasmodic dysphonia. *Ann Otol Rhinol Laryngol*. 2007;116:695–701.
22. Eadie T, Rajabzedeh R, Isetti D, et al. The effect of information and severity on perception of speakers with adductor spasmodic dysphonia. *Amer JSpeech-Lang Path*. 2017;26:327–341. https://doi.org/10.1044/2016_AJSLP-15-0191.
23. Shipp T, Izdebski K, Reed C, et al. Intrinsic laryngeal muscle activity in a spastic dysphonia patient. *J Speech Hear Dis*. 1985;50:54–59.
24. Dumay N, Content A. Searching for syllabic coding units in speech perception. *J Memory Lang*. 2012;66:680–694.
25. Liss JM, Spitzer S, Caviness JN, et al. Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *J Acoust Soc America*. 1998;104:2457–2466.
26. Bennett S, Weinberg B. Acceptability ratings of normal, esophageal, and artificial larynx speech. *J Speech Lang Hear Res*. 1973;16:608–615.
27. O'Brian S, Packman A, Onslow M, et al. Is listener comfort a viable construct in stuttering research? *J Speech Lang Hear Res*. 2003;46:503–509.
28. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psych Bull.*. 1979;86:420.
29. Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. 892. Upper Saddle River, NJ: Pearson/Prentice Hall; 2009.
30. Sapienza CM, Walton S, Murry T. Acoustic variations in adductor spasmodic dysphonia as a function of speech task. *J Speech Lang Hear Res*. 1999;42:127–140.
31. Prosek RA, Runyan CM. Temporal characteristics related to the discrimination of stutterers' and nonstutterers' speech samples. *J Speech Lang Hear Res*. 1982;25:29–33.
32. Prosek RA, Vreeland LL. The intelligibility of time-domain-edited esophageal speech. *J Speech Lang Hear Res*. 2001;44:525–534.
33. Klopfenstein M. Relationship between acoustic measures and speech naturalness ratings in Parkinson's disease: A within-speaker approach. *Clin Ling Phon*. 2015;29:938–954. https://doi.org/10.3109/02699206.2015.1081293.
34. Nagle K, Eadie T, Yorkston K. Everyday listeners' impressions of speech produced by individuals with adductor spasmodic dysphonia. *J Comm Dis*. 2015;58:1–13. https://doi.org/10.1016/j.jcomdis.2015.07.001.
35. Kreiman J, Gerratt BR, Kempster GB, et al. Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *J Speech Hear Res*. 1993;36:21–40.
36. Kreiman J, Gerratt BR, Berke GS. The multidimensional nature of pathologic vocal quality. *J Acoust Soc Amer*. 1994;96:1291–1302.
37. Kent RD. Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *Amer J Speech Lang Pathol*. 1996;5:7–23.
38. Eadie TL, Kapsner-Smith M. The effect of listener experience and anchors on judgments of dysphonia. *J Speech Lang Hear Res*. 2011;54:430–447.
39. Chan KM, Yiu EM. The effect of anchors and training on the reliability of perceptual voice evaluation. *J Speech Lang Hear Res*. 2002;45:111–126. https://doi.org/10.1044/1092-4388(2002/009.
40. Cannito MP, Burch AR, Watts C, et al. Disfluency in spasmodic dysphonia: A multivariate analysis. *J Speech Lang Hear Res*. 1997;40:627–641.
41. Susca M, Healey EC. Listener perceptions along a fluency−disfluency continuum: A phenomenological analysis. *J Fluency Dis*. 2002;27:135–161.
42. Eadie T, Doyle PC. Classification of dysphonic voice: acoustic and auditory-perceptual measures. *Journal of Voice*. 2005;19(1):1–14.