JSLHR

Research Note

Acoustic Model of Perceived Overall Severity of Dysphonia in Adductor-Type Laryngeal Dystonia

Daniel P. Buckley,[a,b] Manuel Diaz Cadiz,[a] Tanya L. Eadie,[c] and Cara E. Stepp[a,b,d]

Purpose: This study is a secondary analysis of existing data. The goal of the study was to construct an acoustic model of perceived overall severity of dysphonia in adductory laryngeal dystonia (AdLD). We predicted that acoustic measures (a) related to voice and pitch breaks and (b) related to vocal effort would form the primary elements of a model corresponding to auditory-perceptual ratings of overall severity of dysphonia.
Method: Twenty inexperienced listeners evaluated the overall severity of dysphonia of speech stimuli from 19 individuals with AdLD. Acoustic features related to primary signs of AdLD (hyperadduction resulting in pitch and voice breaks) and to a potential secondary symptom of AdLD (vocal effort, measures of relative fundamental frequency) were computed from the speech stimuli. Multiple linear regression analysis was applied to construct an acoustic model of the overall severity of dysphonia.
Results: The acoustic model included an acoustic feature related to pitch and voice breaks and three acoustic measures derived from relative fundamental frequency; it explained 84.9% of the variance in the auditory-perceptual ratings of overall severity of dysphonia in the speech samples.
Conclusions: Auditory-perceptual ratings of overall severity of dysphonia in AdLD were related to acoustic features of primary signs (pitch and voice breaks, hyperadduction associated with laryngeal spasms) and were also related to acoustic features of vocal effort. This suggests that compensatory vocal effort may be a secondary symptom in AdLD. Future work to generalize this acoustic model to a larger, independent data set is necessary before clinical translation is warranted.

Laryngeal dystonia (LD), also known as spasmodic dysphonia, is a focal neurological dystonia that affects the laryngeal musculature (Blitzer et al., 2018). Individuals may have adductory LD (AdLD), abductory LD, or a mixed presentation of adductor/abductor LD. In AdLD, the vocal folds hyperadduct during purposeful speech because of involuntary spasms in the intrinsic muscles of vocal fold adduction (Murry, 2014). Individuals may also experience voice and pitch breaks due to these laryngeal muscle spasms occurring during voicing. With this disrupted neurological input during vocal fold adduction and the subsequent increase in laryngeal tension during a spasm, individuals with AdLD often experience a secondary increase of strain and vocal effort (Nagle et al., 2015; Shoffel-Havakuk et al., 2019). This increase of vocal effort is likely compensatory, in response to the unreliable nature of the vocal system.

The assessment and diagnosis of LD comes from an interdisciplinary team of providers, most often including the otolaryngologist, speech-language pathologist, and neurologist (Stewart et al., 1997). A history of the voice problem, visualization of vocal fold vibration via laryngeal videostroboscopy, auditory-perceptual evaluation, and a neurological evaluation including laryngeal electromyography are often the main assessments used for developing a diagnosis of LD (Langeveld et al., 2000). Other disorders, including primary muscle tension dysphonia, may present similarly, and thus, LD can be a diagnosis of exclusion in some instances. A comprehensive evaluation is often needed to assess for disorder-specific qualities, such as increased perception of spasms with voiced-loaded sentences (Erickson, 2003; Roy et al., 2007) and differentially improved symptoms

[a]Department of Speech, Language, and Hearing Sciences, Boston University, MA
[b]Department of Otolaryngology–Head and Neck Surgery, Boston University School of Medicine, MA
[c]Department of Speech and Hearing Sciences, University of Washington, Seattle
[d]Department of Biomedical Engineering, Boston University, MA
Correspondence to Cara E. Stepp: cstepp@bu.edu

Editor-in-Chief: Bharath Chandrasekaran
Editor: Jack J. Jiang
Received November 22, 2019
Revision received January 28, 2020
Accepted May 19, 2020
https://doi.org/10.1044/2020_JSLHR-19-00354

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

during innate relative to learned vocal behaviors (Guiry et al., 2019). Differential diagnosis may even be difficult for experts without the use of strict classification guidelines, due to reliance on perceptual judgments (Ludlow et al., 2018) as well as patient-reported symptoms (Shoffel-Havakuk et al., 2019). Given this tenuous path to diagnosis, the need for objective measures in the diagnostic assessment of LD remains a priority. In addition to diagnostic needs, objective measures that are sensitive to the severity of signs and symptoms in dysphonia are also imperative in order to assess treatment efficacy.

Given the signs and symptomatology of AdLD, speech acoustics may provide an ideal modality for providing objective assessment. However, current methods are generally insufficient, perhaps due to a focus on measures of voice quality that are not specific to AdLD and measures that relate to its primary signs only. Studies have found low levels of sensitivity identifying AdLD using single acoustic measures such as jitter, voicing length (Moerman et al., 2015), frequency shifts, and aperiodic segments (Yanagida et al., 2018). Furthermore, there is not a currently known measure that reliably correlates to both its primary and secondary perceptual features. Cepstral–spectral acoustic measures, such as cepstral peak prominence and low-high spectral ratio (Awan et al., 2010), may be used in concert to provide information on the degree of overall severity of the dysphonia but are unlikely to be sensitive to the disorder-specific qualities of AdLD. For instance, the Cepstral Spectral Index of Dysphonia, an acoustic index of dysphonia severity, is only weakly correlated with listener ratings of overall severity of dysphonia in AdLD (i.e., explaining 20%–46% of the variance; Roy et al., 2014). Current methods of analyzing voice/pitch breaks, a primary feature of hyperadduction in those with AdLD where there is the cessation of voicing during an expected voiced segment, appear more promising (Cannito et al., 2012; Cimino-Knight & Sapienza, 2001; Izdebski, 1984; Sapienza et al., 1999, 2000; Siemons-Lühring et al., 2009; Yanagida et al., 2018) but typically require manual processing and thus are time-consuming. For example, a method used for calculating the percentage of voice breaks in speech involves manually identifying each phonatory break, determining each break's duration, calculating the sum of the duration of the phonatory breaks in a sample, and then dividing that value by the sum of the duration of voicing in the given sample (Sapienza et al., 1999; Yanagida et al., 2018). This technique can be automated for simple stimuli. For instance, the Multi-Dimensional Voice Program (Kay Elemetrics, 1993) includes this calculation, but the simplistic approach to automation limits its validity to sustained phonation (i.e., isolated vowels). Overall, due to these issues, quantifying voice breaks in speech stimuli is laborious in the clinical setting. It remains a promising goal, however: when combined with a global measure of voice quality (cepstral peak prominence), these measures were able to explain 62% of the variance in listeners' ratings of voice roughness (corresponding with hyperadduction in speakers with AdLD; Cannito et al., 2012).

Since the auditory-perceptual qualities present in AdLD are likely to stem from both primary (frequent spasms) and secondary (vocal effort) symptoms, an acoustic metric that incorporates both may be critical for diagnostic specificity. The acoustic measure relative fundamental frequency (RFF) may be a useful adjunct to acoustic measures focused on primary AdLD symptoms. RFF describes the changes in fundamental frequency that occur in the 10 cycles of voicing before and after a voiceless consonant, normalized to more steady-state values. RFF values are lower in individuals thought to have increased laryngeal tension—those with muscle tension dysphonia (Heller Murray et al., 2017; Roy et al., 2016; Stepp et al., 2010) and Parkinson's disease (Goberman & Blomgren, 2008; Stepp, 2013). Further support for RFF as an acoustic correlate for laryngeal tension comes from evidence that RFF values in individuals with muscle tension dysphonia normalize after therapy (Stepp et al., 2011), that RFF values decrease with increased vocal effort in typical speakers (Lien et al., 2015; McKenna et al., 2016), and that RFF values are correlated with the perception of vocal effort in individuals with AdLD (Eadie & Stepp, 2013). Combining measures of voice/pitch breaks with RFF may result in an objective, multidimensional acoustic measure sensitive to both the primary and secondary features of AdLD.

The purpose of this study was to build on previous work indicating that listener perceptions' of voice in AdLD are related to acoustic features of voice and pitch breaks (Cannito et al., 2012). Specifically, our goal was to construct an acoustic model for the perception of overall severity of dysphonia that included automated acoustic features of voice and pitch breaks as well as acoustic correlates of vocal effort (RFF values). Given the extent to which both primary and secondary features appear to impact overall severity, we predicted that a multidimensional approach of acoustic analysis incorporating both voice and pitch breaks and vocal effort would relate strongly to the perception of overall severity of dysphonia in individuals with AdLD. An acoustic metric sensitive to both the primary and secondary features of AdLD may assist in its diagnosis and assessment of its treatment response.

Method

Speech recordings from 19 adult speakers with AdLD and auditory-perceptual judgments from 20 adult listeners were used in this study. These production (RFF) and perception (overall severity of dysphonia) data have been previously reported in the studies of Eadie et al. (2007) and Eadie and Stepp (2013). The current methods incorporate new acoustic analyses.

Speech Stimuli

Speech stimuli were recorded from 19 adults (9 males, 10 females)[1] who had received a diagnosis of AdLD from a laryngologist based on case history, auditory-perceptual assessment of voice, videolaryngostroboscopy evaluation,

[1]Gender information was not collected.

and fine-wire laryngeal electromyography; all received botulinum toxin injections regularly for symptom management and were recorded at the end of their treatment cycle. The mean age of speakers was 58.2 years (range: 37–80 years). Speakers were deliberately selected to represent voices across a range of overall severity of dysphonia, since overall severity is strongly related to the hyperadduction or "overpressure" feature that characterizes AdLD (Cannito et al., 2012).

Speakers were recorded in a quiet environment with low levels of ambient noise while reading the first paragraph of the Rainbow Passage (Fairbanks, 1960), a sentence loaded with voiceless consonants ("He saw half a shape mystically cross a path about fifty or sixty steps from his sister Kathy's house"; Dedo & Shipp, 1980), and a sentence loaded with voiced consonants ("Early one morning a man and a woman ambled along a one-mile lane running near Rainy Island Avenue"; Dedo & Shipp, 1980). Speech samples were recorded with a headset microphone (AKG C420, Harman International Industries, Inc.) routed to an audio recorder (Tascam DAP1, TEAC Corporation) at a sampling rate of 44.1 kHz.

Listeners

Twenty healthy individuals (3 males, 17 females)[2] participated as listeners. The mean age of listeners was 25.7 years (range: 19–41 years). Listeners were native speakers of American English and did not report any speech, language, or voice disorders. Listeners passed a hearing screening at 40 dB HL or better for octave frequencies from 250 to 4000 Hz. They had no prior experience with or exposure to voice disorders. All listeners provided informed consent in compliance with the University of Washington Human Subjects Committee.

Stimuli Preparation and Listening Procedure

The second sentence of the Rainbow Passage ("The rainbow is a division of white light into many beautiful colors"; includes both voiced and unvoiced phonemes) was extracted for each speaker for use as stimuli in the auditory-perceptual task. Each stimulus was peak-normalized using sound-editing software. Peak normalization was performed in order to limit differences in the perceived loudness across samples. Normalization was achieved using Sony Soundforge 7.0, sound-editing software.

Before the rating tasks, listeners were provided with a definition of the overall severity of dysphonia and were familiarized with the auditory-perceptual dimension. Overall severity was defined as "a comprehensive measure of how 'good' or 'poor' the voice sample is judged to be by the listener" (Eadie & Doyle, 2002). Listeners were exposed to the voices of one male and one female speaker with AdLD (neither included in the study) in order to familiarize them with the disorder.

[2]Gender information was not collected.

Stimuli were presented to listeners in random order via a custom software program using headphones set at a comfortable loudness level. The software included a 100-mm Visual Analog Scale on which listeners were asked to record their perceptual judgments. One end of the scale (0) was anchored as *normal* and the other end (100) was anchored as *severe*. Each listener was presented with stimuli from all speakers; approximately 25% of stimuli were repeated in order to assess intrarater reliability. Group means of listener ratings for the overall severity of dysphonia were computed per speaker.

Acoustic Data Analysis

Measures Designed to Capture Voice and Pitch Breaks

Two acoustic features aimed at capturing voice and pitch breaks in running speech were computed using the sentences from each speaker loaded with voiced consonants. The sentences loaded with voiced consonants were used in order to elicit symptomatic speech for analysis (i.e., those with the greatest potential for voice and pitch breaks). These acoustic features were based on an instantaneous fundamental frequency estimator algorithm called "Halcyon" (Azarov et al., 2016). Halcyon was implemented using open source MATLAB scripts (Petrovsky, n.d.). Further processing of the output from Halcyon was also accomplished via custom MATLAB scripts.

Our goal was to identify fast transitions between different periodic and/or aperiodic patterns during voiced segments, since these transitions should occur during both pitch and voice breaks. Halcyon provides estimates of fundamental frequency of an input signal using a multirate sampling framework. This method has shown superior time–frequency resolution (Azarov et al., 2016) when compared to other widely used fundamental frequency estimators, such as autocorrelation, Auditory Sawtooth Waveform Inspired Pitch Estimator—Prime (Camacho, 2012), YIN (de Cheveigne & Kawahara, 2002), and Robust Algorithm for Pitch Tracking (Talkin, 1995), which is why it was employed for detection of fast transitions rather than conventional fundamental frequency estimation methods.

Stimuli were analyzed using an automated signal envelope thresholding approach to determine which time frames were voiced. Only voiced segments were considered in further analysis. This allowed for the detection of fast transitions in fundamental frequency in existing voiced regions (pitch breaks) as well as discontinuities in fundamental frequency caused by the removal of unvoiced segments (voice breaks). Stimuli were then processed by Halcyon, collecting a range of possible fundamental frequency candidates between 30 and 400 Hz over 1-ms sampling windows. These fundamental frequency estimates were then resampled to correspond to the original acoustic signal using linear interpolation.

Using the estimated fundamental frequency as a function of time, we evaluated the frequency content of that time-varying signal. The fundamental frequency contour was visualized as a "spectrogram" (i.e., the frequency characteristics of the contour as a function of time), with rapid fundamental frequency excursions having correspondingly high-frequency energy. We then time-averaged each of these spectrograms to generate an aggregated spectrum-like representation, denoted as *meanSgram.* Similarly, we calculated the total dispersion (standard deviation) of each spectrogram across time, which we denoted as *stdSgram.* From each of these representations, the energy above 1000 Hz was summed (effectively high-pass filtering the information), resulting in two composite measures, *meanSgram-HP* and *stdSgram-HP,* in which *HP* refers to the resulting measure having been high-pass filtered. The threshold of 1000 Hz was chosen empirically: after the meanSgram and stdSgram for all samples were calculated and visually inspected, this value appeared to best capture the fast changes associated with samples with more voice and pitch breaks.[3]

These two measures were designed to increase in value when the changes in the fundamental frequency were occurring more often and/or more quickly. The motivation of this process can be explained by a two-case comparison presented in Figure 1. Two speech samples with low and high occurrence of voice and pitch breaks (as determined by a voice-specializing speech-language pathologist) are presented in Figures 1A and 1B, respectively. Using Halcyon, time waveforms of the fundamental frequency are obtained. These fundamental frequency waveforms are transformed into spectrograms using 25-ms Hamming windows with 90% overlap (shown in Figures 1C and 1D). As illustrated in these examples, the individual with more voice and pitch breaks (Figure 1D) results in more sporadic jumps on the frequency axis than the individual with fewer voice and pitch breaks (see Figure 1C). To capture this, the energy is aggregated, calculating the mean and total dispersion (standard deviation) over the time axis of these spectrograms. This collapses the spectrogram into "spectrum-like" representations *meanSgram* and *stdSgram,* respectively (shown in Figures 1E and 1F). This operation aggregates all irregular events in the fundamental frequency trace and illustrates them as prominent high-energy content in the higher frequency range. This is quantified by calculating the total normalized sum over *meanSgram* and *stdSgram* above 1000 Hz (colored areas in Figures 1E and 1F) to determine *meanSgram-HP* and *stdSgram-HP* values. Note that in the two examples in Figure 1, both the *meanSgram-HP* and *stdSgram-HP* are higher in the speech sample with the high occurrence of voice and pitch breaks (0.148 and 0.464, respectively) than in the speech sample with the low occurrence of voice and pitch breaks (0.054 and 0.286, respectively).

Measures Designed to Capture Vocal Effort

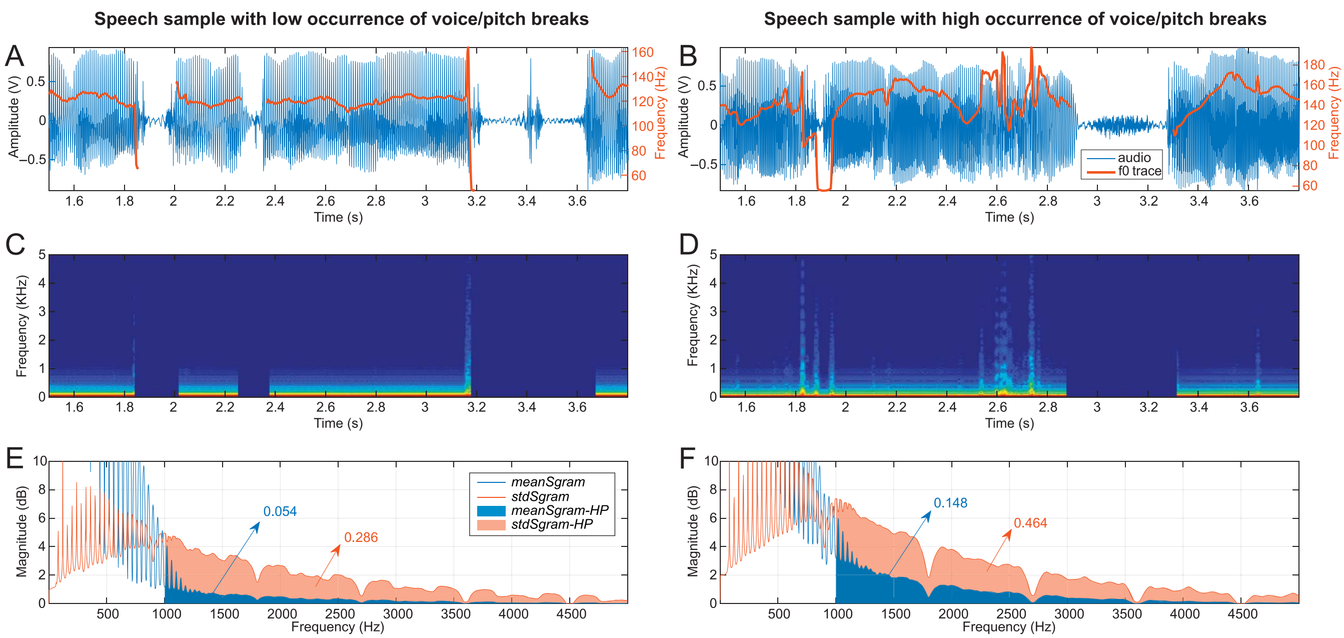
RFF values were computed for each speaker using the sentences from each speaker loaded with unvoiced consonants. The sentences loaded with unvoiced consonants were used in order to assess instances in which RFF could be calculated—those in which there was a voiceless consonant surrounded by voiced segments. Manual RFF analysis was consistent with previous work (e.g., Heller Murray et al., 2016; Lien et al., 2014; Stepp, 2013). For all participants, nine instances were used from the sentence "He saw half a shape mystically cross a path about fifty or sixty steps from his sister Kathy's house": "he saw," "half a," "a shape," "cross a," "a path," "path about," "or sixty," "sister Kathy," and "Kathy." The automated method of calculating RFF was not possible, as it currently requires isolated instances of vowel–consonant–vowel utterances (i.e., /ɒfɒ, ifi, ufu/; Vojtech et al., 2019), which were not elicited from our speech-level stimuli. Thus, manual calculations were completed.

The final author (C. E. S.) visualized each RFF instance in Praat and localized the 10 vocal cycles preceding the voiceless consonant (termed as "offset cycles") and the 10 vocal cycles following the voiceless consonant (termed as "onset cycles"). The instantaneous fundamental frequency of each offset and onset cycle was computed as the inverse of the period, identified using the pulse function of Praat (Boersma & Weenink, 2008). Each fundamental frequency was subsequently normalized in semitones (ST) relative to the associated vocal cycles nearest to the midpoint of the voiced segment: Offset Cycle 1 and Onset Cycle 10. All available RFF instances per speaker were used to compute an average value of each offset and onset cycle. Some productions could not be used to reliably determine RFF due to glottalization, insufficient vocal cycles to achieve steady state, or voicing during the obstruent. Thus, across speakers, at least one RFF instance was available for all speakers, with an average of 5.7 instances used for each participant's offset averages and 4.6 instances used for each participant's onset averages; this is consistent with a previous work utilizing manual RFF analysis of running speech in which an average of 2.4–9.8 usable instances were available for averaging (Heller Murray et al., 2016; Stepp, 2013; Stepp et al., 2011).

These average RFF values were used to compute six RFF features for each speaker: *RFFoff10* and *RFFon1* (the two values closest to the transition between voiced and unvoiced segments), Δ*RFFoff10-9* and Δ*RFFoff10-5* (measures aimed to capture the slope of change in offset values by subtracting from RFF Offset Cycle 10 values the values of Offset Cycles 9 and 5, respectively), and Δ*RFFon1-2* and Δ*RFFoff1-6* (measures aimed to capture the slope of change in onset values by subtracting from RFF Onset Cycle 1 values the values of Onset Cycles 2 and 6, respectively). Our choice of measures aimed at capturing the

[3]The 1000-Hz effective sampling rate of the 1-ms processing window used in Halcyon to estimate the fundamental frequency and the linear interpolation to resample the fundamental frequency contour results in low-frequency subharmonics below 1000 Hz. These can be seen in Figures 1E and 1F. These are not related to the number of voice and pitch breaks. Our choice to incorporate a 1000-Hz cutoff for the parameters meanSgram-HP and stdSgram-HP was informed by the presence of these harmonics. The 1000-Hz cutoff appeared to best capture the fast changes associated with samples with more voice and pitch breaks and also allowed for the removal of the processing-based harmonics.

Figure 1. Example of the analysis of two participants with adductory laryngeal dystonia. Segments of the original speech sample (audio) and respective fundamental frequency waveform estimates (f0 trace) for an individual with (A) low occurrence of voice and pitch breaks and (B) high occurrence of voice and pitch breaks. Panels C and D show the spectrogram representations of the fundamental frequency contours from Panels A and B, respectively. Panels E and F illustrate the aggregation over time of the previous spectrograms: aggregation of the mean (*meanSgram,* in blue) and total dispersion (*stdSgram,* in orange). The final feature calculation is illustrated by the filled-in areas in the frequency range above 1000 Hz. Arrows of the respective colors indicate the final value of the features *meanSgram-HP* (in blue) and *stdSgram-HP* (in orange). *meanSgramHP* = the high-pass filtered time average of each "spectrogram" of the fundamental frequency contour; *stdSgramHP* = the high-pass filtered standard deviation of each "spectrogram" of the fundamental frequency contour.



slope of change in RFF values was based on examination of the current data set and previous RFF literature (e.g., Heller Murray et al., 2017, 2016; Lien et al., 2014, 2015), in which Offset Values 1 through 5 and Onset Values 6 through 10 are typically relatively stable.

The investigator (C. E. S.) reevaluated approximately 10% of RFF stimuli approximately 18 months after the initial evaluation, and a second trained researcher (L. B.) also evaluated approximately 10% of the RFF stimuli. Intrarater and interrater reliability of the RFF analysis was assessed with Pearson *r* using the reevaluated data sets, resulting in $r = .93$ and $r = .94$, respectively.

Statistical Analysis

All statistical analyses were completed using Minitab Statistical Software, Version 17 (Minitab, Inc.). Significance for all statistical testing was set a priori at $p < .05$.

The mean Pearson product–moment correlations between original and repeated samples were computed to assess intrarater reliability of listeners' ratings of overall severity, yielding an average $r = .71$ ($SD = 0.34$); this value is similar those previously reported for overall severity of dysphonia, which have ranged from $r = .57$ to $r = .87$ (Eadie & Doyle, 2005; Eadie et al., 2010; Zraick et al., 2011). The interrater reliability of listeners' ratings was assessed using Cronbach's alpha, which was .98; this value is similar to previously reported values, which ranged from .97 to .99 (Eadie & Doyle, 2005; Eadie et al., 2010).

The model was constructed by conducting a stepwise multiple linear regression using the two acoustic features designed to capture voice and pitch breaks (*meanSgram-HP* and *stdSgram-HP*) as well as the six RFF features as independent variables. The overall severity of dysphonia was the dependent variable. The resulting regression was compared to the mean ratings of overall severity of dysphonia, resulting in R^2 and R^2_{adj} values. The R^2 reflects the proportion of the variance explained by the regression, whereas the R^2_{adj} adjusts the R^2 based on the number of factors in the regression model (i.e., the greater the number of factors in the analysis, the smaller the R^2_{adj} value; Abu-Bader, 2016). The model was run with the alpha for entrance and removal set to .05, .10, and .20, all of which resulted in the same final model. Pearson correlation coefficients were computed between all independent and dependent variables to provide descriptive information about relationships among variables.

Results

The overall severity of dysphonia ratings of the 19 speakers ranged from 6.2 to 85.9 mm ($M = 45.6$ mm, $SD = 22.2$ mm), indicating that the speakers represented a full range. Likewise, RFF values varied in the group. For instance *RFFoff10* values ranged from −3.51 to 1.76 ST (M = −1.17 ST, SD = 1.49 ST) and *RFFon1* values ranged from −0.80 to 4.60 ST (M = 2.55 ST, SD = 1.48 ST).

The stepwise regression model analyzed the relationship between the two acoustic features designed to capture voice and pitch breaks, the six RFF features, and the auditory-perceptual ratings of overall severity of dysphonia. Table 1 details Pearson correlation coefficients between all independent and dependent variables. Table 2 includes the full initial model. The final model (see Table 3) accounted for R^2 = 84.9% of the variance in the auditory-perceptual ratings. The associated R^2_{adj} was 80.6%. The relationship between the model output and the auditory-perceptual ratings is shown in Figure 2. The associated regression model is shown in Equation 1.

$$\begin{array}{c}\text{Overall Severity}\\ \text{of Dysphonia}\end{array} = \begin{array}{l} 51.41 + 90.4 \times \mathit{meanSgram\text{-}HP} - \\ 20.10 \times \Delta \mathit{RFFoff\,10\text{-}9} - 9.19 \times \\ \Delta \mathit{RFFon1\text{-}6} + 7.18 \times \ \mathit{RFFoff\,10} \end{array} \quad (1)$$

Discussion

Our results show that listener perceptions of overall severity of dysphonia in speakers with AdLD are associated with various acoustic features; specifically, 84.9% of the variance in the auditory-perceptual evaluation of overall severity was accounted for using a single acoustic feature designed to capture voice and pitch breaks and three RFF features designed to capture vocal effort. This finding expands on other studies that have similarly reported that acoustic features of voice and pitch breaks are strong correlates of the level of dysphonia in AdLD. For instance, Cannito et al. (2012) found that three measures of voice and pitch break features and cepstral peak prominence were able to explain 61.6% of the variance in auditory-perceptual ratings of roughness, which was strongly related to hyperadduction in speakers with AdLD.

The inclusion of RFF metrics in this model suggests that RFF may offer additional prediction of the overall severity of dysphonia in AdLD. Based on the prior work suggesting that RFF is sensitive to increased laryngeal tension and vocal effort (Lien et al., 2015; McKenna et al., 2016; McKenna & Stepp, 2018), this suggests that overall severity of dysphonia in AdLD is a result of both primary (laryngeal spasms) and secondary (compensatory vocal effort) symptoms. In individuals with AdLD, compensatory effort may be characterized by hyperfunctional behaviors needed to overcome adductory spasms. This may even include strategies to maintain airflow through the glottis, resulting in a breathy voice, that is, in accordance with what Izdebski (1984) has proposed as a two-factor model of AdLD. Given the strong relationship between our acoustic model and the auditory-perceptual ratings of overall severity, these results show potential for the development of a multidimensional clinical acoustic metric for assessing AdLD.

The use of the acoustic measure relating to vocal effort, RFF, was completed in our study via a manual method. Due to the type of stimuli gathered, manual calculations of RFF were required. As previously mentioned, manual methods of calculation have limited clinical use due to time and resource constraints. In response to this, recent work has sought to automate RFF analysis (Lien et al., 2017; Vojtech et al., 2019). This method has been shown to be reliable with the correct stimuli (i.e., /ɒfɒ/) but cannot yet be used on running speech. Because of these barriers, future research should be conducted that uses acoustic stimuli compatible with the automated method of RFF calculation, which would allow for our acoustic model's transition to the clinical evaluation of AdLD. Alongside this, the development of future automated RFF algorithms that allow for RFF analysis in running speech may be most useful for analyzing a larger set of data across broader clinical, linguistic, and speech contexts.

Table 1. Pearson correlation coefficients between independent and dependent variables.

Variable	Overall severity of dysphonia	*meanSgram-HP*	*stdSgram-HP*	*ΔRFFoff10-5*	*ΔRFFoff10-9*	*ΔRFFon1-2*	*ΔRFFon1-6*	*RFFoff10*
**meanSgram-HP*	**.70**							
stdSgram-HP	**.64**	**.95**						
ΔRFFoff10-5	−.14	.13	.17					
**ΔRFFoff10-9*	−.21	.14	.11	**.73**				
ΔRFFon1-2	**−.50**	−.26	−.22	.18	−.04			
**ΔRFFon1-6*	**−.59**	−.28	−.17	−.26	−.11	**.78**		
**RFFoff10*	−.13	.10	.15	**.98**	**.77**	.12	.19	
RFFon1	**−.63**	−.40	−.27	.24	−.14	**.69**	**.94**	.19

Note. Bolded values indicate statistically significant correlations (p > .05). Dependent variables included in the final stepwise regression model (see Table 3) are indicated with asterisks in the left-most column. *meanSgramHP* = the high-pass filtered time average of each "spectrogram" of the fundamental frequency contour; *stdSgramHP* = the high-pass filtered standard deviation of each "spectrogram" of the fundamental frequency contour; Δ*RFFoff10-5* = the difference between relative fundamental frequency Offset Cycles 10 and 5; Δ*RFFoff10-9* = the difference between relative fundamental frequency Offset Cycles 10 and 9; Δ*RFFon1-2* = the difference between relative fundamental frequency Onset Cycles 1 and 2; Δ*RFFon1-6* = the difference between relative fundamental frequency Onset Cycles 1 and 6; *RFFoff10* =relative fundamental frequency Offset Cycle 10; *RFFon1* = relative fundamental frequency Onset Cycle 1.

Table 2. Summary of variables in the initial regression model.

Source	*df*	Standardized coefficient	95% CI	*F*	*p*	VIF
Regression	8	n/a	n/a	7.49	.002	n/a
**meanSgram-HP*	1	10.70	[–11.90, 33.20]	1.11	.317	**14.56**
stdSgram-HP	1	1.48	[–19.83, 22.80]	0.02	.880	**12.98**
ΔRFFoff10-5	1	–1.80	[–32.30, 28.70]	0.02	.898	**26.65**
**ΔRFFoff10-9*	1	–17.10	[–27.76, –6.43]	12.76	.005	3.25
ΔRFFon1-2	1	1.35	[–8.63, 11.33]	0.09	.769	2.85
**ΔRFFon1-6*	1	–9.20	[–32.10, 13.70]	0.80	.391	**14.99**
**RFFoff10*	1	13.40	[–17.90, 44.70]	0.91	.362	**28.02**
RFFon1	1	–5.97	[–27.34, 15.41]	0.39	.548	**13.05**

Note. Variance inflation factor (VIF) values over 10 are bolded. Dependent variables included in the final stepwise regression model (see Table 3) are indicated with asterisks in the left-most column. *df* = degrees of freedom; CI = confidence interval; n/a = not applicable; *meanSgramHP* = the high-pass filtered time average of each "spectrogram" of the fundamental frequency contour; *stdSgramHP* = the high-pass filtered standard deviation of each "spectrogram" of the fundamental frequency contour; Δ*RFFoff10-5* = the difference between relative fundamental frequency Offset Cycles 10 and 5; Δ*RFFoff10-9* = the difference between relative fundamental frequency Offset Cycles 10 and 9; Δ*RFFon1-2* = the difference between relative fundamental frequency Onset Cycles 1 and 2; Δ*RFFon1-6* = the difference between relative fundamental frequency Onset Cycles 1 and 6; *RFFoff10* = relative fundamental frequency Offset Cycle 10; *RFFon1* = relative fundamental frequency Onset Cycle 1.

Of the potential eight independent variables, only four variables were included in the final stepwise regression model: a measure intended to capture voice and pitch breaks (*meanSgram-HP*), a measure of RFF slope in the transition from a voiced to unvoiced segment (Δ*RFFoff10-9*), a measure of RFF slope in the transition from an unvoiced to a voiced segment (Δ*RFFon1-6*), and the RFF value closest to the transition from a voiced to an unvoiced segment (*RFFoff10*). A simplistic interpretation would be to conclude that these features, unlike the other potential features, are providing information specific to the overall severity of dysphonia in AdLD. However, inspection of the correlation coefficients among the potential predictor variables (see Table 1) and the variance inflation factors for predictor variables when all are included (see Table 2) suggests a more complex interpretation. For instance, when estimated independently, both measures intended to capture voice and pitch breaks (*meanSgram-HP* and *stdSgram-HP*) have statistically significant relationships with overall severity of dysphonia (r = .70 and r = .64, respectively; see Table 1). However, their inclusion together in the regression model (see Table 2) leads to inappropriately high variance inflation factors for both variables (14.56 and 12.98), likely driven by the very high correlation between the two measures (r = .95). Thus, it is unsurprising that only *meanSgram-HP* is included in the final model. Likewise, although somewhat less obvious, the relationships among RFF measures are also likely influencing the predictors selected for the final model. When estimated independently, the following measures showed statistically significant relationships with overall severity of dysphonia: Δ*RFFon1-2*, Δ*RFFon1-6*, and *RFF on1*. Of note, all of these variables reflect onset RFF values and slopes, and only one of these (Δ*RFFon1-6*) is included in the final model. As might be expected, these measures are not independent but instead show statistically significant correlations among one another, ranging from r = .69 to r = .94. Likewise, RFF offset values and slopes are also interrelated, with statistically significant correlations among one another, ranging from r = .73 to r = .98. In summary, although only four of the eight potential measures are included in the model, their inclusion is likely driven by complex relationships within measures rather than "superiority" of one measure over another.

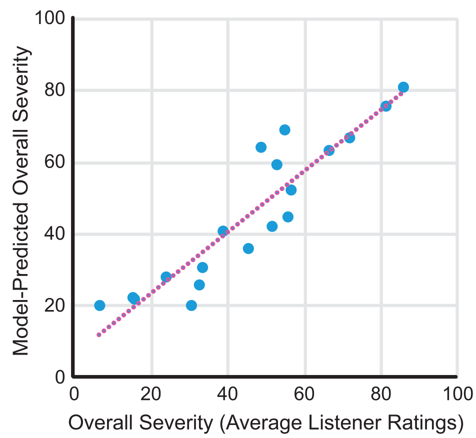
Although the current results are promising, they should be interpreted with caution. Any acoustic model developed relies on the relevance and quality of the dependent

Table 3. Summary of variables in the final stepwise regression model.

Source	*df*	Standardized coefficient	95% CI	*F*	*p*	VIF
Regression	4	n/a		19.70	< .001	n/a
meanSgram-HP	1	12.89	[7.66, 18.13]	27.92	< .001	1.12
ΔRFFoff10-9	1	–16.16	[–24.58, –7.74]	16.95	.001	2.90
ΔRFFon1-6	1	–13.27	[–19.03, –7.52]	24.45	< .001	1.36
RFFoff10	1	10.72	[2.13, 19.30]	7.17	.018	3.01

Note. *df* = degrees of freedom; CI = confidence interval; VIF = variance inflation factor; n/a = not applicable; *meanSgramHP* = the high-pass filtered time average of each "spectrogram" of the fundamental frequency contour; Δ*RFFoff10-9* = the difference between relative fundamental frequency Offset Cycles 10 and 9; Δ*RFFon1-6* = the difference between relative fundamental frequency Onset Cycles 1 and 6; *RFFoff10* = relative fundamental frequency Offset Cycle 10.

Figure 2. A scatter plot of the model fit as a function of auditory-perceptual evaluation of the overall severity of dysphonia across all 19 participants with adductory laryngeal dystonia. The dotted line indicates the line of best fit for each relationship.



measure used—here, auditory-perceptual ratings of the overall severity of dysphonia. Listener perceptions may be sensitive to a myriad of factors, including those unrelated to the voice disorder being studied, such as features of the listener and the scale (Barsties & De Bodt, 2015). Whereas individual ratings of voice quality may be highly variable, reducing their value in clinical evaluations, use of averaged responses from large groups of listeners in research designs may overcome this concern (Shrivastav et al., 2005). Here, we employed the mean responses from 20 reliable listeners, which likely mitigates concern over using these ratings as the dependent variable for our model. Furthermore, the overall severity of dysphonia has been shown to be a relatively robust, gestalt measure of voice quality (e.g., De Bodt et al., 1997; Solomon et al., 2011), and it has been shown to best characterize the primary perceptual feature (hyperadduction) in speakers with AdLD (Cannito et al., 2012). Future work may expand on this by incorporating the perceptual evaluation of expert listener speech-language pathologists and laryngologists who may refine the perceived degree of overall severity in this population.

In addition, future work should also consider how common treatment, such as botulinum toxin A (BTX), affects perceived severity of AdLD. For example, Cannito et al. (2012) found that after BTX, speakers with AdLD were best characterized by breathy voices (and hypoadduction) more than hyperadduction. However, in their study, the only significant acoustic measure associated with breathiness was a measure of smoothed cepstral peak prominence, accounting for only about 30% of the variance in auditory-perceptual ratings. Any objective index of AdLD must therefore consider how AdLD speech characteristics may change after BTX. Since the increased vocal effort in AdLD is thought to be compensatory to the primary sign of laryngeal spasms, we predict that, in most patients, a positive response to BTX would be captured by changes in RFF metrics. However, some patients with AdLD may require behavioral intervention even after BTX to aid in active unlearning of compensatory behaviors (Murry & Woodson, 1995). In these patients, we anticipate that RFF metrics will only improve after the compensatory behaviors have been addressed.

Finally, a core limitation of this study is that the small sample size necessitated that the acoustic model was derived from and tested on the same data set. Because of this, the model is likely to be overfit to the presentation of AdLD within our specific sample group. In order to account for the likely heterogeneity in presentation of the general population and to increase the probability of external validity, further research should be completed that uses a larger, independent sample. This further work may either validate, refute, or enhance our current multidimensional acoustic model of AdLD.

Conclusion

An acoustic model of overall severity of dysphonia for AdLD was constructed. The model included an acoustic feature designed to capture primary signs of AdLD (pitch and voice breaks) as well as RFF measures designed to capture a secondary symptom of AdLD (vocal effort). The output of the model was significantly associated with auditory-perceptual ratings of overall severity of dysphonia, with R^2 = 84.9% and R^2_{adj} = 80.6%. Although these results are promising, evaluation of the model on the same data sets that were used for model construction limits the generalization of findings. Further evaluation in a larger, independent test set, as well as studies that investigate differences due to treatment, such as BTX, is essential before clinical translation is warranted.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grant R01DC015570 (awarded to C. E. S.). The authors thank Leah Bowen for performing relative fundamental frequency analysis to calculate interrater reliability and Devon Sawin for assistance with perceptual data analysis.

References

Abu-Bader, S. H. (2016). *Advanced and multivariate statistical methods for social science research*. Oxford University Press.

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*(9), 742–758. https://doi.org/10.3109/02699206.2010.492446

Azarov, E., Vashkevich, M., & Petrovsky, A. (2016). Instantaneous pitch estimation algorithm based on multirate sampling. *2016 IEEE International Conference on Acoustics, Speech and Signal*

Processing Proceedings (pp. 4970–4974). https://doi.org/10.1109/ICASSP.2016.7472623
Barsties, B., & De Bodt, M. (2015). Assessment of voice quality: Current state-of-the-art. *Auris, Nasus, Larynx, 42*(3), 183–188. https://doi.org/10.1016/j.anl.2014.11.001
Blitzer, A., Brin, M. F., Simonyan, K., Ozelius, L. J., & Frucht, S. J. (2018). Phenomenology, genetics, and CNS network abnormalities in laryngeal dystonia: A 30-year experience. *Laryngoscope, 128*(Suppl. 1), S1–S9. https://doi.org/10.1002/lary.27003
Boersma, P., & Weenink, D. (2008). *Praat: Doing phonetics by computer* (Version 5.0.20) [Computer program]. http://www.praat.org/, date last viewed 4/3/2012
Camacho, A. (2012). *On the use of auditory models' elements to enhance a sawtooth waveform inspired pitch estimator on telephone-quality signals*. Paper presented at the 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA). https://doi.org/10.1109/ISSPA.2012.6310450
Cannito, M. P., Doiuchi, M., Murry, T., & Woodson, G. E. (2012). Perceptual structure of adductor spasmodic dysphonia and its acoustic correlates. *Journal of Voice, 26*(6), 818. e5–818.e13. https://doi.org/10.1016/j.jvoice.2012.05.005
Cimino-Knight, A. M., & Sapienza, C. M. (2001). Consistency of voice produced by patients with adductor spasmodic dysphonia: A preliminary investigation. *Journal of Speech, Language, and Hearing Research, 44*(4), 793–802. https://doi.org/10.1044/1092-4388(2001/062)
De Bodt, M. S., Wuyts, F. L., Van de Heyning, P. H., & Croux, C. (1997). Test–retest study of the GRBAS scale: Influence of experience and professional background on perceptual rating of voice quality. *Journal of Voice, 11*(1), 74–80. https://doi.org/10.1016/S0892-1997(97)80026-4
de Cheveigne, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America, 111*(4), 1917–1930. https://doi.org/10.1121/1.1458024
Dedo, H. H., & Shipp, T. (1980). *Spastic dysphonia: A surgical and voice therapy treatment program*. College Hill Press.
Eadie, T. L., & Doyle, P. C. (2002). Direct magnitude estimation and interval scaling of naturalness and severity in tracheoesophageal (TE) speakers. *Journal of Speech, Language, and Hearing Research, 45*(6), 1088–1096. https://doi.org/10.1044/1092-4388(2002/087)
Eadie, T. L., & Doyle, P. C. (2005). Classification of dysphonic voice: Acoustic and auditory-perceptual measures. *Journal of Voice, 19*(1), 1–14. https://doi.org/10.1016/j.jvoice.2004.02.002
Eadie, T. L., Kapsner, M., Rosenzweig, J., Waugh, P., Hillel, A., & Merati, A. (2010). The role of experience on judgments of dysphonia. *Journal of Voice, 24*(5), 564–573. https://doi.org/10.1016/j.jvoice.2008.12.005
Eadie, T. L., Nicolici, C., Baylor, C., Almand, K., Waugh, P., & Maronian, N. (2007). Effect of experience on judgments of adductor spasmodic dysphonia. *Annals of Otology, Rhinology & Laryngology, 116*(9), 695–701. https://doi.org/10.1177/000348940711600912
Eadie, T. L., & Stepp, C. E. (2013). Acoustic correlate of vocal effort in spasmodic dysphonia. *Annals of Otology, Rhinology and Laryngology, 122*(3), 169–176. https://doi.org/10.1177/000348941312200305
Erickson, M. L. (2003). Effects of voicing and syntactic complexity on sign expression in adductor spasmodic dysphonia. *American Journal of Speech-Language Pathology, 12*(4), 416–424. https://doi.org/10.1044/1058-0360(2003/087)
Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed., p. 127). Harper and Row.
Goberman, A. M., & Blomgren, M. (2008). Fundamental frequency change during offset and onset of voicing in individuals with Parkinson disease. *Journal of Voice, 22*(2), 178–191. https://doi.org/10.1016/j.jvoice.2006.07.006
Guiry, S., Worthley, A., & Simonyan, K. (2019). A separation of innate and learned vocal behaviors defines the symptomatology of spasmodic dysphonia. *Laryngoscope, 129*(7), 1627–1633. https://doi.org/10.1002/lary.27617
Heller Murray, E. S., Hands, G. L., Calabrese, C. R., & Stepp, C. E. (2016). Effects of adventitious acute vocal trauma: Relative fundamental frequency and listener perception. *Journal of Voice, 30*(2), 177–185. https://doi.org/10.1016/j.jvoice.2015.04.005
Heller Murray, E. S., Lien, Y. A. S., Van Stan, J. H., Mehta, D. D., Hillman, R. E., Noordzij, J. P., & Stepp, C. E. (2017). Relative fundamental frequency distinguishes between phonotraumatic and non-phonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 60*(6), 1507–1515. https://doi.org/10.1044/2016_JSLHR-S-16-0262
Izdebski, K. (1984). Overpressure and breathiness in spastic dysphonia: An acoustic (LTAS) and perceptual study. *Acta Oto-Laryngologica, 97*(3–4), 373–378. https://doi.org/10.3109/00016488409131003
Kay Elemetrics. (1993). *Multi-Dimensional Voice Program (MDVP) Model 4305: Operations manual.*
Langeveld, T. P., Drost, H. A., Frijns, J. H., Zwinderman, A. H., & Baatenburg de Jong, R. J. (2000). Perceptual characteristics of adductor spasmodic dysphonia. *Annals of Otology, Rhinology & Laryngology, 109*(8 Pt. 1), 741–748. https://doi.org/10.1177/000348940010900808
Lien, Y. A. S., Gattuccio, C. I., & Stepp, C. E. (2014). Effects of phonetic context on relative fundamental frequency. *Journal of Speech, Language, and Hearing Research, 57*(4), 1259–1267. https://doi.org/10.1044/2014_JSLHR-S-13-0158
Lien, Y. A. S., Michener, C. M., Eadie, T. L., & Stepp, C. E. (2015). Individual monitoring of vocal effort with relative fundamental frequency: Relationships with aerodynamics and listener perception. *Journal of Speech, Language, and Hearing Research, 58*(3), 566–575. https://doi.org/10.1044/2015_JSLHR-S-14-0194
Lien, Y. A. S., Murray, E. S. H., Calabrese, C. R., Michener, C. M., Van Stan, J. H., Mehta, D. D., Hillman, R. E., Noordzij, J. P., & Stepp, C. E. (2017). Validation of an algorithm for semi-automated estimation of voice relative fundamental frequency. *Annals of Otology, Rhinology & Laryngology, 126*(10), 712–716. https://doi.org/10.1177/0003489417728088
Ludlow, C. L., Domangue, R., Sharma, D., Jinnah, H. A., Perlmutter, J. S., Berke, G., Sapienza, C., Smith, M. E., Blumin, J. H., Kalata, C. E., Blindauer, K., Johns, M., Hapner, E., Harmon, A., Paniello, R., Adler, C. H., Crujido, L., Lott, D. G., Bansberg, S. F., . . . Stebbins, G. (2018). Consensus-based attributes for identifying patients with spasmodic dysphonia and other voice disorders. *JAMA Otolaryngology—Head & Neck Surgery, 144*(8), 657–665. https://doi.org/10.1001/jamaoto.2018.0644
McKenna, V. S., Heller Murray, E. S., Lien, Y. A. S., & Stepp, C. E. (2016). The relationship between relative fundamental frequency and a kinematic estimate of laryngeal stiffness in healthy adults. *Journal of Speech, Language, and Hearing Research, 59*(6), 1283–1294. https://doi.org/10.1044/2016_JSLHR-S-15-0406
McKenna, V. S., & Stepp, C. E. (2018). The relationship between acoustical and perceptual measures of vocal effort. *The Journal of the Acoustical Society of America, 144*(3), 1643. https://doi.org/10.1121/1.5055234
Moerman, M., Martens, J. P., & Dejonckere, P. H. (2015). Multidimensional assessment of strongly irregular voices such as

in substitution voicing and spasmodic dysphonia: A compilation of own research. *Logopedics, Phoniatrics, Vocology, 40*(1), 24–20. https://doi.org/10.3109/14015439.2014.936497

Murry, T. (2014). Spasmodic dysphonia: Let's look at that again. *Journal of Voice, 28*(6), 694–699. https://doi.org/10.1016/j.jvoice.2014.03.007

Murry, T., & Woodson, G. E. (1995). Combined-modality treatment of adductor spasmodic dysphonia with botulinum toxin and voice therapy. *Journal of Voice, 9*(4), 460–465. https://doi.org/10.1016/S0892-1997(05)80211-5

Nagle, K. F., Eadie, T. L., & Yorkston, K. M. (2015). Everyday listeners' impressions of speech produced by individuals with adductor spasmodic dysphonia. *Journal of Communication Disorders, 58,* 1–13. https://doi.org/10.1016/j.jcomdis.2015.07.001

Petrovsky, A. (n.d.). Halcyon. Retrieved March 25, 2017, from http://dsp.tut.su/halcyon.html

Roy, N., Fetrow, R. A., Merrill, R. M., & Dromey, C. (2016). Exploring the clinical utility of relative fundamental frequency as an objective measure of vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 59*(5), 1002–1017. https://doi.org/10.1044/2016_JSLHR-S-15-0354

Roy, N., Mauszycki, S. C., Merrill, R. M., Gouse, M., & Smith, M. E. (2007). Toward improved differential diagnosis of adductor spasmodic dysphonia and muscle tension dysphonia. *Folia Phoniatrica et Logopedica, 59*(2), 83–90. https://doi.org/10.1159/000098341

Roy, N., Mazin, A., & Awan, S. N. (2014). Automated acoustic analysis of task dependency in adductor spasmodic dysphonia versus muscle tension dysphonia. *Laryngoscope, 124*(3), 718–724. https://doi.org/10.1002/lary.24362

Sapienza, C. M., Walton, S., & Murry, T. (1999). Acoustic variations in adductor spasmodic dysphonia as a function of speech task. *Journal of Speech, Language, and Hearing Research, 42*(1), 127–140. https://doi.org/10.1044/jslhr.4201.127

Sapienza, C. M., Walton, S., & Murry, T. (2000). Adductor spasmodic dysphonia and muscular tension dysphonia: Acoustic analysis of sustained phonation and reading. *Journal of Voice, 14*(4), 502–520. https://doi.org/10.1016/S0892-1997(00)80008-9

Shoffel-Havakuk, H., Marks, K. L., Morton, M., Johns , M. M., III., & Hapner, E. R. (2019). Validation of the OMNI vocal effort scale in the treatment of adductor spasmodic dysphonia. *Laryngoscope, 129*(2), 448–453. https://doi.org/10.1002/lary.27430

Shrivastav, R., Sapienza, C. M., & Nandur, V. (2005). Application of psychometric theory to the measurement of voice quality using rating scales. *Journal of Speech, Language, and Hearing Research, 48*(2), 323–335. https://doi.org/10.1044/1092-4388(2005/022)

Siemons-Lühring, D. I., Moerman, M., Martens, J. P., Deuster, D., Müller, F., & Dejonckere, P. (2009). Spasmodic dysphonia, perceptual and acoustic analysis: Presenting new diagnostic tools. *European Archives of Oto-Rhino-Laryngology, 266*(12), 1915–1922. https://doi.org/10.1007/s00405-009-0995-7

Solomon, N. P., Helou, L. B., & Stojadinovic, A. (2011). Clinical versus laboratory ratings of voice using the CAPE-V. *Journal of Voice, 25*(1), e7–e14. https://doi.org/10.1016/j.jvoice.2009.10.007

Stepp, C. E. (2013). Relative fundamental frequency during vocal onset and offset in older speakers with and without Parkinson's disease. *The Journal of the Acoustical Society of America, 133*(3), 1637–1643. https://doi.org/10.1121/1.4776207

Stepp, C. E., Hillman, R. E., & Heaton, J. T. (2010). The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research, 53*(5), 1220–1226. https://doi.org/10.1044/1092-4388(2010/09-0234)

Stepp, C. E., Merchant, G. R., Heaton, J. T., & Hillman, R. E. (2011). Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 54*(5), 1260–1266. https://doi.org/10.1044/1092-4388(2011/10-0274)

Stewart, C. F., Allen, E. L., Tureen, P., Diamond, B. E., Blitzer, A., & Brin, M. F. (1997). Adductor spasmodic dysphonia: Standard evaluation of symptoms and severity. *Journal of Voice, 11*(1), 95–103. https://doi.org/10.1016/S0892-1997(97)80029-X

Talkin, D. (1995). A Robust Algorithm for Pitch Tracking (RAPT). In W. B. Kleijn & K. K. Paliwal (Eds.), *Speech coding and synthesis* (pp. 495–518). Elsevier.

Vojtech, J. M., Segina, R. K., Buckley, D. P., Kolin, K. R., Tardif, M. C., Noordzij, J. P., & Stepp, C. E. (2019). Refining algorithmic estimation of relative fundamental frequency: Accounting for sample characteristics and fundamental frequency estimation method. *The Journal of the Acoustical Society of America, 146*(5), 3184–3202. https://doi.org/10.1121/1.5131025

Yanagida, S., Nishizawa, N., Hashimoto, R., Mizoguchi, K., Hatakeyama, H., Homma, A., & Fukuda, S. (2018). Reliability and validity of speech evaluation in adductor spasmodic dysphonia. *Journal of Voice, 32*(5), 585–591. https://doi.org/10.1016/j.jvoice.2017.06.022

Zraick, R. I., Kempster, G. B., Connor, N. P., Thibeault, S., Klaben, B. K., Bursac, Z., Thrush, C. R., & Glaze, L. E. (2011). Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *American Journal of Speech-Language Pathology, 20*(1), 14–22. https://doi.org/10.1044/1058-0360(2010/09-0105)