Original Article

Annals of Otology, Rhinology & Laryngology
1–12

DOI: 10.1177/00034894251388829
journals.sagepub.com/home/aor
Sage

Observation of Laryngeal Postures via High-Speed Videoendoscopy in Adductor Laryngeal Dystonia

Katherine L. Marks, PhD[1,2], Jenny Vojtech, PhD[1], Manuel E. Díaz Cádiz, ME[1], Laura E. Toles, PhD[3], Taylor F. Feaster, BS[1], Andrew Wrede[1], Maanya Baranwal[1], Anastasia Haubrich[1], Misora Ito[1], Sophie Li[1], Bonnie Little[1], Samantha Lonergan, BS[1], Franklin Phan[1], Giavanna Siracusano[1], Alex Estrada, BS[1], Jose M. Rojas, BS[1], Daniel P. Buckley, MS[1,4], Cara L. Sauder, PhD[5], Andrew Keltz, MS[6], Michael Lerner, MD[6], Pavan Mallur, MD[7], Gregory Grillone, MD[4], Pieter Noordzij, MD[4], Lauren F. Tracy, MD[4], Ted Mau, MD[3], and Cara E. Stepp, PhD[1,4,8]

Abstract

Objective: Adductor laryngeal dystonia (AdLD) is a neurological disorder characterized by spasms of the laryngeal muscles during speech, resulting in acoustic discontinuities. High-speed videoendoscopy (HSV) has been used to capture supraglottic compression and visual obstructions of the vocal folds during sustained phonation in speakers with AdLD. The purpose of this study was to investigate the co-occurrence of these laryngeal postures via HSV and acoustic discontinuities in speakers with and without AdLD during connected speech.

Methods: Speakers with (n = 16) and without (n = 16) AdLD read sentences while undergoing trans nasal laryngoscopy. Microphone and HSV signals were simultaneously recorded. Acoustic discontinuities and laryngeal postures via HSV were independently manually labeled. A multinomial logistic regression was performed to examine the relationship between the presence of acoustic discontinuities—labeled as phonatory break, creak, or frequency shift—and that of view of vocal fold obstruction and supraglottic compression. Additional fixed factors included in the regression model were group (AdLD, control), sex, and the interactions of group × obstruction and group × supraglottic compression.

Results: Factors that were significantly associated with the odds of an acoustic discontinuity included group, vocal fold obstruction, supraglottic compression, and the interaction of group × vocal fold obstruction. There was not a significant interaction effect of group on supraglottic compression events.

Conclusions: These findings suggest that acoustic discontinuities co-occur with supraglottic compression and vocal fold obstructions. Understanding this relationship may lead to more effective voice assessment methods for individuals with AdLD.

Keywords

laryngoscopy, vocal cords, speech, dystonia, spasm

[1]Department of Speech, Language, and Hearing Sciences, Boston University, MA, USA
[2]Department of Otolaryngology, Emory University School of Medicine, Atlanta, GA, USA
[3]Department of Otolaryngology—Head and Neck Surgery, University of Texas Southwestern Medical Center, Dallas, USA
[4]Department of Otolaryngology—Head and Neck Surgery, Boston University School of Medicine, MA, USA
[5]Department of Speech and Hearing Sciences, University of Washington, Seattle, USA
[6]Department of Otolaryngology Surgery, Yale School of Medicine, Greenwich, CT, USA
[7]Department of Otolaryngology, Beth Israel Deaconess Medical Center, Boston, MA, USA
[8]Department of Biomedical Engineering, Boston University, MA, USA

Corresponding Author:
Katherine L. Marks, PhD, CCC-SLP, Department of Otolaryngology, Emory University School of Medicine, 550 Peachtree Street NW, Floor 11, Atlanta, GA 30308, USA.
Email: Katie.marks@emory.edu

Introduction

Adductor laryngeal dystonia (AdLD) is a neurological voice disorder in which involuntary spasms of the laryngeal muscles occur during speech, specifically on voiced sounds or during transitions between voiced and voiceless sounds. Laryngeal spasms are associated with irregularities in voice fundamental frequency—known as *discontinuities*—and manifest acoustically as phonatory breaks, frequency shifts, or creak.[1-4] Phonatory breaks, often called voice breaks, occur as an absence of fundamental frequency (f_o) in what would typically be a voiced segment (ie, not between words or during voiceless consonants).[1] Frequency shifts, also called pitch breaks, are abrupt changes in f_o in either direction that occur within 50 ms. Historically, aperiodicity was manually labeled in AdLD speech, defined as irregular timing of the glottal pulses.[2] Sapienza's figures of aperiodicity examples consisted of segments in which the f_o had no pattern. However, in our initial study investigating these manual labels in AdLD speech,[5] we observed fewer instances of segments that were truly aperiodic (ie, no pattern in f_o), but rather irregularly spaced pulses that were not without pattern. Instead of labeling segments of aperiodicity, we instead adopted the term "creak," as an umbrella term for several types of f_o irregularities. Keating et al,[6] describes the subtypes of creak as (1) prototypical creak: low rate of f_o, irregular f_o, and constricted glottis resulting in low glottal airflow; (2) vocal fry: low f_o with a constricted glottis but not necessarily irregular; (3) a multiply pulsed voice of alternating longer and shorter pulses or period doubling; (4) aperiodic voice: f_o is so irregular that there is no periodicity and no perceived pitch; (5) nonconstricted creak: low and irregular without a constricted glottis but rather a spread glottis with higher airflow, and (6) tense/pressed voice: the glottis is constricted, but the f_o is neither low nor irregular. We used the term creak to indicate any of these types of irregularities in f_o.[5] It should be noted that instances of phonatory breaks, frequency, shifts, and creak are not mutually exclusive. For example, creak can occur directly after a phonatory break, and a change from modal to creak f_o may meet the definition of a frequency shift if it occurs within 50 ms.

In AdLD speech, laryngeal spasms are presumed to co-occur with compensatory hyperfunction, as speakers inadvertently tense other muscles to compensate for the spasms. However, it is unclear whether laryngeal behaviors like supraglottic compression are compensatory or are a manifestation of the spasm itself. AdLD is task specific: spasms occur during speech but not during innate vocal behaviors like crying or laughing, and spasms are more likely to occur during connected speech than sustained phonation. Laryngeal spasms are thought to occur during vocal fold adduction in AdLD and are more likely to occur during voiced sounds or transitions to voiced sounds.[7-9] Currently, no clinically feasible instrumental measures have been validated for specificity in detecting features of AdLD. Although some acoustic signals may be helpful in differentiating AdLD from a more common voice disorder called muscle tension dysphonia (MTD), these measures require specific stimuli,[9,10] complex calculations,[11] or inaccessible software.[5,12] The only reliable method for identifying features of voice that are unique to AdLD is the manual labeling of acoustic discontinuities,[1-4,13] a time-consuming process that requires a trained technician to manually discern these events from the acoustic signal. Though it is the most reliable method, it is not clinically practical due to the amount of time that it takes. However, the manual labeling methodology can be used as a gold standard comparison for researchers who seek to validate new measures.[5]

Videostroboscopic imaging of the laryngeal mechanism is one of the primary methods used in voice clinics to detect or rule out pathology;[14] however, the technique cannot capture laryngeal spasms due to its limited framerate (30 frames per second),[15-19] which is too slow for the rapid timescale of these spasms.[20] As such, laryngeal spasms have been associated with relatively slow, aberrant postures such as vocal fold hyperadduction and supraglottic compression (eg, false vocal fold constriction and laryngeal constriction).[21-25] However, evidence of hyperfunctional laryngeal posturing has been observed via endoscopy in both AdLD and MTD.[21] The similarity of these patterns in MTD suggests that traditional laryngoscopy with or without videostroboscopy is insufficient for differentiating between AdLD and MTD. Clinical diagnoses of AdLD are thus performed by combining auditory-perceptual assessments and acoustic signal analyses with videostroboscopic findings, providing a more comprehensive diagnostic approach.[3,26,27]

Laryngeal high-speed videoendoscopy (HSV) has been applied in laboratory settings to study the supraglottic and vibratory aspects of voice production—such as the factors contributing to glottic insufficiency—as well as components related to vocal tremor at frame rates at or exceeding 4000 frames per second.[28] Specific to AdLD, HSV during sustained phonation has revealed evidence of supraglottic compression[28] and irregularities in vibratory components, such as oscillatory breaks and micromotions (ie, abrupt, minute aperiodic abductory motions of 1 vocal fold edge compared to the entire mass of the vocal folds) observed in the high speed digital imagining waveforms recorded with a rigid endoscope during sustained phonation.[29] However, because AdLD speech is more symptomatic in connected speech than sustained vowels, studying the vocal postures that occur during connected speech is ecologically important.

Traditionally, HSV has used rigid endoscopy, which is limited to sustained phonation, to capture cycle-by-cycle vocal fold motion. However, a flexible endoscope has been used during HSV to capture connected speech,[30-33] typically using somewhat slower frame rates. For example, a single case study validated a method of temporal segmentation for

videos recorded at 4000 frames per second.[33] Though not a main finding, the authors observed visual obstruction of the vocal folds by the epiglottis, leading to the removal of image frames in which the vocal folds were out of view. A subsequent study[32] used the same methodology to detect visual obstructions of the glottis by the epiglottis, arytenoid cartilage(s), false vocal folds, and/or laryngeal constriction in 4 speakers with AdLD and 3 speakers without voice disorders. Other studies have used even slower frame rates to quantify slower motions of abduction and adduction and glottal angles.[34-38] Although 500 to 1000 frames per second is too slow to capture the rapid features of the mucosal wave, such as micromotions, during vocal fold vibration, it is sufficient for capturing laryngeal postures, which occur on a timescale of 104 to 227 ms.[36]

More recently, supraglottic laryngeal maneuvers have been studied using HSV in 6 speakers with AdLD and 6 controls.[31] Specifically, automated detection of vocal fold obstruction was applied to identify segments that contained obstruction. Three labelers manually identified the supraglottic tissues that were responsible for the obstructions. They also calculated the occurrence and duration of the obstructions. Seven different types of obstruction were noted in their work: epiglottic, arytenoid, other supraglottal tissues, instances in which the vocal folds were out of endoscopic view, mediolateral compression, anteroposterior compression, and sphincteric compression. There were statistically greater percentages of obstructed frames between speakers with AdLD and controls. The obstructions also occurred for longer duration in speakers with AdLD than controls. The authors used an analysis that ranked variables in the order of their ability to distinguish between speaker groups and found that the most discriminative variable was the number of obstructions.[31] Their preliminary work offers some insight into the pathophysiology of laryngeal spasms in a small sample size; however, further work is needed in a larger sample.

No study to date has examined laryngeal postures in speakers with AdLD relative to the acoustic discontinuities that co-occur during laryngeal spasms during connected speech. Building on the few related studies in the literature,[29,31] the purpose of this study was to describe the laryngeal postures that co-occur with acoustic discontinuities in speakers with and without AdLD during connected speech. We hypothesized that the presence of acoustic discontinuities would be statistically related to the presence of supraglottic compression and obstruction of the view of the vocal folds, henceforth referred to as vocal fold obstructions.

Methods

As approved by the Boston University (BU) Institutional Review Board (# 2625) and the University of Texas Southwestern Medical Center (UTSW; #STU-2023-0965), participants in the study included 16 speakers diagnosed with AdLD and 16 speakers without known voice problems, henceforth referred to as "controls." All participants with AdLD had been formally diagnosed by a board-certified otolaryngologist based on consensus criteria[39] and participated in the study when they were in a symptomatic state per patient report. For speakers who were receiving botulinum toxin (BoNT) treatment, they were recorded just prior to their next routinely scheduled injection. All control speakers and 9 of 16 speakers with AdLD participated at BU in a laboratory setting, whereas 7 of 16 speakers with AdLD participated at UTSW in a clinical setting. Participants with AdLD were matched with controls based on sex and age within 5 years, as described in Table 1. Speakers in the control group reported no history or presence of voice or speech disorder, no current report of voice changes, no previous diagnosis of a voice disorder, and no previous participation in voice therapy. Those participating in the university laboratory setting (n = 25) received an aerosolized nasal decongestant without a topical anesthetic, and those participating in the clinical setting (n = 7) received an aerosolized nasal decongestant with topical lidocaine, approximately 10 minutes before the endoscopy.

Participants were seated in an examination chair in a sound-treated booth during simultaneous acoustic and HSV recordings. In both settings, an ear-mounted directional microphone was placed approximately 7 cm off-center from the lips. A flexible laryngeal endoscope (at BU: PENTAX Medical, Model FNL-7RP3, 2.4 mm; at UTSW: Olympus, Model ENF-P4, 3.6 mm) was passed through the nasal cavity to capture laryngoscopic images. The endoscope was attached to a camera (at BU: FASTCAM Mini AX100l, Model 540K-C-16GB; at UTSW: Phantom V311 [Vision Research]), both set to 256 × 256 pixels, and a steady xenon light source (at BU: 300-W KayPentax Model 7162B; at UTSW: 400-W Titan 400E). Video images were acquired at 1000 frames per second at BU and 500 frames per second at UTSW. HSV image acquisition was synchronized with microphone recordings. Participants read 2 sentences at a time for a total of 7 partition recordings. An eighth partition recording was made in which participants produced vowel + voiceless consonant + vowel syllable strings for future analysis. That partition was not analyzed in the current study. The total duration of nasoendoscopy was approximately 8 minutes, with the scope remaining in place for the entirety of the recordings.

Participants read aloud 14 unfamiliar complex sentences that were designed to elicit laryngeal spasms via syntactic and linguistic complexity[40] and allow for future analysis of relative fundamental frequency.[34] In a study by Erickson,[40] complex, center-embedded sentences comprised of mostly voiced sounds evoked the most acoustic signs of AdLD (phonatory breaks, frequency shifts, and aperiodicity).

Table 1. Participant Demographics.

	Control		AdLD	
Group	Sex (count)	Age (mean)	Sex (count)	Age (mean)
Female	11	64.5	11	64.5
Male	5	65.2	5	62.2

Therefore, we used complex, center-embedded sentences. For example, in the sentence, "The mushy banana that Annie feeds Nan amuses Macy every day," "Annie feeds Nan" is centered, increasing the linguistic complexity. Eight sentences were loaded with primarily voiced sounds and 6 sentences were loaded with primarily voiceless sounds. Table 2 displays sentence stimuli.

The acoustic files were digitized at 44.1 kHz. Following methods of Ludlow, Sapienza and colleagues[1-4,13] and our prior work,[5] trained technicians manually labeled acoustic independently labeled each file in Praat,[41] then met and reviewed signals to form consensus labels for any label in which there was disagreement. Acoustic discontinuity labels included instances of phonatory breaks, frequency shifts, and creak.[5] Based on the work of Sapienza and colleagues,[1-4,13] phonatory breaks were defined as an absence of voicing lasting 50 ms or longer in what would typically be a connected voice signal, and frequency shifts were defined as shifts of 50 Hz or more in the fundamental frequency (f_o) within 50 ms. Labels of creak were based on the definition of creak as an umbrella term by Keating et al[6]: low and irregular f_o, multiply pulsed voice, aperiodic voice, or tense/pressed voice. Though sometimes they occurred sequentially, the acoustic labels of phonatory breaks, frequency shifts, and creak did not overlap with one another. To account for pauses, labels were not made between words.[2] Labelers had the ability to listen to samples in addition to viewing the acoustic signal and spectrogram.

Independent from the acoustic labeling, 2 technicians used a custom script in MATLAB (version 9.3; The MathWorks, Natick, MA) to manually identify laryngeal postures of visual obstruction of the vocal folds and supraglottic compression in the HSV data during each of the 14 sentences spoken by each participant. The 14 sentences had been transcribed (and transcriptions were corrected by researchers if a word was misread, added, or omitted), and a p2fa forced text grid aligner[42] implemented in MATLAB displayed text boxes of words and syllables, as displayed in Figure 1 in the ARPABET and Word rows under the acoustic signal. Though not time aligned with the video playback, the labelers had the ability to listen to the auditory output of the samples as they were labeling features of the video.

Each technician was trained using 6 sample participants labeled by the senior authors, who then met with the senior authors to form consensus labels. Greater than 85% agreement with the senior authors was required before the trainees labeled real data. Postures were categorized as obstruction or supraglottic compression events, as visually identified from the HSV image frames (Figure 1). Frames were labeled as visual obstructions if more than two-thirds of the vocal folds were covered by the epiglottis or arytenoid cartilages such that underlying compression of the vocal folds could not be assessed. The direction (anteroposterior or mediolateral) and degree of compression (mild, moderate, or severe) were classified based on how much of the true vocal folds were covered by false vocal folds (ie, lateral constriction) and/or anteroposterior constriction ($<1/3$ was mild, 1/3-2/3 was moderate, and $>2/3$ was severe), a modification to the supraglottic activity section of the Voice-Vibratory Assessment with Laryngeal Imaging for high-speed endoscopy.[43] Labels of obstructions and compression were mutually exclusive: if a frame was obstructed by the epiglottis or arytenoid angle or was out of laryngoscope view such that compression could not be judged, the compression label was not made. Following their independent labeling of these postures, the technicians convened to review the video and labels, reaching consensus on any labels for which they disagreed.

A multinomial logistic regression was performed to examine the relationship between the presence of acoustic discontinuities—labeled as break, creak, frequency shift, or no discontinuity—and that of the laryngeal postures of vocal fold obstruction and supraglottic compression. For the purposes of this study, obstruction and supraglottic compression events were treated as binary variables (ie, present or not, regardless of degree and direction). Additional fixed factors included in the regression model were group (AdLD and control), and the interactions of group $\times$ obstruction and group $\times$ supraglottic compression. Log-likelihood ratio tests were used to examine the effects of the predictors on acoustic discontinuities at a significance level of $P < .05$. Post hoc Chi-square tests of independence were performed to examine significant interaction effects at a significance level of $P < .05$. Odds ratios were calculated based on model coefficients. Estimated marginal means were used to compare the probabilities of an acoustic event occurring between groups.

Results

Figure 2 illustrates the group descriptive data for the proportion of: A. observed acoustic discontinuity, B. obstruction, and C. supraglottic compression events when categorized by voicing status for each phoneme (ie, voiced or unvoiced). Across groups, these events occurred more frequently during the production of voiced phonemes, with supraglottic compression being observed most often (voiced: $63.7 \pm 21.7\%$, unvoiced: $49.6 \pm 22.6\%$), followed by obstruction events (voiced: $33.1 \pm 7.5\%$, unvoiced:

Table 2. Sentence Stimuli.

Voiced phoneme-loaded sentences	Voiceless phoneme-loaded sentences
The leafy tree where the dew shimmered was where she bravely married the man.	A shiny key with a true sheen opens a spiffy shoe case in her home.
The mushy banana that Annie feeds Nan amuses Macy every day.	The cheeky kid kicked his squeaky chew toy, so Sheila put the sticky tape to the tip of it.
My good buddy Goofy that used 80 shiny new shoe bins delivered wine 1 day.	Should she see Kasey put her feet on the sofa tonight, Fifi can tell Pete to pay a hefty fee for his sin.
The beefy meat that the wee teen enjoyed was with a goofy new fan of New England.	We feasted 2 cheeky teens sushi and tofu stew when she came home Saturday.
We ate messy eggs in New Orleans on greasy oatmeal in their shiny metal bowls.	Sixty-two soups found their place at her table with 2 fishy sides at her Sushi cart.
The girl Lisa that Mandy meets in the busy scene dreamed of owning blue shimmery nails again.	Tracy will pay for Lucy to fetch treats from the seaside with Stacey and Pete.
When we gave Vicky a key to the shack, the knife on the beam was a major deal.	
Murphy sees Mia near the new bin in Lisa's big, blue sailboat at dawn on Monday morning.	

31.1 ± 7.9%) then acoustic discontinuities (voiced: 14.1 ± 8.2%, unvoiced: 12.3 ± 7.8%).

The overall binomial logistic regression model was found to be statistically significant ($\chi^2(5)=722$, $P<.001$; Table 3). Table 4 displays predictions of log odds of an acoustic discontinuity per model coefficient. Results are described according to each model predictor below.

Group

Group was a significant predictor of acoustic discontinuities ($\chi^2(1)=23.57$; $P<.001$): acoustic discontinuities were 1.67 times more likely to occur in the AdLD group compared to the control group. The predicted probability of an acoustic discontinuity occurring was higher in the AdLD group (*Pr*=.18, 95% CI [0.17, 0.19]) than in the control group (*Pr*=.14, 95% CI [0.13, 0.15]).

Obstruction and Supraglottic Compression Events

Both the occurrence of visual obstruction events ($\chi^2(1)=217.25$; $P<.001$) and supraglottic compression events ($\chi^2(1)=165.05$; $P<.001$) were significant predictors of acoustic discontinuities. The presence of an obstruction event increased the odds of an acoustic discontinuity by 4.68 times, whereas a compression event increased these odds by 2.69 times.

Interaction Effects: Group With Obstruction and Supraglottic Compression Events

The interaction between group and visual obstruction events was statistically significant in predicting acoustic discontinuities ($\chi^2(1)=13.43$, $P<.001$; see Figure 3A). In this context, the presence of an obstruction event during an acoustic discontinuity was less likely to occur in the AdLD group compared to the control group (OR=0.62). When no obstruction event was present, the probability of an acoustic event occurring was higher in the AdLD group (*Pr*=.12, 95% CI [0.11, 0.13]) than in the control group (*Pr*=.07, 95% CI [0.06, 0.08]; see Figure 4A). However, when an obstruction event occurred, the probabilities of an acoustic event co-occurring were similar for AdLD (*Pr*=.28, 95% CI [0.26, 0.30]) and control (*Pr*=.27, 95% CI [0.24, 0.30]) groups.

The interaction between group and supraglottic compression events was not statistically significant (*P*=.804; see Figure 3B), and the odds of a supraglottic compression event occurring during an acoustic discontinuity were relatively similar between the 2 groups (OR=1.03). In the absence of a supraglottic compression event, the probability of an acoustic event occurring was higher in the AdLD group (*Pr*=.12, 95% CI [0.11, 0.13]) than in the control group (*Pr*=.09, 95% CI [0.08, 0.10]; see Figure 4B). When a supraglottic compression event occurred, the probability of an acoustic event was also higher in AdLD (*Pr*=.27, 95% CI [0.25, 0.29]) compared to controls (*Pr*=.22, 95% CI [0.20, 0.24]).

Discussion

This study was the first of our knowledge to describe the co-occurrence of acoustic discontinuities and laryngeal postures via HSV during connected speech in speakers with and without AdLD. Our results build on the work of Patel et al[29] and Naghibolhosseini et al[31] by describing the laryngeal postures that co-occur with acoustic discontinuities in speakers

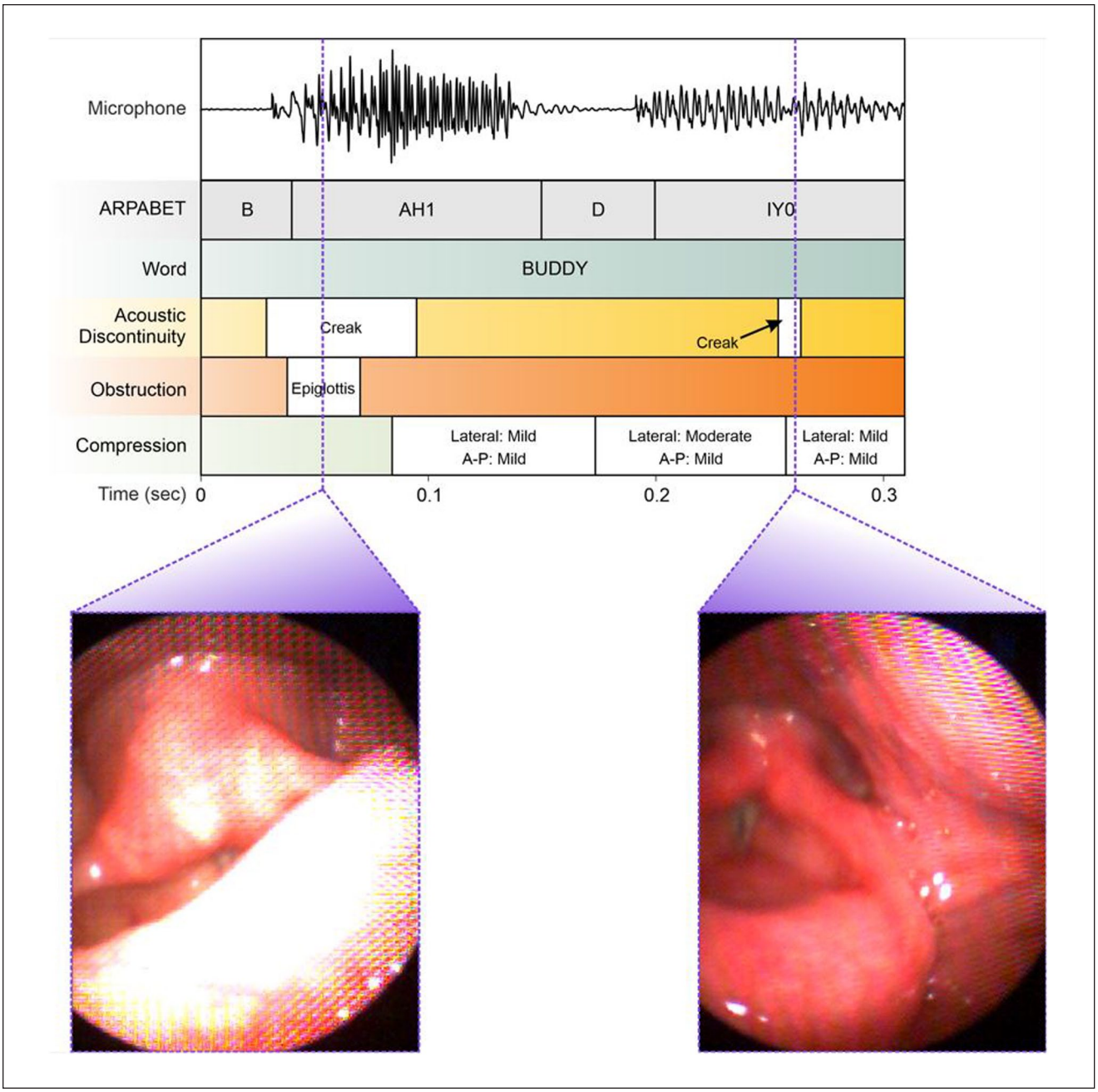

Figure 1. Example of technician-labeled time-aligned high-speed video (HSV) segments of the larynx. Top panel displays microphone signal, with ARPABET phoneme transcription and word underneath. Technicians labeled acoustic discontinuities (creak, frequency shifts, and phonatorty breaks) in the yellow tier, obstruction of the view of the vocal folds (epiglottis and arytenoids) in the orange tier, and compression (lateral, anterior-posterior as mild, moderate, or severe). Time-aligned corresponding HSV frames displayed below, with purple dotted lines indicating the instance of the frame in view. The left image is an example of the frame associated with a co-occurrence of creak and obstruction of the view by the epiglottis. The right image is an example of a frame associated with the co-occurrence of creak with mild lateral and anteroposterior supraglottic compression.

with AdLD and controls during connected speech. We hypothesized that the presence of acoustic discontinuities would be statistically related to the presence of supraglottic compression and visual obstructions of the vocal folds. Results confirmed this hypothesis in that both obstruction and compression postures observed via HSV were statistically related to the presence of an acoustic discontinuity, regardless of group.

As expected, there was a statistically significant interaction effect of group and obstruction events on the presence of

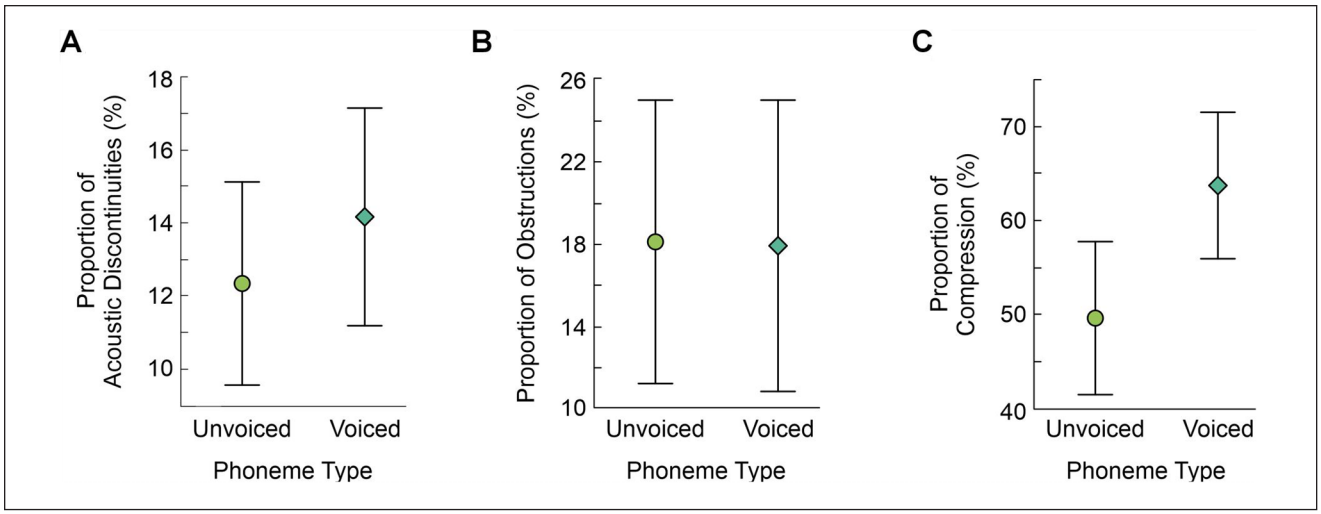

Figure 2. Descriptive results of the proportion in which an (A) acoustic discontinuity, (B) obstruction, or (C) compression event occurred by phoneme type: unvoiced displayed as circles and voiced displayed as diamonds.

Table 3. Likelihood Ratio Test for Acoustic Discontinuity Predictors.

Predictor	χ^2	*df*	*P*
Group	23.57	1	<.001
Obstruction event	217.25	1	<.001
Compression event	165.05	1	<.001
Group × obstruction event	13.43	1	<.001
Group × compression event	0.06	1	.804

acoustic discontinuities. However, the predicted probabilities of acoustic events co-occurring with obstruction events in AdLD and control groups were surprisingly close, suggesting the acoustic discontinuities may not actually reflect pathological events. Rather, obstruction events are assumed to occur either phonemically or in response to some kind of perturbation of the laryngeal system. Regardless of group, when the laryngeal system is not in an easy state, an acoustic discontinuity occurs. So, although we would not expect laryngeal spasms to occur in the control group, both groups appear to react to perturbations in the laryngeal system as evidenced by the presence of obstruction with acoustic discontinuities.

Interestingly, the key difference in acoustic discontinuities between groups emerged not when obstruction events occurred, but when they did not. Speakers with AdLD were significantly more likely to experience acoustic events in the *absence* of obstruction events compared to controls. In unperturbed laryngeal states, we would not expect control participants to have acoustic discontinuities, and our results support this. However, acoustic discontinuities occurred in the AdLD group even when the laryngeal system was not perturbed. This finding suggests that acoustic discontinuities still occur in AdLD even when there are not obvious laryngeal behaviors associated with them.

Contrary to our hypothesis, there was no statistically significant interaction effect of group and supraglottic compression events on the presence of acoustic discontinuities. However, it is important to note that by definition of the obstruction events, the view of the vocal folds was obstructed such that underlying compression could not be assessed. Therefore, it is possible that differences in supraglottic compression events existed when they occurred at the same time as obstruction events but were unable to be captured via laryngeal endoscopy and therefore unquantifiable.

Based on what was observable, the unexpected result of similar co-occurrence of acoustic discontinuities and supraglottic compression by group may be attributed to several factors. First, the presence of the endoscope in the nasal cavity and pharynx during speech could have influenced participants' vocal postures. In the control group, the scope was placed without a nasal topical anesthetic; it is not surprising that these individuals exhibited increased supraglottic compression as a physiological response to the sensory input of the scope (or a perturbed laryngeal system). Only the participants with AdLD who enrolled at a clinic received topical anesthetic; however, a *post hoc* analysis revealed there were no statistically significant differences in obstructions or compression between the those in the AdLD group who did and did not receive the topical anesthetic. The extended duration of the HSV procedure (approximately 8 minutes) may have induced discomfort (another example of a perturbed laryngeal system), potentially contributing to atypical voice productions even in controls. Our findings in controls align with previous studies reporting supraglottic compression during endoscopic examination in individuals without voice disorders.[44-47] However, Naghibolhosseini et al[31] did find statistic differences in sphincteric compression between 6 speakers with AdLD and 6 controls. It was not reported whether their

Table 4. Model Coefficients Predicting Log Odds of an Acoustic Discontinuity.

Predictor	Estimate	SE	*z*	*P*	Odds ratio
Intercept	−3.05	0.07	−40.73	<.001	0.05
Group (AdLD:Control)	0.52	0.11	4.85	<.001	1.67
Obstruction event (1:0)	1.54	0.10	15.22	<.001	4.69
Compression event (1:0)	0.99	0.08	11.98	<.001	2.69
Group × obstruction event (AdLD:Control × 1:0)	−0.48	0.13	−3.68	<.001	0.62
Group × compression event (AdLD:Control × 1:0)	0.03	0.11	0.25	.80	1.03

Note. Parentheses indicate binary comparisons, with second variable listed as reference level ("0" for absence or control group).

participants received topical anesthetic or how long the endoscope was in place.

In the current study, it is possible that some participants with AdLD who received botiumlinum toxin (BoNT) treatment scheduled their repeat BoNT injections in a duration that prevented them from becoming severely symptomatic; however, they were asked to schedule the study visit just prior to their next injection to capture their most symptomatic state, and all participants self-reported being in a symptomatic state. The mean age of participants in each group was approximately 64 years, so it is possible some participants in either group may have exhibited presbylarynx. This was also a limitation of the Naghibolhosseini et al[31] study. Therefore, it is possible that compensatory hyperfunction existed in those with presbylarynx related to glottic insufficiency. In future work, participant age cut off could be lower to avoid such interactions; however, given that AdLD is a rare disorder that occurs most frequently in middle aged and older adults, this may pose a challenge for achieving large enough sample sizes. Alternatively, glottal closure could be systematically assessed in future related work. Finally, the presence of the scope may have acted as a sensory trick (geste antagoniste) for speakers with AdLD, potentially reducing symptomatic speech patterns.[48,49] This phenomenon, although not quantified in our study, could have influenced our results and merits further investigation.

Other limitations to the study include no explicit procedure in the manual labeling to account for gagging induced by the scope. However, gagging was very rarely observed during the procedure. The sentences used in the study were specifically designed to allow for analysis of relative fundamental frequency, an acoustic measure, in future related work. They were also designed to invoke spasms based on syntactic complexity and phoneme specificity.[40] Because the sentences included vowels other than /i/, supraglottic compression caused by production of other vowels were possible and not accounted for during manual labeling procedures. As such, future investigations are warranted.

The results of this study offer preliminary insights into the physiological manifestations that co-occur with acoustic discontinuities in speakers with and without AdLD. Group, vocal fold obstructions, supraglottic compression, and the interaction of group × vocal fold obstruction were all statistically related to the presence of an acoustic discontinuity. These findings highlight several areas for future research, which may quantify the direction and extent of supraglottic compression associated with discontinuities, delineate physiological features associated with the subtypes of acoustic discontinuities, and investigate differences between AdLD and MTD. Investigating differences between voiced and voiceless phoneme-loaded sentences may offer additional insights into distinguishing laryngeal postures. Future research would additionally benefit from larger sample sizes, longitudinal designs, and standardized endoscopic procedures to address potential confounding factors. Further investigation is needed to determine whether certain types of acoustic discontinuities, such as phonatory breaks, are uniquely associated with visual obstructions and compression. Pairing detailed auditory-perceptual analysis with both labels of acoustic discontinuities and laryngeal postures may also validate perceptual ratings with underlying physiology in AdLD. Moreover, investigation of laryngeal postures and/or acoustic discontinuity pattern or timing analysis may provide greater insights into the underlying pathophysiology of laryngeal spasms. Such investigations could enhance our understanding of the underlying mechanisms contributing to AdLD, potentially improving diagnostic accuracy, and informing more targeted treatment approaches.

Conclusion

Our study reveals important differences in the manifestation of acoustic discontinuities between individuals with AdLD and those with typical voices. Speakers with AdLD demonstrated a higher likelihood of aberrant laryngeal postures coinciding with acoustic discontinuities compared to controls. Our findings suggest that acoustic discontinuities characterized by phonatory breaks, sudden shifts in fundamental frequency, or creak may be physiologically linked to vocal fold obstruction and/or instances of supraglottic compression. This relationship appears to be more pronounced in AdLD, highlighting the complex interplay between laryngeal physiology and acoustic outcomes in this disorder. Further research is warranted to explore the direction and extent of supraglottic compression associated with specific types of acoustic discontinuities. Such investigations could

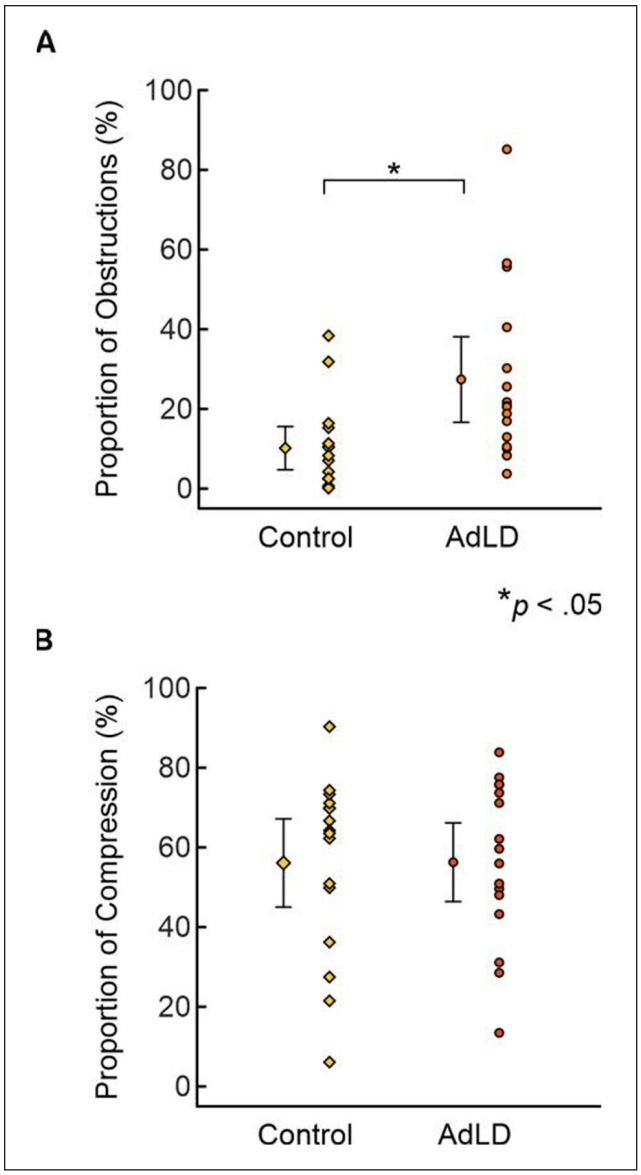

Figure 3. Proportion of (A) obstruction of the view of the vocal folds and (B) supraglottic compression events by group. The proportions (or frequencies) of obstruction and compression events were calculated by averaging each participant's event rates. Since both variables were binary and participants produced the same stimuli, this was done by dividing the number of events by the total number of observations (phonemes) for each participant. The mean of controls (yellow diamonds) with 95% confidence intervals are displayed on the left of each panel with the distribution of individual participant means to the right, and the mean of the AdLD group (orange circles) with 95% confidence intervals are displayed on the right of each panel with the distrubtion of individual participant means to the right.

enhance our understanding of the underlying mechanisms of AdLD, potentially leading to more comprehensive assessments and tailored interventions for individuals with this challenging voice disorder.

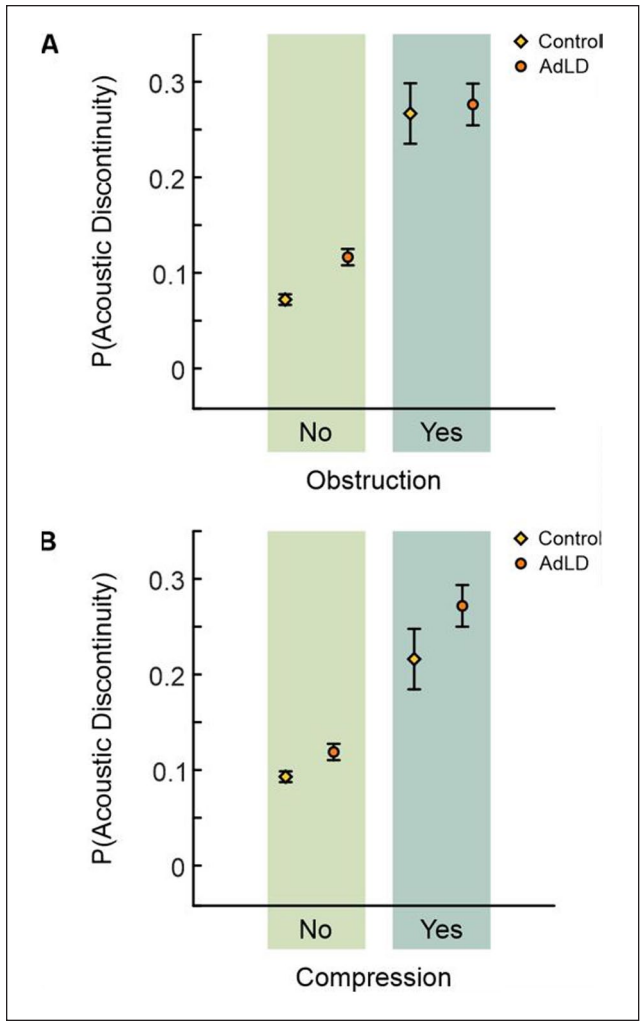

Figure 4. Likelihood of acoustic discontinuity based on group and an (A) obstruction event or a (B) supraglottic compression event. Since both variables were binary and participants produced the same stimuli, this was done by dividing the number of events by the total number of observations (phonemes) for each participant. The mean of controls (yellow diamonds) with 95% confidence intervals are displayed on the left of each panel, and the mean of the AdLD group (orange circles) with 95% confidence intervals are displayed on the right of each panel. The *x*-axis indicates the binary of whether an obstruction or compression event occurred (light green shaded panel) or not (dark green shaded panel).

Acknowledgments

The authors acknowledge the lab managers, research fellows, and clinical recruitment assistants who assisted in the recruitment, screening, and scheduling of participants.

Author Note

Portions of the contents of this manuscript were presented at the International Conference on Voice Physiology and Biomechanics, July 22, 2024, Erlangen, Germany.

Ted Mau is now affiliated to Department of Otolaryngology - Head and Neck Surgery, Stanford University, Redwood, CA, USA.

ORCID iDs

Katherine L. Marks https://orcid.org/0000-0001-5483-3794
Laura E. Toles https://orcid.org/0000-0002-2665-3283
Cara L. Sauder https://orcid.org/0000-0001-9218-3737
Michael Lerner https://orcid.org/0000-0002-8715-4533
Pavan Mallur https://orcid.org/0000-0001-8086-0534
Ted Mau https://orcid.org/0000-0002-1804-6707
Cara E. Stepp https://orcid.org/0000-0002-8045-252X

Ethical Considerations

This study was approved by the Boston University (BU) Institutional Review Board (# 2625) and the University of Texas Southwestern Medical Center (UTSW; #STU-2023-0965).

Consent to Participate

As approved by the Boston University (BU) Institutional Review Board (# 2625) and the University of Texas Southwestern Medical Center (UTSW; #STU-2023-0965), all participants provided informed consent prior to participating in the study. All participants provided written informed consent prior to participating in the study, as governed by the BU and UTSW Institutional Review Boards.

Consent for Publication

Not applicable.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institute on Deafness and Other Communication Disorders (DC015570 awarded to Dr. Cara Stepp and DC020349 awarded to Dr. Katherine Marks), Boston University Sargent College Student Research Grant, and the American Speech Language Hearing Foundation Speech Science Grant (awarded to Dr. Marks).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

As voice recordings are protected health information, data is not publicly available.

References

1. Ludlow CL, Naunton RF, Sectary SE, Schulz GM, Hallett M. Effects of botulinum toxin injections on speech in adductor spasmodic dysphonia. *Neurology*. 1988;38(8):1220-1220. doi:10.1212/WNL.38.8.1220
2. 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(1):127-140. doi:10.1044/jslhr.4201.127
3. Sapienza CM, Murry T, Brown WS. *Variations in Adductor Spasmodic Dysphonia: Acoustic Evidence*. Vol 12. 1998:214-222.
4. Sapienza CM, Walton S, Murry T. Adductor spasmodic dysphonia and muscular tension dysphonia: acoustic analysis of sustained phonation and reading. *J Voice*. 2000;14(4):502-520.
5. Marks KL, Stepp CE, Feaster TF, Baker S, Díaz-Cádiz ME, Doyle PC. Spectral aggregate of the high-passed fundamental frequency and its relationship to the primary acoustic features of Adductor Laryngeal Dystonia. *J Speech Lang Hear Res*. 2022;65(11):4085-4095. doi:10.1044/2022_JSLHR-22-00157
6. Keating P, Garellek M, Kreiman J. Acoustic properties of different kinds of creaky voice. *ICPhS*. 2015;1:2-7.
7. Ludlow CL, Yamashita T, Schulz GM, Deleyiannis FWB. Abnormalities in long latency responses to superior laryngeal nerve stimulation in adductor spasmodic dysphonia. *Ann Otol Rhinol Laryngol*. 1995;104(12):928-935. doi:10.1177/000348949510401203
8. Roy N, Gouse M, Mauszycki SC, Merrill RM, Smith ME. Task specificity in adductor spasmodic dysphonia versus muscle tension dysphonia. *Laryngoscope*. 2005;115(2):311-316. doi:10.1097/01.mlg.0000154739.48314.ee
9. Roy N, Mazin A, Awan SN. Automated acoustic analysis of task dependency in adductor spasmodic dysphonia versus muscle tension dysphonia. *Laryngoscope*. 2014;124(3):718-724. doi:10.1002/lary.24362
10. Houtz DR, Roy N, Merrill RM, Smith ME. Differential diagnosis of muscle tension dysphonia and adductor spasmodic dysphonia using spectral moments of the long-term average spectrum. *Laryngoscope*. 2010;120(4):749-757. doi:10.1002/lary.20741
11. Roy N, Awan SN, Jennings S, Jensen J, Merrill RM. Adductor Laryngeal Dystonia versus Muscle Tension Dysphonia: examining the utility of automated acoustic analysis to detect task dependency as a distinguishing feature. *J Speech Lang Hear Res*. 2024;67(10):3612-3630. doi:10.1044/2024_JSLHR-24-00104
12. Marks KL, Díaz Cádiz ME, Toles LE, et al. Automated creak differentiates Adductor Laryngeal Dystonia and Muscle Tension Dysphonia. *Laryngoscope*. 2023;133(10):2687-2694. doi:10.1002/lary.30588
13. Sapienza CM, Cannito MP, Murry T, Branski R, Woodson G. Acoustic variations in reading produced by speakers With Spasmodic Dysphonia pre-Botox injection and within early stages of post-Botox injection. *J Speech Lang Hear Res*. 2002;45(5):830-843. doi:10.1044/1092-4388(2002/067)
14. Patel RR, Awan SN, Barkmeier-Kraemer J, et al. Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function. *Am J Speech Lang Pathol*. 2018;27(3):887-905. doi:10.1044/2018_AJSLP-17-0009
15. Deliyski DD, Petrushev PP, Bonilha HS, Gerlach TT, Martin-Harris B, Hillman RE. Clinical implementation of laryngeal

high-speed videoendoscopy: challenges and evolution. *Folia Phoniatr Logop*. 2008;60(1):33-44. doi:10.1159/000111802
16. Mehta DD, Hillman RE. Current role of stroboscopy in laryngeal imaging. *Curr Opin Otolaryngol Head Neck Surg*. 2012;20(6):429-436.
17. Olthoff A, Woywod C, Kruse E. Stroboscopy versus high-speed glottography: a comparative study. *Laryngoscope*. 2007;117(6):1123-1126. doi:10.1097/MLG.0b013e318041f70c
18. Patel R, Dailey S, Bless D. Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders. *Ann Otol Rhinol Laryngol*. 2008;117(6):413-424. doi:10.1177/000348940811700603
19. Kunduk M, Doellinger M, McWhorter AJ, Lohscheller J. Assessment of the variability of vocal fold dynamics within and between recordings with high-speed imaging and by phonovibrogram. *Laryngoscope*. 2010;120(5):981-987. doi:10.1002/lary.20832
20. Hillel AD. The study of laryngeal muscle activity in normal human subjects and in patients with Laryngeal Dystonia using multiple fine-wire electromyography. *Laryngoscope*. 2001;111(S97):1-47.
21. Leonard R, Kendall K. Differentiation of spasmodic and psychogenic dysphonias with phonoscopic evaluation. *Laryngoscope*. 1999;109(2):295-300.
22. Parnes SM, Lavorato AS, Myers EN. Study of spastic dysphonia using videofiberoptic laryngoscopy. *Ann Otol Rhinol Laryngol*. 1978;87(3):322-326. doi:10.1177/000348947808700305
23. Woodson GE, Zwirner P, Murry T, Swenson M. Use of flexible fiberoptic laryngoscopy to assess patients with spasmodic dysphonia. *J Voice*. 1991;5(1):85-91.
24. Woodson E, Zwirner S, Murry T, Swenson MR. Functional assessment of patients with Spasmodic Dysphonia. *J Voice*. 1992;6(4):338-343.
25. Zwirner P, Murry T, Swenson MJ, Woodson GE. Acoustic changes in Spasmodic Dysphonia after botulinum toxin injection. *J Voice*. 1991;5(1):78-84.
26. Bangayan P, Long C, Alwan AA, Kreiman J, Gerratt BR. Analysis by synthesis of pathological voices using the Klatt synthesizer. *Speech Commun*. 1997;22(4):343-368. doi:10.1016/S0167-6393(97)00032-0
27. Zwirner P, Murry T, Woodson GE. Perceptual-acoustic relationships in spasmodic dysphonia. *J Voice*. 1993;7(2):165-171.
28. Parker LA, Kunduk M, Fink DS, McWhorter A. Reliability of high-speed videoendoscopic ratings of Essential Voice Tremor and Adductor Spasmodic Dysphonia. *J Voice*. 2019;33(1):16-26. doi:10.1016/j.jvoice.2017.10.009
29. Patel RR, Liu L, Galatsanos N, Bless DM. Differential vibratory characteristics of Adductor Spasmodic Dysphonia and Muscle Tension Dysphonia on high-speed digital imaging. *Ann Otol Rhinol Laryngol*. 2011;120(1):21-32.
30. Woo P, Baxter P. Flexible fiber-optic high-speed imaging of vocal fold vibration: a preliminary report. *J Voice*. 2017;31(2):175-181.
31. Naghibolhosseini M, Henry TM, Zayernouri M, Zacharias SR, Deliyski DD. Supraglottic laryngeal maneuvers in adductor laryngeal dystonia during connected speech. *J Voice*. Published online August 30, 2024. doi:10.1016/j.jvoice.2024.08.009.
32. Yousef AM, Deliyski DD, Zacharias SRC, Naghibolhosseini M. Detection of vocal fold image obstructions in high-speed videoendoscopy during connected speech in Adductor Spasmodic Dysphonia: a convolutional neural networks approach. *J Voice*. 2024;38(4):951-962. doi:10.1016/j.jvoice.2022.01.028
33. Naghibolhosseini M, Deliyski DD, Zacharias SRC, de Alarcon A, Orlikoff RF. Temporal segmentation for laryngeal high-speed videoendoscopy in connected speech. *J Voice*. 2018;32(2):256.e1-256.e12. doi:10.1016/j.jvoice.2017.05.014
34. McKenna VS, Murray ESH, Lien YAS, Stepp CE. The relationship between relative fundamental frequency and a kinematic estimate of laryngeal stiffness in healthy adults. *J Speech Lang Hear Res*. 2016;59(6):1283-1294. doi:10.1044/2016_JSLHR-S-15-0406
35. Diaz-Cadiz M, McKenna VS, Vojtech JM, Stepp CE. Adductory vocal fold kinematic trajectories during conventional versus high-speed videoendoscopy. *J Speech Lang Hear Res*. 2019;62(6):1685-1706. doi:10.1044/2019_JSLHR-S-18-0405
36. Dailey SH, Kobler JB, Hillman RE, et al. Endoscopic measurement of vocal fold movement during adduction and abduction. *Laryngoscope*. 2005;115(1):178-183. doi:10.1097/01.mlg.0000150701.46377.df
37. Park Y, Wang F, Díaz-Cádiz M, Vojtech JM, Groll MD, Stepp CE. Vocal fold kinematics and relative fundamental frequency as a function of obstruent type and speaker age. *J Acoust Soc Am*. 2021;149(4):2189-2199. doi:10.1121/10.0003961
38. Vojtech JM, Cilento DD, Luong AT, et al. Acoustic identification of the voicing boundary during intervocalic offsets and onsets based on vocal fold vibratory measures. *Appl Sci*. 2021;11(9):3816. doi:10.3390/app11093816
39. 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(8):657-665. doi:10.1001/jamaoto.2018.0644
40. Erickson ML. Effects of voicing and syntactic complexity on sign expression in Adductor Spasmodic Dysphonia. *Am J Speech Lang Pathol*. 2003;12(4):416-424. doi:10.1044/1058-0360(2003/087)
41. Boersma P, Weenik D. Praat: doing phonetics by computer (Version 5.3. 39); 2013.
42. Yuan J, Liberman M. Speaker identification on the SCOTUS corpus. *J Acoust Soc Am*. 2008;123(5):3878.
43. Poburka BJ, Patel RR, Bless DM. Voice-vibratory assessment with laryngeal imaging (VALI) form: reliability of rating stroboscopy and high-speed videoendoscopy. *J Voice*. 2017;31(4):513.e1-513.e14. doi:10.1016/j.jvoice.2016.12.003

44. Behrman A, Dahl LD, Abramson AL, Schutte HK. Anterior-posterior and medial compression of the supraglottis: signs of nonorganic dysphonia or normal postures? *J Voice*. 2003;17(3):403-410.
45. Mayerhoff RM, Guzman M, Jackson-Menaldi C, et al. Analysis of supraglottic activity during vocalization in healthy singers: supraglottic activity in singers. *Laryngoscope*. 2014;124(2):504-509. doi:10.1002/lary.24310
46. Stager SV, Bielamowicz SA, Regnell JR, Gupta A, Barkmeier JM. Supraglottic activity. *J Speech Lang Hear Res*. 2000;43(1):229-238. doi:10.1044/jslhr.4301.229
47. Stager SV, Neubert R, Miller S, Regnell JR, Bielamowicz SA. Incidence of supraglottic activity in males and females: a preliminary report. *J Voice*. 2003;17(3):395-402.
48. Khosravani S, Mahnan A, Yeh IL, et al. Laryngeal vibration as a non-invasive neuromodulation therapy for spasmodic dysphonia. *Sci Rep*. 2019;9(1):17955. doi:10.1038/s41598-019-54396-4
49. Young VVN, Kidane J, Gochman GE, Bracken DJ, Ma Y, Rosen CA. Abnormal laryngopharyngeal sensation in adductor laryngeal dystonia compared to healthy controls. *Laryngoscope*. 2023;133(9):2271-2278. doi:10.1002/lary.30462