The Laryngoscope

Automated Creak Differentiates Adductor Laryngeal Dystonia and Muscle Tension Dysphonia

Katherine L. Marks, PhD; Manuel E. Díaz Cádiz, MS; Laura E. Toles, PhD; Daniel P. Buckley, MS; Lauren F. Tracy, MD; J. Pieter Noordzji, MD; Gregory A. Grillone, MD; Cara E. Stepp, PhD

Objective: The purpose of this study was to determine whether automated estimates of vocal creak would differentiate speakers with adductor laryngeal dystonia (AdLD) from speakers with muscle tension dysphonia (MTD) and speakers without voice disorders.

Methods: Sixteen speakers with AdLD, sixteen speakers with MTD, and sixteen speakers without voice disorders were recorded in a quiet environment reading aloud a standard paragraph. An open-source creak detector was used to calculate the percentage of creak (% creak) in each of the speaker's six recorded sentences.

Results: A Kruskal-Wallis one-way analysis of variance revealed a statistically significant effect of group on the % creak with a large effect size. Pairwise Wilcoxon tests revealed a statistically significant difference in % creak between speakers with AdLD and controls as well as between speakers with AdLD and MTD. Receiver operating characteristic curve analyses indicated that % creak differentiated AdLD from both controls and speakers with MTD with high sensitivity and specificity (area under the curve statistics of 0.94 and 0.86, respectively).

Conclusion: Percentage of creak as calculated by an automated creak detector may be useful as a quantitative indicator of AdLD, demonstrating the potential for use as a screening tool or to aid in a differential diagnosis.

Key Words: acoustics, creak, laryngeal dystonia, muscle tension dysphonia, speech-language pathology, voice disorders.

Level of Evidence: 3

***Laryngoscope*, 00:1–8, 2023**

INTRODUCTION

Laryngeal dystonia (LD) is a neurological condition in which the intrinsic muscles of the larynx involuntarily spasm during phonation.[1] The most common type of LD is adductor laryngeal dystonia (AdLD),[2] in which spasms occur in the muscles that adduct the vocal folds during phonation, resulting in auditory perceptual features of strain, roughness, asthenia, and vocal fry.[3–5] Laryngeal spasms also result in related LD speech discontinuities, such as phonatory breaks, frequency shifts, and aperiodicity[6–9] or creak.[10] These symptoms of AdLD negatively impact communication effectiveness,[11] participation, and quality of life for individuals with the disorder.[12,13]

Unfortunately, several barriers exist that make it difficult for individuals with AdLD to receive appropriate treatment in a timely manner. Individuals seeking treatment for AdLD report seeing 3–4 providers over 4–5 years before receiving an accurate diagnosis and appropriate treatment options.[14] This is, in part, because primary care physicians, general otolaryngologists, and even neurologists may not be familiar with the primary signs of the disorder. In fact, it can be difficult even for laryngologists to diagnose,[14] as the primary signs of AdLD may present similarly to those of muscle tension dysphonia (MTD): a functional disorder characterized by excessive perilaryngeal musculoskeletal activity during phonation that occurs in the absence of concurrent pathology.[15–17] Moreover, AdLD is diagnosed almost exclusively using auditory-perceptual judgments, leading to subjective bias from raters.[18] AdLD is task-specific, in that the frequency and severity of AdLD signs may vary depending on the demands of the voice task.[18–20] As such, task-specific stimuli have been used to improve the discrimination of the signs of AdLD, for example, using sustained vowels and connected speech,[20] voiced and voiceless phoneme-loaded sentences,[21,22] or whispered speech and connected speech.[21,23–25] For individuals with AdLD, there is a difference in signs of LD between these types of stimuli, whereas, for individuals with MTD, the signs are consistent between stimuli. Thus, task specificity is used to help clinicians determine a differential diagnosis. However, voice evaluations do not always include specialized stimuli; task-specific stimuli are often only employed when AdLD is already suspected.[21]

From the Department of Speech, Language, and Hearing Sciences (K.L.M., M.E.D.C., D.P.B., C.E.S.), Boston University, Boston, Massachusetts, U.S.A.; Department of Otolaryngology – Head and Neck Surgery (L.E.T.), University of Texas Southwestern Medical Center, Dallas, Texas, U.S.A.; Department of Otolaryngology – Head and Neck Surgery (D.P.B., L.F.T., J.P. N., G.A.G., C.E.S.), Boston University School of Medicine, Boston, Massachusetts, U.S.A.; and the Department of Biomedical Engineering (C.E.S.), Boston University, Boston, Massachusetts, U.S.A.

Editor's Note: This Manuscript was accepted for publication on January 11, 2023.

Contents of this manuscript were presented at the Fall Voice Conference, San Francisco, CA, USA on October 7, 2022.

Send correspondence to
Katherine L. Marks, PhD, CCC-SLP, 677 Beacon St. Boston, MA 02115.
Email: kmarks6@bu.edu

DOI: 10.1002/lary.30588

No clinically feasible quantitative measures exist that are sensitive and specific to the primary signs of AdLD (i.e., laryngeal spasms) that do not require specific stimuli (vowels compared to voiced words or selected voiced words).[26,27] Manually identified (subjective) instances of LD discontinuities are sensitive and specific to AdLD,[6–9,25] but due to the time-consuming nature of the methodology, these measures are not clinically feasible. In a series of studies, Sapienza and colleagues manually identified instances of phonatory breaks, frequency shifts, and aperiodicity in the acoustic signals of speakers with AdLD.[6–9] They found that the most frequently occurring (%) type of discontinuity in LD speakers in vowels was aperiodicity, which they defined as a segment consisting of non-repetitive cycles.[6,8,9] Interestingly, in two of their studies, they found aperiodicity to be the least common LD discontinuity type in 15 voiced words selected from the rainbow passage,[8,9,28] but in another study using the same methods and stimuli, aperiodicity was the predominant event produced by individuals with AdLD both pre- and post-Botox during reading.[7] In a stepwise discriminant function analysis comparing speakers with AdLD to speakers without voice disorders, the number of manually identified aperiodic segments accounted for the greatest proportion of the variance in a model that differentiated between individuals with AdLD before treatment and controls.[7]

Cepstral-spectral acoustic measures are commonly used to evaluate voice disorders. One such measure is the Cepstral Spectral Index of Dysphonia (CSID), which has been found to vary with overall severity across voice disorders.[20,29,30] Roy, Mazin, and Awan[20] found that the CSID calculated as the difference in CSID from connected speech and vowels differentiated between AdLD and MTD groups with 67% sensitivity and 64% specificity, indicating acceptable discrimination. This finding was comparable to the discriminative validity of subjective auditory-perceptual evaluations found by the same group.[31] In another study, the long-term average spectrum, calculated from an all-voiced sentence, also demonstrated acceptable discrimination between AdLD and MTD.[32] Although these measures can be used to differentiate AdLD and MTD, they are not *specific* to the primary signs of LD; the CSID and LTAS have been used to describe other types of voice disorders as well.[33–36] Moreover, the CSID relied on task-specificity (i.e., the difference in CSID of connected speech and vowels), and the LTAS relied on an all-voiced sentence to maximize the potential symptoms elicited in speakers with AdLD.

Our recent work aimed to establish concurrent validity of a new acoustic measure, spectral aggregate of the high-passed fundamental frequency contour (SAHf_o) for speakers with AdLD by comparing SAHf_o to the % LD discontinuities.[10] Our methodology for manually labeling the discontinuities was similar to that of Sapienza and others[6–9,37]; however, through our labeling, we noticed many instances that did not strictly meet the criteria of either aperiodicity or frequency shifts but were not typical modal phonation. Instead of aperiodicity, we adopted the umbrella term "creak," based on Keating et al.'s description of the acoustic features of different types of creaky voices.[38] Keating et al. defined the prototypical creaky voice as having a low fundamental frequency (f_o), irregular f_o, and a constricted glottis with a small peak opening, long closed phase, and low glottal airflow. Out of the manual labels examined (phonatory breaks, frequency shifts, and creak), we found that creak was the most observed label in the speakers with AdLD. However, this observation alone is not sufficient to support the use of creak clinically, because creak occurs even in speakers with typical voices.[39] Further, in typical speakers, creak can be linguistically driven,[40] suggesting that the specific stimuli spoken may be a factor in how much creak is present. This finding and the associated limitations led to our current research question: In a new dataset, can creak differentiate between the connected speech of individuals with AdLD from individuals with muscle tension dysphonia and controls with typical voices?

As an alternative to the time-intensive task of manually labeling instances of creak, creak detectors have been developed to automatically detect instances of creak in running speech.[39,41–48] One such approach captures the glottal pulse duration associated with a creaky voice,[47,49,50] which considers the physiological and acoustic features associated with a creak. This methodology has been improved upon with several iterations.[47–49,51] Drugman, Kane, and Gobl[41] eventually implemented a neural network that outperformed their original model. The algorithm for the neural network creak detector is available open-source [Covarep] (v1.3.2), and the neural network version of the creak detector has now been considered the current state of the art for creak detection.[52] Automatic creak detection could be used to provide a readily available and fast option for assessment in AdLD.

Specifically, the automated creak detector[53] is the result of an artificial neural network model that was trained to detect the following acoustic features associated with creak: (1) H2−H1 and f_o creak, which characterizes the strong presence of secondary residual peaks often found in creaky voice; (2) residual peak prominence, which is meant to characterize each excitation peak in the time domain; (3) power peak parameters, which highlight the amplitude variation within individual pulses; (4) inter-pulse similarity, which is used to discriminate glottal pulses corresponding to creaky voice from unvoiced regions; (5) intra-frame period, which was designed to help differentiate creaky voice from other voiced regions; and (6) additional acoustic features: energy norm, power standard deviation, and ZeroXrate, which were included to avoid false positives in unvoiced and silent regions. From their visual analysis of the acoustic speech signals, Drugman, Kane, and Gobl[53] described three creaky voice patterns: highly irregular temporal characteristics, fairly regular temporal characteristics with strong excitation peaks, and fairly regular temporal characteristics without strong secondary excitations.

The purpose of this study was to investigate an open-source creak detector as a potential outcome measure for AdLD. Our first hypothesis was that creak would differentiate speakers with AdLD from speakers without voice disorders with high sensitivity and specificity using running speech stimuli typically recorded during clinical voice evaluations. Our second hypothesis was that creak would differentiate speakers with AdLD from speakers with MTD with high sensitivity and specificity using the same stimuli. Results could implicate creak as a potential screening tool for AdLD that would not require the task-specific stimuli that are often only employed when AdLD is already suspected. Our third hypothesis was that the amount of speech used to calculate the percentage (%) creak would affect the discrimination performance of the measure. Given that the presence of creak in speakers with typical voices is tied to specific linguistic contexts,[40] our goal was to understand how much speech stimuli are necessary to minimize this variability—specifically, is the amount of speech typically collected in clinical voice evaluations sufficient? Findings could potentially support the discriminative validity of creak as an automated acoustic outcome measure for AdLD and determine the amount of connected speech stimuli required for creak to be a reliable indicator of AdLD.

METHODS

Participants

Participants in this study included 16 individuals with AdLD, 16 with MTD, and 16 without voice disorders, as approved by the Boston University Institutional Review Board (# 2625) and the University of Texas Southwestern Medical Center (#STU-2022-0388). Patients with AdLD and MTD were all diagnosed via comprehensive voice evaluation by a board-certified otolaryngologist. Patients with AdLD were diagnosed based on consensus criteria from Ludlow et al., (2018): intermittent glottal stops or vowel breaks on voiced sentences; strained-strangled, effortful, tight voice quality; patient report of speaking effort; symptoms reduced during whisper; typical structure and symmetry of the vocal folds at rest; intermittent vocal fold or arytenoid hyperadduction. Individuals who did not meet all criteria were excluded from the study. Patients with AdLD were matched with patients with MTD and controls based on age and sex. Patients with MTD were diagnosed if they demonstrated consistent supraglottic compression throughout the video stroboscopic evaluation with no evidence of concurrent vocal fold pathology or neurological condition. Individuals without voice disorders were volunteers who reported no history of speech, voice, language, or hearing disorders. All participants were English speakers. Demographics[1] for each group are outlined in Table I.

Acoustic Analysis

Audio recordings were made for each participant as they read the same first six sentences of the Rainbow Passage[28] in a quiet clinical environment using a head-mounted microphone placed approximately 3–7 cm off-center from the lips. Files were digitized at 44.1 kHz. The audio signal was viewed in Praat and stimuli were cleaned by removing repetitions of words (approximately 10 words total). An automated creak detector that is available open-source [Covarep][52] (v1.3.2) was used in MATLAB[54] as in Drugman, Kane, and Gobl,[53] which was trained to employ a combination of acoustic features to detect at least three patterns found in creaky voice: highly irregular temporal characteristics, fairly regular temporal characteristics with strong excitation peaks, and fairly regular temporal characteristics without strong secondary excitations.[53]

The same stimuli (first six sentences of the Rainbow Passage[28]) were analyzed for all participants. In a custom MATLAB script, non-voiced audio segments were removed before the samples were input to the creak detector, as creak is only expected during voiced segments. The non-voiced segments were removed by applying a simple threshold procedure based on a 35-ms window of analysis that calculates the envelope of the voiced signal (max-pooling). The creak detector resampled the audio signals for each sentence to 16 kHz before it calculated a set of signal features. These features were used to yield a binary output for the presence of a creak at each time point, employing an Artificial Neural Network. From the result, we calculated the percentage (%) of creak per sentence, which is defined as the total duration of creak in the stimuli divided by the total duration of the voice segments for each sentence.

Statistical Analysis

Nonparametric analyses were performed due to the unequal variance of % creak in patients with AdLD and MTD and individuals without voice disorders. First, the % creak of the first paragraph of the Rainbow Passage[28] was compared among AdLD, MTD, and control groups via a Kruskal-Wallis one-way analysis of variance (ANOVA) implemented using *R* statistical software,[55] with $\alpha = 0.05$. *Post hoc* pairwise Wilcoxon tests were used to compare each group against the other. Next, three separate Receiver-Operator Characteristic (ROC) curve analyses were used to assess the sensitivity and specificity of % creak in differentiating each group from the other. ROC curves are plots of the sensitivity versus $1 - \text{specificity}$. The area under the curve (AUC) is the two-dimensional area underneath the ROC curve that summarizes the diagnostic accuracy of a test, with 0 indicating a perfectly inaccurate test and 1 indicating a perfectly accurate test.[56] The maximum positive and associated negative likelihood ratios were also calculated (LR+ and LR−, respectively). LR+ is the ratio between the probability of a positive test result for the presence of the disorder and the probability of a positive test result for the absence of a disorder. LR− is the ratio between the probability of a negative test result, indicating the presence of the disease, and the probability of a negative test result indicating the absence of a disorder.[20] Next, the sensitivity of creak was maximized between AdLD and MTD to find a threshold that could be used as a screening tool, and corresponding likelihood ratios were calculated.

To determine the amount of running speech required for creak to reliably differentiate speakers with AdLD from speakers with MTD, a custom MATLAB script was used to simulate all possible combinations of the six sentences of the first paragraph of the Rainbow Passage.[28] For example, for a one-sentence condition, any one of the six sentences could be used; for a two-sentence condition, any combination of pairs of sentences could be used (e.g., sentences 1 and 2, sentences 2 and 3, or sentences 1 and 3), and so forth. For each combination, speaker-averaged % creak was computed and used in ROC analysis to compute the associated AUC. The results were used to compare the AUC for discriminating between AdLD and MTD per number of sentences.

RESULTS

The mean % creak values and 95th% confidence intervals from the first paragraph of the Rainbow Passage[28] are shown in Figure 1 per group. The ANOVA revealed a statistically significant effect of group on the

[1]Gender information was not available for all participants.

TABLE I.
Demographics.

Group		AdLD	MTD	Control
Age, mean (SD), years		55 (17)	56 (16)	55 (17)
Age, Min–Max, years		20–75	19–73	19–76
Sex (*n*)	Female	12	12	12
	Male	4	4	4
Race (*n*)	Asian	1	0	0
	Black	0	6	1
	White	15	4	14
	Native American/Alaskan	0	1	0
	More than one race	0	1	1
	Unknown/Not reported	0	4	0
Ethnicity (*n*)	Hispanic or Latino	1	2	0
	Not Hispanic or Latino	14	6	16
	Not reported	1	8	0
Foreign language accent (*n*)		3	7	0
Overall severity of dysphonia, mean (SD)		35.6 (19.2)	27.7 (18.7)	4.8 (3.6)
Overall severity of dysphonia, Min–Max		0–79	4.0–82.6	0–11.7

Note: The mean, standard deviation (SD), and range of age are provided for each group. The AdLD group consists of individuals with adductor laryngeal dystonia, the MTD group consists of individuals with muscle tension dysphonia, and the Control group consisted of individuals without voice disorders. Sex, race, and ethnicity, as well as presence of a foreign language accent, are described per group as the number of participants (*n*). Gender information was not available for all participants. The overall severity of dysphonia mean, SD, and range is reported per group, as evaluated by a voice-specialized speech-language pathologist using the Consensus Auditory-Perceptual Evaluation of Voice.

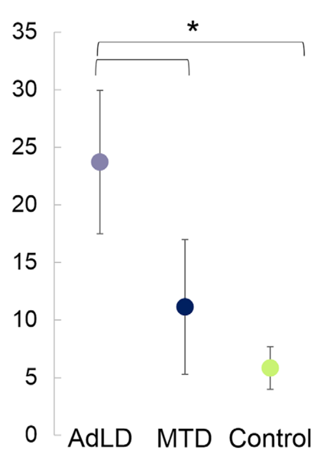

Fig. 1. Creak was calculated from the entire six sentences of the first paragraph of the Rainbow passage. Mean % creak and 95% confidence intervals are plotted per group: adductor laryngeal dystonia (AdLD) in light purple, muscle tension dysphonia (MTD) in dark blue, and Control in green. [Color figure can be viewed in the online issue, which is available at www.laryngoscope.com.]

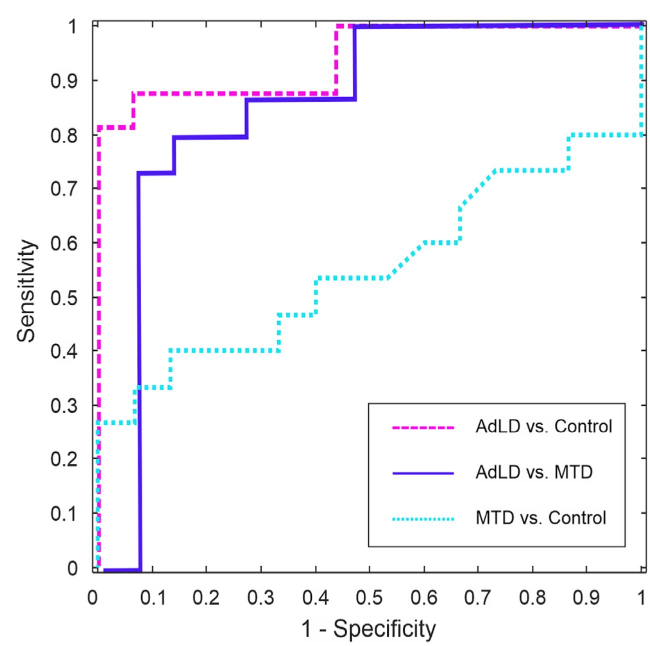

Fig. 2. Receiver-operator characteristic curve plot illustrating the sensitivity and 1 − specificity of % creak in differentiating speakers with adductor laryngeal dystonia (AdLD) from Controls (magenta dashed line), AdLD from muscle tension dysphonia (MTD; solid dark blue line), and MTD from Controls (turquoise dotted line). [Color figure can be viewed in the online issue, which is available at www.laryngoscope.com.]

mean % creak, X^2 (3, $N = 48$) = 18.14, $p > 0.05$, with a large effect size ($\eta^2 = 0.36$). *Post hoc* pairwise Wilcoxon tests revealed statistically significant differences between AdLD and MTD groups with a medium effect size ($p < 0.05$; $r = 0.45$) and AdLD and control groups with a large effect size ($p > 0.05$; $r = 0.66$), but not between MTD and control groups ($p = 0.61$). Results of the three ROC curve analyses are displayed in Figure 2 and detailed in Table II. The area under the curve (AUC) for % creak between the AdLD and Control groups was .94 (dashed magenta line in Fig. 2). The AUC for % creak between the AdLD and MTD groups was 0.86 (blue solid line in Fig. 2). The AUC for % creak between the MTD and

TABLE II. Receiver Operating Characteristic Curve Results.			
Statistic	AdLD vs. Controls	MTD vs. Controls	AdLD vs. MTD
AUC	0.94	0.55	0.86
LR+	13.0	5.0	11.0
LR−	0.14	0.71	0.29
Sensitivity	0.87	0.33	0.73
Specificity	0.93	0.93	0.93
Optimal threshold	10.18	10.18	17.61

Note: Statistics listed from three receiver-operator characteristic curves are listed, which compared speakers with adductor laryngeal dystonia (AdLD) to control speakers without voice disorders (Controls), speakers with muscle tension dysphonia (MTD) and Controls, and speakers with AdLD and MTD. Statistics include area under the receiver-operator characteristic curve (AUC), the maximum positive likelihood ratio (LR+), and the associated negative likelihood ratio (LR−) values, sensitivity, specificity, and thresholds for discrimination.

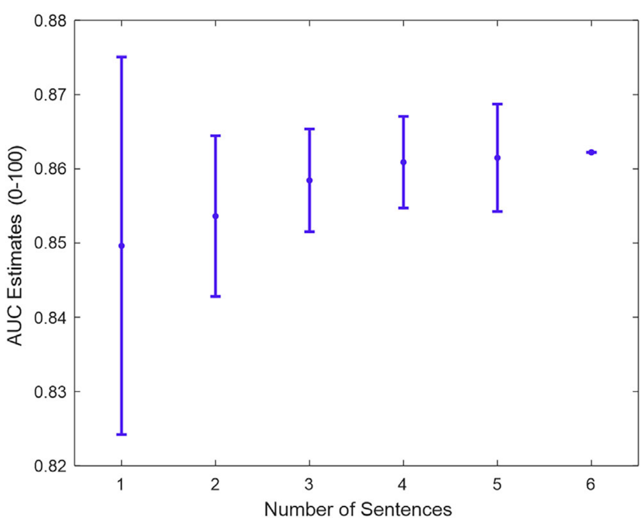

Fig. 3. Mean and 95% confidence intervals of the areas under the area under the curves (AUCs) for % creak in differentiating speakers with adductor laryngeal dystonia (AdLD) from speakers with muscle tension dysphonia (MTD) are plotted as a function of the number of sentences from the first paragraph of the Rainbow Passage used. For example, for a one-sentence condition, any one of the six sentences could be used; for a two-sentence condition, any combination of pairs of sentences could be used (e.g., sentences 1 and 2, sentences 2 and 3, or sentences 1 and 3), and so forth. [Color figure can be viewed in the online issue, which is available at www.laryngoscope.com.]

Control groups was .55 (turquoise dotted line in Fig. 2). The maximum LR+ (11) in differentiating AdLD and MTD was achieved with a threshold of 17.6%. This threshold was associated with a specificity of 0.73, sensitivity of 0.93, and LR− of 2.9. When maximizing the sensitivity in differentiating AdLD and MTD, the threshold was 6.5% (LR+ = 2 and LR− = 0).

The AUC values were recalculated as a function of the number of sentences of speech stimuli to determine the amount of connected speech needed for creak to reliably differentiate AdLD and MTD. Figure 3 illustrates the results comparing the AUC per number of sentences of the Rainbow Passage.[28] When using only one sentence, an AUC ranged from 0.82 to 0.87. Combinations of two sentences yield AUCs ranging from 0.84 to 0.87. Combinations of three, four, and five sentences yielded AUCs ranging from 0.85 to 0.87, and the combination of all 6 sentences yielded an AUC of 0.87, as reported earlier.

DISCUSSION

This study investigated an open-source creak detector as a potential outcome measure for AdLD. We predicted that creak would differentiate speakers with AdLD from speakers without voice disorders with high sensitivity and specificity. Our prediction was confirmed by an area under the ROC characteristic curve (AUC) of 0.94 in speakers with AdLD and controls, which indicated that % creak had outstanding discrimination in differentiating speakers with AdLD from speakers without voice disorders with good sensitivity and outstanding specificity.[56] We also predicted that creak would differentiate speakers with AdLD from speakers with MTD, which was confirmed by an AUC of 0.87, indicating that % creak had excellent discrimination in differentiating speakers with AdLD from speakers with MTD.

These findings provide preliminary evidence that % creak detected via automated creak detection has discriminative validity in differentiating AdLD from MTD. This differentiation is particularly important, as AdLD is often misdiagnosed as MTD, and the two can be difficult even for experts to differentially diagnose,[14] even despite the task-specific nature of AdLD.[19,21] Currently, task specificity is often used when diagnosing AdLD; for example, whispered speech, singing, or speaking in falsetto is often less symptomatic than connected speech at typical pitch and loudness.[57] Connected speech with increased linguistic and motor complexity (i.e., rapid articulatory adjustments) is also more symptomatic than sustained vowels[19,20,22,25,58] and predominantly voiced sentences are more symptomatic than predominantly voiceless sentences.[22] In 2018, consensus-based attributes for identifying individuals with laryngeal dystonia were established by a Delphi panel of 13 experts.[21] However, clinically, the stimuli published in the supplement[21] are typically only used when laryngologists already suspect the presence of laryngeal dystonia.

In the current study, we employed ubiquitous stimuli of the Rainbow Passage,[28] which is one stimulus referenced by ASHA for measuring acoustics in clinical voice assessment.[59] Because clinics vary in the amount of speech they collect from the Rainbow Passage, it was important to determine the amount of connected speech needed for creak to reliably discriminate AdLD from MTD. The average AUC varied modestly across the number of sentences used; however, when only one sentence was used, some single sentences resulted in AUCs as low as 0.82, which is substantially lower than the AUC for all six sentences combined (0.87). Nevertheless, by using only two sentences, the range of potential AUCs is substantially narrowed, with performance commensurate with the AUC using all six sentences. This result is similar to findings by Barsties and Maryn[60] who found that consistency of overall severity judgments improved from stimuli of 17 syllables to 35.5 syllables, but was comparable between 35.5 syllables and 93 syllables.

Based on our findings, our recommendation is for creak to be calculated over at least two sentences for this result to generalize. Thus, should a creak detector be incorporated as a clinical screening tool, it would theoretically be relatively quick to implement (i.e., less than 10 s) during a voice evaluation, particularly if that stimulus is already being collected as part of a typical voice evaluation. The optimal threshold of 17.6% indicates that % creak that is >17.6% indicates AdLD with high sensitivity and specificity even amongst a pool of speakers that are typically easily confused (MTD and AdLD). When sensitivity was maximized, the threshold became 6.5%. If a creak detector were to be implemented as a screening tool, a threshold of 6.5% creak could be used to identify individuals who may need a further workup for an AdLD differential diagnosis. It is important to note, however, that this study employed a creak detector that relies on MATLAB, a program that is not accessible in many clinics. Practically, the creak detector would need to be incorporated into a user-friendly interface before implementation as a screening tool would be feasible in clinical practice. This application could be addressed in future related work.

In comparison to other quantitative acoustic measures that have shown discriminative validity in differentiating AdLD from MTD, % creak had greater sensitivity (0.73) and specificity (0.93) than both the CSID (0.67 sensitivity and 0.64 specificity)[20] and LTAS (0.61 sensitivity and 0.68 specificity).[32] Of note, the CSID discrimination was based on task specificity (the difference in CSID between connected speech and a sustained vowel) and the LTAS was calculated from a predominantly voiced sentence, whereas % creak was calculated from a mixed phoneme passage that is commonly collected as part of a standard voice evaluation. The positive likelihood ratio (LR+) statistic was greater for % creak compared to CSID and LTAS. A test is considered highly diagnostic if the value is >10. In differentiating AdLD from MTD, the LR+ of % creak was 11. This means that those testing positive by % creak were 11 times more likely to have AdLD. Comparatively, the LR+ of CSID was 1.88[20] and the LR+ of LTAS was 1.92.[32] When interpreting the negative likelihood ratio (LR−), a value <0.10 is indicative that a person who tests negative likely does not have the disorder of interest. The LR− in the current study was 0.29, which does not meet this criterion, but is lower than the LR− for CSID (0.52)[20] and LTAS (0.44).[32] In all, % creak appears to have the best diagnostic accuracy of the automated quantitative measures that have been studied in differentiating AdLD and MTD.

This study was limited to previously recorded audio samples; additional work is needed to prospectively investigate automatic % creak in speakers with AdLD and MTD, with improved distribution of overall severity. Although the overall severity of speakers in this study was comparable between groups, the severity ratings were positively skewed in both patient groups. Moreover, these ratings were provided by only a single voice-specialized speech-language pathologist. Future studies may formally address the relationship between a creak and overall severity by using more than one rater. Inadvertently, it was observed that seven of the 16 speakers in the MTD group of this study were judged to have a foreign language accent, although this was true for only three speakers in the AdLD group and zero speakers in the control group. Therefore, future work is warranted to investigate the interaction of foreign language accents on % creak in speakers of English with AdLD and MTD.

Although the use of the Rainbow Passage is common, we recognize that not all clinics employ a reading passage during voice evaluations. Similar results could be expected from other reading stimuli, such as the CAPE-V sentences,[61] as long as at least two sentences are analyzed. An important remaining question is how the creak detector will differentiate between speakers with AdLD and speakers with MTD in spontaneous speech. We hypothesize that creak may be less discriminative in spontaneous speech, based upon the way typical speakers use creaky voices to communicate attitude and affective states.[53,62] Creak is likely to vary with levels of formality of speaking[63] (i.e., spontaneous speech may be more casual than reading a passage).

Finally, further psychometric evaluation of % creak is needed before it is employed as an outcome measure for AdLD. Based on the results of manual identification of creak in our previous work,[10] we would expect the creak detector to identify a higher percentage of creak in voiced phoneme-loaded sentences (which are typically more symptomatic in speakers with AdLD) than in voiceless phoneme-loaded sentences (which are typically less symptomatic in speakers with AdLD). However, further work is needed to determine whether the creak detector would similarly differentiate between voiced and voiceless phoneme-loaded sentences. Responsiveness may also be investigated by comparing the connected speech of individuals with AdLD before and after botulinum toxin injections. However, there is evidence to suggest that % creak may not actually change after an injection; interestingly, Sapienza et al.[7] found that % aperiodicity in selected words from a reading passage did not statistically change from pre- to post-Botox injection. Further investigation is warranted using connected speech to determine whether automated % creak changes after Botox injections. Additional psychometric analyses could provide clinically relevant information, such as the minimal % creak that is clinically important (i.e., the minimum clinically important difference).

CONCLUSION

Automated % creak derived via an open-source creak detector was sensitive and specific in differentiating speakers with AdLD from speakers with MTD and controls with high sensitivity and specificity, demonstrating discriminant validity of % creak as a potential outcome measure for AdLD. To reliably differentiate AdLD from MTD, at least two sentences of connected speech are recommended to be used as stimuli. Percent creak has the potential to be used as a screening tool to identify patients who may need a further workup for AdLD. Use of a creak detector does not require specific stimuli and thus could be easily implemented into a comprehensive voice evaluation.

FUNDING INFORMATION

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).

CONFLICT OF INTEREST

Cara E. Stepp has received consulting fees from Altec, Inc./Delsys, Inc., companies focused on developing and commercializing technologies related to human movement. Stepp's interests were reviewed and managed by Boston University in accordance with its conflict-of-interest policies. The other authors have declared that no competing financial or nonfinancial interests existed at the time of this study.

BIBLIOGRAPHY

1. Blitzer A. Spasmodic dysphonia and botulinum toxin: experience from the largest treatment series. *Eur J Neurol*. 2010;17(Suppl 1):28-30. https://doi.org/10.1111/j.1468-1331.2010.03047.
2. Tisch SH, Brake HM, Law M, Cole IE, Darveniza P. Spasmodic dysphonia: clinical features and effects of botulinum toxin therapy in 169 patients—an Australian experience. *J Clin Neurosc Off J Neurosurg Soc Australas*. 2003;10(4):434-438. https://doi.org/10.1016/s0967-5868(03)00020-1.
3. Langeveld TP, Drost HA, Zwinderman AH, Frijns JH, De Jong RJB. Perceptual characteristics of adductor spasmodic dysphonia. *Ann Otol Rhinol Laryngol*. 2000;109(8):741-748.
4. Cannito MP, Doiuchi M, Murry T, Woodson GE. Perceptual structure of adductor spasmodic dysphonia and its acoustic correlates. *J Voice*. 2012; 26(6):818. e5-818. e13.
5. Zwirner P, Murry T, Woodson GE. Perceptual-acoustic relationships in spasmodic dysphonia. *J Voice*. 1993;7(2):165-171.
6. Sapienza CM, Murry T, Brown WS. Variations in adductor spasmodic dysphonia: acoustic evidence. *J Voice*. 1998;12(2):214-222. https://doi.org/10.1016/s0892-1997(98)80041-6.
7. Sapienza CM. Acoustic variations in reading produced by speakers with spasmodic dysphonia pre-botox injection and within early stages of botox injection. *J Speech Lang Hear Res*. 2002;45(5):830-843.
8. 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.
9. 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. https://doi.org/10.1016/S0892-1997(00)80008-9.
10. Marks KL, Feaster TF, Baker S, Díaz-Cádiz ME, Doyle PC, Stepp CE. 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.
11. Bender BK, Cannito MP, Murry T, Woodson GE. Speech intelligibility in severe adductor spasmodic dysphonia. *J Speech Lang Hear Res*. 2004; 47(1):21-32. https://doi.org/10.1044/1092-4388(2004/003).
12. Isetti D, Xuereb L, Eadie TL. Inferring speaker attributes in adductor spasmodic dysphonia: ratings from unfamiliar listeners. *Am J Speech Lang Pathol*. 2014;23(2):134-145. https://doi.org/10.1044/2013_AJSLP-13-0010.
13. Braden MN, Johns MM III, Klein AM, Delgaudio JM, Gilman M, Hapner ER. Assessing the effectiveness of botulinum toxin injections for adductor spasmodic dysphonia: clinician and patient perception. *J Voice*. 2010;24(2):242-249.
14. Creighton FX, Hapner E, Klein A, Rosen A, Jinnah HA, Johns MM. Diagnostic delays in spasmodic dysphonia: a call for clinician education. *J Voice*. 2015;29(5):592-594.
15. Hillman RE, Stepp CE, Van Stan JH, Zañartu M, Mehta DD. An updated theoretical framework for vocal hyperfunction. *Am J Speech Lang Pathol*. 2020;29(4):2254-2260.
16. Oates J, Winkworth A. Current knowledge, controversies and future directions in hyperfunctional voice disorders. *Int J Speech Lang Pathol*. 2008; 10(4):267-277.
17. Colton RH, Casper JK, Leonard RJ. *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*. Lippincott Williams & Wilkins; 2006:498.
18. Ludlow C. Management of the spasmodic dysphonias. In: Rubin JS, ed. *Diagnosis and Treatment of Voice Disorders*. Igaku-Shoin; 1995:436-454.
19. 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. https://doi.org/10.1097/01.mlg.0000154739.48314.ee.
20. 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.
21. 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.
22. 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. https://doi.org/10.1044/1058-0360(2003/087).
23. Aronson AE, Brown JR, Litin EM, Pearson JS, Spastic dysphonia. I. Voice, neurologic, and psychiatric aspects. *J Speech Hear Disord*. 1968;33(3): 203-218.
24. Aronson AE, Brown JR, Litin EM, Pearson JS. Spastic dysphonia. II. Comparison with essential (voice) tremor and other neurologic and psychogenic dysphonias. *J Speech Hear Disord*. 1968;33(3):219-231.
25. Ludlow CL, Connor NP. Dynamic aspects of phonatory control in spasmodic dysphonia. *J Speech Lang Hear Res*. 1987;30(2):197-206.
26. Yanagida S, Nishizawa N, Hashimoto R, et al. Reliability and validity of speech evaluation in adductor spasmodic dysphonia. *J Voice*. 2018;32(5): 585-591. https://doi.org/10.1016/j.jvoice.2017.06.022.
27. Moerman M, Martens JP, Dejonckere P. Multidimensional assessment of strongly irregular voices such as in substitution voicing and spasmodic dysphonia: a compilation of own research. *Logoped Phoniatr Vocol*. 2015; 40(1):24-29. https://doi.org/10.3109/14015439.2014.936497.
28. Fairbanks G. *Voice and Articulation Drillbook, Vol. 2*. Harper and Row; 1960.
29. Awan SN, Roy N, Jetté ME, Meltzner GS, Hillman RE. Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: comparisons with auditory-perceptual judgements from the CAPE-V. *Clin Linguist Phon*. 2010;24(9):742-758. https://doi.org/10.3109/02699206.2010.492446.
30. Awan SN, Roy N, Dromey C. Estimating dysphonia severity in continuous speech: application of a multi-parameter spectral/cepstral model. *Clin Linguist Phon*. 2009;23(11):825-841. https://doi.org/10.3109/02699200903242988.
31. Roy N, Mauszycki SC, Merrill RM, Gouse M, Smith ME. Toward improved differential diagnosis of adductor spasmodic dysphonia and muscle tension dysphonia. *Folia Phoniatr Logop*. 2007;59(2):83-90.
32. 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.
33. Awan SN, Roy N, Zhang D, Cohen SM. Validation of the cepstral spectral index of dysphonia (CSID) as a screening tool for voice disorders: development of clinical cutoff scores. *J Voice*. 2016;30(2):130-144. https://doi.org/10.1016/j.jvoice.2015.04.009.
34. Gillespie AI, Gartner-Schmidt J, Lewandowski A, Awan SN. An examination of pre- and posttreatment acoustic versus auditory perceptual analyses of voice across four common voice disorders. *J Voice*. 2017;32(2):169-176. https://doi.org/10.1016/j.jvoice.2017.04.018.
35. Frøkjær-Jensen B, Prytz S. Evaluation of speech disorders by means of long-time-average-spectra. *Annu Rep Institut Phon Univ Copenhagen*. 1974;8: 227-237.

36. Tanner K, Roy N, Ash A, Buder EH. Spectral moments of the long-term average spectrum: sensitive indices of voice change after therapy? *J Voice*. 2005;19(2):211-222.
37. Roy N, Whitchurch M, Merrill RM, Houtz D, Smith ME. Differential diagnosis of adductor spasmodic dysphonia and muscle tension dysphonia using phonatory break analysis. *Laryngoscope*. 2008;118(12):2245-2253. https://doi.org/10.1097/MLG.0b013e318184577.
38. Keating PA, Garellek M, Kreiman J. Acoustic properties of different kinds of creaky voice. *ICPhS*. 2015;1:2-7.
39. Dallaston K, Docherty G. *Estimating the prevalence of creaky voice: A fundamental frequency-based approach*; 2019:532-536.
40. Davidson L. The versatility of creaky phonation: segmental, prosodic, and sociolinguistic uses in the world's languages. *Wiley Interdiscipl Rev Cogn Sci*. 2021;12(3):e1547.
41. Drugman T, Kane J, Gobl C. Data-driven detection and analysis of the patterns of creaky voice. *Comput Speech Lang*. 2014;28(5):1233-1253.
42. Martin P. Automatic detection of voice creak. *6th International Conference on Speech Prosody*; ISCA Archive. 2012:43-46.
43. Murton O, Shattuck-Hufnagel S, Choi J-Y, Mehta DD. Identifying a creak probability threshold for an irregular pitch period detection algorithm. *J Acoust Soc Am*. 2019;145(5):EL379-EL385.
44. Narendra N, Rao KS. Automatic detection of creaky voice using epoch parameters. *Interspeech*. 2015;1:2347-2351.
45. Tavi L, Alumäe T, Werner S. Recognition of creaky voice from emergency calls. *Interspeech*. 2019;1:1990-1994.
46. White H, Penney J, Gibson A, Szakay A, Cox F. Optimizing an automatic creaky voice detection method for Australian English speaking females. *Interspeech*. 2021;1:1384-1388.
47. Ishi CT, Sakakibara KI, Ishiguro H, Hagita N. A method for automatic detection of vocal fry. *IEEE Trans Audio Speech Lang Process*. 2008;16(1):47-56. https://doi.org/10.1109/tasl.2007.910791.
48. Vishnubhotla S, Espy-Wilson CY. Automatic detection of irregular phonation in continuous speech. *9th Int Conf Spoken Lang Process*. 2006;1:949-952.
49. Kane J, Drugman T, Gobl C. Improved automatic detection of creak. *Comput Speech Lang*. 2013;27(4):1028-1047. https://doi.org/10.1016/j.csl.2012.11.002.
50. Blomgren M, Chen Y, Ng ML, Gilbert HR. Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *J Acoust Soc Am*. 1998;103(5):2649-2658.
51. Drugman T, Kane J, Gobl C. Resonator-based creaky voice detection. *Interspeech*. 2012;1:1592-1595.
52. Degottex G, Kane J, Drugman T, Raitio T, Scherer S. *COVAREP—A collaborative voice analysis repository for speech technologies*. IEEE; 2014:960-964.
53. Drugman T, Kane J, Gobl C. Data-driven detection and analysis of the patterns of creaky voice. arXiv preprint arXiv:2006.00518. 2020; 1–39.
54. Mathworks I. *MATLAB: The Language of Technical Computing: Computation, Visualization, Programming. (R2021a). Version 9.10.0*. Natick, MA: Mathworks; 2021.
55. R Core Team. *R: A Language and Environment for Statistical Computing. Version 4.1.2*. Vienna, Austria: R Foundation for Statistical Computing; 2021.
56. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 2010;5(9):1315-1316.
57. Cannito MP, Woodson GE, Murry T, Bender B. Perceptual analyses of spasmodic dysphonia before and after treatment. *Arch Otolaryngol Head Neck Surg*. 2004;130(12):1393-1399.
58. Bloch CS, Hirano M, Gould WJ. Symptom improvement of spastic dysphonia in response to phonatory tasks. *Ann Otol Rhinol Laryngol*. 1985;94(1):51-54.
59. 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.
60. Barsties B, Maryn Y. The influence of voice sample length in the auditory-perceptual judgment of overall voice quality. *J Voice*. 2017;31(2):202-210.
61. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18(2):124-132.
62. Yanushevskaya I, Gobl C, Chasaide AN. Voice quality and f0 cues for affect expression: implications for synthesis. *Interspeech*. 2005;1:1849-1852.
63. Kane J, Pápay K, Hunyadi L, Gobl C. On the use of creak in Hungarian spontaneous speech. *Citeseer*. 2011;XVII:1014-1017.