

Reliability of High-speed Videoendoscopic Ratings of Essential Voice Tremor and Adductor Spasmodic Dysphonia

*Lindsey A. Parker, *Melda Kunduk, †Daniel S. Fink, and ‡Andrew McWhorter, *Baton Rouge and ‡New Orleans, Louisiana, and †Aurora, Colorado

Summary: Objectives. The main objectives of this study were to (1) examine intrarater and inter-rater reliabilities in perceptual ratings of vocal fold vibratory patterns and supraglottic characteristics for essential vocal tremor and adductor spasmodic dysphonia (AdSD) using high-speed videoendoscopy (HSV), and (2) to investigate the specificity of the parameters in differentiating these two voice disorders.

Methods. HSV recordings of 34 cases diagnosed with essential vocal tremor, AdSD, or AdSD with vocal tremor were evaluated blindly by two voice speech pathologists. The two raters examined all HSV video segments twice across nine supraglottic and vocal fold vibratory characteristics for inter-rater and intrarater reliabilities. A separate consensus rating was then developed, with the results analyzed to explore differentiation.

Results. Raters demonstrated moderate intrarater reliability with mean Spearman's rho correlation coefficients of 0.68 (rater 1) and 0.73 (rater 2). Moderate inter-rater reliability for the two raters was seen across all parameters with a mean Cohen's kappa coefficient of 0.51. Raters showed higher intrarater and inter-rater reliabilities for supraglottic parameters. Only the presence of tremor differentiated between the two voice disorders in cases with a consistent diagnosis.

Conclusions. The high level of concomitance between vocal tremor and AdSD may affect subjective perceptual analysis of supraglottic and vocal fold vibratory patterns. Results indicate similar global involvement of supraglottic laryngeal structures for both vocal tremor and AdSD.

Key Words: High-speed videoendoscopy–Essential voice tremor–Adductor spasmodic dysphonia–Inter-rater reliability–Intrarater reliability.

INTRODUCTION

Establishing improved reliability for the endoscopic features of vocal function is key to identifying characteristics that may differentiate between voice disorders. This is particularly important for two voice disorders, essential vocal tremor (EVT) and adductor spasmodic dysphonia (AdSD). EVT is a neurologic voice disorder characterized by the perception of an unsteady, tremulous voice with involuntary fundamental frequency modulations caused by near rhythmic oscillations of muscle movements during phonation.^{1–3} AdSD is a neurologic, focal laryngeal dystonia that presents with a strained, strangled vocal quality, irregular voice breaks, and marked effort during phonation as a result of a spasm of the laryngeal muscles.^{4–8} Clinicians and doctors often find them difficult to reliably distinguish because of overlapping auditory-perceptual characteristics, their aperiodic voice signals, and high co-occurrence rates.^{4,9–11}

Misdiagnosis between EVT and AdSD or with other functional voice disorders such as muscle tension dysphonia^{12–14} may delay appropriate treatments for patients. Patients with

spasmodic dysphonia wait an average of 4.5 years to receive an accurate diagnosis after initially seeing a physician.¹⁵ In the same patient interviews, approximately 25% of the subjects reported being prescribed ineffective medications or treatments other than botulinum toxin, such as anxiety or reflux medications, to attempt to treat their symptoms because of misdiagnosis.¹⁵ A reliable diagnosis will ensure the timely delivery of correct treatment techniques for these similar-sounding voice disorders.

The differential diagnoses of EVT and AdSD is based on subjective auditory perception, visual evaluation of endoscopic laryngeal imaging, patient report, and a medical case history. Presentations of EVT and AdSD often present with similar perceptual hallmarks, with full voice stoppages documented for severe EVT cases and instances of tremor co-occurring with AdSD.^{4,9} Recent studies report EVT accompanying AdSD between 26.0% and 54.4% of all cases.^{11,16} These concomitant cases often demonstrate an irregular tremor similar in rate to EVT¹⁰ and periodic fluctuations in pitch and loudness during sustained phonation.^{5,13,17,18} Such cases can make it much more difficult to reliably diagnose these disorders.

Despite similarities in auditory-perceptual presentation, clinical diagnosis of EVT and AdSD still often relies most heavily upon auditory-perceptual evaluations.^{10,11,13,19} Several studies have assessed the reliability of auditory-perceptual evaluations of EVT and AdSD.^{20,21} Chhetri et al²¹ found 64% exact agreement between listeners rating the severity of cases with AdSD alone. Johnson et al²⁰ examined the perceptual reliability between diagnostic cases of AdSD, EVT, AdSD with tremor, and abductor spasmodic dysphonia. Although Johnson et al discovered that raters could reliably differentiate EVT or AdSD from other voice disorders over the telephone, when asked to provide a differential diagnosis

Accepted for publication October 11, 2017.

From the *Department of Communication Sciences and Disorders, Louisiana State University, Baton Rouge, Louisiana; †Department of Otolaryngology, University of Colorado School of Medicine, Aurora, Colorado; and the ‡Department of Otolaryngology-Head and Neck Surgery, Louisiana State University Health Sciences Center, New Orleans, Louisiana.

Address correspondence and reprint requests to Melda Kunduk, Department of Communication Sciences and Disorders, Louisiana State University, 68 Hatcher Hall, Field House Drive, Baton Rouge, LA 70803. E-mail: mkunduk@lsu.edu

Journal of Voice, Vol. 33, No. 1, pp. 16–26

0892-1997

© 2017 The Voice Foundation. Published by Elsevier Inc. All rights reserved.

<https://doi.org/10.1016/j.jvoice.2017.10.009>

between the two disorders, results indicated poor reliability ($\kappa = 0.20\text{--}0.33$).²⁰

Direct laryngeal imaging of phonation is essential for reliable differential diagnosis of voice disorders. Complementing auditory-perceptual, aerodynamic, and acoustic analysis, laryngeal imaging techniques can detect clinically significant findings impossible by indirect means.^{22–26} Reliably judging the voice and the vocal fold movement, however, can be challenging for clinical assessment.^{27–29} Subjective auditory and visual-perceptual judgments can be biased because of a lack of standardized, well-defined criteria for reliably identifying and rating laryngeal endoscopic parameters for voice disorders.³⁰

Studies examining the differential diagnosis of voice disorders using videostroboscopy^{27,31–33} and high-speed videoendoscopy (HSV)^{34–36} often show high reliability differentiating normal from disordered cases. However, ratings for individual dynamic laryngeal and vocal fold vibratory parameters are often found to have poor to fair inter-rater reliability.^{27,30,37–40} Common reported parameters like glottal closure (0.38–0.83),^{27,30,33,40–44} mucosal wave (0.50–0.74),^{30,40–45} and vocal fold regularity (0.32–0.96),^{27,30,40,41,44} for example, demonstrate wide variance in agreement across studies. Despite the need for consistent and reproducible ratings, recent reviews of the literature indicate that studies examining voice therapy outcomes also under-reported rater reliability for videostroboscopy. In the limited studies on voice outcomes that do report the reliability, the findings are often deemed to have poor reliability or insufficient methodology.^{38,39} To improve reliability outcomes, researchers have emphasized training protocols, consensus ratings, minimal experience requirements, and rating forms. Further research, however, is required to determine whether the mixed outcomes seen across individual parameters may be attributed to insufficient experience and training or human perceptual and technical limitations.

With videostroboscopy derived from an indirect, acoustic-based signal, technical limitations arise in its ability to reliably diagnose aperiodic voice disorders. New techniques using simulated stroboscopy, which extracts the vibratory phase directly from HSV data, show promise for better reliability over traditional videostroboscopy for benign lesions.⁴⁵ HSV allows for examination across each individual glottal cycle and lets clinicians observe laryngeal spasms, phonatory breaks, and rapid laryngeal movements such as supraglottic constriction,^{23,25,46} features that are often attributed to EVT and AdSD. In a recent comparison of vocal fold vibratory parameters between stroboscopy and HSV for clinical use, the inclusion of HSV changed the diagnosis in 7% of cases.⁴⁷ Although previous studies have sought to characterize AdSD against other voice disorders,^{4,8,26} identify risk factors,⁵ and examine tremor severity,^{37,48} no studies have used HSV to examine reliability or to identify potential parameters between AdSD and EVT.^{4,8,19,23,29,49}

The goals of the present study were twofold: first, to assess the reliability of the raters' judgments of supraglottic and vocal fold vibratory characteristics obtained with HSV; and second, to examine whether any of the parameters could differentiate between EVT and AdSD. It was hypothesized that (1) the raters would achieve a higher reliability for parameters with supraglottic activity during phonation compared with true vocal fold

(TVF) vibratory patterns, (2) reliability would increase in cases where the diagnosis clearly indicated either EVT or AdSD with no diagnosed concomitance, and (3) predicted parameters with higher reliability would distinguish between EVT and AdSD.

METHODS

Data collection

HSV data were compiled to conduct this retrospective analysis. The HSV video recordings were obtained over a period of 5 years during routine clinical examinations of patients. The same otolaryngologist determined a clinical differential diagnosis: EVT, AdSD, or AdSD with vocal tremor for each case. The selection criteria for HSV videos included an unobstructed view of the arytenoids, the false vocal folds (FVFs), and the TVFs during sustained phonation. Thirty-four HSV recordings met these criteria. The distribution across the three diagnostic categories was as follows: AdSD ($n = 16$), EVT ($n = 8$), AdSD with vocal tremor ($n = 10$).

All video cases were performed using a 70° rigid laryngoscope (Model 9106; KayPENTAX, Montvale, NJ) and paired with an HSV system (Model 9700, KayPENTAX) and a 300-W cold light source (CLV-U20; Olympus America, Center Valley, PA). During the collection procedure, the patients were instructed to sustain the vowel /i/ with a comfortable pitch and loudness. Each video recording was digitally stored in a database at an uncompressed 8-bit monochrome grayscale with a pixel resolution of 120 × 256 pixels and a sampling rate of 2000 frames per second (fps). Selected video segments included both initiation and sustained phonation. No audio files were used during the rating of supraglottic and vocal fold vibratory features for any of the 34 cases in the study to ensure that raters were rating only the parameters related to the supraglottic and vocal fold vibratory behaviors. No perceptual or additional identifying information was included in the presentation of the videos for the ratings. Raters were blind to the diagnosis for all the cases included in the study.

Parameter selection

Following a review of the literature for supraglottic and vocal fold vibratory characteristics identified with either EVT or AdSD, nine parameters were initially selected. The parameters were to be rated by their presence or absence. These included four supraglottic parameters: *rhythmic supraglottic oscillation*, *arytenoid tremor*, *supraglottic constriction*, and *sustained FVF involvement*; and five vocal fold vibratory parameters: *TVF vibratory disruption*, *aperiodicity*, *right-left (R/L) TVF asymmetry*, *anterior-posterior (A/P) TVF asymmetry*, and *phase asymmetry*. Of these supraglottic parameters, the two identified with EVT were arytenoid tremor and rhythmic supraglottic oscillation.^{2,3,9,10,50,51}

The HSV videos from the 70° rigid laryngoscope captured a tight frame around the larynx and did not include all supraglottic structures. Although tremor along the tongue, palate, and posterior pharyngeal wall can be present in EVT, this presence was not visible within the identified frames of the HSV videos for the current study and was therefore excluded. However, rhythmic supraglottic oscillation patterns were evaluated. Two

other parameters, supraglottic constriction and sustained involvement of supraglottic structures such as the sustained FVF involvement during phonation, are linked more closely with AdSD.^{4,5,10} Although Sulica and Louis¹⁰ also assert some involvement of global laryngeal structures during phonation for EVT, the presence of these characteristics was reported less frequently.

Depending on the severity of the presentation, disruptions of the TVF vibratory behavior can potentially be present in all three diagnostic categories in the present study: EVT, AdSD, and AdSD with vocal tremor.^{4,9} However, given EVT's physiological presentation with rhythmic movement of the laryngeal structures, it was hypothesized that EVT would show less aperiodicity, TVF asymmetry, and phase asymmetry; whereas the potential effects of the intermittent hyperadduction of vocal folds in AdSD would likely show increased aperiodicity and asymmetry.^{3,8,35,52} From these nine dichotomous parameters examining presence or absence, subcategories were created for a total of 21 potential variables (see Appendix for a complete list of parameters and their definitions). These subcategories were added to indicate the location (right, left, or both TVFs) and the direction of movement (lateral, A/P direction) for the affected structure, and the degree of involvement (absent, intermittent, or consistent).

Subjective video analysis

The thirty-four HSV cases were rated by two speech-language pathologists specializing in voice disorders. Both raters had prior experience in evaluating laryngeal endoscopic examinations. The two raters had different levels of clinical experience: a doctoral candidate with 3 years of experience with voice and a PhD speech-language pathologist specializing in voice disorders. Before beginning the video analysis, the raters participated in two training sessions where two separate HSV cases not included in the current study were presented. With these training samples used as a visual anchor, the characteristics were identified, discussed, and defined between the raters before judging the study's HSV data. The definition set for each of the parameters made during the training session was available as a written anchor for each rater during their review of the recordings. For intrarater reliability, both raters completed a second round of ratings for all 34 cases.

The selected HSV video segments for each case were randomized and observed in a sequential order by the raters. Two playback rates were used in the presentation of the recordings to accommodate judgments for supraglottic and TVF characteristics. First, the raters viewed an individual case at a 200-fps playback rate to judge the identified supraglottic structures. These observations were followed by a slow playback rate of 20 fps to judge the vocal fold vibratory function. The raters assessed each case based on the identified parameters, scoring first for the absence or the presence and then scoring the subcategories along a scaled rating for the location, direction, and degree of the involvement as appropriate.

Consensus rating for differentiation

Following the reliability ratings, a separate consensus rating between the two raters was established to produce a standard

score for each case across all of the identified characteristics. This consensus score was used to analyze the specificity of the parameters in differentiating between EVT and AdSD. During the development of the consensus rating, each of the 34 individual cases were reviewed and discussed between the two speech-language pathologists to come to an agreed upon score across each of the parameters.

Diagnosis reliability rating

Given the high level of documented concomitant cases of EVT and AdSD, two otolaryngologists conducted a secondary analysis of diagnostic reliability, including the same otolaryngologist who made the initial diagnoses. The otolaryngologists reviewed each of the 34 cases using auditory-perceptual evaluation in addition to laryngeal videostroboscopy for confirmation to make their diagnosis. Cases where the two laryngologists agreed upon a diagnosis were further examined.

Statistical analysis

Statistical analysis was conducted using *SPSS 22.0* (SPSS, Chicago, IL) and *MedCalc 15.11.4* (MedCalc Software, Ostend, Belgium). Inter-rater reliability for the dichotomous parameters measuring presence or absence was calculated using the Cohen's kappa coefficient (κ). An alternative, weighted kappa can be used for categorical data with an ordinal structure to take into consideration scaled disagreements between raters along with the degree of magnitude of the potential disagreement.⁵³ Thus, a quadratic weighted Cohen's kappa was chosen for the subcategory parameters based on location and severity.

Intrarater reliability was assessed using the Spearman's rank correlation coefficient after both raters completed a second round of ratings for all 34 cases. Spearman's rank measures the strength of the association between two sets of ratings for an individual rater, standardized on a -1 to 1 scale where 1 is a direct correlation and -1 reflects an inverse correlation. Analysis for the discrimination of parameters between the three diagnostic categories was conducted via Pearson chi-square analysis and descriptive statistics. The observed level of statistical significance was set by alpha ($P < 0.05$).

RESULTS

Inter-rater reliability

Inter-rater reliability using kappa and weighted kappa statistics met a moderate mean agreement ($\kappa = 0.51$) for all 21 parameters. The Landis and Koch scale for assessing the magnitude of agreement provided additional interpretation for the findings (where 0.01–0.20 equals slight agreement, 0.21–0.40 equals fair agreement, 0.41–0.60 equals moderate agreement, 0.61–0.80 equals substantial agreement, and 0.81–1.0 is considered an almost perfect agreement).⁵⁴ Table 1 shows the reliability distribution for the parameters between the raters. The raters demonstrated substantial agreement (.61–.80) for 7 of the 21 total parameters (33%), with high reliability noted for the dichotomous presence or absence of rhythmic supraglottic oscillation. Ratings for 9 of the remaining 14 parameters reached moderate agreement (43%), including the dichotomous presence

TABLE 1.
Inter-rater Reliability via the Cohen's Kappa

		R1 vs R2 (n = 34)
Supraglottic parameters		
Arytenoid tremor	Presence of arytenoid tremor	0.541
	Location of arytenoid tremor	0.366*
Rhythmic supraglottic oscillation	Presence of rhythmic supraglottic oscillation	0.686
	Degree of rhythmic supraglottic oscillation	0.714*
	Location of rhythmic supraglottic oscillation	0.689*
Supraglottic constriction	Presence of supraglottic constriction	0.403
	Degree of lateral supraglottic constriction	0.771*
	Degree of A/P supraglottic constriction	0.657*
Constant supraglottic activity	Presence of sustained FVF involvement	0.287
	Degree of lateral FVF involvement	0.687*
	Location of lateral FVF involvement	0.474*
	Degree of A/P supraglottic activity	0.419*
TVF vibratory parameters		
TVF vibratory disruption	Presence of TVF vibratory disruption	0.541
	Degree of TVF vibratory disruption	0.762*
Aperiodicity	Presence of aperiodicity	0.454
	Degree of aperiodicity	0.420*
R/L TVF asymmetry	Presence of R/L TVF asymmetry	0.266
	Location of R/L TVF asymmetry	0.226*
A/P TVF asymmetry	Presence of A/P TVF asymmetry	0.457
	Location of A/P TVF asymmetry	0.388*
Phase asymmetry	Presence of phase asymmetry	0.481

* Weighted Cohen's kappa, Cohen's kappa correlation significant at a *P* level of <0.05 (two-tailed).

or absence of arytenoid tremor, TVF vibratory disruption, supraglottic constriction, aperiodicity, A/P TVF asymmetry, and phase asymmetry. The raters demonstrated fair inter-rater reliability for the presence of sustained FVF involvement and the presence of R/L TVF asymmetry (Table 1).

Intrarater reliability

Findings calculated from the Spearman's rank correlation coefficient indicated a moderate mean intrarater reliability for both raters (rater 1 [R1]: $r = 0.68$, and rater 2 [R2]: $r = 0.73$, respectively). Table 2 shows the intrarater reliability statistics for each rater, comparing judgments across all 34 cases from two separate rating intervals. The level of magnitude was interpreted based on a scale with 0.2 being minimal agreement, 0.5 being moderate agreement, and 0.8 being strong agreement.⁵⁵ The raters exhibited higher levels of agreement for supraglottic parameters (R1 mean $r = 0.74$ and R2 mean $r = 0.73$). Both raters demonstrated a greater mean intrarater agreement for supraglottic characteristics compared with TVF vibratory characteristics, with the exception of TVF vibratory disruption.

Diagnosis reliability

Findings from the comparison of diagnostic ratings between the two laryngologists indicated good inter-rater agreement ($r = 0.68$) for clinical diagnosis. Descriptive statistics indicated 61% exact agreement in the diagnosis of the 34 cases between the three diagnostic categories: EVT, AdSD, and AdSD with vocal tremor. Although the laryngologists only outright differed on 6% of the cases between AdSD or EVT, 33% of cases revealed a discrepancy

in determining a concomitant diagnosis. The laryngologist who determined the initial diagnosis met fair agreement ($r = 0.40$) for intrarater reliability. Descriptive analysis found 44% exact agreement against the earlier diagnosis made years previously. The laryngologist changed to or from a diagnosis of AdSD with vocal tremor 41% of the time.

Further analysis on cases where the two otolaryngologists agreed on the diagnosis provided insight into whether a consistent diagnosis produced any changes in reliability across the parameters for the two raters. This resulted in 12 cases that met the criteria (AdSD, $n = 5$; EVT, $n = 7$). The mean results of a second analysis for inter-rater reliability using the Cohen's kappa and a quadratic weighted Cohen's kappa, and intrarater reliability via the Spearman's rank are provided in Table 3. Both raters demonstrated reduced agreement for TVF vibratory parameters with this narrowed number of cases.

Parameter differentiation

The consensus rating was obtained between the two raters following the reliability test to investigate the occurrence of the parameters in the three diagnostic categories: EVT, AdSD, and AdSD with vocal tremor. From this consensus rating, six out of the nine initial parameters were selected to determine if these six parameters distinguished between EVT and AdSD. Arytenoid tremor showed up in less than 10% of the cases for the HSV videos and was therefore excluded. Likewise, two parameters, constant FVF involvement and with R/L TVF asymmetry, were also discarded. This was due to poor inter-rater reliability ($\kappa = 0.266$) for R/L TVF asymmetry, and a high occurrence of

TABLE 2.
Intrarater Reliability via the Spearman's Rho

		R1	R2
Supraglottic parameters			
Arytenoid tremor	Presence of arytenoid tremor	0.698	0.685
	Location of arytenoid tremor	0.740	0.707
Rhythmic supraglottic oscillation	Presence of rhythmic supraglottic oscillation	0.868	0.827
	Degree of rhythmic supraglottic oscillation	0.773	0.873
	Location of rhythmic supraglottic oscillation	0.814	0.816
Supraglottic constriction	Presence of supraglottic constriction	0.633	0.597
	Degree of lateral supraglottic constriction	0.751	0.842
	Degree of A/P supraglottic constriction	0.762	0.697
Constant supraglottic activity	Presence of sustained FVF involvement	0.769	0.557
	Degree of lateral FVF involvement	0.912	0.913
	Location of lateral FVF involvement	0.480	0.407
	Degree of A/P supraglottic activity	0.620	0.738
TVF vibratory parameters			
TVF vibratory disruption	Presence of TVF vibratory disruption	0.812	0.795
	Degree of TVF vibratory disruption	0.823	0.850
Aperiodicity	Presence of aperiodicity	0.296	0.647
	Degree of aperiodicity	0.616	0.769
R/L TVF asymmetry	Presence of R/L TVF asymmetry	0.581	0.757
	Location of R/L TVF asymmetry	0.581	0.711
A/P TVF asymmetry	Presence of A/P TVF asymmetry	0.626	0.742
	Location of A/P TVF asymmetry	0.626	0.659
Phase asymmetry	Presence of phase asymmetry	0.471	0.768

Note: Spearman's correlation coefficient significant at a *P* level of <0.05 (two-tailed).

TABLE 3.
Comparison of Reliability With and Without Concomitant Cases

Parameters	Inter-rater Reliability		Intrarater Reliability			
	R1 vs R2 (n = 34)	R1 vs R2 (n = 12)	R1 (n = 34)	R1 (n = 12)	R2 (n = 34)	R2 (n = 12)
Supraglottic mean	0.558	0.533	0.735	0.840	0.721	0.683
TVF vibratory mean	0.444	0.398	0.604	0.444	0.731	0.744
Total mean reliability	0.510	0.440	0.679	0.670	0.744	0.709

Note: Cohen's kappa correlation significant at a *P* level of <0.05 (two-tailed).

constant FVF involvement (90%) regardless of diagnosis. Thus, the presence or the absence of six parameters, rhythmic supraglottic oscillation, supraglottic constriction, TVF vibratory disruption, aperiodicity, A/P TVF asymmetry, and phase asymmetry, was examined for differentiation between the groups. Of the combined 34 cases analyzed in the present study from the consensus rating, none of the means for the selected parameters met the observed significance level ($P < 0.05$) (Table 4).

Analysis of cases where both otolaryngologists agreed on the diagnosis (N:12) demonstrated the presence of rhythmic supraglottic oscillation ($\chi^2 = 5.182$, $P = .023$) separated AdSD from EVT (Table 5). None of the other parameters produced significant levels of differentiation between the two voice disorders. A frequency distribution for rhythmic supraglottic oscillation comparing the full 34 cases against the reduced number ($n = 12$) can be seen in Table 5. Removing cases with inconsistent diagnoses

TABLE 4.
Differentiation for Initial and Nonconcomitant Cases

Parameters	Initial Cases (n = 34)	Nonconcomitant Cases (n = 12)
Rhythmic supraglottic oscillation	4.415	5.182*
Supraglottic constriction	0.900	0.343
TVF vibratory disruption	1.054	2.743
Aperiodicity	0.895	0.069
Phase asymmetry	4.200	1.029
A/P asymmetry	3.745	1.656

Note: Calculated with a Pearson chi-square analysis.

* Correlation significant at a *P* level of <0.05 (two-tailed).

TABLE 5.
Chi-square Frequency Distribution for the Presence of Rhythmic Supraglottic Oscillation

Parameters	Rhythmic Supraglottic Oscillation (n = 34)		Rhythmic Supraglottic Oscillation (n = 12)*	
	Absent (%)	Present (%)	Absent (%)	Present (%)
EVT	(1) 12.5	(7) 87.5	(1) 14.3	(6) 85.7
AdSD	(8) 50.0	(8) 50.0	(4) 80.0	(1) 20.0
AdSD with tremor	(2) 20.0	(8) 80.0	—	—

Note: Calculated with a Pearson chi-square analysis.

* Observed significance level at $P < 0.05$ (two-tailed).

showed the presence of rhythmic supraglottic oscillation in six out of the seven (85.7%) subjects diagnosed with EVT. A single case of AdSD remained attributed to rhythmic supraglottic oscillation in the consensus rating.

DISCUSSION

The first objective of the present study examined rater reliability in the judgment of laryngeal supraglottic and TVF parameters using HSV. The results indicated stronger inter-rater and intrarater agreement identifying supraglottic characteristics. Subtle, intermittent TVF vibratory patterns proved more difficult to detect given the aperiodic fluctuations seen across glottal cycles. Although reliability did generally improve for supraglottic parameters when concomitant cases were removed from analysis, both raters' reliability decreased in detecting the vocal fold vibratory features investigated in the present study. The second objective, through a consensus rating, examined whether any of the parameters differentiated between EVT, AdSD, and cases of AdSD with EVT when the data were collected with HSV. Supraglottic parameters that were reported in the literature to identify AdSD, such as supraglottic constriction and sustained FVF involvement, however, failed to show any predominance across diagnoses. The results suggest that these features were present in both EVT and AdSD in the current study and did not differentiate these diagnostic groups.

Inter-rater reliability

Although the raters met an adequate, moderate inter-rater reliability⁵⁴ across 76% of their judgments, divergent patterns in agreement emerged for the 21 parameters. The current findings support the initial hypothesis, with the raters reaching the highest level of agreement with supraglottic parameters compared with those focused on vocal fold vibratory movement. The exception to this was TVF vibratory disruption, which met substantial agreement similar to observed levels for other supraglottic parameters. Specifically, the raters demonstrated the highest agreement for parameters that either had more definitive movement patterns across glottal cycles (arytenoid tremor and rhythmic supraglottic oscillation) or could readily be confirmed by pausing the video via a frame-by-frame review (TVF vibratory disruption). Lowell et al³⁷ found a similar moderate-to-substantial inter-rater reliability (mean correlations ranging from $r = 0.480$ – 0.759 post training) for visual-perceptual judgments of tremor in pharyngeal and laryngeal structures using a tremor rating scale.

Inconsistent or mild presenting parameters proved more difficult to judge perceptually in HSV data. Given the nature of these voice disorders, this finding may reflect in the lower reliability for TVF vibratory parameters. The four of the five TVF vibratory characteristics, as well as supraglottic constriction, exhibited poorer reliability in the judgment of their presence and absence in the HSV data. Concerning supraglottic constriction, other laryngeal movements may have impacted the raters' judgments. The presence of intermittent, inconsistent constriction of the FVFs at times mimicked lateral supraglottic constriction. Raters noted this unusual constriction to be found in a small number of cases across all three diagnoses with no discernible pattern. The importance of inconsistent and transient TVF vibratory characteristics during sustained phonation for differential diagnosis may be better highlighted in future studies with the quantification of HSV data, which is the strength of this laryngeal imaging tool. Objective quantification may identify transient patterns across glottal cycles with greater accuracy and consistency than human perception.

Overall, the raters showed fair-to-moderate inter-rater agreement judging the presence or the absence of TVF vibratory parameters ($\kappa = 0.266$ – 0.481). Rosen²⁷ found similar low-to-adequate inter-rater reliability evaluating TVF symmetry (intraclass correlation coefficient [ICC] = 0.54) and TVF periodicity (ICC = 0.37–0.46) for normal and disordered voices. Using simulated stroboscopy derived from HSV, Powell et al⁴⁵ reported low agreement for left-right phase asymmetry (ICC = 0.35). In contrast, Bonilha et al⁵⁶ found higher agreement for left-right phase asymmetry (ranging between $r = 0.75$ and 1.00 within one scalar level) when judging normal voices. However, unlike the current study, the authors did not report an exact agreement or examine dysphonic speakers. Differences in methodology for analyzing agreement, such as examining differences at the scalar level rather than exact agreement, likely elevated the inter-rater agreement compared with the current study.

Both the study design and what is being measured can affect the strength of rater agreement analysis. In the current study, the raters evaluated supraglottic and TVF vibratory features with aperiodic, dysphonic voices. Several studies using videostroboscopy found higher reliability distinguishing normal controls from voice disorders³² or examining parameters for normal voices.⁵⁶ However, even in cases where raters determine a diagnosis using videoendoscopy and HSV rather than identify individual

parameters, previous studies still show mixed-to-poor reliability for dynamic rater judgments.^{29,45}

Despite the use of visual anchors during training and the provision of textual anchors during the ratings, some differences in parameter definitions and judgments of severity persisted between the raters. Although most dichotomous parameters met moderate inter-rater reliability, the presence of sustained FVF involvement and the presence of R/L TVF asymmetry both showed poor agreement for the present study. However, the raters demonstrated greater internal agreement for both sustained FVF involvement ($r = 0.759$ [R1] and $r = 0.557$ [R2]) and R/L TVF asymmetry ($r = 0.581$ [R1] and $r = 0.757$ [R2]) compared with between-rater agreement (sustained FVF involvement $\kappa = 0.287$, R/L TVF asymmetry $\kappa = 0.266$). Discussion during the consensus rating that followed the initial reliability judgments revealed divergent internal reference points for the magnitude of severity, particularly for the mildest presentations of parameters. Poburka et al⁴¹ found greater reliability judging stroboscopy and HSV in their recent validation of the Voice-Vibratory Assessment With Laryngeal Imaging (VALI) form.⁴¹ Incorporating visual anchors into the rating form itself may increase inter-rater reliability for TVF parameters across examinations. However, variation in vibratory patterns between glottic cycles in phonation segment in HSV data might prove to be difficult to summarize in cases like AdSD and EVT.

Intrarater reliability

Similar to the findings for inter-rater reliability, raters showed greater internal consistency judging parameters that displayed consistent supraglottic movements across glottal cycles. Although neither rater met strong (≥ 0.80) agreement, R1 and R2 demonstrated moderate reliability for supraglottic parameters (R1: mean $r = 0.74$, R2: mean $r = 0.73$). Differences in experience level and clinical practice suggest that expert raters may be able to better account for TVF characteristics that appear inconsistently across phonation. R2 (expert) exhibited similar mean reliability between supraglottic ($r = 0.73$) and TVF ($r = 0.72$) parameters. R1 (intermediate) showed more variability rating TVF vibratory parameters ($r = 0.54$) compared with supraglottic parameters ($r = 0.74$). Bonilha et al⁵⁶ found similar differences in ability based on experience level and clinical practice for laryngeal evaluations. Whereas raters who conducted frequent evaluations showed no difference in their ratings for still (kymography) and dynamic evaluations (HSV), raters who conducted fewer laryngeal evaluations in their daily work noted the presence of TVF asymmetries more often using still imaging techniques over dynamic ones.⁵⁶ Thus, the frequency of clinical practice with laryngeal imaging may impact the rater's ability to detect subtle TVF characteristics.

Examining the findings, both raters demonstrated higher mean internal consistency compared with mean inter-rater reliability across parameters. Previous studies report similar gaps within and between groups for perceptual judgments of laryngeal evaluations.^{27,37} The differences seen in the current study, in part caused by diverging internal reference points and rater experience, share commonalities with reliability challenges seen for auditory-perceptual evaluations of voice.^{37,57,58} Although training,

the use of external anchor points, and the study methodology can help close these perceptual gaps for individual raters,⁵⁸ increasing inter-rater reliability can be a challenge. Although training and visual anchors promoted internal consistency for rating tremor location and severity, Lowell et al³⁷ found no significant benefit for inter-rater reliability. How well these gaps may be bridged between raters must be further explored for dynamic, visual-perceptual judgments.

Reliability for cases with an agreed diagnosis

Contrary to the second hypothesis, the inter-rater reliability of TVF parameters did not improve when concomitant cases were removed. Although it was thought that removing cases of AdSD with vocal tremor might improve the inter-rater reliability, focusing on cases with a diagnosis of only EVT or AdSD did not improve the agreement between the two raters when examining vocal fold characteristics. This finding may be due to differences in experience or internal reference points between the raters that were consistent across all diagnostic cases. The reduction in the number of cases may have also affected the outcomes reported. Despite all the efforts to improve reliability and careful choice of parameters based on the literature review on their presence/absence in AdSD and EVT, this study's results mostly demonstrated that subjective perceptual evaluations of the dynamic supraglottic and vocal fold vibratory parameters in HSV data lacked a strong agreement between raters.

Examining the selected cases using audio recordings and videostroboscopy, the two otolaryngologists reached moderate inter-rater agreement ($r = 0.68$) through a combination of auditory and visual ratings. The intrarater reliability of otolaryngologist who made the original diagnosis is comparable with the existing literature for differential diagnosis. Preliminary unpublished findings by Ludlow⁵¹ reported a mean poor reliability for auditory-perception ($r = 0.30$) similar to the poor auditory rater agreement ($\kappa = 0.20$ – 0.33) seen by Johnson et al²⁰ when making a differential diagnosis, and moderate reliability ($r = 0.50$) for nasoendoscopy discriminating between AdSD and EVT. The otolaryngologists differed the most between their judgments of a concomitant diagnosis. Such variance reflects the inherent challenges in standardizing assessment between raters given varying definitions, internal reference points, and clinical experience from person to person and from clinic to clinic.

Parameter differentiation

Against the expected hypothesis, none of the main parameters identified in the present study differentiated between the three diagnoses. Even the presence of rhythmic supraglottic oscillation did not discriminate between cases of EVT from cases of AdSD or AdSD with vocal tremor. Further analysis found a significant difference between cases of EVT and AdSD only when excluding cases that failed to reach a consistent agreement in diagnosis.

Examining the frequency distribution of rhythmic supraglottic oscillation in the reduced subjects ($n = 12$) showed that the consensus rating indicated tremor to be present in 85% of cases diagnosed with EVT (six of the seven). In the case that presented no rhythmic supraglottic oscillation in the video segment,

the source of the tremor may have been located outside of the larynx. Although the distribution of tremor most commonly affects laryngeal structures in individuals with EVT, tremor may also occur (40%) in velopharyngeal, pulmonary, and oral structures.⁴⁸ Although evident in an auditory recording, such oscillations would not be detected visually with a 70° rigid laryngoscope, given the restricted viewpoint to the larynx and the vocal folds. Reviewing the mild, single case of AdSD where the raters determined rhythmic supraglottic oscillation to be present, the video displayed a mild rocking of the larynx that mimicked the tremulous movements associated with rhythmic oscillation. None of the other parameters proved to separate the reduced cases.

Future objective studies with HSV quantifying amplitude and frequency modulation activity at the TVF level and rhythmic supraglottic activity may aid our understanding of tremor types and variation for when it co-occurs with AdSD. Patel *et al*⁸ visually identified the number of oscillatory breaks, motion irregularities, and micromotions using the extracted HSV motion waveforms and the glottal cycle montage in AdSD and muscle tension dysphonia. They reported that the motion irregularity (defined as “segments consisting of nonrepetitive vibratory cycles in the extracted vocal fold area waveform” was the only significant predictor of AdSD. Given the variability of parameters across glottal cycles, using static image analysis of laryngeal HSV with kymography may reveal distinctions that perceptual evaluations of dynamic movement alone cannot; however, this method may not allow the investigation of vibratory patterns along the full length of vocal folds. Furthermore, given the subtleties and variations in the presentation of these disorders, objective quantification of laryngeal parameters may allow for a more precise identification, offering consistency toward the discovery of the most pertinent parameters.

Given that none of the hypothesized parameters differentiated between the diagnoses, further study is warranted to determine if these results hold true with larger sample sizes. How well we can truly separate these factors given the variability, concomitance, and accuracy of diagnoses for EVT and AdSD should also be investigated in future studies.

Limitations

The inconsistency in presentation of certain parameters across glottal cycles within the same phonation segment may have reduced rater agreement. Because of the aperiodic fluctuations of the TVFs and the variation in supraglottic movements from cycle to cycle for these voice disorders, 1- to 2-second sections were selected. Raters may have been more likely to miss the presentation of intermittent vibratory and supraglottic features. The results of the present study suggest that perceptual judgments using dynamic laryngeal HSV in cases of aperiodic voice disorders may not adequately capture the fluctuating vibratory patterns that can vary across glottal cycles. Using static composite analysis of glottal cycles might promote higher inter-rater reliability for subjective judgments,^{8,26} or future studies

employing objective quantification of the vibratory patterns are needed.

The development of quantitative and objective image processing tools for HSV imaging techniques may have better clinical applications to aid in the differential diagnosis of these voice disorders. There is a long history examining the image segmentation of the glottal area waveform to examine vocal fold kinematics.⁵⁹ Given the variability of parameter presentation, the effects of training, and potential perceptual limitations with dynamic video, objective measures may offer more reliable outcomes for vocal fold vibratory characteristics.

The use of a textual rating form alone without the use of visual anchors in the rating form itself or used in parallel during the ratings may have also impacted reliability. Providing a visual anchor for each parameter built into the rating form and not just provided during training may promote increased reliability for future studies.⁴¹

The current limitations of HSV using a rigid endoscope restrict both the image recorded during data collection and the voice sample to a sustained vowel /i/. Other characteristics not captured within the analysis frame may still impact these voice disorders and contribute to a potential diagnosis. Similarly, although current collection methods allow for adequate perception of vocal tremor, sustained phonation may not be best suited for eliciting the characteristics of AdSD. Vocal tremor is most clearly perceived during sustained phonation compared with connected speech.⁶⁰ Research also shows that listeners perceive more voice modulation with vocal tremor audio samples with the vocal tract shape for /i/ vs /a/.⁵² However, previous studies indicate the clearest perception of AdSD on voiced consonant sentences.⁶¹ Although the raters observed an intermittent hyperadduction of the supraglottic laryngeal structures characteristic of AdSD in the video segments for this current study, the use of a sustained vowel may not allow for the best differentiation of movement patterns from EVT, especially in milder cases.

Clinical implications

HSV allows visualization of the individual glottal cycles for EVT and AdSD where stroboscopy cannot. However, subjective perceptual judgments alone without the use of visual anchors during ratings may not provide a reliable parameter identification between raters. Rater experience, concomitant characteristics, and the intermittent presentations of parameters for aperiodic voice disorders create challenges in consistent ratings for dynamic video examinations. Thus, HSV is a viable examination tool for aperiodic and severe voices but ought to be employed in conjunction with other assessments used in clinical practice, including auditory and acoustic measures, the patient’s physical presentation, and relevant medical history. Continued focus on standard methodologies to close these gaps in disagreement between raters is required. To increase precision, analysis of HSV data with objective quantification measures may provide researchers and clinicians with new insights for differentiation.

APPENDIX. PARAMETER DEFINITIONS

Supraglottic Parameters

Arytenoid tremor	Absent = no arytenoid tremor present Present = consistent "wobbly" rhythmic motion of arytenoid structures during phonation
Location of arytenoid tremor	0 = No tremor 1 = Lateral rhythmic arytenoid cartilage motion, pulling apart and coming back toward each other along the horizontal axis 2 = Anterior rhythmic arytenoid cartilage motion, which looks like the arytenoids are bobbing along a vertical axis 3 = Consistent, rhythmic movement of the arytenoids along both axis
Rhythmic supraglottic oscillation	Absent = no rhythmic oscillation during phonation Present = the presence of rhythmic compression of the supraglottic structures during a period of phonation in the middle of the glottal cycle
Degree of rhythmic supraglottic oscillation	0 = No rhythmic oscillation 1 = Consistent presence of lateral or anterior-posterior rhythmic compression of the supraglottic structures during phonation in the middle of the glottal cycle 2 = Occasional lateral or anterior-posterior rhythmic compression of the supraglottic structures during a period of phonation in the middle of the glottal cycle
Location of rhythmic supraglottic oscillation	0 = No rhythmic oscillation 1 = Lateral movement (along the horizontal axis) 2 = Anterior-posterior movement (along the vertical axis) 3 = Both movements combined
TVF vibratory disruption	Absent = Continuous TVF vibration during phonation Present = Disruption of the TVF vibratory movement during the glottal cycle viewed via slowed down playback
Degree of TVF vibratory disruption	0 = Continuous TVF vibration 1 = Disruption of the TVF vibratory pattern without complete cessation during the glottal cycle, limited movement 2 = Brief cessation of the TVFs with no vibratory movement during the glottal cycle of less than five frames 3 = Prolonged cessation of the TVFs with no vibratory movement during the glottal cycle of less than five frames
Supraglottic constriction	Absent = no supraglottic constriction during voice initiation Present = supraglottic constriction during the onset of vibratory motion of the TVFs following respiration; glottal attack
Degree of lateral supraglottic constriction	0 = No supraglottic constriction during voice initiation 1 = Compression of the FVFs and supraglottic structures covering one-third of the TVF width 2 = Compression of the FVFs and supraglottic structures covering two-thirds of the TVF width 3 = Complete covering of the TVFs via compression of the FVFs and supraglottic structures
Degree of anterior-posterior supraglottic constriction	0 = No supraglottic constriction during voice initiation 1 = Compression of the supraglottic structures covering one-third of the TVF length 2 = Compression of the supraglottic structures covering two-thirds of the TVF length 3 = Complete covering of the TVFs via compression of supraglottic structures
Sustained FVF involvement	Absent = no sustained FVF involvement during phonation Present = sustained involvement of the FVFs during phonation. The FVFs are seen to be pressing in and covering over part or all of the TVFs and remain in such a position.
Degree of sustained FVF involvement	0 = No FVF involvement 1 = FVF compression covering one-third of the width of the TVFs 2 = FVF compression covering two-thirds of the width of the TVFs 3 = Complete involvement of the FVFs covering over the whole of the TVFs

(Continued)

Supraglottic Parameters

Location of lateral sustained FVF involvement	0 = No FVF involvement 1 = Greater right FVF involvement 2 = Greater left FVF involvement 3 = Equal right and left FVF involvement
Degree of anterior-posterior sustained FVF involvement	0 = No supraglottic activity during sustained phonation 1 = Compression of the supraglottic structures covering one-third of the TVF length 2 = Compression of the supraglottic structures covering two-thirds of the TVF length 3 = Complete covering of the TVFs via compression of supraglottic structures
TVF parameters	
Aperiodicity	Absent = The TVFs come together with regular precision and consistent duration for the glottal cycles, approximately >80% regular Present = the TVFs do not come together with regular precision or any consistency during glottal cycles, approximately <20% regularity.
R/L TVF asymmetry	Absent = Both the left and the right TVFs come together with the same pattern, in sync, during the glottal cycle. Present = horizontal asymmetry between the right and the left TVFs coming together in uneven patterns during the glottal cycle
Location of R/L TVF asymmetry	0 = Symmetric motion between R and L TVFs 1 = Right TVF asymmetry 2 = Left TVF asymmetry 3 = Asymmetry is presented for both R and L TVFs.
A/P TVF asymmetry	Absent = Both the anterior and posterior portions of each TVF come together with the same pattern top and bottom, in sync together during the glottal cycle. Present = vertical asymmetry along a single vocal fold, where the anterior and posterior portions of each TVF are out of sync during the glottal cycle
Location of A/P TVF asymmetry	0 = Symmetric motion along the same fold 1 = Right TVF asymmetry 2 = Left TVF asymmetry 3 = Asymmetry is presented for both R and L TVFs along the same fold.
Phase asymmetry	Absent = The TVFs consistently come together and meet at the midline of the glottis. Present = The TVFs come together past the midline; one vocal fold overlapping onto the other during the glottal cycle.

REFERENCES

- Kreiman J, Gabelman B, Gerratt BR. Perception of vocal tremor. *J Speech Lang Hear Res.* 2003;46:203–214.
- Warrick P, Dromey C, Irish J, et al. The treatment of essential voice tremor with botulinum toxin A: a longitudinal case report. *J Voice.* 2000;14:410–421.
- Lester RA, Barkmeier-Kraemer J, Story BH. Physiologic and acoustic patterns of essential vocal tremor. *J Voice.* 2013;27:422–432.
- Ludlow CL, Adler CH, Berke GS, et al. Research priorities in spasmodic dysphonia. *Otolaryngol Head Neck Surg.* 2008;139:495–505.
- Tanner K, Roy N, Merrill RM, et al. Case-control study of risk factors for spasmodic dysphonia: a comparison with other voice disorders. *Laryngoscope.* 2012;122:1082–1092.
- Murry T. Spasmodic dysphonia: let's look at that again. *J Voice.* 2014;28:694–699.
- Orbelo DM, Duffy JR, Hughes Borst BJ, et al. Differences in botulinum toxin dosing between patients with adductor spasmodic dysphonia and essential voice tremor. *J Voice.* 2014;28:123–127.
- Patel RR, Liu L, Galatsanos N, et al. Differential vibratory characteristics of adductor spasmodic dysphonia and muscle tension dysphonia on high-speed digital imaging. *Ann Otol Rhinol Laryngol.* 2011;120:21–32.
- Lundy DS, Roy S, Xue JW, et al. Spastic/spasmodic vs. tremulous vocal quality: motor speech profile analysis. *J Voice.* 2004;18:146–152.
- Sulica L, Louis ED. Clinical characteristics of essential voice tremor: a study of 34 cases. *Laryngoscope.* 2010;120:516–528.
- Patel AB, Bansberg SF, Adler CH, et al. The Mayo Clinic Arizona spasmodic dysphonia experience: a demographic analysis of 718 patients. *Ann Otol Rhinol Laryngol.* 2015;124:859–863.
- Roy N, Mauszycki SC, Merrill RM, et al. Toward improved differential diagnosis of adductor spasmodic dysphonia and muscle tension dysphonia. *Folia Phoniatr Logop.* 2007;59:83–90.
- Barkmeier JM, Case JL, Ludlow CL. Identification of symptoms for spasmodic dysphonia and vocal tremor: a comparison of expert and nonexpert judges. *J Commun Disord.* 2001;34:21–37.
- Whurr R, Lorch M. Review of differential diagnosis and management of spasmodic dysphonia. *Curr Opin Otolaryngol Head Neck Surg.* 2016;24:203–207.
- Creighton FX, Hapner E, Klein A, et al. Diagnostic delays in spasmodic dysphonia: a call for clinician education. *J Voice.* 2015;29:592–594.
- White L, Klein A, Hapner E, et al. Co-prevalence of tremor with spasmodic dysphonia: a case-control study: co-morbid tremor in spasmodic dysphonia. *Laryngoscope.* 2011;121:1752–1755.
- Tanner K, Roy N, Merrill RM, et al. Spasmodic dysphonia: onset, course, socioemotional effects, and treatment response. *Annals of Otolaryngology, Rhinology & Laryngology.* 2011;120:465–473.
- Tanner K, Roy N, Merrill RM, et al. Risk and protective factors for spasmodic dysphonia: a case-control investigation. *J Voice.* 2011;25:e35–e46.
- Barkmeier-Kraemer J, Story B. Conceptual and clinical updates on vocal tremor. *ASHA Leader.* 2010;15:16–19.
- Johnson DM, Hapner ER, Klein AM, et al. Validation of a telephone screening tool for spasmodic dysphonia and vocal fold tremor. *J Voice.* 2014;28:711–715.
- Chhetri DK, Merati AL, Blumin JH, et al. Reliability of the perceptual evaluation of adductor spasmodic dysphonia. *Ann Otol Rhinol Laryngol.* 2008;117:159–165.

22. Mehta DD, Deliyski DD, Hillman RE. Commentary on why laryngeal stroboscopy really works: clarifying misconceptions surrounding Talbot's law and the persistence of vision. *J Speech Lang Hear Res.* 2010;53:1263–1267.
23. Deliyski DD, Hillman RE. State of the art laryngeal imaging: research and clinical implications. *Curr Opin Otolaryngol Head Neck Surg.* 2010;18:147–152.
24. Mehta DD, Hillman RE. Current role of stroboscopy in laryngeal imaging. *Curr Opin Otolaryngol Head Neck Surg.* 2012;20:429–436.
25. Deliyski DD, Petrushev PP, Bonilha HS, et al. Clinical implementation of laryngeal high-speed videoendoscopy: challenges and evolution. *Folia Phoniatr Logop.* 2008;60:33–44.
26. Tsuji DH, Hachiya A, Dajer ME, et al. Improvement of vocal pathologies diagnosis using high-speed videolaryngoscopy. *Int Arch Otorhinolaryngol.* 2014;18:294–302.
27. Rosen CA. Stroboscopy as a research instrument: development of a perceptual evaluation tool. *Laryngoscope.* 2005;115:423–428.
28. van Gogh CD, Verdonck-de Leeuw IM, Boon-Kamma BA, et al. The efficacy of voice therapy in patients after treatment for early glottic carcinoma. *Cancer.* 2006;106:95–105.
29. Mendelsohn AH, Remacle M, Courey MS, et al. The diagnostic role of high-speed vocal fold vibratory imaging. *J Voice.* 2013;27:627–631.
30. Nawka T, Konerding U. The interrater reliability of stroboscopy evaluations. *J Voice.* 2012;26:812, e811–812.e810.
31. Poburka BJ. A new stroboscopy rating form. *J Voice.* 1999;13:403–413.
32. Peppard RC, Bless DM. A method for improving measurement reliability in laryngeal videostroboscopy. *J Voice.* 1990;4:280–285.
33. Sulter AM, Schutte HK, Miller DG. Standardized laryngeal videostroboscopic rating: differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency, and age. *J Voice.* 1996;10:175–189.
34. Yamauchi A, Imagawa H, Yokonishi H, et al. Evaluation of vocal fold vibration with an assessment form for high-speed digital imaging: comparative study between healthy young and elderly subjects. *J Voice.* 2012;26:742–750.
35. Patel RR, Dailey S, Bless D. Comparison of high-speed digital imaging with stroboscopy for laryngeal imaging of glottal disorders. *Annals of Otolaryngology & Laryngology.* 2008;117:413–424.
36. Inwald EC, Dollinger M, Schuster M, et al. Multiparametric analysis of vocal fold vibrations in healthy and disordered voices in high-speed imaging. *J Voice.* 2011;25:576–590.
37. Lowell SY, Kelley RT, Busekroos L, et al. The effect of anchors on reliability of endoscopic tremor ratings. *Laryngoscope.* 2017;127:411–416.
38. Bonilha HS, Focht KL, Martin-Harris B. Rater methodology for stroboscopy: a systematic review. *J Voice.* 2015;29:101–108.
39. Focht KL, Martin-Harris B, Bonilha HS. Stroboscopic parameters reported as voice outcome measures in patients treated for laryngeal cancer: a systematic review. *J Med Speech Lang Pathol.* 2013;21.
40. Poburka BJ, Bless DM. A multi-media, computer-based method for stroboscopy rating training. *J Voice.* 1998;12:513–526.
41. 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;31:513.e1–513.e14.
42. Hillel AT, Johns MM 3rd, Hapner ER, et al. Voice outcomes from subligamentous cordectomy for early glottic cancer. *Ann Otol Rhinol Laryngol.* 2013;122:190–196.
43. Chang J, Fang T-J, Yung K, et al. Clinical and histologic predictors of voice and disease outcome in patients with early glottic cancer. *Laryngoscope.* 2012;122:2240–2247.
44. Dejonckere PH. Clinical implementation of a multidimensional basic protocol for assessing functional results of voice therapy. A preliminary study. *Rev Laryngol Otol Rhinol (Bord).* 2000;121:311–313.
45. Powell ME, Deliyski DD, Hillman RE, et al. Comparison of videostroboscopy to stroboscopy derived from high-speed videoendoscopy for evaluating patients with vocal fold mass lesions. *Am J Speech Lang Pathol.* 2016;25:576–589.
46. Kunduk M, Yan Y, McWhorter AJ, et al. Investigation of voice initiation and voice offset characteristics with high-speed digital imaging. *Logoped Phoniatr Vocol.* 2006;31:139–144.
47. Zacharias SRC, Deliyski DD, Gerlach TT. Utility of laryngeal high-speed videoendoscopy in clinical voice assessment. *J Voice.* 2017;doi:10.1016/j.jvoice.2017.05.002.
48. Hemmerich AL, Finnegan EM, Hoffman HT. The distribution and severity of tremor in speech structures of persons with vocal tremor. *J Voice.* 2017;31:366–377.
49. Colton RH, Casper JK, Leonard R. *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment.* 4th ed. Baltimore, MD: Lippincott Williams & Wilkins; 2011.
50. Anand S, Shrivastav R, Wingate JM, et al. An acoustic-perceptual study of vocal tremor. *J Voice.* 2012;26:811, e811–817.
51. Ludlow CL. *Recent Understanding in Spasmodic Dysphonia.* Phoenix, AZ: National Spasmodic Dysphonia Association; 2012.
52. Lester RA, Story BH. The effects of physiological adjustments on the perceptual and acoustical characteristics of simulated laryngeal vocal tremor. *J Acoust Soc Am.* 2015;138:953–963.
53. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol.* 2012;8:23–34.
54. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33.
55. Ferguson C. An effect size primer: a guide for clinicians and researchers. *Prof Psychol Res Pr.* 2009;40:532–538.
56. Bonilha HS, Deliyski DD, Gerlach TT. Phase asymmetries in normophonic speakers: visual judgments and objective findings. *Am J Speech Lang Pathol.* 2008;17:367–376.
57. Kreiman J, Gerratt BR, Precoda K. Listener experience and perception of voice quality. *J Speech Hear Res.* 1990;33:103–115.
58. Kreiman J, Gerratt BR, Ito M. When and why listeners disagree in voice quality assessment tasks. *J Acoust Soc Am.* 2007;122:2354–2364.
59. Woo P. Objective measures of laryngeal imaging: what have we learned since Dr. Paul Moore. *J Voice.* 2014;28:69–81.
60. Lederle A, Barkmeier-Kraemer J, Finnegan E. Perception of vocal tremor during sustained phonation compared with sentence context. *J Voice.* 2012;26:668, e661–668.e669.
61. Cannito MP, Chorna LB, Kahane JC, et al. Influence of consonant voicing characteristics on sentence production in abductor versus adductor spasmodic dysphonia. *J Voice.* 2014;28:394, e313–322.